Value Determinant Visualization Tool - US Housing Market

Final Report
Team 74

[redacted]

Georgia Institute of Technology
04/21/2023

1 INTRODUCTION

Increasing housing prices has been a topic of concern over the last few decades. In recent years, we have experienced significant fluctuations in the housing market as global economic policies created shifts in the financial market. One notable economic factor is interest rates. There has been an increase in homes ownership as buyers hope to take advantage of lower rates, subsequently home prices have increase due to lower inventory. Other than interest rates, what other economic factors contribute to the pricing fluctuations. Platforms, such as Zillow only provide tools to return listing prices. How can we deconstruct the pricing of homes into various features to allow home buyers to make more informed decisions?

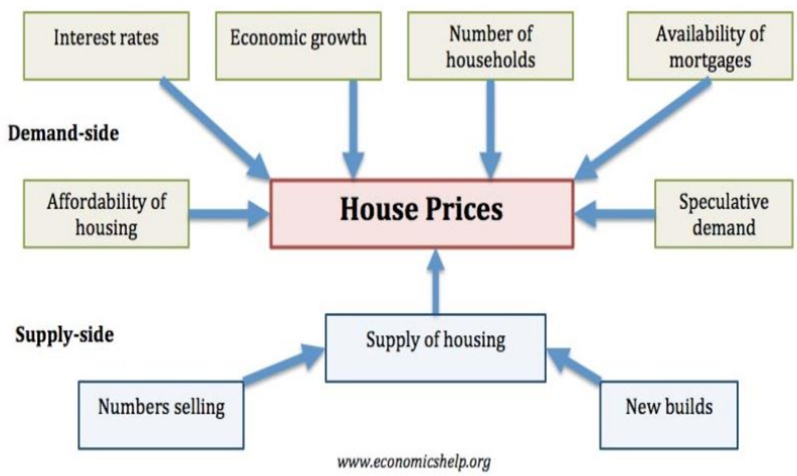

Figure 1: Common Factors Contributing to House Prices

2 PROBLEM DEFINITION

Our goal is to create an interactive tool to measure the weight of macroeconomic factors on housing prices across multiple regions the United States. One of the main measures for housing prices used today is the House Price Index (HPI). HPI's limitations is using data only from Fannie Mae and Freddie Mac and does not include loans and rates from other sources. Our approach will use historical data to create our own housing price index using an analysis from economic trends. The incorporation of novel visualizations to illustrate this updated analysis will also provide more easily ingest-able insights as opposed to more typical report formats. The target audience are those involved in the US real estate market including real estate/rental agencies, home buyers, government agencies, mortgage lenders, and investment firms. The risks associated with this approach are availability bias affecting housing prices and selection bias from data acquisition. Another risk is the limited amount of open-source data that is available concerning home price data, a key required input data set for this analysis to properly run. The payoff from our approach is a current snapshot of housing prices to allow for rapid application.

3 LITERATURE SURVEY

In the past few years, homes have become even more unaffordable, as the rate of increase has multiplied. In the Netherlands, from 1900-1985, housing prices grew an average of 0.23% per year[5]. However, from 1985 to 2010 housing prices have grown an average of 7.15% per year. This increase can be explained by macro-economic determinants such as: lumber costs, labor costs, GDP per capital, unemployment rates and housing supply [1][5]. Linear regression analysis is commonly used to determine the weight of each factor [12].

Pasharde and Savva focuses on the Cyprus housing market; they determined that the cost of labor and housing material had the highest correlation to price changes, whereas interest and loan rate had the lowest correlation [13]. However due to scarcity of data available, there may be unaddressed bias in their findings. Tripathi also using regression analysis, concluded that factors such as: rent, income, GDP and real exchange rates were the most influential factors, but agrees that interest rate had near zero impact. [17]

On a global scale, the housing market can be simplified into three categories: supply factors, demand factors and structural and institutional factors [7]. As the supply of homes decrease and income remains stagnant, prices will continue to rise. Institutional tax breaks and lowered mortgage and interest rates contribute to an increase in property investment leading to more demand. Vargas-Silva examines the impact of monetary policy shocks on the US housing market [18]. He concludes that there is an overall negative response to restrictive monetary policy on the housing market. One shortcoming Nang highlights is the effect of monetary policy and structural changes on short term trends in the data, leading the over valuation and shifts in equilibrium [7].

Housing valuation can also be explained with the hedonic price method, which incorporates Rosen's theory of supply and demand [9][15]. Rosen's demand theory is based on how the characteristics of the consumer influence their purchasing decisions. Although the consumer's purchasing power may be indirectly affected by economic conditions. Additional analysis also shows that lowering interest rates led to the higher demands [14]. Calhoun discusses the OFHEO house price indexes based on mortgage transactions received from Freddie Mac and Fannie Mae [2]. Corelogic, a mortgage company created their own HPI to make determination on home valuations [4]. House price indexes are widely used within the U.S., and we will create our own pricing index for our analysis.

We need to also consider micro-level factors from a consumer's perspective when looking at home valuation factors [16]. On a micro economic scale, home valuation is further influences by factors such as home size, home age, demography, location, population, and number of renovations [3][6][8]. On a local level, crime rate, foreclosure rate and elderly population may play a greater role in housing price fluctuations [11].

Lastly, Kang creates a weighted regression analysis using geographical data from urban homes in Boston to create a deep learning model to predict appreciation rate [10]. In our survey, the common method of analysis used were correlation coefficient and linear regression. Shortcomings range from the scarcity in data, undetected outliers from short term trends, and pricing determinants not included in the valuation model.

4 PROPOSED METHOD

4.1 Innovations

(1) **Dynamic Parameter Inputs -** The majority of the literature we reviewed hyper-focused on a small geographic location when performing their analysis. While this enables the analysis to be very detailed, it also means the results can only be applied to that local area. Our analysis expands to across multiple years, and a multitude of different regions in the US, that can be selected at the user's discretion.
(2) **House Price Index Forecasting -** Implementing an autoregressive integrated moving average (ARIMA) model using existing data from state and federal housing price index to evaluate trends. The model will forecast the HPI value up to ten years, allowing users to anticipate shifts in the housing market. Though housing price prediction may be impacted other factors, the ARIMA model is one option to make informed purchasing and selling decisions.
(3) **Feature Selection -** The incorporation of a regression model when selecting the region and time span we want to investigate gives the user the choice when performing feature selection. Instead of trying to squeeze 10-20 features that predict housing price onto a cluttered graph, this step allows the user to dynamically show only the most important features that affect housing price for a given set of inputs.

4.2 Intuition

We are looking to evaluate the various features individually have on housing prices and develop a model that can be used across multiple regions, while displaying it in an interactive visualization. Our leveraging of features such as HPI, aims to address the shortcomings previously addressed in the literature survey. Our model will allow for a wider range of pricing estimates outside of standard single region analysis as we will be using both local and federal data.

4.3 Technical Overview

- **Data -** Our data is sourced from a mixture of both federal level and local data. Our federal level data is mainly sourced from Federal Reserve Economic Data (FRED), a source of US economic data hosted and made publicly available by the Federal Reserve Bank of St. Louis. This federal data does not vary from region-to-region, but still may be significant for a given region.

  We also used local data, which, unlike federal data, tended to vary from region-to-region. For these sources of data, we still used sources hosted by the federal government, but used data sources that were segmented by region (i.e. state-level) like unemployment rate from BLS Beta Labs (beta.bls.gov), and median household income from the the US Census (census.gov).

- **Inputs -** Curated features have been loaded into our scripting environment, cleaned, and normalized. Our tool utilizes a customization step that allows the end-user to specify two specific ranges that filter the data being investigated: timespan and state.

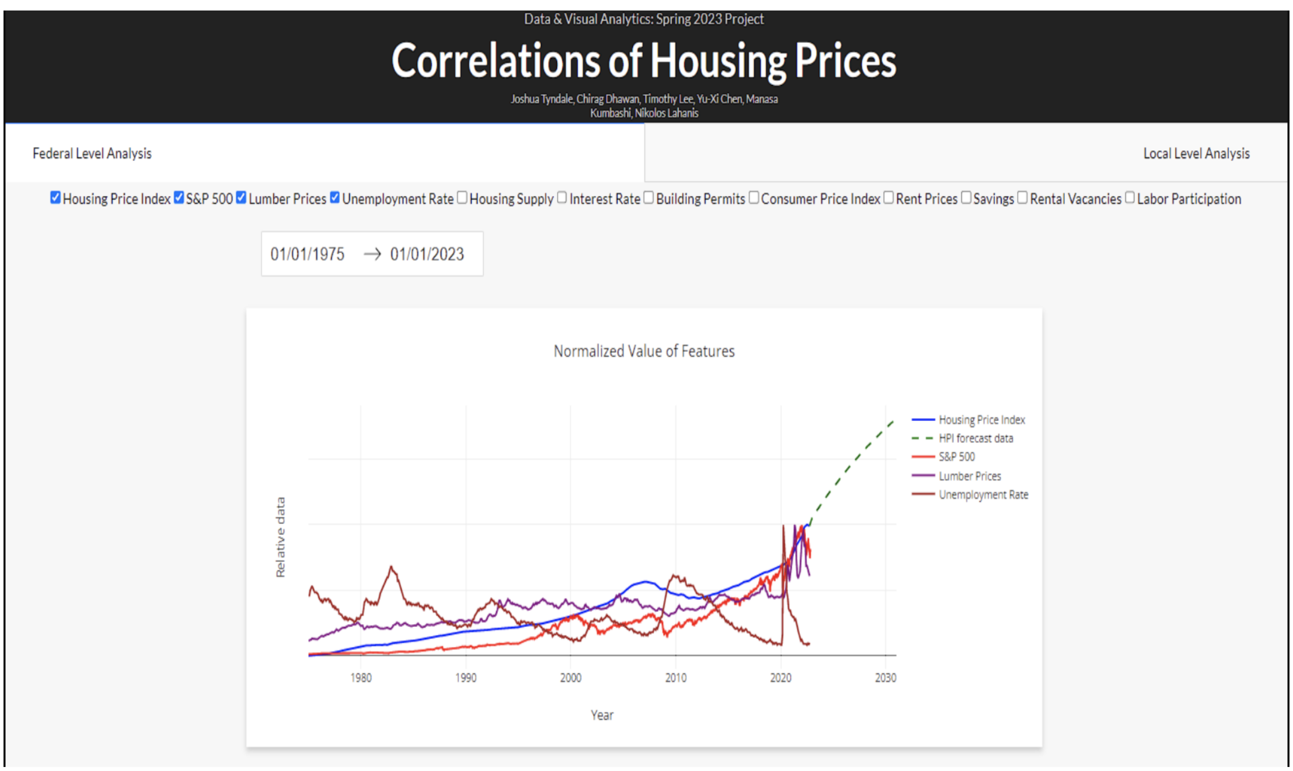

Figure 2: Federal Level Plot

- **Feature Selection-** Our team believes the best way to solve this is through a statistical method known as regression. By implementing a regression model over this data set, we can identify the features that have the strongest predictability associated with housing price by investigating their coefficients. By choosing the N features with the largest absolute coefficients, we can inherently perform feature selection, and only visualize the most important features to house prices for a given state and timespan.
- **Output -** Now that we have finalized on filtering and normalizing all our data sets, with only the most significant features chosen for the range of values specified by the input user. To perform the visualization, we are using Dash, a framework that is part of the larger Plotly visualization library in Python, to visualize the narrowed data set and show the end user the most significant features that correlate to housing price based on their input range.
- **Usability -** Our visualization interface will display two tabs highlighting either Federal Level Analysis or Local Level Analysis. In both levels the graph will be populated by features implemented in our model, the user will then be able to see real time changes as they select which features to include. The user also the ability to narrow down the time frame to the exact dates they want to view. In the local level tab, there is a scroll down menu populated by all fifty states, each with its own visualization of the selected features. The main priority our visualization interface is the provide the user a tool that is not only aesthetically palatable, but also analytically reliable.

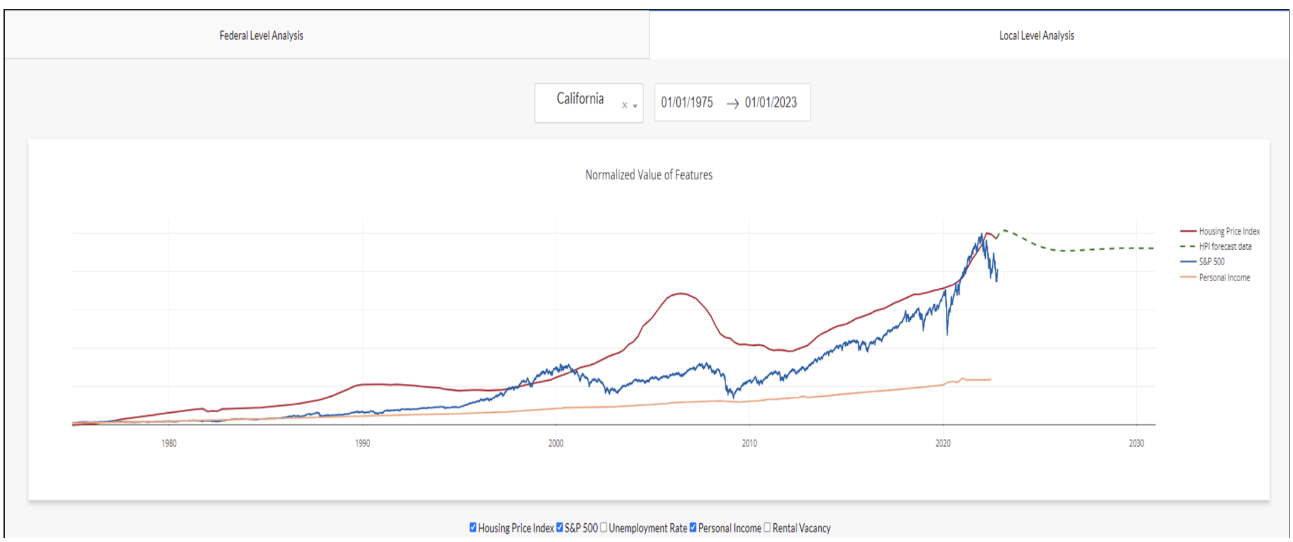

Figure 3: Local Level Plot

5 EXPERIMENTS & EVALUATION

5.1 Questions for Experiment

1) Which features are most/least significant and do they align with the findings from our literature survey?

2) How does the local level analysis compare to the federal level? Specifically, which states deviates from the common trends?

3) What methods can we apply to make our user interface more user friendly?

5.2 Experimental Observations

1) Regression Analysis and Correlation Coefficients-

In both federal and state level analysis tabs, our tool displays a regression analysis bar chart and coefficients table. Our observations were conducted with all features selected at the Federal level in both Figure 4 and 5. In Figure 4, we have observed that Lumber had the highest regression coefficient at 1.2099 and interest rates had the lowest -0.3784.

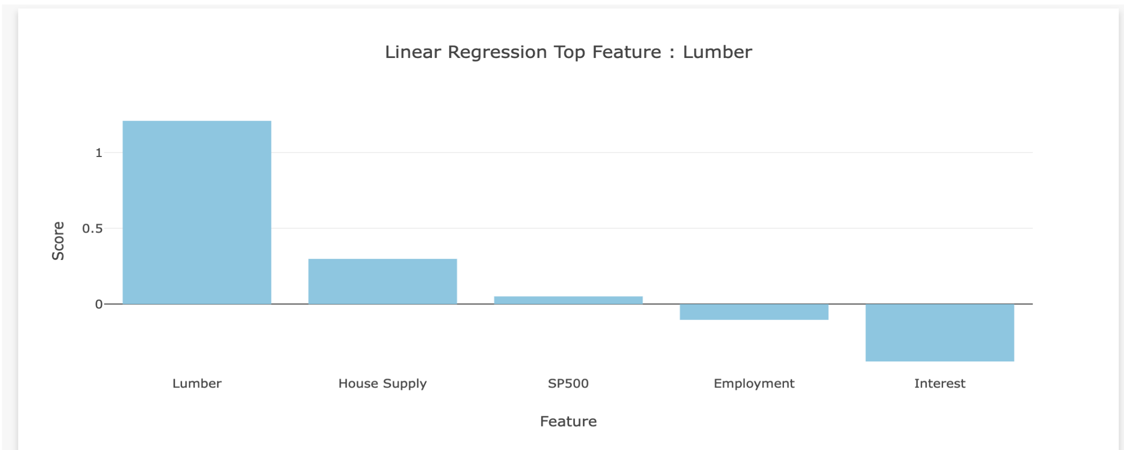

Figure 4: Federal Regression Analysis

In Figure 5, we observed that Rent Prices, S&P 500, and Lumber Prices with a correlation coefficient of 0.9748, 0.9293 and 0.8649 respectively are the top three correlated features. Whereas, Savings, Interest Rates and Unemployment rate with a correlation of -0.8588, -0.7241 and -0.3675 respectively are the three lowest correlated features.

Data_Set_1	Data_Set_2	Correlation_Coefficient
Housing Price Index	Rent Prices	0.9748
Housing Price Index	S&P 500	0.9293
Housing Price Index	Lumber Prices	0.8649
Housing Price Index	Consumer Price Index	0.7086
Housing Price Index	Labor Participation	0.6947
Housing Price Index	Rental Vacancies	0.4008
Housing Price Index	Building Permits	-0.0013
Housing Price Index	Housing Supply	-0.0773
Housing Price Index	Unemployment Rate	-0.3675
Housing Price Index	Interest Rate	-0.7241
Housing Price Index	Savings	-0.8588

Figure 5: Federal Level Coefficients

2) Federal vs State- At the Federal level, our HPI forecast steadily increases over our ten year time frame. The forecast ends at January, 2031 with a coefficient of 180.915. When compared to the various states forecast on the local level. We observed similarities of HPI growth in most states, some of which indicate a higher level of increase compared to Federal. In states such Alaska HPI remained at 102.4792 over the time frame, and

California HPI begins at 101.4434 on January 2023 ends at 92.0689 on January, 2031. We observed that they deviated from the common trend. These states displayed either stagnant growth or a decreasing HPI pattern.

3) User Interface Design- While designing our tool, we considered implementing design concepts discussed in week 4 lectures, such as, Color Scheme, Font, and Gestalt's 8 Laws of Grouping. Our charts initially, were designed with darker backgrounds, which we adjusted to white and light grey. We adjusted colors of line chart based on the order they are presented on the y-axis, from top to bottom. The color palette of each feature was chosen using an open-source color palette generator. While the user adding in the filters for features, time span and region, we decided to incorporate the mouse-over function to highlights the exact coefficient of a feature at a given time.

6 CONCLUSION & DISCUSSION

When comparing our findings to the research conducted in the literature survey, we found that our results are in line with theirs. Rental prices and building material, such as Lumber are still the most influential features. On the other hand, unemployment rates and interest rates near zero or negative influence. We have also found that although most states follow the Federal level growth trend, some states will not expect to see an increase in HPI. The results from our tool provides users reliable information to make better and more informed decisions in the home buying and selling process.

6.1 Limitations & Future Extensions

One of our main limiting factor was finding relevant data sets during our exploratory analysis phase. However, because of this limitation, we expanded on our idea to cover both federal and state level housing price. Given more time to continue to work on this project, we have identified several areas to improve upon. We plan on implementing a statistical method to identify whether the feature demonstrates correlation or causation. By primarily using Python and Plotly, we were limited in visual customization as well as feature loading time, which can be improved by using JavaScript and D3. Lastly, we plan on exploring alternative regression models to compare results to improve overall model accuracy.

6.2 Work Distribution

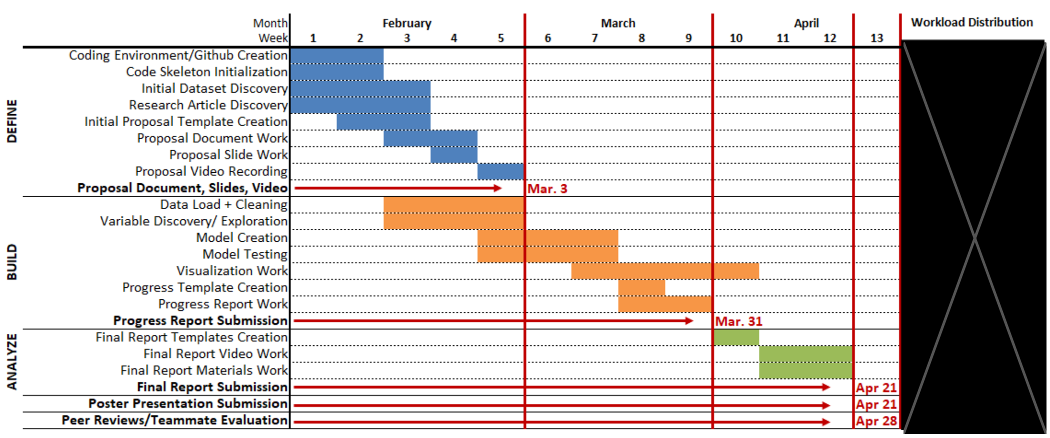

Figure 6: Timeline & Work Distribution
**all team members have evenly distributed the workload for this project*

REFERENCES

[1] Bernardina Algieri. “House price determinants: Fundamentals and underlying factors”. In: *Comparative Economic Studies* 55 (2013), pp. 315–341.

[2] Charles A Calhoun. “OFHEO house price indexes: HPI technical description”. In: *Office of Federal Housing Enterprise Oversight* 20552 (1996), pp. 1–15.

[3] Dennis Capozza et al. “Determinants of Real House Price Dynamics”. In: (Nov. 2002). DOI: 10.3386/w9262.

[4] CoreLogic. *Special Report: The Role of Housing in the Longest Economic Expansion.* 2021. URL: https://www.corelogic.com/wp-content/uploads/sites/4/2021/05/report-Special-Report-The-Role-of-Housing-in-the-Longest-Economic-Expansion.pdf.

[5] Martijn Droes, Alex van de Minne, et al. “Do the determinants of house prices change over time? Evidence from 200 years of transactions data”. In: (2016).

[6] Gabriele Galati, Federica Teppa, and Rob JM Alessie. “Macro and Micro Drivers of House Price Dynamics: An Application to Dutch Data”. In: (2011).

[7] Ms Nan Geng. *Fundamental drivers of house prices in advanced economies.* International Monetary Fund, 2018.

[8] Edward Glaeser and Joseph Gyourko. “The Economic Implications of Housing Supply”. In: *Journal of Economic Perspectives* 32.1 (Feb. 2018), pp. 3–30. DOI: 10.1257/jep.32.1.3. URL: https://www.aeaweb.org/articles?id=10.1257/jep.32.1.3.

[9] Shanaka Herath and Gunther Maier. “A FRAMEWORK FOR ANALYSING HOUSE PRICES USING TIME, SPACE AND QUALITY CRITERIA”. In: *17th Annual European Real Estate Society Conference in Milan, Italy* (May 2010).

[10] Yuhao Kang et al. “Understanding house price appreciation using multi-source big geo-data and machine learning”. In: *Land Use Policy* 111 (2021), p. 104919. ISSN: 0264-8377. DOI: https://doi.org/10.1016/j.landusepol.2020.104919. URL: https://www.sciencedirect.com/science/article/pii/S0264837719316746.

[11] Wei-Shong Lin et al. “Effects of socioeconomic factors on regional housing prices in the USA”. In: *International Journal of Housing Markets and Analysis* 7.1 (2014), pp. 30–41.

[12] Sharmila Muralidharan et al. “Analysis and prediction of real estate prices: a case of the Boston housing market”. In: *Issues in Information Systems* 19.2 (2018), pp. 109–118.

[13] Panos Pashardes and Christos Savva. “Factors Affecting House Prices in Cyprus: 1988-2008”. In: *Econ. Policy Rev.* 3 (Jan. 2009).

[14] Gregory D Sutton, Dubravko Mihaljek, and Agne Subelyte. “Interest rates and house prices in the United States and around the world”. In: (2017).

[15] Vivian W. Y. Tam et al. “Effects of locations, structures and neighbourhoods to housing price: an empirical study in Shanghai, China”. In: *International Journal of Construction Management* 22.7 (2022), pp. 1288–1307. DOI: 10.1080/15623599.2019.1695097. eprint: https://doi.org/10.1080/15623599.2019.1695097. URL: https://doi.org/10.1080/15623599.2019.1695097.

[16] Maude Toussaint-Comeau and Jin Man Lee. “Determinants of housing values and variations in home prices across neighborhoods in cook county”. In: *ProfitWise News and Views* 1 (2018), pp. 1–22.

[17] Sabyasachi Tripathi. ““Macroeconomic Determinants of Housing Prices: A Cross Country Level Analysis” Real Estate Finance, 36:4 (2020), 229-238, Wolters Kluwer.” In: (Feb. 2020).

[18] Carlos Vargas-Silva. “Monetary policy and the US housing market: A VAR analysis imposing sign restrictions”. In: *Journal of Macroeconomics* 30.3 (2008), pp. 977–990. ISSN: 0164-0704. DOI: https://doi.org/10.1016/j.jmacro.2007.07.004. URL: https://www.sciencedirect.com/science/article/pii/S0164070407000997.