C5: MODELING CAPSTONE - FORECASTING THE SEATTLE HOUSING MARKET

August 9, 2023

Andy Klatt
California Lutheran University – Masters In Quantitative Economics

Introduction

Residential real estate is often the largest and most important investment a consumer will make. Sometimes, it is the only long-term investment a middle-class household will make and as such, is vital to a household's economic mobility and wealth generation. Housing supply has not met demand in recent decades in many U.S. geographies, including the Seattle metro area (King County), and resulted in a shrinking middle class and wider inequality. Until the economic crisis of 2007/2008, real estate investments were a reliably appreciating asset, year over year, until a downturn during the Great Recession (GR). In the past 15 years since the Seattle market has seen accelerated growth prompting some to wonder whether another asset appreciation bubble is about to pop.

The Seattle area housing market is of particular interest to me as I have previously bought and sold real estate in the market. It is a fascinating market whose growth has paralleled other regional economic indicators like wealth and jobs. Recently, however, the Seattle housing market has seen an inflection point with YoY growth turning negative for the first time in years.

There are several real estate firms analyzing the U.S. housing market. These include Seattle-based firms like Zillow and Redfin. While many have research teams dedicated to highlighting recent historical trends in the U.S. and the region. These research firms may make broad predictions for the country as a whole or state, but few will publicize forecasts for a particular metro area. This paper intends to do so for the Seattle metro-area which is largely encompassed by King County.

An economic agent in this market, whether they be a buyer (or owner), should have a reasonable expectation of how a real estate asset will appreciate in forthcoming quarters and years so as to take appropriate action to maximize utility and appropriately balance a portfolio of other assets. Forecasting the median home price for King County, which encompasses the job hubs of Seattle and Bellevue, among other suburbs, would provide a short- and medium-term outlook for how the market in aggregate might move so a buyer or seller can make informed decisions in the market. Forecasting inventory, or number of active listings, would give buyers, sellers, and business owners needed context for the market as it represents housing supply which has proven to be a key determinant of demand and price. If inventory was expected to grow buyers might wait to engage in the market until a time when they have more options. Business owners might have to raise wages for employees if costs of living are on the rise.

Key Dependent Variables

Figure 1: Median sale price for residential, single-family homes, by month; King County, WA

Figure 2: Year-over-Year Percent Change in Median Sale Price, by month, for residential, single family homes in King County, WA

Figure 3: Number of active listings, residential single-family homes in King, County, WA

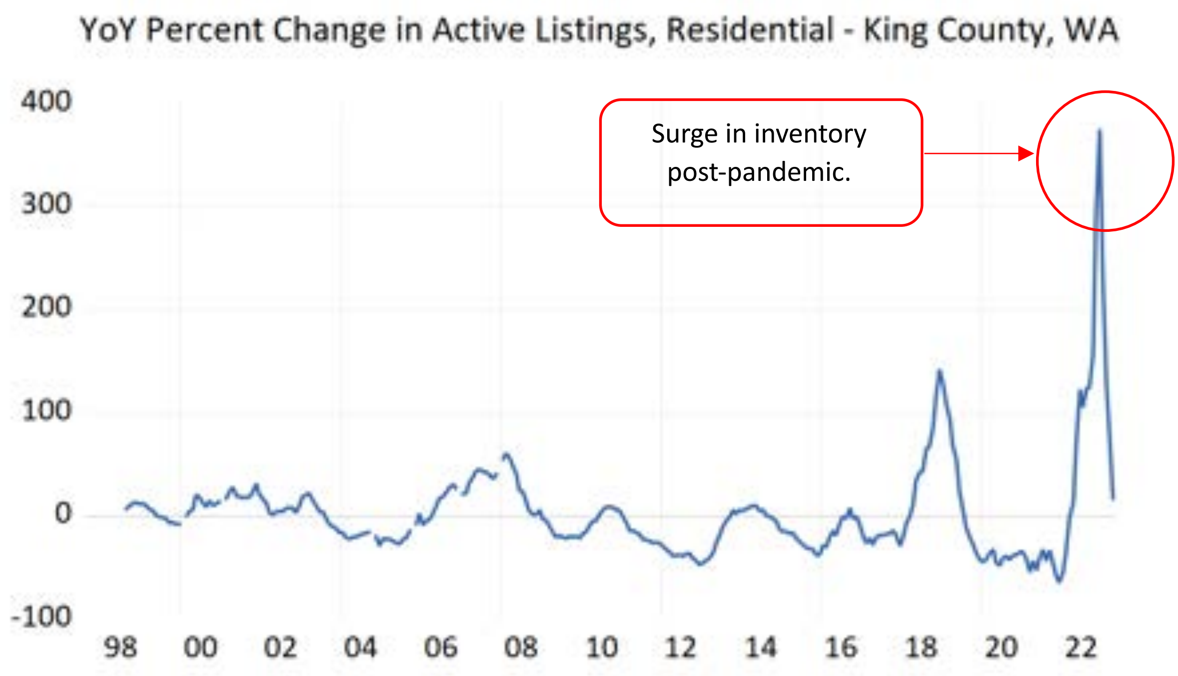


Figure 4: Year over year change in number of active listings, residential in King County, WA

From the above charts we can see that both median sale price and number of active listings show significant trend and seasonality. Median sale price will fluctuate with higher prices in the spring and summer quarters. Median prices overall trend upwards but saw a significant dip after the Great Recession and recovery, from 2008 through 2012. After 2012, median sale price saw several years of steady and accelerated growth, a possible correction in 2019, followed by another short period of accelerated growth. Most recently another possible correction has emerged as growth rates have turned negative.

Real estate data is known to be seasonal as inventory and prices rise in the busy spring and summer quarters. To remove that seasonality, I've differenced the data by looking at a year-over-year percent change. This differenced data also appears to show trending particularly during the Great Recession and recovery. We will need to perform unit root testing to ensure it is stationary.

When visualizing active listings, the seasonality of this data is even more apparent. More listings come on the market during Q2 and Q3. There is a clear trend where listings rose during the Great Recession and have been steadily falling since about 2010. There are a couple interesting large jumps in year-over-year active listings. One is during the year 2018 and the other is during the year 2022. The 2022 jump in listings is likely due to pent up supply during the pandemic when homeowners were sedentary due to pandemic enforced policy measures.

	YoY Percent Change				
	Minimum	**Maximum**	**Average**	**Recent Average (April '21-April '23)**	**Median**
Median Sale Price	-17.29	29.46	6.40	8.66	7.53
Number of Active Listings	-63.63	376.20	3.78	54.78	-2.44

Table 1: Summary Statistics of YoY Percent Change in median sale price and number of Active Listings

From the table above we can see that median sale price has grown on average 6.4% YoY since 1998. The recent average of 8.66% shows that the residential real estate is still seeing above average growth in the last 2 years. The number of active listings is clearly skewed by a few outlier values: the previously noted years of 2018 and 2022. The pandemic likely has something to do with the large 2022 percent change in YoY active listings. Both the years 2020 and 2021 saw depressed levels of active listings. It wasn’t until 2022 when the market started to normalize, and listings come back at the market at normal levels. The percent change then represents 2 years of rebound.

Data

Data Table

No.	Code	Geography	Description	Units	Source	SA or NSA	Frequency	Range	Nobs
1	KING_PMED	King County, WA	Median Sale Price for Single Family Residential Property	USD	NWMLS	NSA	Monthly	Aug 1997 - Apr 2023	308
2	KING_PAVE	King County, WA	Average Sale Price for Residential Property	USD	NWMLS		Monthly	Aug 1997 - Apr 2023	320
3	KING_ACTLIST	King County, WA	Number of Active Listings as of the last day of the month	#, Units	NWMLS	NSA	Monthly	Aug 1998 - Apr 2023	320
4	KING_SOLD	King County, WA	Number of Single Family residential properties sold and closed on during the month	#, Units	NWMLS	NSA	Monthly	Aug 1998 - Apr 2023	320
5	KING_SALES VOLUME	King County, WA	Own calulation: Average Sale Price * Units Sold	USD	NWMLS	NSA	Monthly	Aug 1998 - Apr 2023	320
6	INCPERCAP	King County, WA	Income: Per capita personal income, (USD)	USD	U.S. Bureau of Economic Analysis (BEA): Local Area Personal Income accounts - Table CAINC4, Table CAINC5N	NSA	Annual	1969 - 2021	53
7	WKLYWAGE	King County, WA	Average weekly wage: Private - Private households, (USD, NSA)	USD	U.S. Bureau of Labor Statistics (BLS): Quarterly Census of Employment & Wages (QCEW Formerly ES202)	NSA	Quarterly	2006 Q1 - 2022 Q4	68
8	MRTGRT	United States	PMMS: 30-year FRM - Commitment rate, (% p.a., NSA)	%, Rate	Freddie Mac: Primary Mortgage Market Survey	NSA	Weekly	4/1/71 - 6/1/23	2723
9	SEA_XCPIU	Seattle-Tacoma-Bellevue MSA	CPI: Urban Consumer - All items, (Index 1982-84=100, SA)	Index	U.S. Bureau of Labor Statistics (BLS): Consumer Price Index (CPI) [Series ID = CUURS49DSA0]; Moody's Analytics Adjusted	SA	Bi-Monthly	Dec 1997 - Apr 2023	305
10	SEA_XCPIUAH	Seattle-Tacoma-Bellevue MSA	CPI: Urban Consumer - Housing, (Index 1982-84=100, NSA)	Index	CPI: Urban Consumer - Housing, (Index 1982-84=100, NSA)	NSA	Bi-Monthly	Dec 1997 - Apr 2023	305
11	JOBS_NF	Seattle-Tacoma-Bellevue MSA	Employment: Total Nonfarm, (Ths. #, NSA)	(Ths. #, NSA)	U.S. Bureau of Labor Statistics (BLS): Current Employment Statistics (CES) [Series ID = SMU53426600000000001]	NSA	Monthly	Jan 1990 - Apr 2023	400
12	PrimeHBAgePct	King County, WA	Percent of Total Population (Resident): Total Aged 25-54	% of Pop	U.S. Census Bureau (BOC): Population Estimates, Projections; Moody's Analytics Estimated		Quarterly	1970 Q1 - 2023 Q1	313
13	PrimeHBAgeCnt	King County, WA	Population (Resident): Total Aged 25-54	Total, (Ths.)	U.S. Census Bureau (BOC): Population Estimates, Projections; Moody's Analytics Estimated		Quarterly	1970 Q1 - 2023 Q1	313
14	EDUC_BACH_PCT	King County, WA	ACS: Educational attainment - Percent bachelor's degree or higher, (%)	% of 25+ Pop	U.S. Census Bureau (BOC): American Community Survey (ACS); Moody's Analytics Calculated		Annual	2006 - 2021	26
15	LocalEquityWealth	Seattle-Tacoma-Bellevue MSA	Weighted Index of Certain Local Stock Prices (AMZN, MSFT, BA)	USD	Yahoo! Finance, My Own Deviation for weigted average		Monthly	Jun 1997 - Apr 2023	311
16	HSTOCKQ	King County, WA	Housing stock: Single-family, (Ths., SA)	(Ths., SA)	U.S. Census Bureau (BOC); Moody's Analytics Estimated	SA	Quarterly	1970 Q2 - 2010 Q2	161
17	HSTOCKACS	King County, WA	Housing stock: Single-family, (Ths., SA)	(Ths., SA)	ACS: Units in Structure - Single-family, (Ths.)		Annual	2009-2021	12
18	PERMITS	King County, WA	Permits: Residential - Single-family, (Units # YTD, NSA)	Units #	Permits: Residential - Single-family, (Units # YTD, NSA)	NSA	Monthly	Jan 1998 - Apr 2023	424
19	NPRIME_SHARE	King County, WA	Share of orignations that are non-prime	%, Rate	U.S. Federal Financial Institutions Examination Council (FFIEC): Home Mortgage Disclosure Act (HMDA); Moody's Analytics Calculated		Annual	2005 - 2021	17
20	HSTOCKOVERPRIMPOP	King County, WA	Housing stock over Population of Prime Home Buying Age (25-54)	%, Rate	My own caclulation		Quarterly	1970 Q2 - 2010 Q2	161
21	W$	United States	Real Net Wealth	Millions of 2012 Q3 dollars	Board of Governors, Federal Reserve	NSA	Quarterly	1957 Q1 - 2023 Q1	265

The dependent variable KING_PMED is the median home price for single family residential properties for the geography of King County, Washington. King county that encompasses all of Seattle, Bellevue, and much of their surrounding suburbs. The Seattle-Tacoma-Bellevue MSA is a geography that is approximate to the same area, although somewhat broader as it contains Tacoma in Pierce County. Whenever possible, I tried to find data that was either specific to King County or the Seattle-Bellevue-Tacoma MSA.

The regional real estate data was obtained via the Northwest Multiple Listing Service, the leading resource for real estate data in the state of Washington. The data was manually compiled from individual monthly pdf reports. These reports went back to August 1997. Monthly reports were provided by a trusted real estate agent in the market. I was not able to obtain NWMLS data further back than August 1997. There was only one monthly report (January 2000) that did not provide the level of detail needed for the concepts. That month was imputed using inertial forecasting or averaging the two adjacent months.

Non-NWMLS data was gathered through the aggregator ‘Moody’s Analytics Data Buffet’. Per capita personal income is available at an annual frequency and is available up through 2021.

Freddie Mac’s Primary Mortgage Market Survey (PMMS) for a 30-year fixed mortgage seems to be a broad and widely used indicator of mortgage interest rates. This data is at the weekly frequency.

For price level/inflation measures, I gathered CPI concepts for the U.S. as well as for the Seattle area, both the ‘All Items’ index and housing-specific index. The concepts for the Seattle region had data for every other month but spanned the entire period of the NWMLS data. The in-between months were imputed directly by taking the average of the previous and subsequent month.

Employment was gathered from the Bureau of Labor Statistics (BLS) Current Employment Statistics (CES). The data is contiguous and available at monthly frequency. I gathered both Non-Farm and Total employment. Non-Farm will be the concept I’ll use in model specifications so as not to account for seasonal workers.

Stock Prices for three large local corporations were gathered from Yahoo! Finance. I obtained the adjusted close prices at the beginning of each month for Amazon, Microsoft, and Boeing. According to Puget Sound Business Journal, these are currently the three large publicly traded employers in the area that have also been publicly traded throughout the time period of the NWMLS data. Combined, these three corporations employ nearly 200,000 employees in the metro area of about 3 million. I’ve weighted these stock prices into one aggregate with equal weights per corporation.

Mortgage originations on sub-prime loans: this data only goes back through 2005. However, there is data for the total number of loans for periods going back to 1994. So, I was able to create a concept that is the share of loans that are non-prime for the periods 2005 through 2021. I then used a report from the San Francisco Fed that showed percent of non-prime loans as a share of all loan originations to manually input estimations that matched the trajectory of rest of the country (SF Fed, 2009).

I gathered quarterly population estimates by age group to derive the percentage of the population in the "prime home buying" age group of 24-55. This data is sourced from the U.S. Census Bureau and obtained via Moody's. This population data, however,

Educational attainment is at an annual frequency. I gathered % bachelor's and above. This data is obtained via the Bureau of The Census (BOC) American Community Survey (ACS). In order to fill in missing data before 2006 I back-casted this data using inertial forecasts. Education data, however, was ultimately not leveraged in forthcoming forecasting models.

Housing stock for the region is only available through 2010 Q2 on a quarterly frequency from the DOC. There is, however, housing stock variable available for the 5-year American Community Survey from 2009 through 2021. I took the approach of simply using inertial forecasting to forecast the quarterly data from 2010 Q3 forward to the present. When quartering the ACS this forecast of the quarterly data had a 0.98 correlation with the ACS derived data. So, I simply decided to use the forecasted quarterly census data in my specifications. Housing stock was ultimately not leveraged in forthcoming forecasting models.

Stochastic Properties

The two key dependent variables, median home price and number of active listings, were tested for their stochastic properties in the section below. It is important in any time-series analysis and forecasting effort to determine at what level of differencing a variable becomes stationary with a time in-variant mean and variance. If a variable is not differenced appropriately its prediction errors will have persistent auto-correlation that results and the untransformed variable won't be able to be modeled using inertial or structural methods.

Median Sale Price- KING_PMED

Correlograms showing auto-correlation for median sale price in levels, first seasonal difference are found in the Appendix. Below is a table that summarizes the unit root test of median home sale price and its first seasonal difference.

***** p<.01 ; ** p < .05 ; * p < .10**

	ACF (1)	ADF Unit Root Test Statistic Null = non-Stationary	DF-GLS Unit Root Test Statistic Null = non-Stationary	KPSS Unit Root Test Statistic Null = Stationary
Levels	0.98	1.25	-0.28	4.15***
First Difference	0.253	-2.47	1.638*	0.128
Second Difference	-0.467	14.351	0.29	0.026
First & Seasonal Difference ($\Delta\Delta^{12}$)	-0.05	-3.37**	-3.31***	0.161
I(0), I(1), or I(2)?	I(1)	I(1)	I(1)	I(1)
Seasonal Lags	>6 years	>6 years	>6 years	>6 years

Table 2: Stochastic Properties of Median Sale Price

We can see that median sale price is very likely I(1) data. When testing the data in levels all unit root tests show non-stationarity as expected. When taking the first difference there is weak evidence for stationarity. KPSS shows evidence of stationarity while ADF and DF-GLS tests show weak evidence at the p<0.15 and p< 0.10 levels respectively. However, after taking the second difference an ACF(1) of -0.467 provides evidence of over-differencing. Thus, the median home price data series is likely to be I(1).

When taking the first and seasonal difference the ADF and DF-GLS unit root tests reject the null hypothesis of non-stationarity at the 0.05 level and <0.01 level respectively. This suggests that a year over year growth rate of this dependent variable will also be suitable for structural regressions .

Number of Active Listings - KING_ACTLIST

Correlograms for levels and the first and seasonal difference of KING_ACTLIST are found in the appendix. Below is a table that summarizes the unit root test of median home sale price and its transformations.

***** p<.01 ; ** p < .05 ; * p < .10**

	ACF(1)	ADF Unit Root Test Statistic Null = non-Stationary	DF-GLS Unit Root Test Statistic Null = non-Stationary	KPSS Unit Root Test Statistic Null = Stationary
Levels	0.979	-1.976	-1.542	2.665***
First Difference	0.557	-4.183***	-2.83***	0.45*
First & Seasonal Difference ($\Delta\Delta^{12}$)	0.510	-6.16***	-1.71*	0.03
I(0), I(1), or I(2)?	I(1)	I(1)	I(1)	I(1)
Seasonal Lags	>6 years	>6 years	>6 years	>6 years

Table 3: Stochastic properties of the variable 'Number of Active Listings'

All three unit-root tests show that the Number of Active Listings is made stationary by taking the 1st difference. Both the ADF and DF-GLS tests of first differenced date show strong evidence of stationarity. While the KPSS test shows weak evidence of non-stationarity. Also included as a robustness check is the first and seasonal difference which shows evidence of stationarity. This leads me to the conclusion that KING_ACTLIST is I(1) and YoY growth will be a suitable dependent variable for structural regressions.

Theory

Econometric modeling and analysis aim to appropriately estimate the structure of the data generating process for economic variables by paying specific attention to the direction of causality. For instance, when real income rises demand for single-family houses will subsequently rise and cause an increase in prices as more high-income consumers with will be chasing an assumed same supply of housing. This contrasts with correlation where variables might move together (or opposite), but there is not consensus theory in place for what variable might cause another; perhaps a third variable is responsible for causing both.   Below is the functional form for each structural regression specified in a multi-equation forecast model. The theory that supports that variables inclusion, and the appropriate sign for each variables coefficient are also detailed below.

Median Sale Price- KING_PMED

$$\mathrm{KING\_PMED} = \mathrm{f}\,(\underset{(-)}{\mathrm{KINGACTLIST}}, \underset{(+)}{\mathrm{JOBSNF}}, \underset{(+)}{\mathrm{NPRIME\_SHARE}}, \underset{(+)}{LOCALEQUITYWEALTH}, \underset{(-)}{\mathrm{MRTGRT}})$$

Equation 1: Functional form of median home price with theoretical signs below causal variables

The median sale price for King County, WA single-family homes is a function of the following determinants: the number of active listings, number of jobs, share of non-prime mortgage originations, the wealth of local corporations who employ high income residents, and mortgage rates.

The number of active listings determines the supply of the current housing market. When inventory is low sellers can command higher prices as consumers are given fewer choices. In this regression specification I am lagging active listings two months since houses close at least one month after listing.

The number of non-farm jobs can be a good indicator of the strength of the overall economy. When a community is adding jobs, more of its residents are securing stable income and will be more willing to take on the risk of debt in the form of a mortgage. When they are losing jobs, they will be less willing to commit to buying real estate.

Many Seattle homebuyers will use built up equity that they may have received in the form of employee stock grants to put a down payment down. This is especially the case in Seattle with large employers like Amazon, Microsoft, and Boeing. The higher the stock prices for these large publicly traded corporations will drive up prices.

As mortgage rates rise, it becomes more expensive to borrow money and for potential homebuyers to afford homes. This will decrease overall demand, all other things equal, and cause a median home price to decrease. As seen during the sub-prime mortgage crisis of the 2000's, the larger the share of mortgages that are non-prime, the more buyers who normally wouldn't be able to afford homes will enter the market. This artificially inflated the housing bubble and lead to the Great Recession.

Number of Active Listings - KING_ACTLIST

$$\text{KING\_ACTLIST} = f\,(\underset{(-)}{(\text{KING\_SOLD}/\text{KING\_ACTLIST})}, \underset{(+)}{\text{MRTGRT}}, \underset{(-)}{(\text{INCPERCAP}/\text{KING\_PMED})})$$

Equation 2: Functional form of number of active listings with theoretical signs below causal variables

Active listings can be modeled as a function of the absorption rate (number of home sold relative to those on the market), mortgage rates, and relative unaffordability of homes (median income relative to median home price).

The absorption rate measures the speed at which homes are sold. A higher absorption rate will result in less homes being on the market at any snapshot in time (Dehan, 2023). Thus, this will have a negative causal effect on the number of active listings. For my specification, this effect will be lagged one period as the effect of a high absorption rate last month appears in the subsequent month's inventory.

Mortgage rates, in the case of inventory, have a positive relationship to the number of listings. When mortgage rates have risen relative to a time in the recent past, credit has become tight, less people can afford a loan, and houses sit on the market longer. This will cause the number of active listings to grow simply through lack of turnover. Changes in mortgage rates also take time to show up in inventory data. Change in mortgage rates will be lagged 3 months.

When home prices rise to the point of being unaffordable to the median citizen's income, the market becomes out of reach to many local home buyers. This theory suggests that when homes become unaffordable, to a point, the market may be flush with more inventory and the number of active listings will rise. To that end the term of median income over median home price is added to capture this positive "unaffordability effect" on inventory.

Intervention variables for COVID and the Great Recession were not added to this regression. This will be discussed in more detail later in this paper.

Identities

Median Home price is denominated in nominal dollars, but it will be modeled in real dollars after it has been deflated by the housing CPI for the Seattle metro. Modeling in real terms takes inflation out of equation as a factor for home price. One identity that will be entered into the model will be that real median home price equals the nominal median home price deflated by the appropriate price index. In this case we have a price index that is both germane to the geography and the category of good. This identity can be represented in the equation below:

$$\text{KING\_PMED} = KING\_PMED\$ \; * \frac{C\_SEA\_XCPIUAH}{100}$$

Equation 3: Identity to convert to real dollars median home price denominated in nominal or current dollars.

The same will be done for average home price:

$$\text{KING\_PAVE} = KING\_PAVE\$ \; * \frac{C\_SEA\_XCPIUAH}{100}$$

Equation 4: Identity to convert real dollars average home price denominated in nominal to current dollars.

Real mortgage rates are rates that consider the current rate of inflation. In this case I use inflation of the U.S. and of all items to deflate the nominal mortgage rate.

$$\mathrm{MRTGRT} = MRTGRT\$ + \frac{XCPIU_t - XCPIU_{t-12}}{XCPIU_{t-12}}$$

Equation 5: Real mortgage rate

We can model nominal sales volume by taking the average home price and multiplying it by the number of homes sold.

$$\mathrm{KIN}G\_S\mathrm{ALES}\_V\mathrm{OLUME} = \mathrm{KIN}G\_P\mathrm{AVE} * K\mathrm{INGSOLD}$$

Equation 6: Sales Volume Identity

Supporting Tautological Regression – Average Home Price

A purposeful tautological regression was specified to model average home price based on median home price. The two variables are highly correlated with average home price always slightly higher than median because of the right skew in distribution of home sales. This regression supports the model in providing an output for average home sale price as well as sales volume. Although these are not key dependent variables, they could be provided by this model specification if asked for. Regressions specification along with coefficient and t-stats provided below.

$$\mathrm{KING\_PAVE} = \alpha\ \mathrm{KIN}G\_P\mathrm{MED}$$

####
(###)

Equation 7: Tautological regression for average home price based on median home price

Structural Regressions

I've estimated two structural regressions to model both the median home price for a single-family residential property and the number of active listings. Both of the key dependent variables have been recently volatile due changing economic conditions resulting from the COVID-19 pandemic while also showing long-term trends. I aimed to include variables that resulted in significant coefficients that provided evidence of a causal relationship while also agreeing with theory.

Median Sale Price- KING_PMED

Below is the estimated model for median home price. Estimated coefficients are below variable names and t-stats below them in parentheses. Also included below is a plot for actual values, fitted, and their residuals and correlogram of residuals.

$$\%\Delta_{12}\ \text{KING\_PMED}_t =$$

$$\underset{(-5.47)}{\underset{-0.06}{\alpha_1}}\ \%\Delta_{12}\ \text{KING\_ACTLIST}_{t-2} + \underset{(3.62)}{\underset{0.89}{\alpha_2}}\ \%\Delta_{12}\text{JOBS\_NF} + \underset{(3.78)}{\underset{0.009}{\alpha_3}}\%\ \Delta_{\text{Ann.}}\ \text{LOCALEQUITYWEALTH}_{t-6}$$

$$+ \underset{(3.37)}{\underset{0.09}{\alpha_4}}\ \%\Delta_{12}\text{NPRIME\_SHARE} + \underset{(-1.70)}{\underset{-0.80}{\alpha_5}}\Delta_6\text{MRTGRT}_t$$

Estimation Technique	Least Squares (NW)
Data Frequency	Monthly
Estimation Range	Dec 1998 to April 2023
Net d.f.	287 (i.e. 293-5)
Resid Unit root test	N/A (differenced DV)
Residuals ACF(1)	0.755
Seasonal Dummies	N/A

Equation 8: Estimated model of median home price

Each coefficient can be interpreted as a one unit increase of a predictor on the year-over-year (YoY) growth rate of median home prices. All coefficients are significant at the .05 level apart from the 6-month difference in mortgage rates that is significant at the 0.15 level. The YoY growth of active listings puts downward pressure on prices. When there is more inventory, buyers have more choices and prices will decline. This effect is significant, but with a small magnitude. The YoY growth in jobs has a positive impact on home prices and with a magnitude close to a 1-for-1, in percentage terms, effect. When there are more jobs created in the region there will be more demand for housing and prices will increase. There is a significant, but relatively small effect of the lagged stock price of local large corporations on home prices. Nevertheless, there is evidence in support of my hypothesis that local stock prices of these corporations has a meaningful effect on the price of homes. A lag is necessary here as I estimate that buyers take about 6 months after seeing increases in these stock prices before driving real estate prices higher. This could agree also with when employees receive stock grants in winter and use them for down-payments in real estate's high season of spring and summer. The percentage on non-prime loans is significant and is positive which would mirror the effect of the sub-prime housing bubble in the early-to-mid 2000's. As more non-prime loans were issued, prices rose inflating the housing bubble that ultimately contributed to the Great Financial crisis of 2007-2008. The difference in mortgage rates was found to have a significant effect at the 0.10 level when lagged 6 months. Its negative sign reflects that market's behavior to catch up with rising or falling cost of credit as well as the months lags between engaging in the real estate market and closing on a property. When mortgage rates rise, as they have been recently, it takes some time for home prices to adjust, and buyers are priced out of the market due to the increased cost of credit.

As a robustness check, I tried adding intervention variables for the months of the Great Recession (according to BEA) that helped the model fit values through that period as number of non-prime loans could not completely model the effects. I also tried an intervention variable for the second year after the year of the pandemic (April 2022 – March 2023) which helped to model the housing markets rebound after two years of pandemic constrained market activity. The coefficients of this model with these interventions is in the appendix. However, one can see that they only helped the model and didn't affect the coefficients signs, magnitudes, or significance in any appreciable way.

I chose not to include intervention variables for the year of the pandemic or year subsequent hoping that inventory and jobs properly modeled the market during this time. I did attempt several specifications where I tried to include aspects of the housing stock either through permits or actual housing stock variables but was not able to bring all variables into significance and signs that agreed with theory. I also tried to include other aspects of home-buying consumers like population growth for those in prime home-buying ages and education level of the population. However, I think that jobs seem to an okay job of capturing the effect of these population variables. Finally, I explored real estate wealth's effect via a theory that existing home-owners with increasing home values would "trade-up" and further increase home prices. But perhaps there's a tendency for existing home owners to sell and move away from King County and be replaced with home owners new to the area.

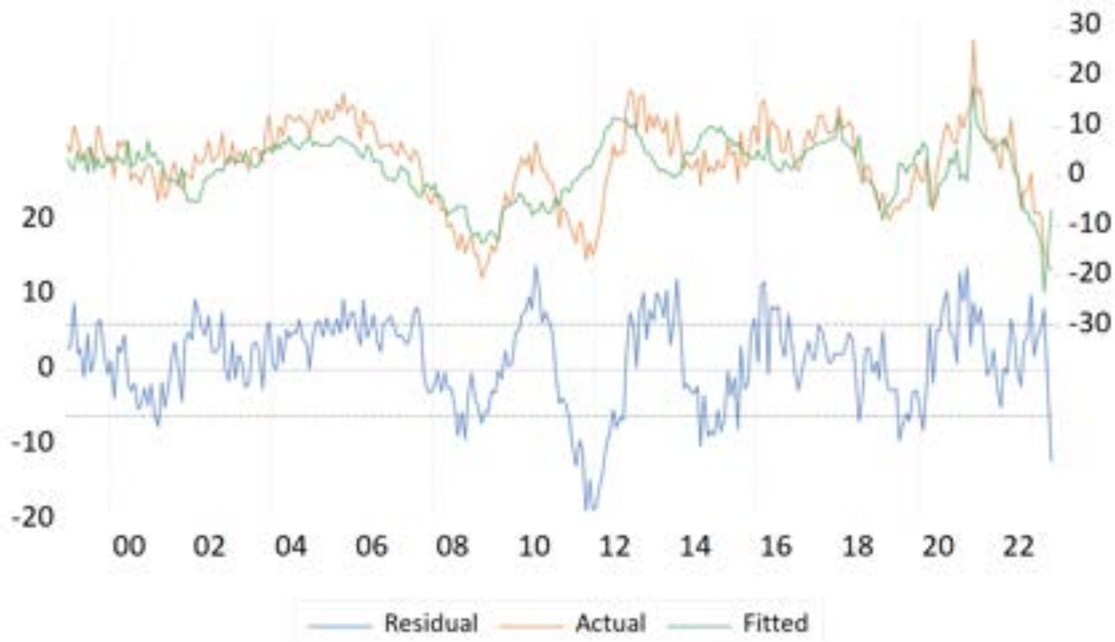
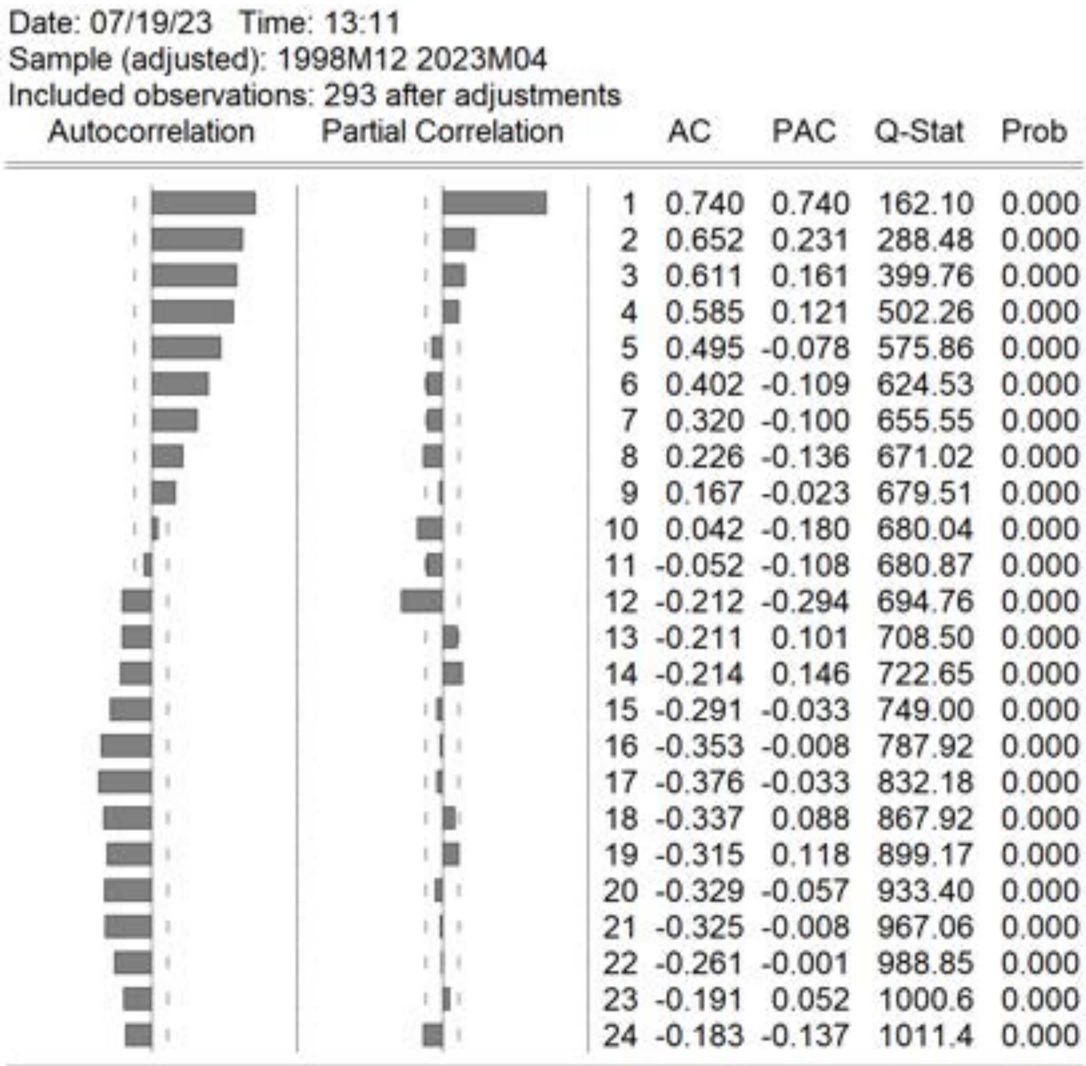


Figure 5: Actual-fitted-residual chart of estimated model of median home price



Date: 07/19/23 Time: 13:11
Sample (adjusted): 1998M12 2023M04
Included observations: 293 after adjustments

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.740	0.740	162.10	0.000
		2	0.652	0.231	288.48	0.000
		3	0.611	0.161	399.76	0.000
		4	0.585	0.121	502.26	0.000
		5	0.495	-0.078	575.86	0.000
		6	0.402	-0.109	624.53	0.000
		7	0.320	-0.100	655.55	0.000
		8	0.226	-0.136	671.02	0.000
		9	0.167	-0.023	679.51	0.000
		10	0.042	-0.180	680.04	0.000
		11	-0.052	-0.108	680.87	0.000
		12	-0.212	-0.294	694.76	0.000
		13	-0.211	0.101	708.50	0.000
		14	-0.214	0.146	722.65	0.000
		15	-0.291	-0.033	749.00	0.000
		16	-0.353	-0.008	787.92	0.000
		17	-0.376	-0.033	832.18	0.000
		18	-0.337	0.088	867.92	0.000
		19	-0.315	0.118	899.17	0.000
		20	-0.329	-0.057	933.40	0.000
		21	-0.325	-0.008	967.06	0.000
		22	-0.261	-0.001	988.85	0.000
		23	-0.191	0.052	1000.6	0.000
		24	-0.183	-0.137	1011.4	0.000

Figure 6: Correlogram of Residuals from estimated regression of Median Home Price

There is a large value for the ACF(1) of the residuals of this regression which begs the question whether the residuals are non-stationary. However, after testing the residuals for a unit root using an ADF test, they showed evidence of being stationary. The machine output for this test can be found in the appendix.

Number of Active Listings - KING_ACTLIST

Below is the estimated model for the number of active listings. Estimated coefficients are below variable names and t-stats below them in parentheses. Also included below is a plot for actual values, fitted, and their residuals and correlogram of residuals.

$$\%\Delta_{12}\ \ln(\mathrm{KING\_ACTLIST_t}) = \beta_1 \left(\frac{KING\_SOLD_{t-1}}{KING\_ACTLIST_{t-1}}\right) + \beta_2\ \Delta_6\ \mathrm{MRTGRT}\ +\ \beta_3 \left(\frac{KING\_PMED_t}{INCPERCAP_t}\right)$$

-3.16	0.86	0.25
(-4.67)	(1.51)	(2.87)

Estimation Technique	Least Squares (NW)
Data Frequency	Monthly
Estimation Range	Feb 1998 to April 2023
Net d.f.	299 (i.e., 302-3)
Resid Unit root test	N/A (differenced DV)
Residuals ACF(1)	0.937
Seasonal Dummies	N/A

Figure 7: Estimated model of number of active listings

This dependent variable is in year over year percent log difference. I decided to take the log due to the volatility in the data. The first predictor is the proportion of sales to active listings last month otherwise known as the absorption rate. This also could also be interpreted as turnover in inventory terms. The higher the absorption rate in a previous period will have a negative effect on the number of active listings in the current period as houses are being bought at a quick pace and not staying on the market very long. The coefficient is significant at the <0.01 level.

The lagged mortgage rate of 6 months is only significant at the 0.15 level. According to theory, as mortgage rates rise the purchasing power of would-be home buyers is diminished resulting in decreased demand. With lower demand houses stay on the market longer and therefore there are more active listings. However, the macroeconomic environment of historic low rates the past 15 years has perhaps dampened the effect of mortgage rates given the data used to specify this regression. Nevertheless, as the current environment sees rising rates, it's important to keep the predictor a part of the regression for forecast validity.

I also attempted to bring in a measure of housing stock relative to population but was not able to get correct signs or significance. As demand has outstripped supply these past decades it's possible housing stock would be a difficult predictor to include in a specification.

In the chart below one will notice a very large shock in 2022-2023. I did attempt intervention variables for this period of COVID recovery or "bounce back", but ultimately decided against it as it made my mortgage variable non-significant.

Figure 8: Actual values, fitted, and residuals for model of number of active listings

Date: 07/19/23 Time: 17:04
Sample (adjusted): 1998M02 2023M04
Included observations: 302 after adjustments

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.937	0.937	267.58	0.000
		2	0.838	-0.319	482.58	0.000
		3	0.727	-0.089	644.86	0.000
		4	0.613	-0.054	760.63	0.000
		5	0.522	0.145	844.88	0.000
		6	0.439	-0.091	904.77	0.000
		7	0.356	-0.107	944.17	0.000
		8	0.273	-0.059	967.38	0.000
		9	0.181	-0.104	977.65	0.000
		10	0.072	-0.202	979.30	0.000
		11	-0.026	0.059	979.51	0.000
		12	-0.117	-0.060	983.84	0.000
		13	-0.167	0.252	992.72	0.000
		14	-0.178	0.060	1002.9	0.000
		15	-0.180	-0.083	1013.2	0.000
		16	-0.173	-0.011	1022.9	0.000
		17	-0.169	-0.002	1032.0	0.000
		18	-0.178	-0.091	1042.3	0.000
		19	-0.195	-0.086	1054.7	0.000
		20	-0.212	-0.039	1069.2	0.000
		21	-0.225	-0.008	1085.8	0.000
		22	-0.222	-0.020	1101.9	0.000
		23	-0.205	0.014	1115.8	0.000
		24	-0.195	-0.113	1128.3	0.000

Figure 9: Correlogram of residuals from the estimated regression for number of active listings

As previously mentioned, despite the high AR(1) shown in the above correlogram of the residuals, an ADF test showed strong evidence of stationarity among the residuals.

I feel as if this is the best regression moving forward to try and model a highly volatile metric. With more time I could attempt other specification methods like an FMOLS specification of YoY Growth.

Shock Analysis

Median Sale Price- KING_PMED

There was a period during the fall/winter/spring of 2011/2012 that saw a particular shocks in the range of 2- to 3-sigma with the largest shock being 3.16 sigma in October of 2011. I currently don't understand the reasons for this period of negative shocks. They can be traced back to a rise in prices one year prior in 2010-2011. I'm unsure of what may have occurred in the market during that time. I will refrain from adding an intervention variable for this 2011-2012 period. I had initially provided step variables to the Great Recession of Post-Covid (Year 2) periods to help model these periods, but their pre-intervention shocks were not large enough to justify intervention variables. As a robustness check I've provided what these shocks were in response to intervention variables. The regression with these variables are in the Appendix.

Number of Active Listing – KING_ACTLIST

There was a period of 2 periods in the winter of 2018/'19 that saw shocks above 2.5 sigma with the highest being 2.9 sigma. I was not sure of the reason for this shock and did not enter an intervention variable. Again, this is a yearlong lag, so winter of '17/'18 must have been one with very low inventory for an as yet unidentified reason.

There was a large positive shock in the second year after the first year of the pandemic especially in the winter of '22/'23. This was likely because the previous winter saw the resurgence of COVID when the Omicron variant was more prevalent. During that winter of '21/'22 number active listings fell below 1,000 for an unprecedented 5 months straight. This was already in a period of low inventory due to COVID and high demand and the resurgence of COVID led people to wait on putting any house on the market in case they were subject to further lockdowns. I've introduced an intervention variable called POSTCOVID_YR2 that attempts to model the resurgence in inventory after COVID began to wane. This actually resulted in my mortgage rate variable to go to only 20% significance level and be respecified down to a 3-month lag rather than 6-month lag.

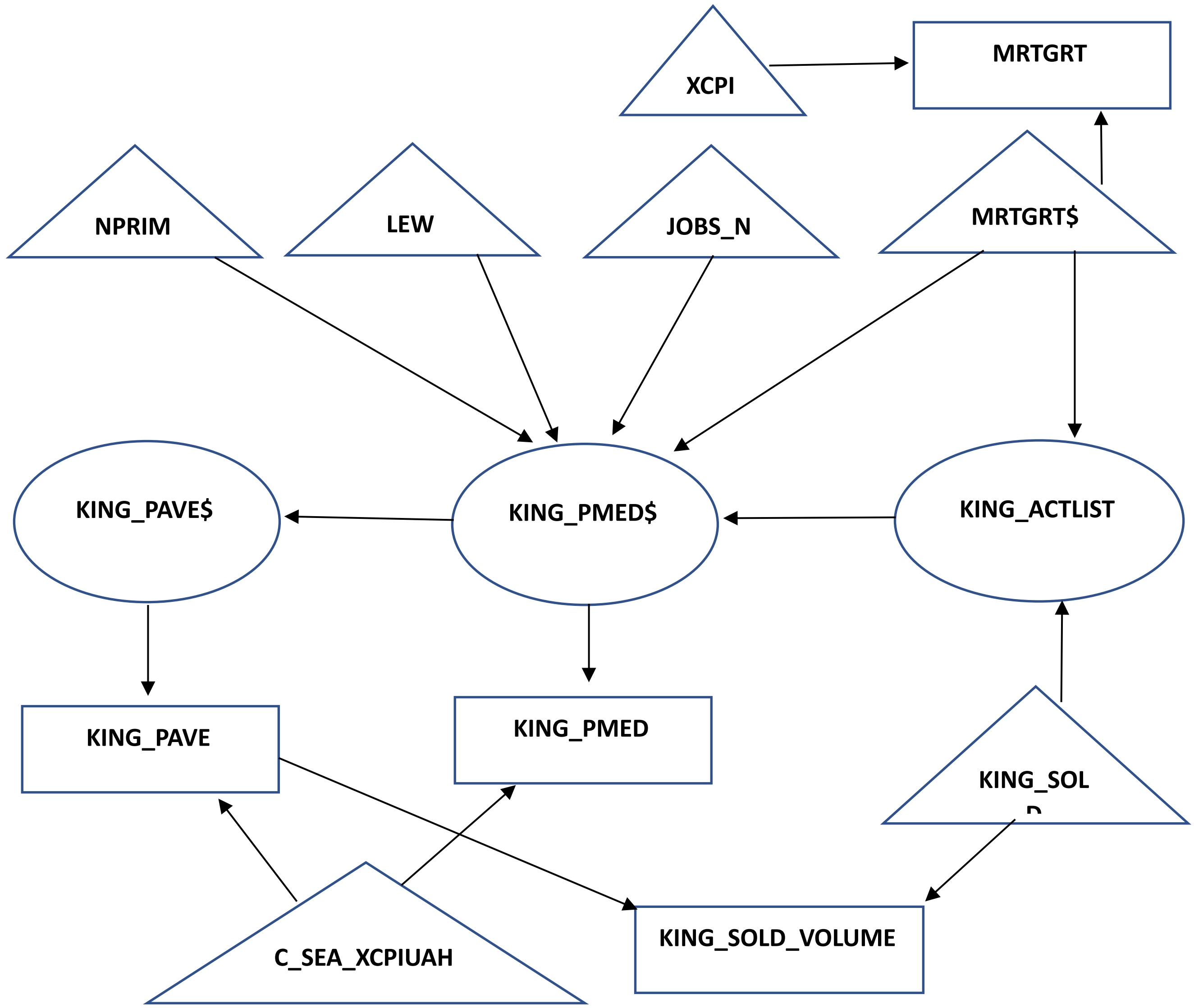
Pre-intervention – Final Model					
	KING_PMED (Final Model)			KING_ACTLIST (Final Model)	
	Largest Sigma Size in Time Period	Time Period		Largest Sigma Size in Time Period	Time Period
Shock #1	-3.2 sigma (2011:10)	Winter 11/12: 2011:10 - 2012:03		2.90 sigma (2018:12)	Winter 18/19: 2018:12 – 2019:01
Shock #2	-1.6 sigma (2008:10)	Great Recession 07/08/09: 2007:12 - 2009:06		5.5 sigma (2022:12)	Post-Covid Yr2 22/23: 2022:04 – 2023:03
Shock #3	1.71 sigma (2022:10)	Post-Covid Year 2 22/23: 2022:04 – 2023:03			

Post Intervention – Robustness Check (Interventions not in Final Model)					
	KING_PMED (Robust Check)			KING_ACTLIST (Robust Check)	
	Largest Sigma Size in Time Period	Time Period		Largest Sigma Size in Time Period	Time Period
Shock #1	-0.47 sigma (2011:10)	Winter 11/12: 2011:10 - 2012:03		4.15 sigma* (2018:12) *No intervention Added.	Winter 18/19: 2018:12 – 2019:01
Shock #2	-1.07 sigma (2008:10)	Great Recession 07/08/09: 2007:12 - 2009:06		3.68 sigma (2022:12)	Post-Covid Yr 2 22/23: 2022:04 – 2023:03
Shock #3	1.19 sigma (2022:10)	Post-Covid Year 2 22/23: 2022:04 – 2023:03			

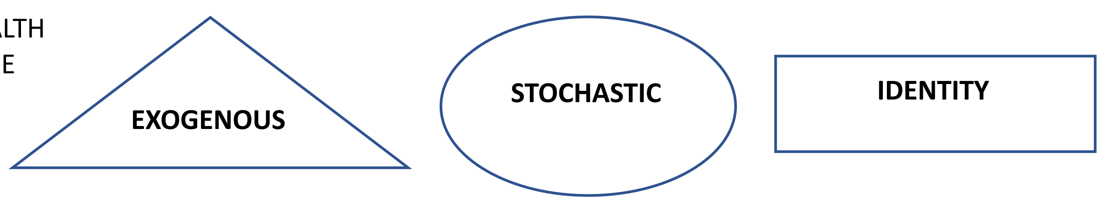
Structural Model

This model aims to forecast two endogenous variables: median home price and number of active listings. Both are instrumental supply and demand indicators of the residential real estate market where dynamics between them are important for where this market finds its equilibrium month-to-month in addition to exogenous variables outside the real estate market that are a part of the regional and national macroeconomy. Average home price is closely related to median home price and is also an endogenous variable of the model, but only insofar as it is so highly correlated with median home price. Including average home price gives a more complete picture of the Seattle real estate market as we're also able to output sales volume by way of derivation.

Model Flowchart



KEY:
LEW = LOCALEQUITYWEALTH
NPRIME = NPRIME_SHARE



Forecasting Exogenous Variables

Exogenous variables used in the structural model were forecasted using ARIMA inertial methods with the exception of NPRIME_SHARE which was subjectively forecast based on history. The number share of sub-prime mortgage originations has not risen above 0.50% since the housing crisis of 2008.

ARIMAs were identified by simply the AR and MA autoregression patterns found in the historical data to forecast data into the future. Sometimes the period of the COVID-19 pandemic was excluded from the period used to identify these patterns. Those variables that were forecast in support of the model include CPI All items, Seattle CPI of Shelter, Local Equity Wealth, Non-Farm Jobs, nominal mortgage rate. Their forecast charts are below.

Charts of Exogenous Variable Forecasts in Support of Structural Model

Figure 10: Nominal Mortgage Rate forecast using ARIMA

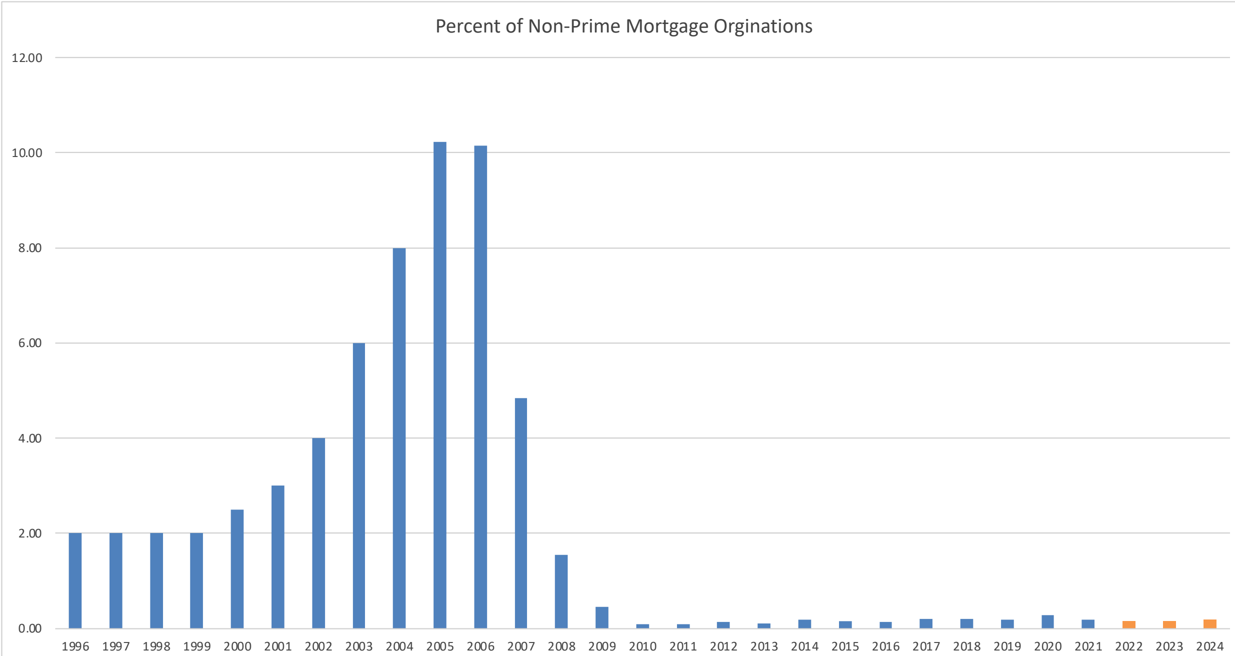


Figure 11: Forecast of NPRIME_SHARE based on safe assumption that levels will remain low. Note the rise of sub-prime mortgage originations in the early 2000's

Figure 12: Seattle Non-Farm Jobs (# Ths.) forecasted using ARIMA

Figure 13: Local Equity Wealth forecasted using ARIMA

Figure 14: Number of Home Sold Forecasted using ARIMA

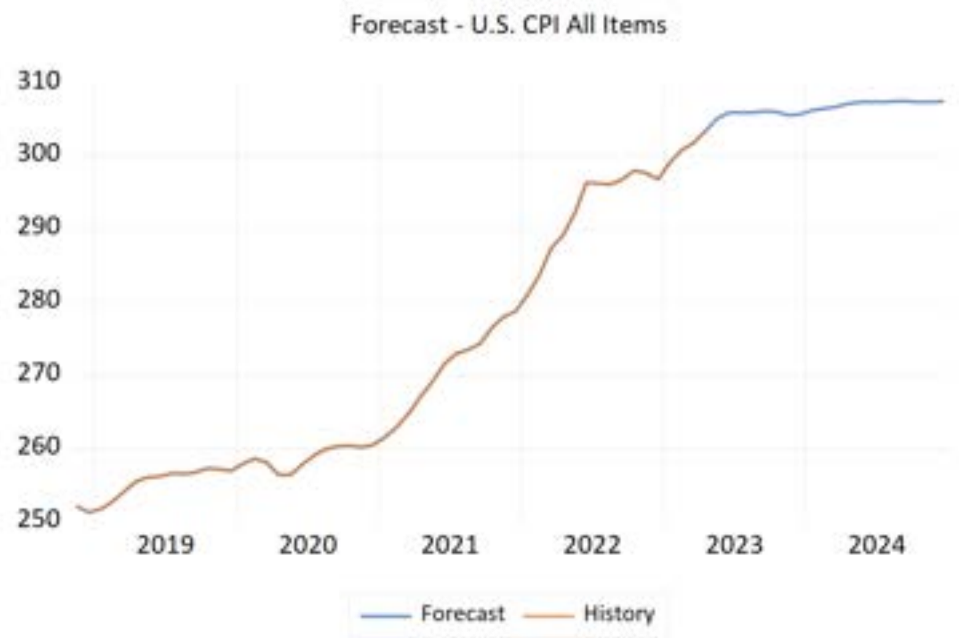


Figure 15: U.S. CPI All Items Forecast Using ARIMA

Figure 16: Seattle Housing CPI Forecast Using ARIMA

In Sample Structural Forecast

The structural model when applied to the in-sample data models the data relatively well until the last twelve months of sample data. It fails to model sharp decline I home prices in the past winter and spring ('22/'23). Instead, it modeled similar prices to the highs that were reached in 2022. This results in large in-sample residuals during these periods. As explained, it was considered to model these residuals using an intervention variable to model the year following the COVID bounce back year of 2022, but ultimately decided not to include intervention variables.

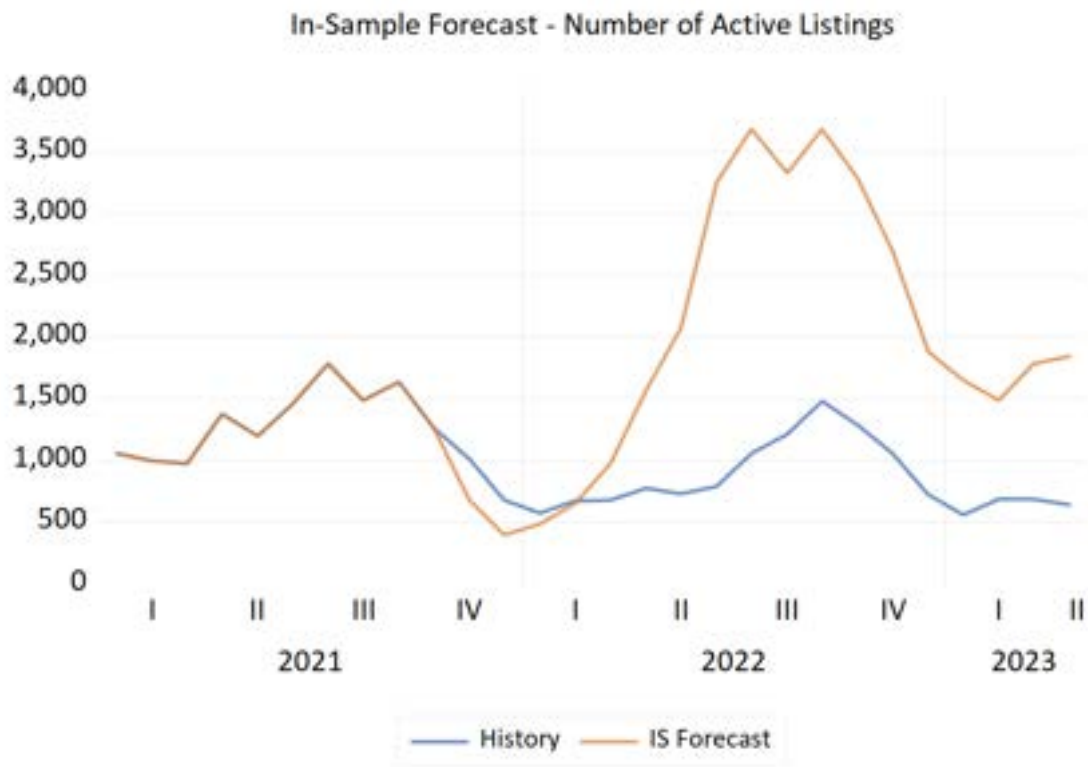


Figure 17: In sample forecast comparison with actual.

We can see that the in-sample forecast of number of active listings was tracking nicely with actuals up until 2022 when it totally missed the inventory surge of 2022 when high prices caused inventory to spike. An intervention variable might have helped for this period but would have caused some coefficients like mortgage rates to lose their signal.

Out of Sample Structural Forecast

Figure 18: Out of sample forecast for structural regression model of median home price

The out of sample forecast based on this model predicts 0 growth for median home price the in 2023, for home prices to resume their historical long-term accelerated growth rate in 2024. Subjectively, this seems like an optimistic forecast. Other forecasting methods discussed in this paper may vary.

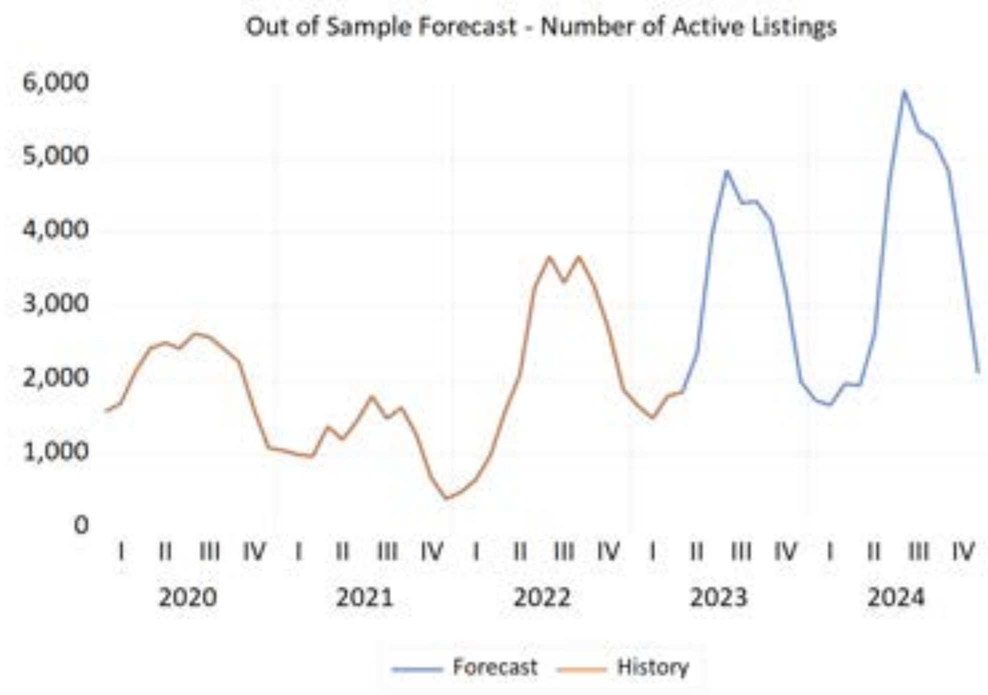


Figure 19: Out of sample forecast for structural regression of number of active listings

The out of sample forecast for number of active listings is similarly optimistic that inventory will pick up in the next few years. This seems increasingly unlikely given the current mortgage rates that are holding down inventory. This regression still may be suffering from some omitted variable bias.

Median Home Price – Impact of Coefficients on Out of Sample Forecast

The forecasted optimistic growth of median home price over the next year and a half can be traced back to the large magnitude of the NF_JOBS variable. We showed that ARIMA methods forecasted the number of jobs in the Seattle metro area to continue to rise at clear positive growth rate. It's my assessment that this is the primary driver of this optimistic forecast. With more time, I might do a few scenarios where the job market would hypothetically have different outcomes at the effect it would have on median home price forecast. Other variables like the stock market, mortgage rates, and inventory will have an effect, but none as noticeable as jobs. What is clear to me is that should the Seattle job market go through some sort of negative shock, we can expect home prices to noticeably fall.

Vector Autoregression (VAR) Modeling

VAR is a multi-variate inertial modeling technique introduced by C. Sims (1980). It is a multivariate inertial model much like ARIMAs are a univariate inertial. Specifying multiple endogenous variables provides a way to analyze the dynamics between variables with the assumption that everything in the macroeconomy affects everything else. VAR is grounded in the notion that the data generating process behind economic variables does not have causality going in one direction, but rather that each variable may be influenced by and/or influencing other variables at different lags. VAR lends itself well to forecasting because it models both muti-variate and inter-temporal dynamics between endogenous variables. Real estate certainly also lends itself well to VAR with demand and supply forces often acting in concert to set prices and inventory.

VAR Specification

My VAR positions both median home price and number of active listings as endogenous variables while considering real mortgage rates, real income per capita, number of jobs, and the post-covid (second year) intervention variable as exogenous. Real estate prices are highly influenced by the inventory available and vice versa. For example, King County has seen that prices remain strong in recent months despite rising mortgage interest rates. This is thought to be due in large part to a lack of supply as no one wants to sell and be forced to trade their low mortgages for a higher one. When supply is low, prices can remain high as homebuyers are left without many choices. So, despite mortgage rates pricing more consumers out of the market which would drive demand and prices down, the low supply effect is keeping prices high as there are still enough (too many) buyers chasing few homes. I do still want to model the exogenous effects of the housing market in this VAR model acknowledging that things like mortgage rates, number of jobs, income, and the emergence of this sector from the pandemic still hold some sway over the market. At times though, they may be overwhelmed by the supply factor.

This VAR as it is specified at order 24 and passes the diagnostics as detailed in the table below. It is stable and specified at a lag length that is optimal according to two lag length selection criteria: FPE and AIC. There doesn't appear to be any large auto-correlation spikes beyond the 24th lag. The number of estimated parameters comes in at 102 which is exactly one-third of the 306 observations. This results in a highly parametrized model, but as we'll see one that performs well on in-sample data.

Endogenous: KING_PMED$, KING_ACTLIST,

Exogenous: MRTGRT$, INCPERCAP, NF_JOBS, POST_COVID_2YR

	Stable (Y/N)	**Estimation Lag Length Used**	**Variable in Levels, Differenced**	**Number of Estimated Parameters - % of Obs**
Estimated VAR	Yes!	24	Endogenous Variables are First & Seasonally Differenced D(x,1,12)	2+(24)(2^2) + 4 = **102** **33% of Obs**
	LR	**FPE**	**AIC**	**HQ**
Lag Length Criteria Results	45	24	24	1

Table 4: VAR Diagnostics

Historical Solve

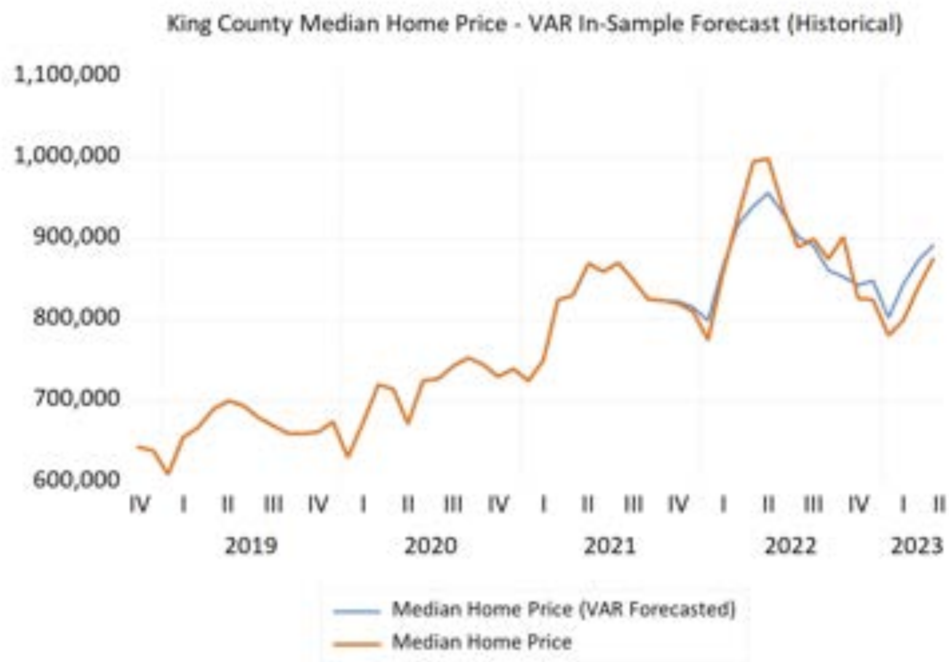


Figure 20: VAR in sample forecast for median home price

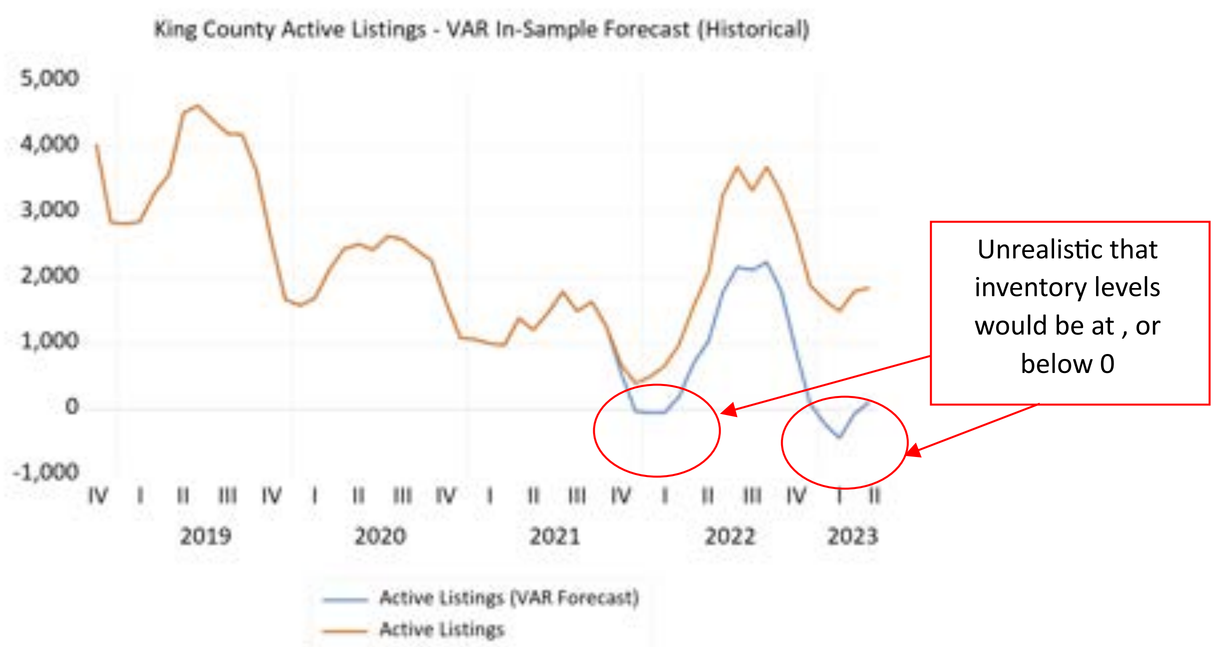


Figure 21: VAR in-sample forecast for number of Active Listings

Forecast Solve

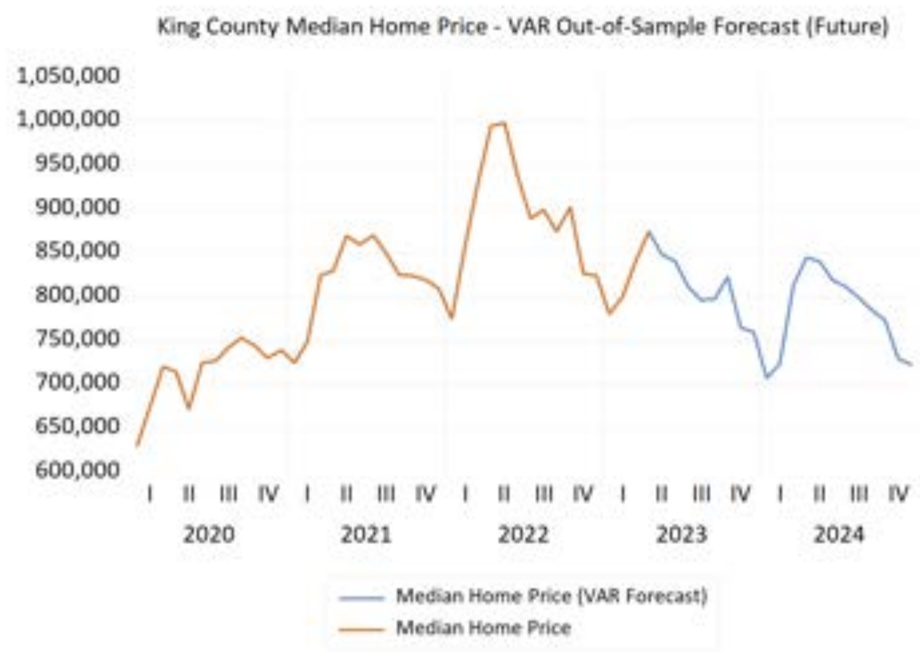


Figure 22: VAR out of sample forecast of median home price

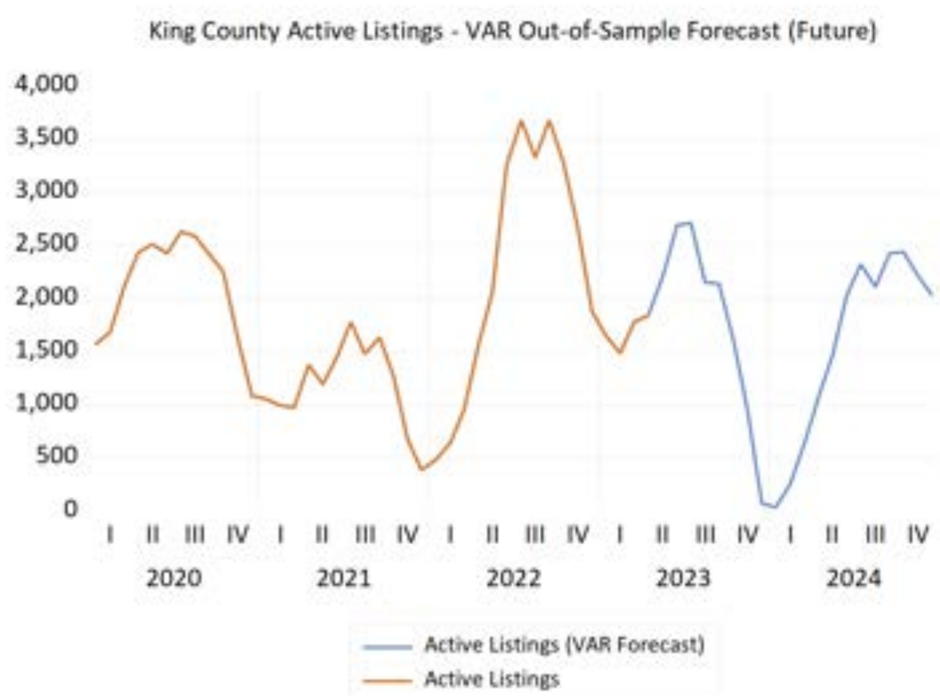


Figure 23: VAR out of sample forecast for number of active listings

VAR Discussion

The in-sample forecast tracks very well for recent history. The out-of-sample forecast for median home price seems natural and plausible. Anecdotally, this somewhat agrees with my subjective forecast, although a little more bearish, that home prices will continue to correct from their post-pandemic highs and temper to flat or moderate growth with rising mortgage rates tamping down demand. However, I thought that these dampened prices would also be influenced by rising inventory in 2024 as pent-up supply would eventually release, similar to post-covid, but in this case induced by interest rates leveling off. The visualized forecast of the number of active listings does not seem realistic, but specifications using KING_SOLD showed no improvement. The in-sample forecast seems to be consistently underestimating active listings. The out-of-sample forecast seems drastically low. Although, I suppose it's possible that high-mortgage rates will continue to reduce supply especially in the off-season months, but it's hard to imagine supply being this low (near 0) in the '23/'24 winter as the VAR forecast suggests.

Subjective Forecast

Subjective forecasting can provide a competing forecast, comparison, or portion of an ensemble forecast that can serve the purpose of incorporating intuition that might not be captured in data or econometric methods. When subject matter experts have spent sufficient time in a particular area or dataset their intuition may be just as informative as sophisticated models and should be considered together with other methods for a complete forecast.

Nationally, many real estate analysts highlight the slowdown that has occurred in the past four quarters, while also bringing to attention a recent surge in prices that is thought to be supported by low inventory levels. Mortgage rates remain high. The Fed has signaled that they will likely raise interest rates one more time this year which means that mortgage rates won't be coming down until next year, at the latest, leaving many buyers on the sidelines and sellers unwilling to trade their low mortgages for recent historical highs. Amid fears of a recession earlier in the year, there were well publicized tech layoffs and hiring freezes in the Seattle region which has seen job growth slowing. Putting all this together, I would expect that the Seattle housing market will continue to go through a correction with negative year-over-year growth rates throughout the rest of 2023 but turn positive 2024. The counterbalancing forces of low inventory and low demand will result in YoY growth leveling off to moderate/average YoY growth by the end of 2024. This subjective forecast is depicted visually below.

Figure 24: Subjective Forecast for Median Home Price

Given high mortgage rates the rest of the year, I would suspect that inventory levels would continue to see YoY declines of about -20%. However, assuming a less restrictive monetary policy in 2024, sellers may take this is a sign to begin to move any inventory that has built up over this past rate hiking cycle. Number of active listings will still be historically low as the market stabilizes after its 2023 correction, but I would expect moderate positive YoY growth in 2024. Job growth has shown to be resilient. Assuming it remains strong in 2024, the pent-up supply will release as sellers who have been waiting for mortgage rates to stabilize will put their properties on the market. Below is a visual of this subjective forecast.

Figure 25: Subjective forecast of number of active listings

ARIMA Modeling – Median Home Price

ARIMA modeling uses the past dynamics and auto-correlations within a univariate data series to forecast the series into out of sample periods. Similar to VAR, it is an inertial modeling technique that leverages that is a-theoretical because the specification on AR and MA coefficients don't necessarily have to adhere to any bespoke theory as in structural regression. VAR uses the dynamics of multiple endogenous variables whereas ARIMA only leverages the one variable.

For median home price, a first and seasonal difference was taken first to make the data stationary, a pre-requisite to ARIMA modeling. The transformed data series was then modeled with an AR term at lag 1, an MA term at lag 1, and a seasonal MA term at lag 12. Using compact ARIMA notation this specification can be identified as (1,1,1) + (0, 1, 1)^{12} and expressed in a mathematical fashion below:

$$\Delta_{12}\Delta\ \text{KING\_PMED\$} = \alpha_1\ \Delta_{12}\Delta\ \text{KING\_PMED\$}_{\,t-1} + \beta_1\ \varepsilon_{t-1} + \beta_2\ \varepsilon_{t-12} + \varepsilon_t$$

Equation 9: ARIMA specification for median home price

No intervention variables were added to this ARIMA specification. When I did try a specification modeling the COVID bounce back of 2022 with a step there was no marked improvement in the in-sample error.

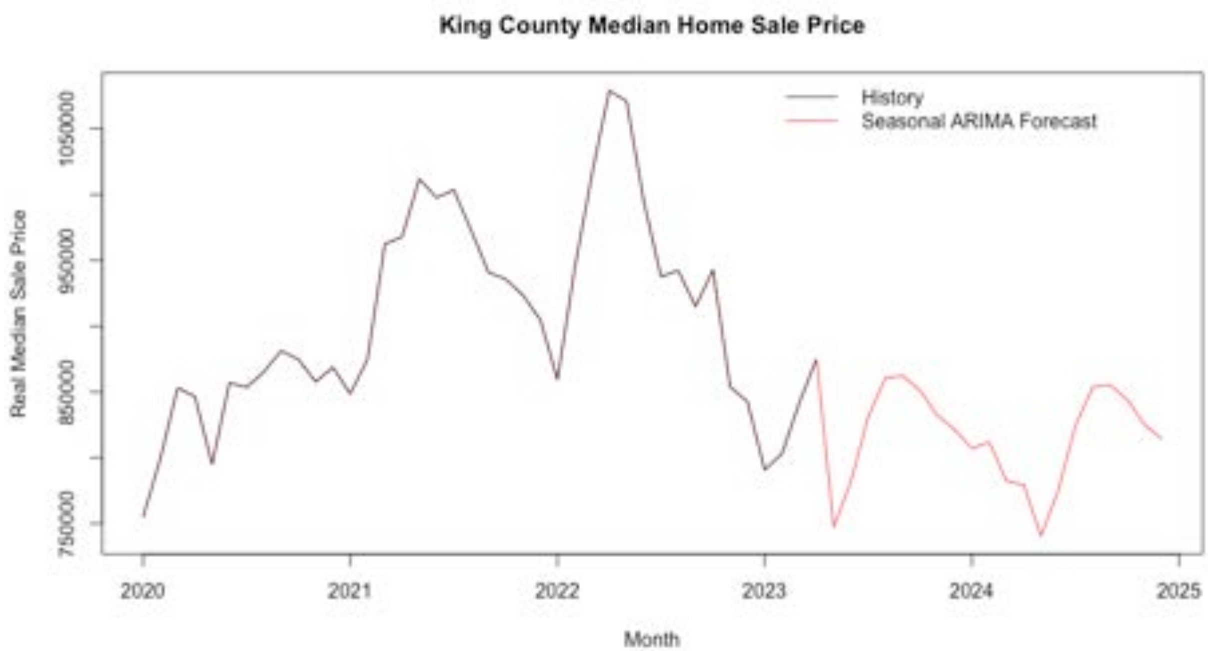


Figure 26: Median Home Price, ARIMA out of sample forecast

The ARIMA method's out of sample forecast aligns more with the VAR and subjective forecast that we can expect 2023 to provide negative growth relative to the highs of 2022 and be seen as a correction to the growth of recent years. Unlike, my subjective forecast, it predicts relatively flat growth over 2024, similar to the VAR. This forecast certainly seems plausible but given how resilient the economy has been in 2023 strong job growth is in the local Seattle economy, I would believe that it is somewhat pessimistic. It does properly capture the seasonality of the data.

Out of Sample Forecast- Methods Comparison

We now have 4 different forecasts using 4 different forecast methods: structural regressions, VAR, subjective, and ARIMA. Each offers a different angle of attacking the problem of forecasting median home price. Some methods like ARIMA and VAR use inertial patterns in the data, while structural regressions took a lot of time and brainpower to specify in a way that incorporated appropriate economic variables and theory. Lastly, subjective forecast incorporates my intuition about where the market may move in the near future given the likely macroeconomic environment and is less reliant on what may be lagging effects and estimates of the data generating process. Each method incorporates a different set of information and that is apparent in the variety of outcomes we each out of sample forecast.

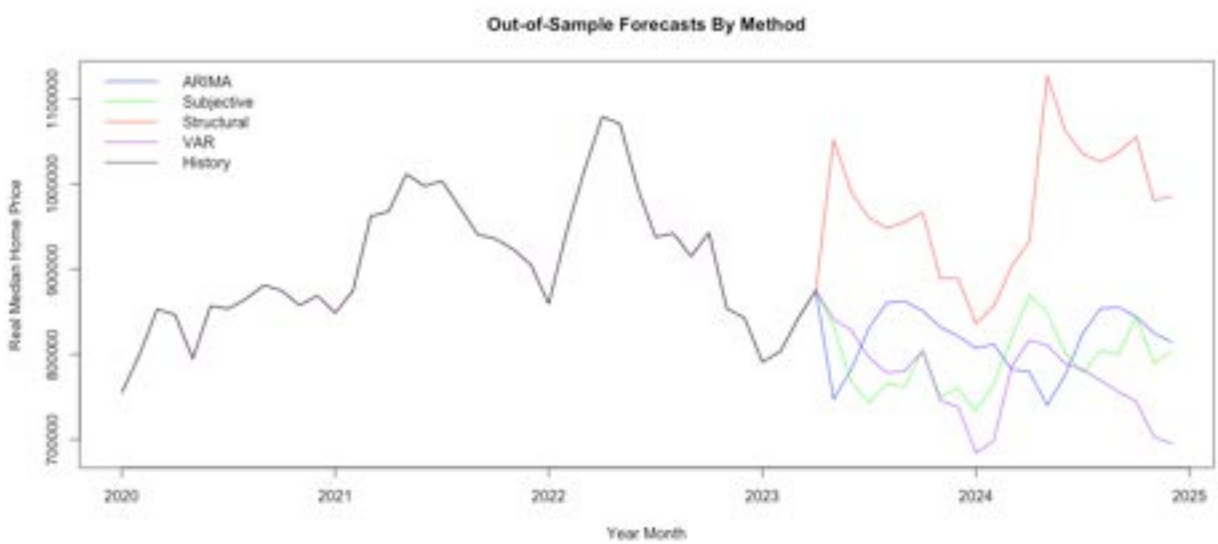


Figure 27: Comparison of out-of-sample forecast by forecast method.

The structural regression forecast is the most optimistic, projecting flat growth in 2023 with a positive trend through 2024. Essentially, the structural model forecasts Seattle will continue on its accelerated growth after a reset in 2023. The subjective, VAR and ARIMA forecasts all seem to be telling a similar story with varying degrees of pessimism. All three project a 2023 correction that erases most of the gains of the past few years and then stagnant growth through 2024. The ARIMA is the most bullish of the three flavors but has a smoothness that seems less plausible. After all, ARIMA it is a relative "weak learner" only incorporating the patterns of the data without any reference to exogenous variables, theory, or intuition. VAR appears to be the most bearish of these three where seasonal highs continue to decline in the next two years. One might almost be able to say the VAR forecasts a downward trend in the median home sale which indicates the negative growth is here to stay of the near term. My subjective forecast falls somewhere in the middle of ARIMA and VAR. The subjective takes the position that 2023 will be where the market bottoms out and 2024 growth, while not on the same growth trajectory as the past decade, will offer homeowners at least long-term average YoY growth in 2024.

In-Sample Forecast Analysis

I can assess each of these forecast methods by using them to forecast recent history and seeing how far off they are at predicting known values. For each in-sample forecast, I'll use an in-sample range from November 2021 through April 2023, my last known observation. These 18 observations represent about 6% of total observations. I'll use two different metrics to assess how far off the prediction is to actual value: squared error and absolute percentage error. After computing these metrics for each in-sample observation I can then summarize each across this in-sample period by taking the mean method The Root Mean Squared Error (RMSE) and Mean Absolute Percent error can then be used as quantitative performance measures for each forecast method.

Where $\widehat{KING\_PMED\$}$ signifies predicted value of each method.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(KING\_PMED\$ - \widehat{KING\_PMED\$})^{2}}$$

Equation 10: Root mean squared error

$$MAPE = \frac{1}{n}\sum_{i=1}^{n} ABS\left(\frac{KING\_PMED\$ - \widehat{KING\_PMED\$}}{KING\_PMED\$}\right)$$

Equation 11: Mean absolute percent error

Figure 28: In-sample forecast method comparison

All three in-sample forecasts track pretty similar to each other. The VAR and Structural in-sample forecasts track so close to history in 2020-2021 that they aren't visible as being distinct. This might actually be a sign of these methods overfitting this period. Only the ARIMA was able to find the signal for the 2022 highs, but it was the VAR that most closely models the come-down from that peak. We can see that structural in-sample forecasted a rise in home prices in 2023 while the actual sale price continued to slide in these most recent months.

In Sample Error Charts

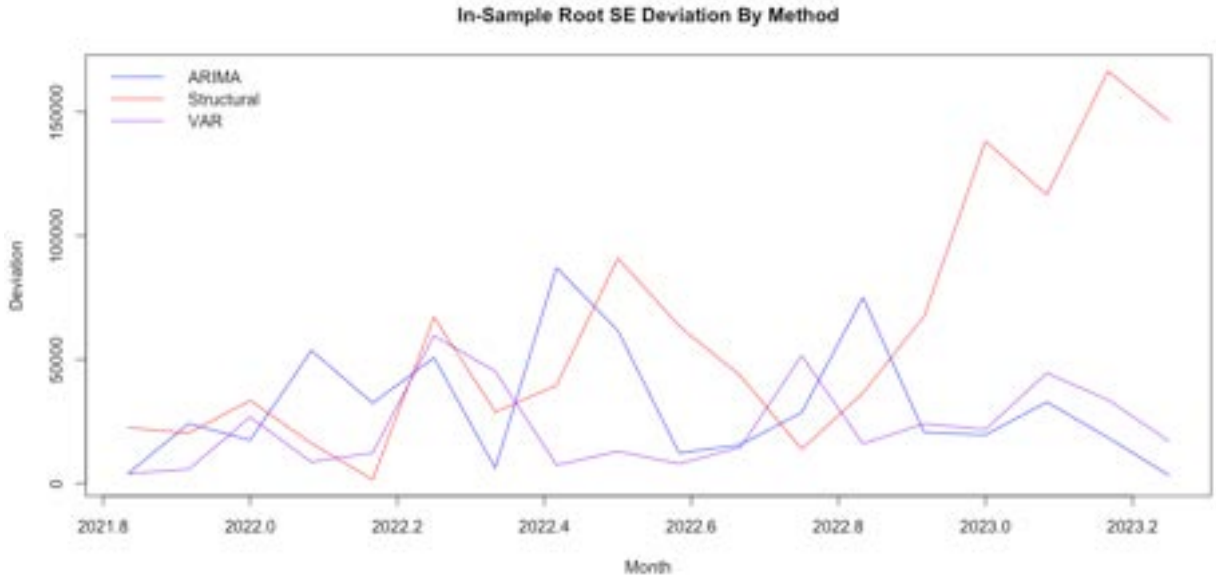


Figure 29: Root Squared Error by Method

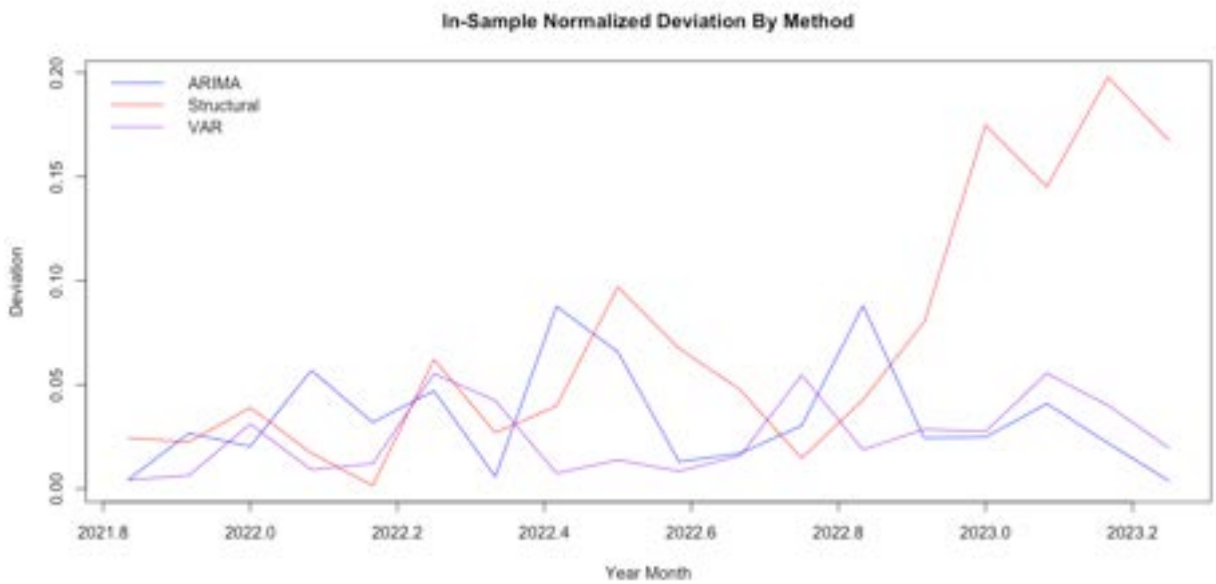


Figure 30: Average Percent Error by Method

We can see from the charts above that both the SE and APE metrics offer very similar analysis of in-sample errors. Both the ARIMA and structural failed relatively to model median home price surge in the middle of 2022 while VAR modeled this period with little error. The structural regression's errors became very large when applied to the most recent months in early 2023. Clearly the structural model could not have predicted the market downturn in 2023. The seed of this miss however can be traced back to its failure to model the 2022 highs that were likely caused by the release of pent-up demand and supply from the pandemic. Again, it was decided not to use an interval variable in the structural regression to model the COVID bounce-back and these in-sample forecast errors reflect one downside of that decision. Ultimately, it's the VAR method that seems to be performing best when applied to within sample observations.

METHOD	RMSE	Equal	Optimal #1*	Optimal #2*
ARIMA	39,385	33.3%	39.8%	26.9%
Structural	78,597	33.3%	13.0%	29.2%
VAR	28,361	33.3%	47.3%	43.9%
COMBINATIONS				
Equal Weight	41,343			
Optimal Weight #1	32,322			
Optimal Weight #2	38,853			

*Rounding is why these don't sum to 1.0

Table 5: Weight for combination forecasts based on RMSE of in sample forecast.

METHOD	MAPE	Equal	Optimal #1*	Optimal #2*
ARIMA	3.3%	33.3%	40.5%	31.1%
Structural	7.0%	33.3%	12.2%	29.5%
VAR	2.5%	33.3%	47.3%	39.4%
COMBINATIONS				
Equal Weight	4.0%			
Optimal Weight #1	3.1%			
Optimal Weight #2	3.8%			

*Rounding is why these don't sum to 1.0

Table 6: Weights for combination forecasts based on MAPE of in-sample

Combination Forecasting

The more information that a forecast model uses the more performant that forecast will be in terms of predicting the short-term outlook of the series. Combining forecasts estimated with different techniques consolidates mutually exclusive information and signal found in each technique's forecast and reduces the variance of the forecast error. Especially when distinct forecasts tell different stories, combining such forecasts will result in a combination forecast (a.k.a. ensemble forecast) that will often perform better than one technique alone. One simple way to combine forecasts is an un-weighted average of all forecast observations:

$$CF = \frac{1}{n}\sum_{i=1}^{n} F_i$$

Equation 12: Equal weighted combination forecast

In the above equation each technique's forecast (F_i) is averaged with an equal weight to produce the combination forecast (CF). However, we can also assign each forecast different weights subjectively or deterministically based on that technique's performance on known observations such as the in-sample range. This is where the RMSE and MAPE in-sample performance measures can be utilized. Instead of assigning equal weights to each technique, we can weight each technique according to the RMSE or MAPE. For instance, we saw that the structural regression performed the worst, in-sample, of the three techniques so we can assign less weight to it for a combination forecast. Optimal #1 weights represent this strategy. These weights were derived by summing the three RMSEs of the three techniques (ARIMA, VAR, Structural) to get a total error share then dividing each RMSE by the total error share to get each method's proportion of total error. Then, we can adjust the equal weight up (down) for those techniques that represent less (more) of the total error share. This strategy can be replicated with MAPE as a performance measure. Optimal #2 takes a similar weighting strategy, but instead of using the in-sample periods, compares data that has been released since the beginning of this project (May, June, July 2023) and uses those (technically) out-of-sample forecasts for these months compared to newly released data to do a similar error analysis as the in-sample range. Optimal #2 thus incorporates the most recent quarter of data to determine the weights for the combination forecast.

Figure 31: Three in-sample combination forecasts applying equal weights and two different optimally derived weights leveraging RMSE as a performance measure

All three weighting strategies forecast the in-sample period very similarly. This is expected because they were derived using this period. Optimal Weight 1 does a slightly better job modeling the past year due in part to its down-weighting of the structural regression that had large errors during time.

Figure 32: Three out of sample combination forecasts applying equal weights and two different optimally derived weights leveraging RMSE as a performance measure

The out-of-sample combination forecast tells a more interesting story. Both the Equal Weight and Optimal Weight 2 are close in their prediction likely because Optimal Weight 2 resulted in a more equal weighting than Optimal Weight 1. This gives validity to equal weights being the best weighting strategy above any more complex weighting system. Optimal Weight 1 offers the most bearish forecast thanks to the down-weighting of the more optimistic structural regression. Optimal 1 tells a story of continued negative growth for real median home price, while the other two combination forecasts show relative flat growth in 2024.

Although they don't make any outlandish claims, combination forecasts seem to offer the most plausible projections. This can be a good thing as they don't fly in the face of consensus forecasts and therefore may be seen as valid if shared among subject matter experts. However, they also don't set the forecast apart and offer any novel predictive insights. Nevertheless, if forecasts are ultimately judged on their accuracy, the reduction in error variance is a priceless of these combination forecasts. I'm relatively confident in these outlooks, particularly Equal Weight. As we saw, Optimal Weight 2 regressed back to equal weights.

Conclusion and Future Work

This project produced a total of seven distinct out-of-sample forecasts for median home price in King County, WA. One subjective, one using a muti-equation model using structural regressions (structural), two using inertial methods (ARIMA & VAR) and three combination forecasts leveraging structural and inertial methods (equal weight, optimal weight 1, optimal weight 2). Of these VAR is my preferred method forecast due to its clear superior performance modeling the in-sample period. The equal weight forecast is my preferred combination forecast despite having the worst in-sample RMSE of the combination forecasts. Much has been written in the forecasting literature of the benefit of equal weights and how they often outperform complex weighting schemes. We saw that Optimal Weight 2 resulted in weights closer to equal than Optimal Weight 1 which offers an early tip that equal weighting will be the best strategy in an uncertain out-of-sample period. In all likelihood, the Optimal Weights 1 & 2 could be overfit to the point that their bias is contributing to increased variance when applied to an uncertain future.

	Methods			Combinations		
	ARIMA	VAR	Structural	Equal Wt	Optimal Wt #1	Optimal Wt #2
RMSE	39,384	28,361	78,597	41,343	32,322	38,852
MAPE	3.4%	2.5%	7.0%	4.0%	3.1%	3.8%

Table 7: Summary of Methods and Combination Forecasts

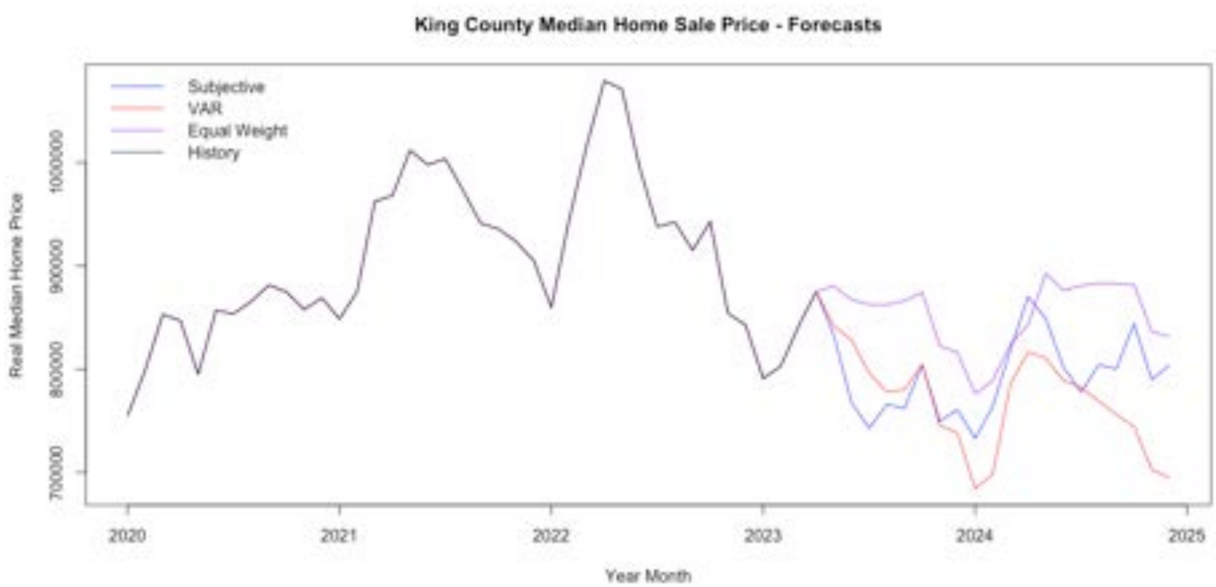


Figure 33: Preferred out-of-sample forecasts: VAR, Equal Weight, Subjective

Equal Weights is the most optimistic of these three preferred forecasts. The Seattle market has shown to be resilient before as in the years after the Great Recession. It would be no surprise if the more optimistic forecast came to fruition. VAR remains the most pessimistic of my preferred forecasts projecting continued negative YoY growth of real median home price. Unsurprisingly, my subjective forecast splits the difference between the other two relative extreme scenarios.

This research and forecasting exercise have been full of lessons learned. Chief among them has been the difficulty and time resources required to specify a multi-equation regression model that agrees with economic theory. The variables that one would think are determinants of real estate like mortgage rates and income have the potential to have their signal overpowered by things like inventory scarcity and a macroeconomic monetary policy that has kept interest rates low, inflating the market, causing unaffordability. Despite these challenges there is still enough econometric basis to produce many sound forecasts and be confident in their accuracy given certain assumptions.

It will be interesting to see what the Seattle housing market does over the next 18 months. Like 2022-2023, I think the lack of inventory that is keeping prices up will have to break at some point in the short term. Will repressed supply and demand cause a surge in prices when it breaks or will a new lower equilibrium form due to exogenous forces like a recession in 2024. Time will tell.

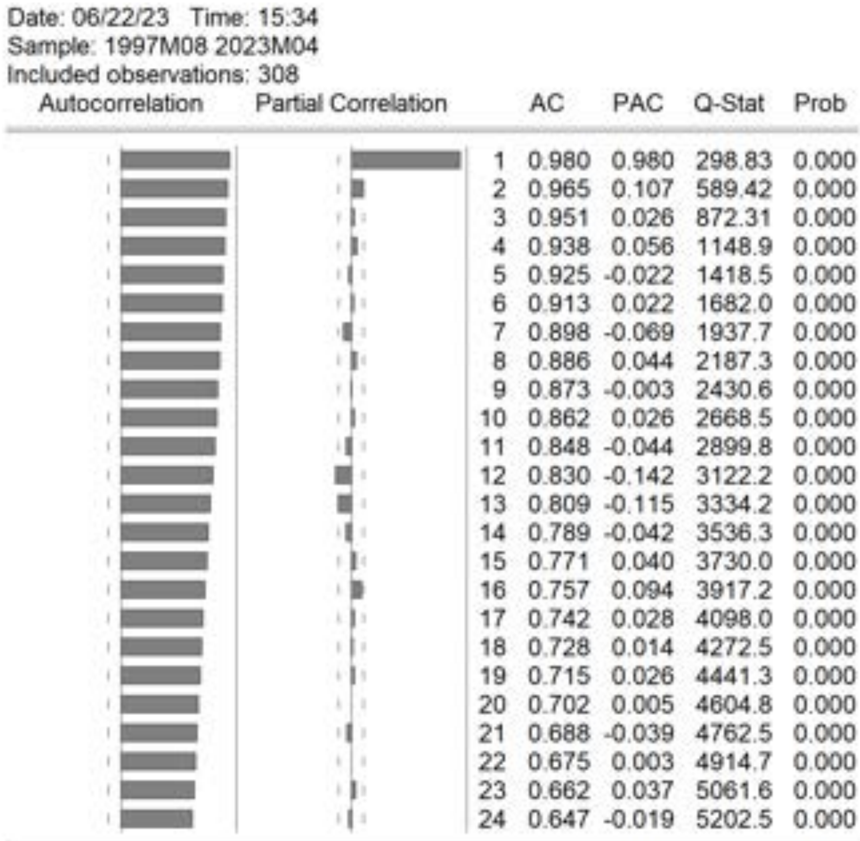
These forecasts and scenarios can offer valuable insights for buyers, sellers, real estate agents and other interested parties in the Seattle residential real estate market. For instance, buyers may take this outlook for the next 18 months as a good time to buy as the market might be experiencing a rare dip. Yet, inventory remains historically low so it may still be hard to find a home that checks all a consumer's boxes. Mortgage rates will also price many out of the market while they remain high relative to recent history. Perhaps this exogenous factor will eventually result in lower prices as constrained supply releases, different from post-Covid surges where a release of suppressed demand and supply worked together to increase prices.

Sellers may see this forecast as reason to hold onto their assets for the short term but should also caution that it's somewhat uncertain whether the market will improve significantly. If the Seattle job market growth slows or even turns negative, we can expect prices to continue to fall. In such a scenario, waiting until 2024 to sell will prove to be a mistake. These are the types of insights that real estate agents should make their clients aware.

If given significant more time and resources, I would do a more thorough literature review on the determinants of real estate and recent success of any peer forecasters in projecting similar markets. I would sink time into the structural regressions and attempt specifications that adhere to ideas I find in the literature. Adding more variables or different transformations would reduce the probability of omitted variable bias.

Appendix 1 – Machine Output

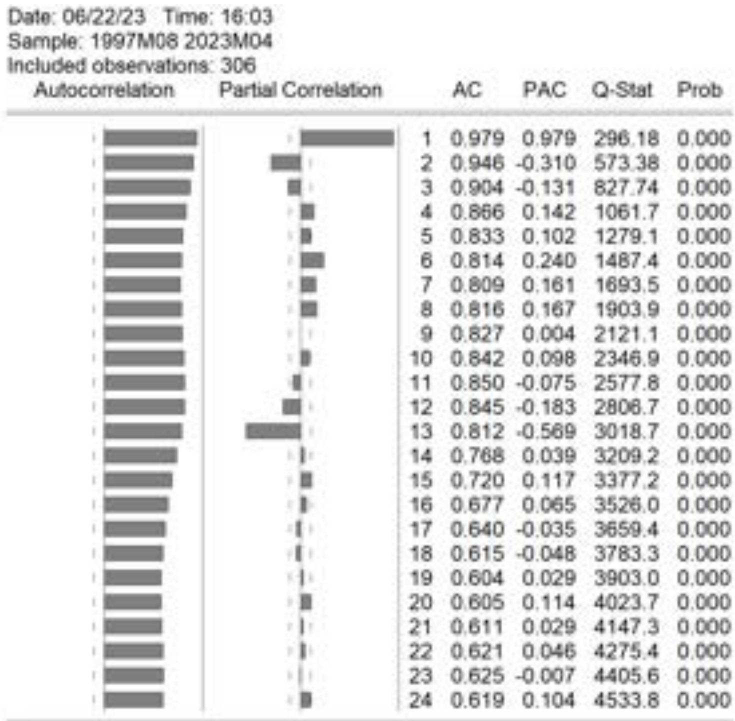
Correlograms of Two KDV in Levels



Date: 06/22/23 Time: 15:34
Sample: 1997M08 2023M04
Included observations: 308

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.980	0.980	298.83	0.000
		2	0.965	0.107	589.42	0.000
		3	0.951	0.026	872.31	0.000
		4	0.938	0.056	1148.9	0.000
		5	0.925	-0.022	1418.5	0.000
		6	0.913	0.022	1682.0	0.000
		7	0.898	-0.069	1937.7	0.000
		8	0.886	0.044	2187.3	0.000
		9	0.873	-0.003	2430.6	0.000
		10	0.862	0.026	2668.5	0.000
		11	0.848	-0.044	2899.8	0.000
		12	0.830	-0.142	3122.2	0.000
		13	0.809	-0.115	3334.2	0.000
		14	0.789	-0.042	3536.3	0.000
		15	0.771	0.040	3730.0	0.000
		16	0.757	0.094	3917.2	0.000
		17	0.742	0.028	4098.0	0.000
		18	0.728	0.014	4272.5	0.000
		19	0.715	0.026	4441.3	0.000
		20	0.702	0.005	4604.8	0.000
		21	0.688	-0.039	4762.5	0.000
		22	0.675	0.003	4914.7	0.000
		23	0.662	0.037	5061.6	0.000
		24	0.647	-0.019	5202.5	0.000

Figure 34: Correlogram of levels of median home sale price



Date: 06/22/23 Time: 16:03
Sample: 1997M08 2023M04
Included observations: 306

	AC	PAC	Q-Stat	Prob
1	0.979	0.979	296.18	0.000
2	0.946	-0.310	573.38	0.000
3	0.904	-0.131	827.74	0.000
4	0.866	0.142	1061.7	0.000
5	0.833	0.102	1279.1	0.000
6	0.814	0.240	1487.4	0.000
7	0.809	0.161	1693.5	0.000
8	0.816	0.167	1903.9	0.000
9	0.827	0.004	2121.1	0.000
10	0.842	0.098	2346.9	0.000
11	0.850	-0.075	2577.8	0.000
12	0.845	-0.183	2806.7	0.000
13	0.812	-0.569	3018.7	0.000
14	0.768	0.039	3209.2	0.000
15	0.720	0.117	3377.2	0.000
16	0.677	0.065	3526.0	0.000
17	0.640	-0.035	3659.4	0.000
18	0.615	-0.048	3783.3	0.000
19	0.604	0.029	3903.0	0.000
20	0.605	0.114	4023.7	0.000
21	0.611	0.029	4147.3	0.000
22	0.621	0.046	4275.4	0.000
23	0.625	-0.007	4405.6	0.000
24	0.619	0.104	4533.8	0.000

Figure 35: Correlogram of the levels of Number of Active Listings

Correlogram of Two KDV First Differenced

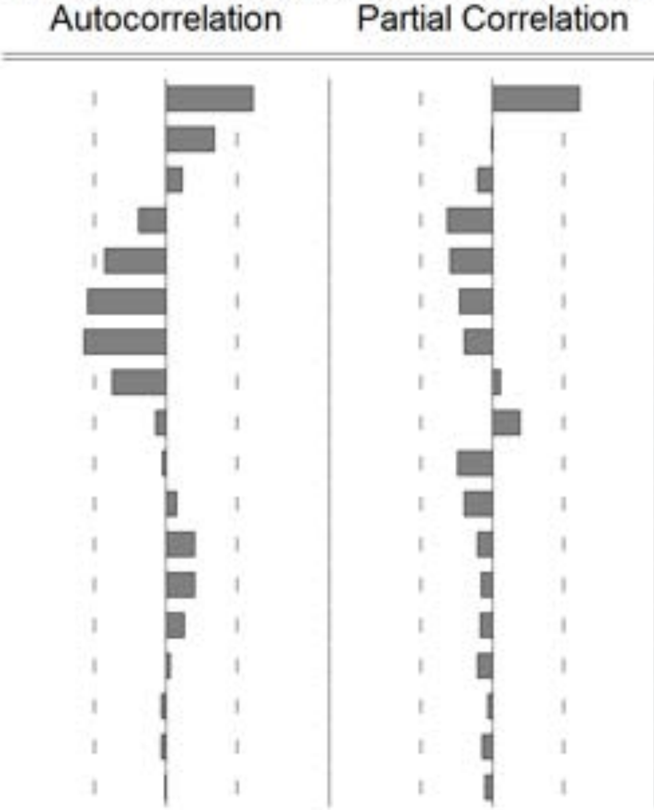
Date: 08/08/23 Time: 13:55
Sample (adjusted): 2021M10 2023M04
Included observations: 19 after adjustments

Autocorrelation | Partial Correlation

	AC	PAC	Q-Stat	Prob
1	0.253	0.253	1.4150	0.234
2	0.092	0.030	1.6132	0.446
3	-0.424	-0.485	6.0956	0.107
4	-0.246	-0.038	7.7095	0.103
5	-0.021	0.220	7.7226	0.172
6	0.144	-0.074	8.3592	0.213
7	-0.085	-0.424	8.6013	0.283
8	-0.137	-0.010	9.2859	0.319
9	-0.389	-0.195	15.320	0.083
10	-0.050	-0.098	15.431	0.117
11	0.025	-0.079	15.461	0.162
12	0.228	-0.078	18.415	0.104
13	0.155	0.009	20.012	0.095
14	0.037	-0.059	20.120	0.126
15	-0.045	-0.068	20.323	0.160
16	-0.020	-0.082	20.378	0.204
17	-0.010	-0.061	20.399	0.254
18	-0.004	-0.236	20.404	0.311

Figure 36: Correlogram of First Difference of KING_PMED

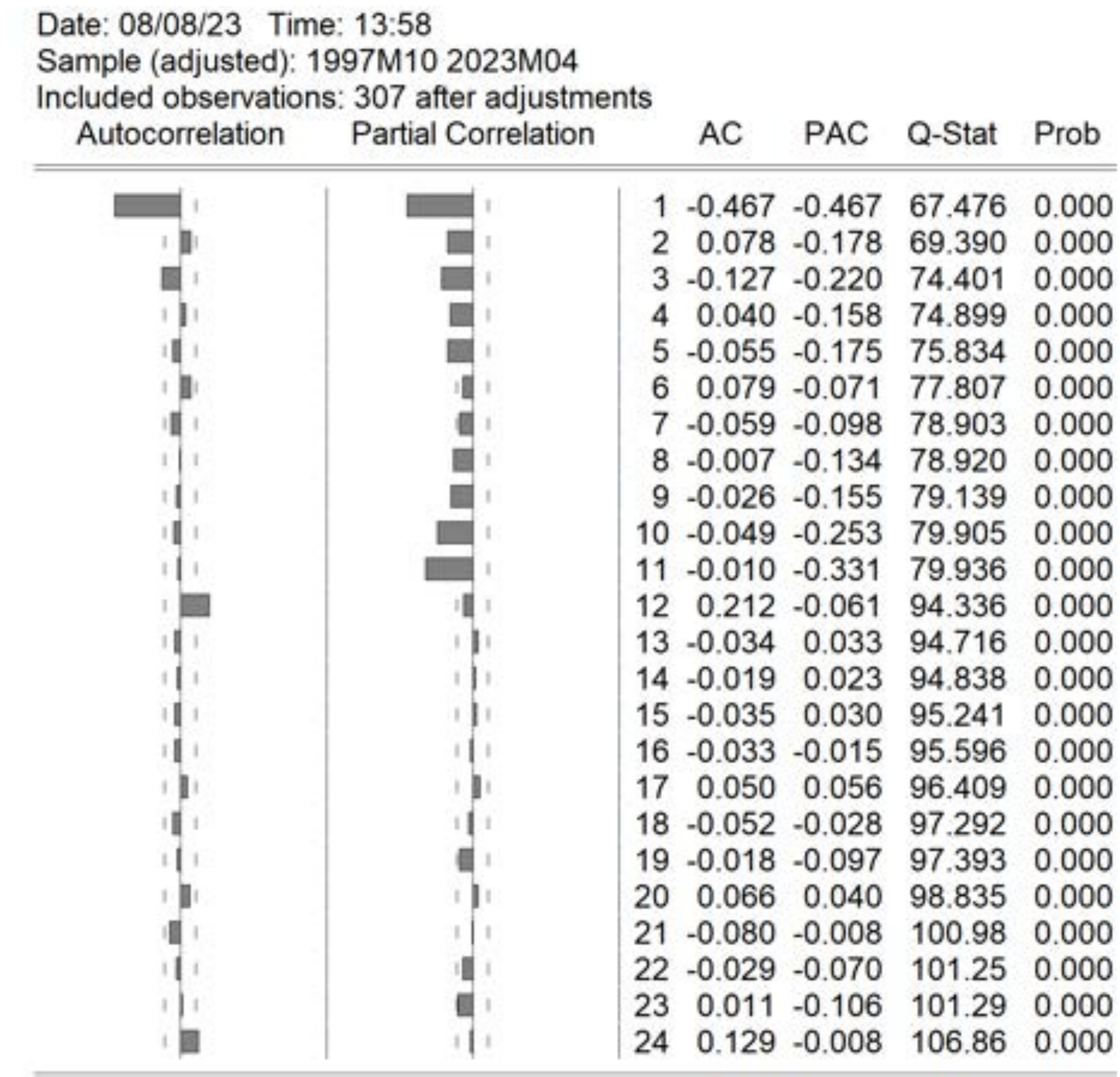
Date: 08/08/23 Time: 14:00
Sample (adjusted): 2021M10 2023M04
Included observations: 19 after adjustments



	AC	PAC	Q-Stat	Prob
1	0.557	0.557	6.8845	0.009
2	0.308	-0.003	9.1178	0.010
3	0.104	-0.097	9.3879	0.025
4	-0.171	-0.280	10.166	0.038
5	-0.387	-0.263	14.438	0.013
6	-0.498	-0.213	22.040	0.001
7	-0.517	-0.177	30.930	0.000
8	-0.336	0.057	35.035	0.000
9	-0.065	0.174	35.204	0.000
10	-0.024	-0.220	35.229	0.000
11	0.073	-0.179	35.493	0.000
12	0.185	-0.095	37.441	0.000
13	0.186	-0.066	39.746	0.000
14	0.120	-0.076	40.886	0.000
15	0.025	-0.097	40.950	0.000
16	-0.027	-0.028	41.045	0.001
17	-0.030	-0.060	41.226	0.001
18	-0.004	-0.046	41.231	0.001

Figure 37: Correlogram of first difference of KING_ACTLIST

Correlogram of KING_PMED Second Differenced

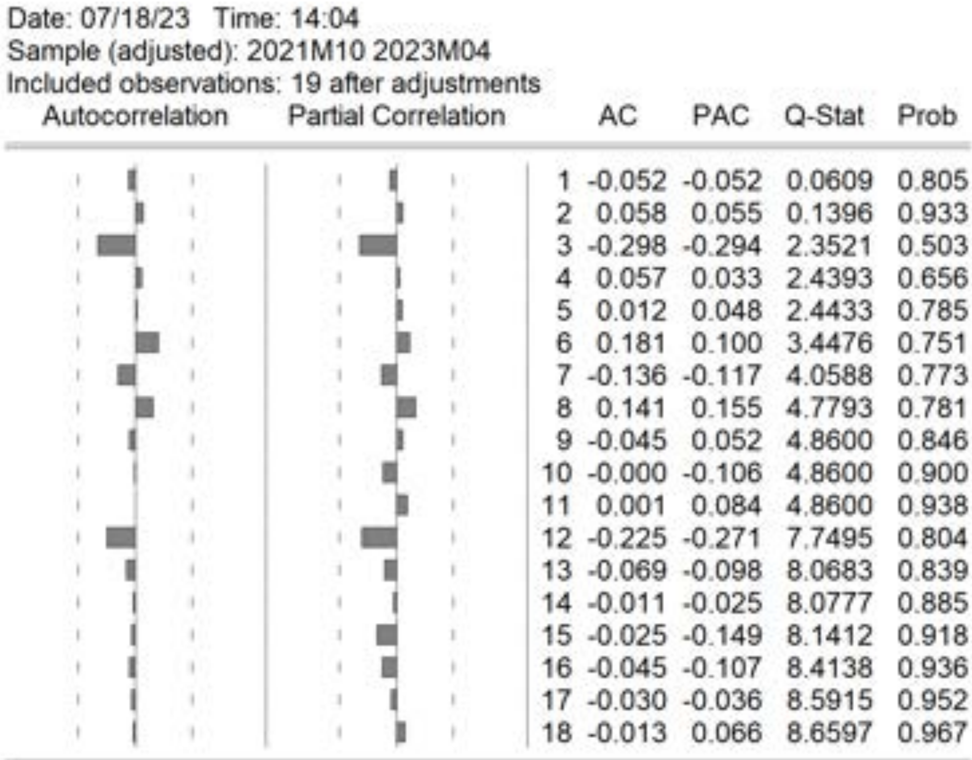


Date: 08/08/23 Time: 13:58
Sample (adjusted): 1997M10 2023M04
Included observations: 307 after adjustments

	AC	PAC	Q-Stat	Prob
1	-0.467	-0.467	67.476	0.000
2	0.078	-0.178	69.390	0.000
3	-0.127	-0.220	74.401	0.000
4	0.040	-0.158	74.899	0.000
5	-0.055	-0.175	75.834	0.000
6	0.079	-0.071	77.807	0.000
7	-0.059	-0.098	78.903	0.000
8	-0.007	-0.134	78.920	0.000
9	-0.026	-0.155	79.139	0.000
10	-0.049	-0.253	79.905	0.000
11	-0.010	-0.331	79.936	0.000
12	0.212	-0.061	94.336	0.000
13	-0.034	0.033	94.716	0.000
14	-0.019	0.023	94.838	0.000
15	-0.035	0.030	95.241	0.000
16	-0.033	-0.015	95.596	0.000
17	0.050	0.056	96.409	0.000
18	-0.052	-0.028	97.292	0.000
19	-0.018	-0.097	97.393	0.000
20	0.066	0.040	98.835	0.000
21	-0.080	-0.008	100.98	0.000
22	-0.029	-0.070	101.25	0.000
23	0.011	-0.106	101.29	0.000
24	0.129	-0.008	106.86	0.000

Figure 38: Correlogram of 2nd Difference of KING_PMED

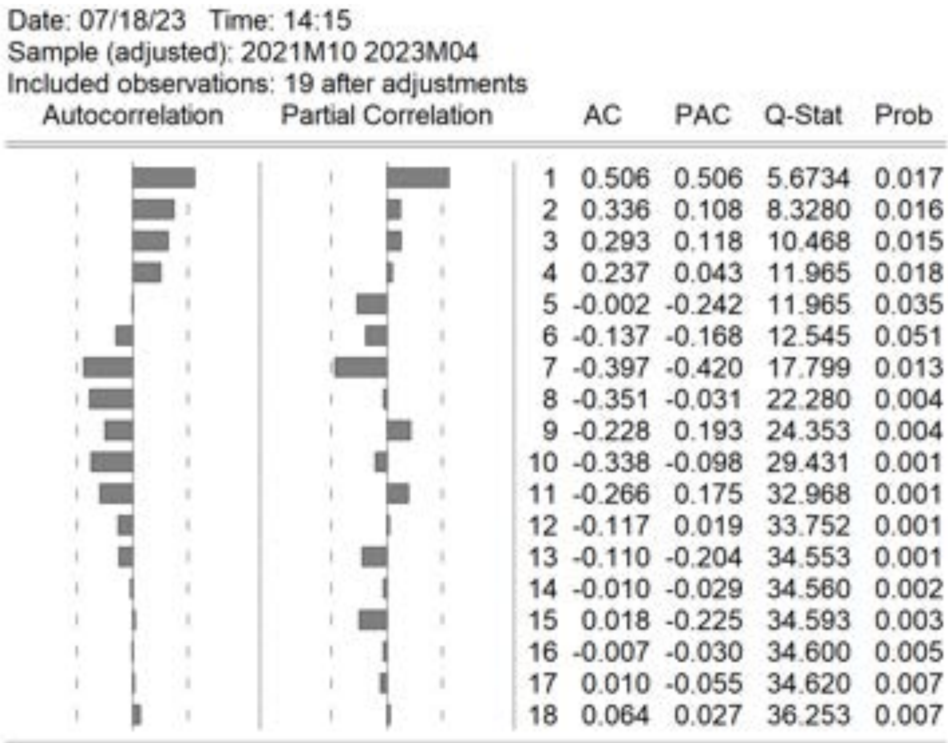
Correlograms of KDVs First & Seasonally Differenced



Date: 07/18/23 Time: 14:04
Sample (adjusted): 2021M10 2023M04
Included observations: 19 after adjustments

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.052	-0.052	0.0609	0.805
		2	0.058	0.055	0.1396	0.933
		3	-0.298	-0.294	2.3521	0.503
		4	0.057	0.033	2.4393	0.656
		5	0.012	0.048	2.4433	0.785
		6	0.181	0.100	3.4476	0.751
		7	-0.136	-0.117	4.0588	0.773
		8	0.141	0.155	4.7793	0.781
		9	-0.045	0.052	4.8600	0.846
		10	-0.000	-0.106	4.8600	0.900
		11	0.001	0.084	4.8600	0.938
		12	-0.225	-0.271	7.7495	0.804
		13	-0.069	-0.098	8.0683	0.839
		14	-0.011	-0.025	8.0777	0.885
		15	-0.025	-0.149	8.1412	0.918
		16	-0.045	-0.107	8.4138	0.936
		17	-0.030	-0.036	8.5915	0.952
		18	-0.013	0.066	8.6597	0.967

Figure 39: Correlogram of first and seasonal difference of Median Home Price



Date: 07/18/23 Time: 14:15
Sample (adjusted): 2021M10 2023M04
Included observations: 19 after adjustments

	AC	PAC	Q-Stat	Prob
1	0.506	0.506	5.6734	0.017
2	0.336	0.108	8.3280	0.016
3	0.293	0.118	10.468	0.015
4	0.237	0.043	11.965	0.018
5	-0.002	-0.242	11.965	0.035
6	-0.137	-0.168	12.545	0.051
7	-0.397	-0.420	17.799	0.013
8	-0.351	-0.031	22.280	0.004
9	-0.228	0.193	24.353	0.004
10	-0.338	-0.098	29.431	0.001
11	-0.266	0.175	32.968	0.001
12	-0.117	0.019	33.752	0.001
13	-0.110	-0.204	34.553	0.001
14	-0.010	-0.029	34.560	0.002
15	0.018	-0.225	34.593	0.003
16	-0.007	-0.030	34.600	0.005
17	0.010	-0.055	34.620	0.007
18	0.064	0.027	36.253	0.007

Figure 40: Correlogram of first and seasonal difference of Number of Active Listings

Unit Root Tests of KDVs Levels and Differenced

Null Hypothesis: D(KING_PMED,1) has a unit root
Exogenous: Constant
Lag Length: 12 (Automatic - based on AIC, maxlag=18)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-2.477249	0.1221
Test critical values:	1% level	-3.452366	
	5% level	-2.871128	
	10% level	-2.571950	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(KING_PMED,2)
Method: Least Squares
Date: 08/08/23 Time: 13:56
Sample (adjusted): 1998M10 2023M04
Included observations: 295 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(KING_PMED(-1),1)	-0.696791	0.281276	-2.477249	0.0138
D(KING_PMED(-1),2)	-0.373656	0.275798	-1.354816	0.1766
D(KING_PMED(-2),2)	-0.371501	0.263771	-1.408420	0.1601
D(KING_PMED(-3),2)	-0.503778	0.247685	-2.033949	0.0429
D(KING_PMED(-4),2)	-0.553698	0.229931	-2.408105	0.0167
D(KING_PMED(-5),2)	-0.606192	0.210651	-2.877706	0.0043
D(KING_PMED(-6),2)	-0.595201	0.189208	-3.145756	0.0018
D(KING_PMED(-7),2)	-0.657957	0.168968	-3.893969	0.0001
D(KING_PMED(-8),2)	-0.711078	0.147880	-4.808474	0.0000
D(KING_PMED(-9),2)	-0.745460	0.129510	-5.756011	0.0000
D(KING_PMED(-10),2)	-0.761342	0.110749	-6.874486	0.0000
D(KING_PMED(-11),2)	-0.600419	0.092626	-6.482221	0.0000
D(KING_PMED(-12),2)	-0.185453	0.067020	-2.767138	0.0060
C	1184.681	1133.691	1.044977	0.2969

R-squared	0.586702	Mean dependent var	139.6610
Adjusted R-squared	0.567582	S.D. dependent var	24100.06
S.E. of regression	15847.84	Akaike info criterion	22.22575
Sum squared resid	7.06E+10	Schwarz criterion	22.40072
Log likelihood	-3264.298	Hannan-Quinn criter.	22.29581
F-statistic	30.68439	Durbin-Watson stat	2.009478
Prob(F-statistic)	0.000000		

Figure 41: ADF Unit Root Test of First Difference of KING_PMED

Null Hypothesis: D(KING_PMED,1) is stationary
Exogenous: Constant
Bandwidth: 5 (Used-specified) using Quadratic Spectral kernel

		LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic		0.128852
Asymptotic critical values*:	1% level	0.739000
	5% level	0.463000
	10% level	0.347000

*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)

Residual variance (no correction)	3.15E+08
HAC corrected variance (Quadratic Spectral kernel)	3.47E+08

KPSS Test Equation
Dependent Variable: D(KING_PMED,1)
Method: Least Squares
Date: 08/08/23 Time: 13:57
Sample (adjusted): 1997M09 2023M04
Included observations: 308 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	2211.039	1013.103	2.182443	0.0298

R-squared	0.000000	Mean dependent var	2211.039
Adjusted R-squared	0.000000	S.D. dependent var	17779.88
S.E. of regression	17779.88	Akaike info criterion	22.41276
Sum squared resid	9.71E+10	Schwarz criterion	22.42487
Log likelihood	-3450.566	Hannan-Quinn criter.	22.41761
Durbin-Watson stat	1.764259		

Figure 42: KPSS unit root test of first difference of KING_PMED

Null Hypothesis: D(KING_PMED,1) has a unit root
Exogenous: Constant
Lag Length: 12 (Automatic - based on AIC, maxlag=18)

		t-Statistic
Elliott-Rothenberg-Stock DF-GLS test statistic		-1.637578
Test critical values:	1% level	-2.572824
	5% level	-1.941903
	10% level	-1.615980

*MacKinnon (1996)

DF-GLS Test Equation on GLS Detrended Residuals
Dependent Variable: D(GLSRESID)
Method: Least Squares
Date: 08/08/23 Time: 13:57
Sample (adjusted): 1998M10 2023M04
Included observations: 295 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
GLSRESID(-1)	-0.229638	0.140230	-1.637578	0.1026
D(GLSRESID(-1))	-0.822885	0.145310	-5.662972	0.0000
D(GLSRESID(-2))	-0.793557	0.145242	-5.463699	0.0000
D(GLSRESID(-3))	-0.894759	0.140568	-6.365333	0.0000
D(GLSRESID(-4))	-0.908642	0.136435	-6.659865	0.0000
D(GLSRESID(-5))	-0.923823	0.130241	-7.093183	0.0000
D(GLSRESID(-6))	-0.870258	0.123563	-7.043056	0.0000
D(GLSRESID(-7))	-0.894460	0.115718	-7.729671	0.0000
D(GLSRESID(-8))	-0.906778	0.107287	-8.451895	0.0000
D(GLSRESID(-9))	-0.903137	0.100364	-8.998624	0.0000
D(GLSRESID(-10))	-0.878699	0.092639	-9.485242	0.0000
D(GLSRESID(-11))	-0.673341	0.084815	-7.938937	0.0000
D(GLSRESID(-12))	-0.216338	0.065352	-3.310327	0.0011

R-squared	0.581320	Mean dependent var	139.6610
Adjusted R-squared	0.563504	S.D. dependent var	24100.06
S.E. of regression	15922.40	Akaike info criterion	22.23191
Sum squared resid	7.15E+10	Schwarz criterion	22.39439
Log likelihood	-3266.206	Hannan-Quinn criter.	22.29697
Durbin-Watson stat	2.020759		

Figure 43: DF-GLS Unit Root Test of First Difference of KING_PMED

Null Hypothesis: D(KING_PMED,2) has a unit root
Exogenous: Constant
Lag Length: 11 (Automatic - based on AIC, maxlag=18)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-14.35139	0.0000
Test critical values:	1% level	-3.452366	
	5% level	-2.871128	
	10% level	-2.571950	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(KING_PMED,3)
Method: Least Squares
Date: 08/08/23 Time: 13:58
Sample (adjusted): 1998M10 2023M04
Included observations: 295 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(KING_PMED(-1),2)	-12.11704	0.844311	-14.35139	0.0000
D(KING_PMED(-1),3)	10.07531	0.804959	12.51654	0.0000
D(KING_PMED(-2),3)	9.077841	0.746676	12.15768	0.0000
D(KING_PMED(-3),3)	7.995549	0.681241	11.73674	0.0000
D(KING_PMED(-4),3)	6.917770	0.607024	11.39620	0.0000
D(KING_PMED(-5),3)	5.843767	0.529319	11.04017	0.0000
D(KING_PMED(-6),3)	4.843854	0.448744	10.79424	0.0000
D(KING_PMED(-7),3)	3.838763	0.368472	10.41806	0.0000
D(KING_PMED(-8),3)	2.841128	0.290040	9.795629	0.0000
D(KING_PMED(-9),3)	1.866069	0.212719	8.772452	0.0000
D(KING_PMED(-10),3)	0.934931	0.137931	6.778265	0.0000
D(KING_PMED(-11),3)	0.229595	0.065194	3.521751	0.0005
C	-442.7459	932.3219	-0.474885	0.6352

R-squared	0.855988	Mean dependent var	0.677966
Adjusted R-squared	0.849860	S.D. dependent var	41270.74
S.E. of regression	15991.53	Akaike info criterion	22.24057
Sum squared resid	7.21E+10	Schwarz criterion	22.40305
Log likelihood	-3267.485	Hannan-Quinn criter.	22.30563
F-statistic	139.6814	Durbin-Watson stat	2.025936
Prob(F-statistic)	0.000000		

Figure 44: ADF unit root test of the 2nd difference of KING_PMED

Null Hypothesis: D(KING_PMED,2) has a unit root
Exogenous: Constant
Lag Length: 17 (Automatic - based on AIC, maxlag=18)

		t-Statistic
Elliott-Rothenberg-Stock DF-GLS test statistic		0.296264
Test critical values:	1% level	-2.572988
	5% level	-1.941925
	10% level	-1.615966

*MacKinnon (1996)

DF-GLS Test Equation on GLS Detrended Residuals
Dependent Variable: D(GLSRESID)
Method: Least Squares
Date: 08/08/23 Time: 13:58
Sample (adjusted): 1999M04 2023M04
Included observations: 289 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
GLSRESID(-1)	0.033678	0.113674	0.296264	0.7673
D(GLSRESID(-1))	-1.999419	0.129321	-15.46090	0.0000
D(GLSRESID(-2))	-2.838029	0.174334	-16.27927	0.0000
D(GLSRESID(-3))	-3.646868	0.237858	-15.33210	0.0000
D(GLSRESID(-4))	-4.288386	0.310007	-13.83319	0.0000
D(GLSRESID(-5))	-4.773125	0.376165	-12.68891	0.0000
D(GLSRESID(-6))	-5.009739	0.428606	-11.68845	0.0000
D(GLSRESID(-7))	-5.103760	0.458505	-11.13131	0.0000
D(GLSRESID(-8))	-5.126680	0.473567	-10.82568	0.0000
D(GLSRESID(-9))	-5.088008	0.479109	-10.61974	0.0000
D(GLSRESID(-10))	-4.979198	0.474847	-10.48589	0.0000
D(GLSRESID(-11))	-4.633846	0.467822	-9.905148	0.0000
D(GLSRESID(-12))	-3.815497	0.451629	-8.448296	0.0000
D(GLSRESID(-13))	-2.834131	0.412145	-6.876532	0.0000
D(GLSRESID(-14))	-1.885650	0.348008	-5.418412	0.0000
D(GLSRESID(-15))	-1.116288	0.258699	-4.315011	0.0000
D(GLSRESID(-16))	-0.576083	0.162041	-3.555156	0.0004
D(GLSRESID(-17))	-0.167518	0.073498	-2.279223	0.0234

R-squared	0.846948	Mean dependent var	46.79931
Adjusted R-squared	0.837347	S.D. dependent var	41614.42
S.E. of regression	16783.22	Akaike info criterion	22.35441
Sum squared resid	7.63E+10	Schwarz criterion	22.58277
Log likelihood	-3212.212	Hannan-Quinn criter.	22.44591
Durbin-Watson stat	2.020049		

Figure 45: DF-GLS unit root test of the 2nd difference of KING_PMED

Null Hypothesis: D(KING_PMED,2) is stationary
Exogenous: Constant
Bandwidth: 5 (Used-specified) using Quadratic Spectral kernel

		LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic		0.025817
Asymptotic critical values*:	1% level	0.739000
	5% level	0.463000
	10% level	0.347000

*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)

Residual variance (no correction)	5.58E+08
HAC corrected variance (Quadratic Spectral kernel)	74617861

KPSS Test Equation
Dependent Variable: D(KING_PMED,2)
Method: Least Squares
Date: 08/08/23 Time: 13:58
Sample (adjusted): 1997M10 2023M04
Included observations: 307 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	130.9446	1350.028	0.096994	0.9228

R-squared	0.000000	Mean dependent var	130.9446
Adjusted R-squared	0.000000	S.D. dependent var	23654.40
S.E. of regression	23654.40	Akaike info criterion	22.98374
Sum squared resid	1.71E+11	Schwarz criterion	22.99588
Log likelihood	-3527.004	Hannan-Quinn criter.	22.98859
Durbin-Watson stat	2.932309		

Figure 46: KPSS unit root test of the 2nd difference of KING_PMED

Null Hypothesis: D(KING_ACTLIST,1) has a unit root
Exogenous: Constant
Lag Length: 18 (Automatic - based on AIC, maxlag=18)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.182695	0.0008
Test critical values:	1% level	-3.452215	
	5% level	-2.871061	
	10% level	-2.571915	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(KING_ACTLIST,2)
Method: Least Squares
Date: 08/08/23 Time: 14:00
Sample (adjusted): 1998M08 2023M04
Included observations: 297 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
D(KING_ACTLIST(-1),1)	-0.634434	0.151681	-4.182695	0.0000
D(KING_ACTLIST(-1),2)	-0.097782	0.150702	-0.648840	0.5170
D(KING_ACTLIST(-2),2)	0.067940	0.148426	0.457739	0.6475
D(KING_ACTLIST(-3),2)	0.133894	0.145075	0.922927	0.3568
D(KING_ACTLIST(-4),2)	0.072466	0.141373	0.512589	0.6086
D(KING_ACTLIST(-5),2)	-0.080611	0.139694	-0.577054	0.5644
D(KING_ACTLIST(-6),2)	-0.085298	0.139421	-0.611799	0.5412
D(KING_ACTLIST(-7),2)	-0.024166	0.139814	-0.172841	0.8629
D(KING_ACTLIST(-8),2)	-0.030097	0.132481	-0.227176	0.8205
D(KING_ACTLIST(-9),2)	-0.065539	0.124443	-0.526661	0.5989
D(KING_ACTLIST(-10),2)	-0.128293	0.116941	-1.097072	0.2736
D(KING_ACTLIST(-11),2)	-0.108191	0.109944	-0.984059	0.3259
D(KING_ACTLIST(-12),2)	0.661117	0.102647	6.440667	0.0000
D(KING_ACTLIST(-13),2)	0.389386	0.100825	3.861998	0.0001
D(KING_ACTLIST(-14),2)	0.205426	0.093727	2.191754	0.0292
D(KING_ACTLIST(-15),2)	0.069933	0.087672	0.797657	0.4258
D(KING_ACTLIST(-16),2)	0.103865	0.081759	1.270387	0.2050
D(KING_ACTLIST(-17),2)	0.196514	0.074098	2.652089	0.0085
D(KING_ACTLIST(-18),2)	0.152056	0.060425	2.516419	0.0124
C	-9.234297	15.72034	-0.587411	0.5574

R-squared	0.795374	Mean dependent var	-0.983165
Adjusted R-squared	0.781338	S.D. dependent var	575.0107
S.E. of regression	268.8823	Akaike info criterion	14.09139
Sum squared resid	20026465	Schwarz criterion	14.34013
Log likelihood	-2072.571	Hannan-Quinn criter.	14.19097
F-statistic	56.66788	Durbin-Watson stat	2.004805
Prob(F-statistic)	0.000000		

Figure 47: ADF unit root test of first difference of KING_ACTLIST

Null Hypothesis: D(KING_ACTLIST,1) has a unit root
Exogenous: Constant
Lag Length: 14 (Automatic - based on AIC, maxlag=18)

		t-Statistic
Elliott-Rothenberg-Stock DF-GLS test statistic		-2.831441
Test critical values:	1% level	-2.572667
	5% level	-1.941881
	10% level	-1.615995

*MacKinnon (1996)

DF-GLS Test Equation on GLS Detrended Residuals
Dependent Variable: D(GLSRESID)
Method: Least Squares
Date: 08/08/23 Time: 14:00
Sample (adjusted): 1998M04 2023M04
Included observations: 301 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
GLSRESID(-1)	-0.283480	0.100119	-2.831441	0.0050
D(GLSRESID(-1))	-0.413668	0.108557	-3.810595	0.0002
D(GLSRESID(-2))	-0.258714	0.111485	-2.320616	0.0210
D(GLSRESID(-3))	-0.198855	0.112314	-1.770532	0.0777
D(GLSRESID(-4))	-0.197586	0.107831	-1.832362	0.0679
D(GLSRESID(-5))	-0.245249	0.102390	-2.395238	0.0173
D(GLSRESID(-6))	-0.255296	0.097607	-2.615539	0.0094
D(GLSRESID(-7))	-0.297281	0.093171	-3.190715	0.0016
D(GLSRESID(-8))	-0.283711	0.088335	-3.211768	0.0015
D(GLSRESID(-9))	-0.289652	0.084355	-3.433730	0.0007
D(GLSRESID(-10))	-0.326057	0.080188	-4.066152	0.0001
D(GLSRESID(-11))	-0.281058	0.076920	-3.653887	0.0003
D(GLSRESID(-12))	0.513822	0.072871	7.051082	0.0000
D(GLSRESID(-13))	0.241102	0.071537	3.370324	0.0009
D(GLSRESID(-14))	0.097925	0.059828	1.636767	0.1028

R-squared	0.781762	Mean dependent var	-0.481728
Adjusted R-squared	0.771079	S.D. dependent var	571.2176
S.E. of regression	273.3033	Akaike info criterion	14.10759
Sum squared resid	21362676	Schwarz criterion	14.29233
Log likelihood	-2108.192	Hannan-Quinn criter.	14.18152
Durbin-Watson stat	1.986420		

Figure 48: DF-GLS unit root test of 1st difference of KING_ACTLIST

Null Hypothesis: D(KING_ACTLIST,1) is stationary
Exogenous: Constant
Bandwidth: 5 (Used-specified) using Quadratic Spectral kernel

		LM-Stat.
Kwiatkowski-Phillips-Schmidt-Shin test statistic		0.045688
Asymptotic critical values*:	1% level	0.739000
	5% level	0.463000
	10% level	0.347000

*Kwiatkowski-Phillips-Schmidt-Shin (1992, Table 1)

Residual variance (no correction)	267764.2
HAC corrected variance (Quadratic Spectral kernel)	518529.2

KPSS Test Equation
Dependent Variable: D(KING_ACTLIST,1)
Method: Least Squares
Date: 08/08/23 Time: 14:00
Sample: 1997M08 2023M04
Included observations: 309

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-13.35275	29.48498	-0.452866	0.6510

R-squared	0.000000	Mean dependent var	-13.35275
Adjusted R-squared	0.000000	S.D. dependent var	518.2987
S.E. of regression	518.2987	Akaike info criterion	15.34221
Sum squared resid	82739140	Schwarz criterion	15.35429
Log likelihood	-2369.372	Hannan-Quinn criter.	15.34704
Durbin-Watson stat	1.198393		

Figure 49: KPSS unit root test of the first difference of KING_ACTLIST

Null Hypothesis: E_KING_PMED_1_RESIDS has a unit root
Exogenous: Constant
Lag Length: 13 (Automatic - based on AIC, maxlag=18)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.580752	0.0002
Test critical values:	1% level	-3.453652	
	5% level	-2.871693	
	10% level	-2.572253	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(E_KING_PMED_1_RESIDS)
Method: Least Squares
Date: 07/27/23 Time: 14:57
Sample (adjusted): 2000M02 2023M04
Included observations: 279 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
E_KING_PMED_1_RESIDS(-1)	-0.274712	0.059971	-4.580752	0.0000
D(E_KING_PMED_1_RESIDS(-1))	-0.260304	0.069534	-3.743564	0.0002
D(E_KING_PMED_1_RESIDS(-2))	-0.114478	0.064595	-1.772232	0.0775
D(E_KING_PMED_1_RESIDS(-3))	0.030726	0.064444	0.476787	0.6339
D(E_KING_PMED_1_RESIDS(-4))	0.201101	0.064049	3.139784	0.0019
D(E_KING_PMED_1_RESIDS(-5))	0.240871	0.065149	3.697239	0.0003
D(E_KING_PMED_1_RESIDS(-6))	0.249240	0.066731	3.734990	0.0002
D(E_KING_PMED_1_RESIDS(-7))	0.240005	0.069099	3.473342	0.0006
D(E_KING_PMED_1_RESIDS(-8))	0.211717	0.071025	2.980896	0.0031
D(E_KING_PMED_1_RESIDS(-9))	0.306296	0.072476	4.226171	0.0000
D(E_KING_PMED_1_RESIDS(-10))	0.186959	0.074742	2.501400	0.0130
D(E_KING_PMED_1_RESIDS(-11))	0.161469	0.073985	2.182473	0.0300
D(E_KING_PMED_1_RESIDS(-12))	-0.257339	0.070488	-3.650830	0.0003
D(E_KING_PMED_1_RESIDS(-13))	-0.182583	0.062752	-2.909605	0.0039
C	0.283896	0.211772	1.340570	0.1812

R-squared	0.370100	Mean dependent var	-0.048122
Adjusted R-squared	0.336696	S.D. dependent var	4.085351
S.E. of regression	3.327253	Akaike info criterion	5.294435
Sum squared resid	2922.641	Schwarz criterion	5.489662
Log likelihood	-723.5737	Hannan-Quinn criter.	5.372750
F-statistic	11.07957	Durbin-Watson stat	1.948633
Prob(F-statistic)	0.000000		

Figure 50: ADF test of residuals for KING_PMED regression showing evidence of stationarity

Null Hypothesis: E_KING_ACTLIST_1_RESIDS has a unit root
Exogenous: Constant
Lag Length: 15 (Automatic - based on AIC, maxlag=18)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.745811	0.0040
Test critical values:	1% level	-3.454443	
	5% level	-2.872041	
	10% level	-2.572439	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(E_KING_ACTLIST_1_RESIDS)
Method: Least Squares
Date: 07/19/23 Time: 17:08
Sample (adjusted): 1999M06 2023M04
Included observations: 270 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
E_KING_ACTLIST_1_RESIDS(-1)	-0.122767	0.032774	-3.745811	0.0002
D(E_KING_ACTLIST_1_RESIDS(-1))	0.261007	0.064192	4.066024	0.0001
D(E_KING_ACTLIST_1_RESIDS(-2))	0.152847	0.066120	2.311659	0.0216
D(E_KING_ACTLIST_1_RESIDS(-3))	0.222899	0.068751	3.242119	0.0013
D(E_KING_ACTLIST_1_RESIDS(-4))	-0.018197	0.064979	-0.280038	0.7797
D(E_KING_ACTLIST_1_RESIDS(-5))	0.102007	0.065377	1.560291	0.1199
D(E_KING_ACTLIST_1_RESIDS(-6))	-0.045523	0.068978	-0.659962	0.5099
D(E_KING_ACTLIST_1_RESIDS(-7))	0.014380	0.067473	0.213127	0.8314
D(E_KING_ACTLIST_1_RESIDS(-8))	0.109673	0.068306	1.605603	0.1096
D(E_KING_ACTLIST_1_RESIDS(-9))	0.341936	0.067921	5.034297	0.0000
D(E_KING_ACTLIST_1_RESIDS(-10))	0.025804	0.070399	0.366538	0.7143
D(E_KING_ACTLIST_1_RESIDS(-11))	0.057344	0.069969	0.819552	0.4132
D(E_KING_ACTLIST_1_RESIDS(-12))	-0.450806	0.073272	-6.152502	0.0000
D(E_KING_ACTLIST_1_RESIDS(-13))	-0.236394	0.076749	-3.080107	0.0023
D(E_KING_ACTLIST_1_RESIDS(-14))	0.150864	0.078011	1.933890	0.0542
D(E_KING_ACTLIST_1_RESIDS(-15))	0.202568	0.080639	2.512031	0.0126
C	0.027883	0.077214	0.361110	0.7183

R-squared	0.393148	Mean dependent var	-0.009468
Adjusted R-squared	0.354771	S.D. dependent var	1.567430
S.E. of regression	1.259057	Akaike info criterion	3.359496
Sum squared resid	401.0616	Schwarz criterion	3.586063
Log likelihood	-436.5320	Hannan-Quinn criter.	3.450476
F-statistic	10.24412	Durbin-Watson stat	2.027919
Prob(F-statistic)	0.000000		

Figure 51: ADF test of residuals for KING_PMED regression showing evidence of stationarity

Structural Regressions Output

Dependent Variable: @PCY(KING_PMED$)
Method: Least Squares
Date: 07/19/23 Time: 14:24
Sample (adjusted): 1998M12 2023M04
Included observations: 293 after adjustments
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 6.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
@PCY(KING_ACTLIST(-2))	-0.060400	0.013574	-4.449773	0.0000
@PCY(JOBS_NF)	0.891789	0.225157	3.960753	0.0001
@PCA(LOCALEQUITYWEALTH(-6))	0.009126	0.002602	3.507147	0.0005
@PCY(NPRIME_SHARE)	0.085979	0.019285	4.458456	0.0000
MRTGRT$-MRTGRT$(-6)	-0.804558	0.492015	-1.635230	0.1031

R-squared	0.445393	Mean dependent var	2.793517
Adjusted R-squared	0.437690	S.D. dependent var	8.136674
S.E. of regression	6.101474	Akaike info criterion	6.471856
Sum squared resid	10721.66	Schwarz criterion	6.534657
Log likelihood	-943.1268	Hannan-Quinn criter.	6.497008
Durbin-Watson stat	0.453546		

Figure 52: Machine output of final KING_PMED regression. No intervention variables.

Dependent Variable: @PCY(LOG(KING_ACTLIST))
Method: Least Squares
Date: 07/27/23 Time: 15:05
Sample (adjusted): 1998M02 2023M04
Included observations: 302 after adjustments
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 6.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
KING_SOLD(-1)/KING_ACTLIST(-1)	-3.160840	0.676417	-4.672914	0.0000
MRTGRT$-MRTGRT$(-6)	0.857099	0.565644	1.515261	0.1308
KING_PMED/INCPERCAP	0.254401	0.088666	2.869219	0.0044

R-squared	0.223219	Mean dependent var	-0.289334
Adjusted R-squared	0.218023	S.D. dependent var	4.660618
S.E. of regression	4.121360	Akaike info criterion	5.680127
Sum squared resid	5078.696	Schwarz criterion	5.716986
Log likelihood	-854.6992	Hannan-Quinn criter.	5.694875
Durbin-Watson stat	0.148708		

Figure 53: Machine output of final KING_ACTLIST regression. No intervention variables.

Dependent Variable: KING_PAVE$
Method: Least Squares
Date: 08/06/23 Time: 16:09
Sample: 1997M08 2023M04
Included observations: 309
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 6.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
KING_PMED$	1.224014	0.004692	260.8523	0.0000

R-squared	0.980565	Mean dependent var	848523.0
Adjusted R-squared	0.980565	S.D. dependent var	170463.6
S.E. of regression	23764.29	Akaike info criterion	22.99299
Sum squared resid	1.74E+11	Schwarz criterion	23.00507
Log likelihood	-3551.416	Hannan-Quinn criter.	22.99782
Durbin-Watson stat	0.691448		

Figure 54: Purposeful tautology regression to model average home price on median home price

Dependent Variable: @PCY(KING_PMED$)
Method: Least Squares
Date: 07/19/23 Time: 14:24
Sample (adjusted): 1998M12 2023M04
Included observations: 293 after adjustments
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 6.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
@PCY(KING_ACTLIST(-2))	-0.067039	0.012160	-5.513227	0.0000
@PCY(JOBS_NF)	0.614869	0.164909	3.728528	0.0002
@PCA(LOCALEQUITYWEALTH(-6))	0.003573	0.002419	1.477071	0.1408
@PCY(NPRIME_SHARE)	0.066427	0.021252	3.125674	0.0020
MRTGRT$-MRTGRT$(-6)	-0.669065	0.443876	-1.507323	0.1328
WINTER_11_12	2.303925	0.834853	2.759678	0.0062
GREATRECESSION	-7.641312	2.280072	-3.351347	0.0009
POSTCOVID_2YR	2.918002	1.962728	1.486707	0.1382

R-squared	0.525462	Mean dependent var	2.793517
Adjusted R-squared	0.513806	S.D. dependent var	8.136674
S.E. of regression	5.673508	Akaike info criterion	6.336416
Sum squared resid	9173.776	Schwarz criterion	6.436899
Log likelihood	-920.2850	Hannan-Quinn criter.	6.376661
Durbin-Watson stat	0.457113		

Figure 55: Robustness check: Machine output from median home price estimated regression with intervention variables

Dependent Variable: @PCY(LOG(KING_ACTLIST))
Method: Least Squares
Date: 07/19/23 Time: 16:11
Sample (adjusted): 1997M12 2023M04
Included observations: 304 after adjustments
HAC standard errors & covariance (Bartlett kernel, Newey-West fixed bandwidth = 6.0000)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
KING_SOLD(-1)/KING_ACTLIST(-1)	-2.561766	0.437108	-5.860721	0.0000
MRTGRT$-MRTGRT$(-3)	0.354306	0.276683	1.280549	0.2013
POSTCOVID_2YR	14.59369	2.001160	7.292613	0.0000

R-squared	0.490651	Mean dependent var	-0.302303
Adjusted R-squared	0.487266	S.D. dependent var	4.648031
S.E. of regression	3.328242	Akaike info criterion	5.252585
Sum squared resid	3334.237	Schwarz criterion	5.289266
Log likelihood	-795.3930	Hannan-Quinn criter.	5.267259
Durbin-Watson stat	0.245318		

Figure 56: Machine output of KING_ACTLIST regression with and intervention variable.

Actual-Fitted-Residual Charts

Figure 57: KING_PMED Actual-Fitted-Residual Chart

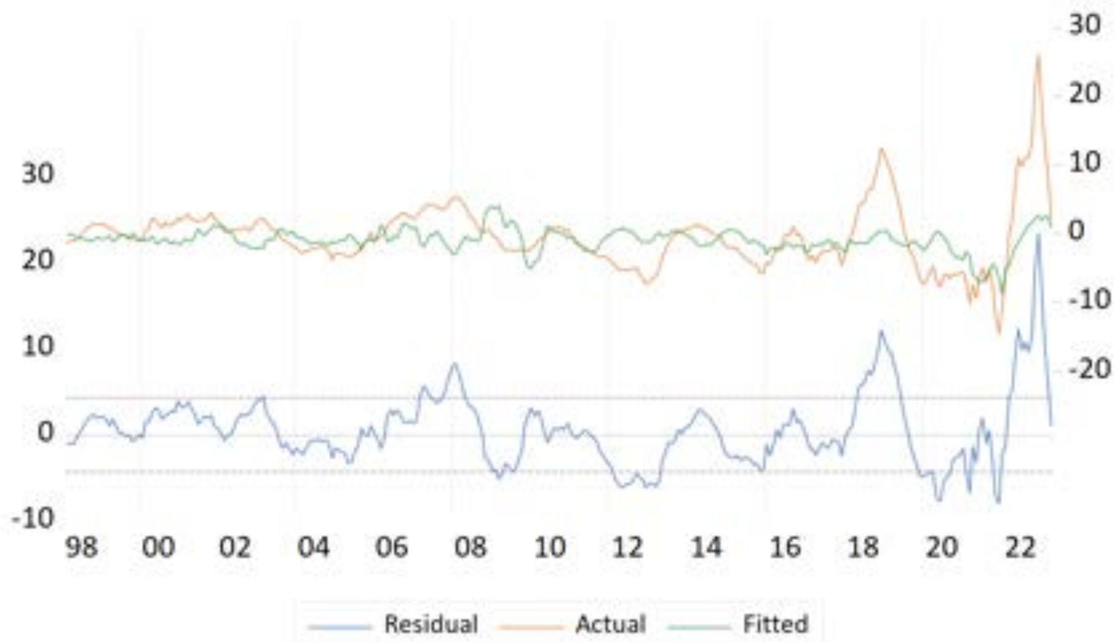


Figure 58: KING_ACTLIST Actual-Fitted Residual Chart

ADF Unit Root Tests for Residuals

Null Hypothesis: E_KING_PMED_1_RESIDS has a unit root
Exogenous: Constant
Lag Length: 13 (Automatic - based on AIC, maxlag=18)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-4.580752	0.0002
Test critical values:	1% level	-3.453652	
	5% level	-2.871693	
	10% level	-2.572253	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(E_KING_PMED_1_RESIDS)
Method: Least Squares
Date: 07/27/23 Time: 14:57
Sample (adjusted): 2000M02 2023M04
Included observations: 279 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
E_KING_PMED_1_RESIDS(-1)	-0.274712	0.059971	-4.580752	0.0000
D(E_KING_PMED_1_RESIDS(-1))	-0.260304	0.069534	-3.743564	0.0002
D(E_KING_PMED_1_RESIDS(-2))	-0.114478	0.064595	-1.772232	0.0775
D(E_KING_PMED_1_RESIDS(-3))	0.030726	0.064444	0.476787	0.6339
D(E_KING_PMED_1_RESIDS(-4))	0.201101	0.064049	3.139784	0.0019
D(E_KING_PMED_1_RESIDS(-5))	0.240871	0.065149	3.697239	0.0003
D(E_KING_PMED_1_RESIDS(-6))	0.249240	0.066731	3.734990	0.0002
D(E_KING_PMED_1_RESIDS(-7))	0.240005	0.069099	3.473342	0.0006
D(E_KING_PMED_1_RESIDS(-8))	0.211717	0.071025	2.980896	0.0031
D(E_KING_PMED_1_RESIDS(-9))	0.306296	0.072476	4.226171	0.0000
D(E_KING_PMED_1_RESIDS(-10))	0.186959	0.074742	2.501400	0.0130
D(E_KING_PMED_1_RESIDS(-11))	0.161469	0.073985	2.182473	0.0300
D(E_KING_PMED_1_RESIDS(-12))	-0.257339	0.070488	-3.650830	0.0003
D(E_KING_PMED_1_RESIDS(-13))	-0.182583	0.062752	-2.909605	0.0039
C	0.283896	0.211772	1.340570	0.1812

R-squared	0.370100	Mean dependent var	-0.048122
Adjusted R-squared	0.336696	S.D. dependent var	4.085351
S.E. of regression	3.327253	Akaike info criterion	5.294435
Sum squared resid	2922.641	Schwarz criterion	5.489662
Log likelihood	-723.5737	Hannan-Quinn criter.	5.372750
F-statistic	11.07957	Durbin-Watson stat	1.948633
Prob(F-statistic)	0.000000		

Figure 59: Unit root test for residuals of median home price structural regression

Null Hypothesis: E_KING_ACTLIST_1_RESIDS has a unit root
Exogenous: Constant
Lag Length: 15 (Automatic - based on AIC, maxlag=18)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-3.745811	0.0040
Test critical values:	1% level	-3.454443	
	5% level	-2.872041	
	10% level	-2.572439	

*MacKinnon (1996) one-sided p-values.

Augmented Dickey-Fuller Test Equation
Dependent Variable: D(E_KING_ACTLIST_1_RESIDS)
Method: Least Squares
Date: 07/19/23 Time: 17:08
Sample (adjusted): 1999M06 2023M04
Included observations: 270 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
E_KING_ACTLIST_1_RESIDS(-1)	-0.122767	0.032774	-3.745811	0.0002
D(E_KING_ACTLIST_1_RESIDS(-1))	0.261007	0.064192	4.066024	0.0001
D(E_KING_ACTLIST_1_RESIDS(-2))	0.152847	0.066120	2.311659	0.0216
D(E_KING_ACTLIST_1_RESIDS(-3))	0.222899	0.068751	3.242119	0.0013
D(E_KING_ACTLIST_1_RESIDS(-4))	-0.018197	0.064979	-0.280038	0.7797
D(E_KING_ACTLIST_1_RESIDS(-5))	0.102007	0.065377	1.560291	0.1199
D(E_KING_ACTLIST_1_RESIDS(-6))	-0.045523	0.068978	-0.659962	0.5099
D(E_KING_ACTLIST_1_RESIDS(-7))	0.014380	0.067473	0.213127	0.8314
D(E_KING_ACTLIST_1_RESIDS(-8))	0.109673	0.068306	1.605603	0.1096
D(E_KING_ACTLIST_1_RESIDS(-9))	0.341936	0.067921	5.034297	0.0000
D(E_KING_ACTLIST_1_RESIDS(-10))	0.025804	0.070399	0.366538	0.7143
D(E_KING_ACTLIST_1_RESIDS(-11))	0.057344	0.069969	0.819552	0.4132
D(E_KING_ACTLIST_1_RESIDS(-12))	-0.450806	0.073272	-6.152502	0.0000
D(E_KING_ACTLIST_1_RESIDS(-13))	-0.236394	0.076749	-3.080107	0.0023
D(E_KING_ACTLIST_1_RESIDS(-14))	0.150864	0.078011	1.933890	0.0542
D(E_KING_ACTLIST_1_RESIDS(-15))	0.202568	0.080639	2.512031	0.0126
C	0.027883	0.077214	0.361110	0.7183

R-squared	0.393148	Mean dependent var	-0.009468
Adjusted R-squared	0.354771	S.D. dependent var	1.567430
S.E. of regression	1.259057	Akaike info criterion	3.359496
Sum squared resid	401.0616	Schwarz criterion	3.586063
Log likelihood	-436.5320	Hannan-Quinn criter.	3.450476
F-statistic	10.24412	Durbin-Watson stat	2.027919
Prob(F-statistic)	0.000000		

Figure 60: Unit root test on the residuals of the KING_ACTLIST regression

VAR Machine Output

Roots of Characteristic Polynomial
Endogenous variables: D(KING_PMED$,1,12) D(KING_ACTLIST,1,12)
Exogenous variables: MRTGRT$ -MRTGRT$(-6) @PCY(INCPERCAP$) @PCY(JOBS_NF) POSTCOVID_2YR C
Lag specification: 1 18
Date: 07/04/23 Time: 14:28

Root	Modulus
0.954495 + 0.167158i	0.969021
0.954495 - 0.167158i	0.969021
0.223066 + 0.934997i	0.961238
0.223066 - 0.934997i	0.961238
0.746873 - 0.604365i	0.960769
0.746873 + 0.604365i	0.960769
-0.921160 + 0.264208i	0.958302
-0.921160 - 0.264208i	0.958302
0.905952 - 0.281040i	0.948542
0.905952 + 0.281040i	0.948542
-0.248692 - 0.898749i	0.932522
-0.248692 + 0.898749i	0.932522
-0.215470 - 0.906157i	0.931423
-0.215470 + 0.906157i	0.931423
-0.664629 - 0.636502i	0.920253
-0.664629 + 0.636502i	0.920253
0.658383 - 0.640653i	0.918643
0.658383 + 0.640653i	0.918643
-0.872438 + 0.274076i	0.914475
-0.872438 - 0.274076i	0.914475
-0.530447 + 0.701748i	0.879673
-0.530447 - 0.701748i	0.879673
0.345098 - 0.784805i	0.857328
0.345098 + 0.784805i	0.857328
-0.630709 + 0.571256i	0.850957
-0.630709 - 0.571256i	0.850957
0.128074 - 0.832653i	0.842446
0.128074 + 0.832653i	0.842446
-0.816673	0.816673
0.472607 + 0.600923i	0.764503
0.472607 - 0.600923i	0.764503
-0.296987 + 0.671886i	0.734596
-0.296987 - 0.671886i	0.734596
0.650883 + 0.318958i	0.724833
0.650883 - 0.318958i	0.724833
-0.682604	0.682604

No root lies outside the unit circle.
VAR satisfies the stability condition.

Figure 61: Stability test of VAR

VAR Lag Order Selection Criteria
Endogenous variables: D(KING_PMED$,1,12) D(KING_ACTLIST,1,12)
Exogenous variables: MRTGRT$-MRTGRT$(-6) @PCY(INCPERCAP$) @PCY(JOBS
Date: 07/04/23 Time: 14:28
Sample: 1997M08 2023M04
Included observations: 232

Lag	LogL	LR	FPE	AIC	SC	HQ
0	-4354.341	NA	7.50e+13	37.62363	37.77220	37.68355
1	-4319.573	67.43960	5.75e+13	37.35838	37.56638*	37.44227*
2	-4313.243	12.16868	5.64e+13	37.33830	37.60572	37.44615
3	-4312.147	2.088068	5.78e+13	37.36333	37.69018	37.49515
4	-4309.274	5.423133	5.84e+13	37.37305	37.75932	37.52883
5	-4307.594	3.143178	5.95e+13	37.39305	37.83875	37.57280
6	-4306.290	2.416781	6.09e+13	37.41629	37.92142	37.62000
7	-4304.169	3.895087	6.20e+13	37.43249	37.99704	37.66017
8	-4300.759	6.201628	6.23e+13	37.43758	38.06156	37.68922
9	-4293.646	12.81522	6.07e+13	37.41075	38.09415	37.68636
10	-4291.364	4.072848	6.16e+13	37.42555	38.16838	37.72513
11	-4282.569	15.54348	5.91e+13	37.38421	38.18647	37.70776
12	-4258.230	42.59245	4.96e+13	37.20888	38.07057	37.55639
13	-4249.621	14.91807	4.77e+13	37.16914	38.09026	37.54062
14	-4246.139	5.972661	4.80e+13	37.17361	38.15415	37.56905
15	-4245.657	0.818381	4.95e+13	37.20394	38.24391	37.62335
16	-4244.575	1.820225	5.08e+13	37.22909	38.32848	37.67246
17	-4243.437	1.892067	5.21e+13	37.25377	38.41259	37.72111
18	-4238.692	7.812854	5.18e+13	37.24735	38.46559	37.73865
19	-4235.471	5.249212	5.22e+13	37.25406	38.53173	37.76933
20	-4229.189	10.12686	5.13e+13	37.23439	38.57148	37.77362
21	-4227.777	2.250767	5.25e+13	37.25670	38.65323	37.81991
22	-4225.488	3.612192	5.34e+13	37.27145	38.72740	37.85862
23	-4212.770	19.84392	4.96e+13	37.19629	38.71167	37.80743
24	-4189.969	35.18517	4.23e+13*	37.03421*	38.60901	37.66931
25	-4189.563	0.619214	4.37e+13	37.06520	38.69942	37.72426
26	-4185.916	5.502263	4.39e+13	37.06824	38.76189	37.75127
27	-4180.415	8.203852	4.34e+13	37.05530	38.80838	37.76230
28	-4178.963	2.139588	4.45e+13	37.07727	38.88978	37.80824
29	-4175.631	4.855290	4.49e+13	37.08302	38.95496	37.83795
30	-4175.369	0.376995	4.65e+13	37.11525	39.04661	37.89415
31	-4170.683	6.664792	4.64e+13	37.10934	39.10013	37.91220
32	-4168.315	3.327258	4.72e+13	37.12341	39.17362	37.95024
33	-4166.720	2.214170	4.83e+13	37.14414	39.25378	37.99493
34	-4166.569	0.207277	5.01e+13	37.17732	39.34638	38.05208
35	-4161.822	6.424409	5.00e+13	37.17088	39.39937	38.06961
36	-4150.111	15.64796	4.70e+13	37.10441	39.39233	38.02710
37	-4148.740	1.809334	4.83e+13	37.12707	39.47441	38.07373
38	-4143.607	6.681348	4.80e+13	37.11730	39.52407	38.08793
39	-4140.726	3.700641	4.87e+13	37.12695	39.59315	38.12154
40	-4138.759	2.493116	4.99e+13	37.14447	39.67010	38.16303
41	-4137.784	1.218366	5.15e+13	37.17055	39.75560	38.21308
42	-4135.274	3.093642	5.24e+13	37.18340	39.82788	38.24989
43	-4133.712	1.898668	5.39e+13	37.20442	39.90832	38.29487
44	-4132.117	1.911900	5.53e+13	37.22514	39.98848	38.33957
45	-4122.753	11.05898*	5.32e+13	37.17890	40.00166	38.31729
46	-4122.232	0.605959	5.52e+13	37.20890	40.09108	38.37125
47	-4122.129	0.118449	5.75e+13	37.24249	40.18410	38.42881
48	-4117.276	5.479935	5.76e+13	37.23514	40.23618	38.44543
49	-4110.505	7.530184	5.67e+13	37.21125	40.27172	38.44551
50	-4106.925	3.920070	5.74e+13	37.21487	40.33476	38.47309
51	-4105.722	1.295616	5.93e+13	37.23899	40.41830	38.52117
52	-4103.540	2.314130	6.08e+13	37.25465	40.49340	38.56081
53	-4101.713	1.905210	6.26e+13	37.27339	40.57156	38.60351
54	-4092.960	8.979858	6.07e+13	37.23241	40.59001	38.58650
55	-4092.522	0.442078	6.33e+13	37.26312	40.68014	38.64117
56	-4084.670	7.783595	6.19e+13	37.22992	40.70637	38.63193
57	-4081.687	2.906152	6.32e+13	37.23868	40.77456	38.66466
58	-4074.833	6.558137	6.25e+13	37.21408	40.80939	38.66403
59	-4070.016	4.526931	6.29e+13	37.20703	40.86176	38.68095
60	-4059.871	9.357643	6.05e+13	37.15406	40.86822	38.65194

* indicates lag order selected by the criterion
LR: sequential modified LR test statistic (each test at 5% level)
FPE: Final prediction error
AIC: Akaike information criterion
SC: Schwarz information criterion
HQ: Hannan-Quinn information criterion

Figure 62: Lag length criteria of VAR

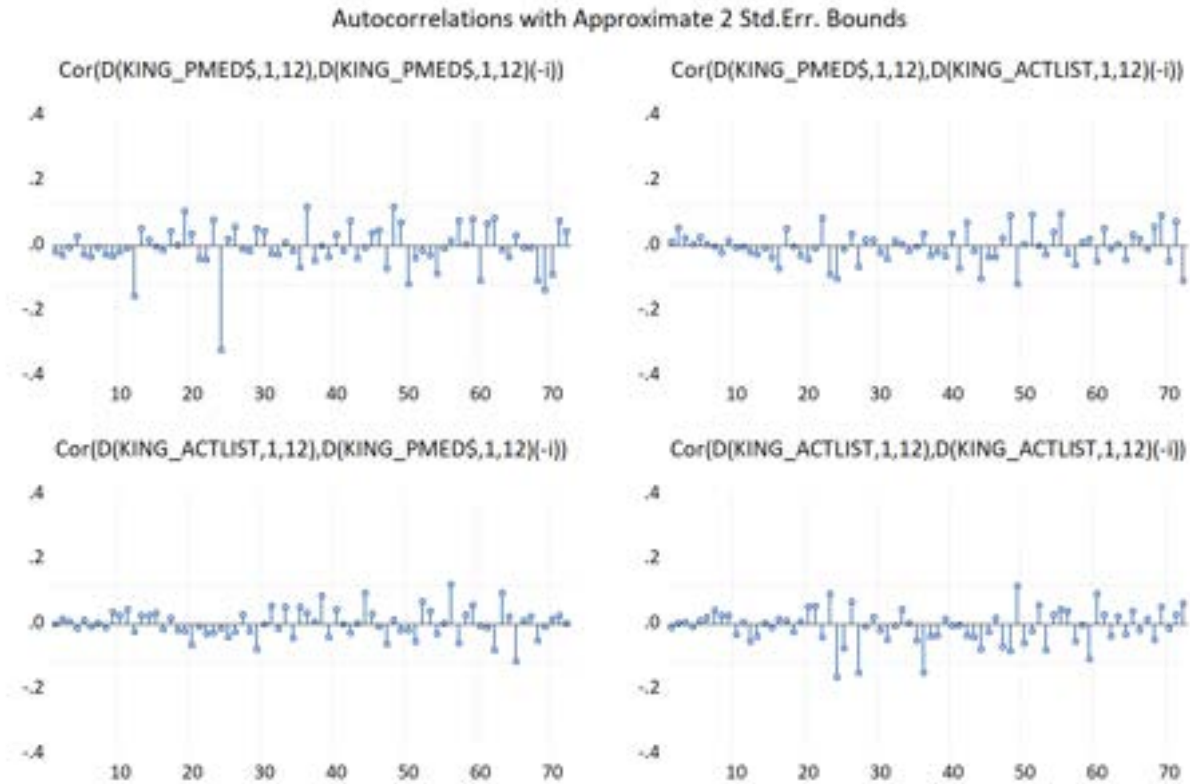


Figure 63: AC Lags of VAR

VAR Granger Causality/Block Exogeneity Wald Tests
Date: 07/20/23 Time: 14:26
Sample: 1997M08 2023M04
Included observations: 268

Dependent variable: D(KING_PMED$,1,12)

Excluded	Chi-sq	df	Prob.
D(KING_ACTLIST,1,12)	52.94090	24	0.0006
All	52.94090	24	0.0006

Dependent variable: D(KING_ACTLIST,1,12)

Excluded	Chi-sq	df	Prob.
D(KING_PMED$,1,12)	25.45519	24	0.3814
All	25.45519	24	0.3814

Figure 64: Block exogeneity test of VAR