

Национальный исследовательский университет

«Высшая школа экономики»

Факультет мировой экономики и мировой политики

Образовательная программа «Мировая экономика»

Отчет о самостоятельной работе

по дисциплине «Эконометрика» на тему

« Исследование рынка недвижимости в Соединенных Штатах Америки»

Выполнена:

Фам Тху Чанг, БМЭ204

Москва 2023

Оглавление

Введение	3
Part 1: Data description	4
Part 2: Analysis of time series' properties	6
Part 2.1: The mortgage rate time series	6
Part 2.2: The house price time series	10
Part 3: Decomposition of time series	14
Part 3.1: The mortgage rate time series	14
Part 3.2: The house price time series	16
Part 4: Autonomous dynamic time series models	23
Part 4.1: Dynamic model for mortgage rates	23
Part 4.2: Dynamic model for house prices	30
Conclusion	Error! Bookmark not defined.

Введение

Во всем мире рынок недвижимости всегда был занят потенциальными покупателями, ищущими доступные дома, в то время как владельцы предлагают свои объявления по соответствующим ценам. В США, за исключением крупнейших коммерческих городов, таких как Нью-Йорк, Лос-Анджелес или Чикаго, не редкость, когда граждане ищут дом в других городах США, которые им удобно добираться до работы или до школы для их детей. Однако после мирового финансового кризиса (МФК) 2007–2009 годов рынок недвижимости в США стал темой для людей во всем мире, учитывая, что именно он был корнем кризиса, который повернул США и весь мир в новом направлении. Теперь общеизвестно, что Великая рецессия с декабря 2007 года по июнь 2009 года, которая была следствием МФК, возникла из-за пузыря на рынке жилья. Обычно понятно, что банки и кредитные учреждения предлагают более низкие ставки по ипотеке, чтобы стимулировать потенциальных покупателей жилья брать у них деньги в долг. Однако в начале кризиса тот факт, что ставки по ипотеке были снижены до нереалистичных уровней, чтобы привлечь заемщиков, которые не смогли бы выплатить свои долги, привел банки к финансовому кризису, который затронул миллионы предприятий в США и, следовательно, по всему миру.

В этом исследовании мы стремимся изучить рынок недвижимости США с момента регистрации данных до мирового финансового кризиса (1987) до конца 2021 года, чтобы увидеть, как цены на жилье и ставки по ипотеке вели себя как в обычное время, так и во время хаоса. Объем нашего исследования, хотя и меньший, остается актуальным, поскольку США только что пережили еще один кризис, который также оказал влияние на рынок недвижимости — пандемию COVID-19. На наш взгляд, жизненно важно оглянуться на историю рынка и определить, как такие кризисы повлияли на показатели рынка жилья, в нашем случае: цены на жилье и ставки по ипотеке. В более широком плане мы стремимся извлечь из нашего исследования полезные выводы как для владельцев недвижимости, так и для тех, кто ищет жилье, чтобы быть в курсе потенциальных рисков на рынке и быть готовыми к заключению лучших сделок.

Для достижения цели в нашем исследовании мы поставили перед собой следующие задачи:

- проанализировать выбранный временной ряд цен на жилье и ставки по ипотеке на основе общих характеристик (автокорреляция, стационарность, тренд, сезонность);
- обнаружить и устранить тренд и сезонность временного ряда;
- выбрать оптимальную модель для прогнозирования будущих значений;
- сформулировать гипотезы и подтвердить их на основе результатов каждой части.

Основная часть нашего исследования состоит из 4 частей, которые в основном соответствуют вышеуказанным целям: часть 1 описывает данные в целом; часть 2 анализирует каждый выбранный временной ряд; часть 3 разлагает временной ряд на тренд, сезонность и остатки; а часть 4 фокусируется на построении оптимальной модели прогноза.

Часть 1: Описание данных

1. Описание данных

No.	Характеристика объекта	Название переменной	Периодичность и количество временных точек	Единицы измерения
1	Проценты, взимаемые по ипотечным кредитам в США	mortgage	Данные собираются ежемесячно (с января по декабрь) в течение 35 лет с 1987 по 2021 год. Таким образом, в ряду имеется $35 * 12 = 420$ временных точек.	процент
2	Цена на жилье в США, рассчитанная по базовому периоду (01.01.2000)	house		индекс*

* Цена дома на 01 января 2000 года является базовой стоимостью (100%), остальные рассчитываются как: $(\text{текущая стоимость} / \text{базовая стоимость}) * 100\%$

2. Источник набора данных:

“U.S. Housing Market Factors” на Kaggle.

URL: https://www.kaggle.com/datasets/faryarmemon/usa-housing-market-factors?select=Monthly_Macroeconomic_Factors.csv

3. Описание выбранных временных рядов

В среднем американцы, желающие купить новый дом, должны были бы занять определенную сумму денег в банке, чтобы купить его. Банки взимали бы с заемщиков определенную сумму денег на основе ставок по ипотеке. Проще говоря, поскольку ставки по ипотеке влияют на спрос людей на жилье, они также были бы связаны с ценами на жилье.

4. Основные гипотезы

Гипотеза 1: В течение 12 месяцев цены на жилье имеют сезонный характер, поскольку они имеют тенденцию снижаться в зимние месяцы и повышаться в летние месяцы.

Обычно домохозяйствам удобнее переезжать летом по следующим причинам:

- благоприятные погодные условия (солнечно и более длинный световой день);
- семьи с детьми, которые находятся на летних каникулах, могут обустроиться до начала нового учебного года осенью;
- особенно в США период каникул с ноября по январь — это напряженное время для семей, которые рассматривают возможность переезда..

Поэтому, поскольку спрос на дома в летние месяцы растет, цены на них также растут.

Гипотеза 2: В течение двух признанных периодов экономических рецессий ставки по ипотечным кредитам испытывали нестабильность с общей тенденцией к снижению.

В сложные экономические времена правительство часто снижает процентные ставки, чтобы помочь людям покупать больше и стимулировать социальные расходы, включая покупки на рынке недвижимости, что приводит к снижению ставок по ипотеке. Кроме того, это также может удержать рынок от замораживания под давлением рецессии.

Гипотеза 3: В течение двух признанных периодов экономических рецессий цены на жилье имели тенденцию к снижению.

Во время Великой рецессии с декабря 2007 по июнь 2009 года в результате мирового финансового кризиса (МФК) спрос на недвижимость снизился; как правило, цены на жилье также снизились. Поэтому цены на жилье во время рецессии, вызванной пандемией COVID-19 с января 2020 года, также имели тенденцию к снижению.

Часть 2: Анализ свойств временных рядов

Часть 2.1: Временной ряд ставок по ипотеке (*mortgage*)

1. Общий график

В целом, из первого наблюдения видно, что временной ряд имеет тенденцию к снижению.

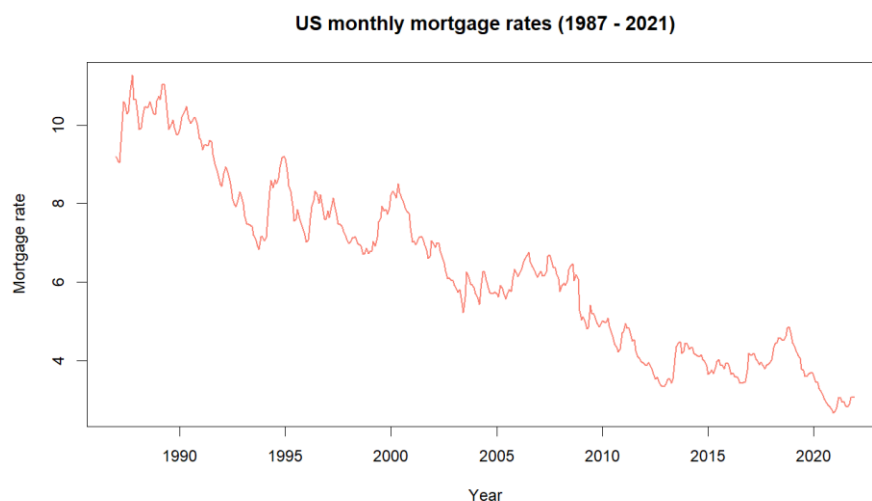


Рисунок 1: График ежемесячных изменений *mortgage rates* в США (1987–2021 гг.)

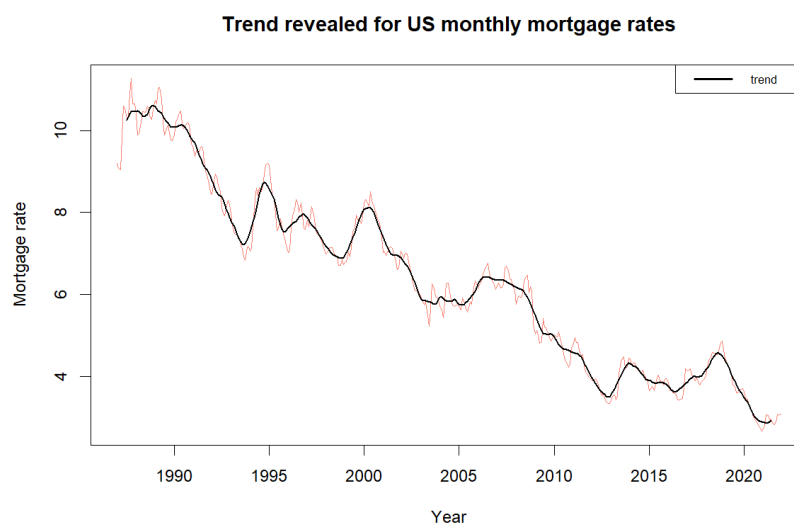


Рисунок 2: График тенденции изменения ежемесячной *mortgage rates* в США (1987–2021 гг.)

На ставку ипотеки сильно влияет процентная ставка, которая является одним из инструментов ФРС для контроля экономического роста и борьбы с инфляцией. Заметно, что в 1987 - 1991 годах ставки ипотеки почти всегда оставались выше 10%. Это можно объяснить тем, что в 1987 году уровень инфляции в США вырос до 4%, ФРС действовала, повышая ставки по фондам (с 5,8% в октябре 1986 года до 7,2% в октябре 1987 года)¹. За этим последовали ипотечные кредиты. Аналогичные события происходили в 1995-х и 2000-х годах. В это время

¹ Jonathan L., 2022. 1987: When mortgage rates last soared this much. *The Orange County Register*. Retrieved from: <https://www.ocregister.com/2022/06/18/1987-when-mortgage-rates-last-soared-this-much/>

быстрые темпы экономического роста сопровождались высоким риском повышения цен на товары и услуги, что привело к повышению процентных ставок ФРС.

После 2000-х годов ставки по ипотеке стали демонстрировать признаки больших колебаний. Относительно высокие ставки по ипотеке, вызванные субстандартной ипотекой² (ипотечные кредиты выдаются без первоначального взноса и без подтверждения дохода) вызвали крах рынка жилья, так как многие заемщики обнаружили, что стоимость их домов ниже ипотеки, и объявили дефолт. Снижение ипотечных ставок в 2008-2009 годах было попыткой ФРС вмешаться и облегчить ситуацию³.

После рецессии в 2013 году ФРС поверила, что экономика в конечном итоге станет достаточно сильной, чтобы справиться с сокращением стимулирования, и повысила процентные ставки⁴, за этим последовал рост ставок по ипотеке. В 2018 году ФРС снова подняла процентные ставки, чтобы бороться с инфляцией⁵. После 2019 года ставки снизились в ответ на пандемию коронавируса, однако с 2021 года, как только экономика начала восстанавливаться, ставки по ипотеке начали расти.

Гипотеза 2: В течение двух признанных периодов экономических рецессий ставки по ипотечным кредитам испытывали нестабильность с общей тенденцией к снижению.

Мы видим, что в период мирового финансового кризиса 2007-2009 гг. и пандемии COVID-19 в 2020 г. ставки по ипотеке, хотя и испытали некоторые потрясения, но в целом имели тенденцию к снижению. Мы уже объясняли выше, что в эти периоды ставки по ипотеке имеют тенденцию к снижению в соответствии с политикой правительства. Поэтому у нас есть основания подтвердить нашу 2-ю гипотезу.

Гипотеза 2: Подвергается

Из этого графика мы не видим никаких признаков сезонности или циклического движения. В то же время видно, что средние ставки по ипотеке и дисперсия ставок по ипотеке не являются постоянными, а меняются со временем. Поэтому мы можем временно предположить, что ряд данных нестационарен.

2. Автокорреляция

² John D.V. Subprime Mortgage Crisis. *Federal Reserve History*. Retrieved from: <https://www.federalreservehistory.org/essays/subprime-mortgage-crisis>

³ Robert R. The Great Recession. *Federal Reserve History*. Retrieved from: <https://www.federalreservehistory.org/essays/great-recession-of-200709>

⁴ Ylan Q.M., 2013. Interest rate spike emerges as concern for Fed. *Washington Post*. Retrieved from: https://www.washingtonpost.com/business/economy/interest-rate-spike-emerges-as-concern-for-fed/2013/06/26/56c521e0-ddac-11e2-948c-d644453cf169_story.html

⁵ Jeff C., 2022. Fed hikes interest rates by 0.75 percentage point for second consecutive time to fight inflation. *CNBC*. Retrieved from: <https://www.cnbc.com/2022/07/27/fed-decision-july-2022-.html>

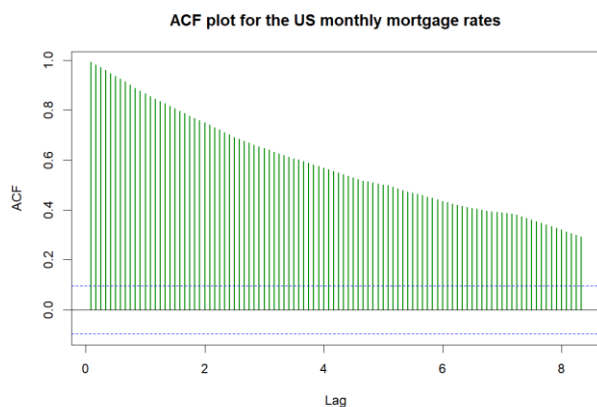


Рисунок 3: График ACF для *mortgage rates*

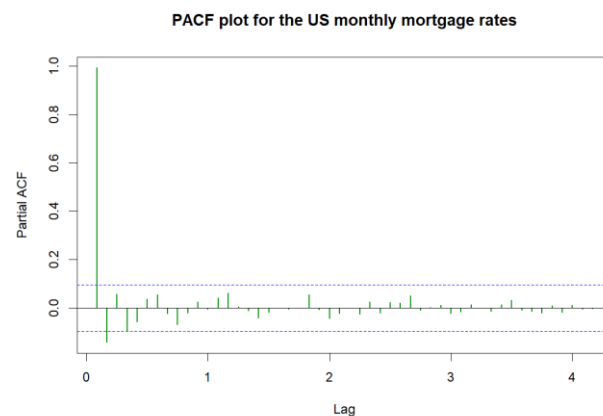


Рисунок 4: PACF для *mortgage rates*

График автокорреляции постепенно уменьшается со всеми пиками выше нулевой границы значимости, что показывает, что цены на жилье имеют высокую автокорреляцию с предыдущим периодом времени и, следовательно, это не белый шум. График PACF показывает, что мы должны учитывать только те первые 2 лага, которые имеют пики за пределами линии значимости, потому что после лага 2 PACF сужается до 0.

В целом, существует тенденция, поскольку график ACF постепенно уменьшается, и нет никаких признаков сезонности, поскольку нет видимых колебаний. Рисунок 4: График в любой период времени.

3. Периодограмма

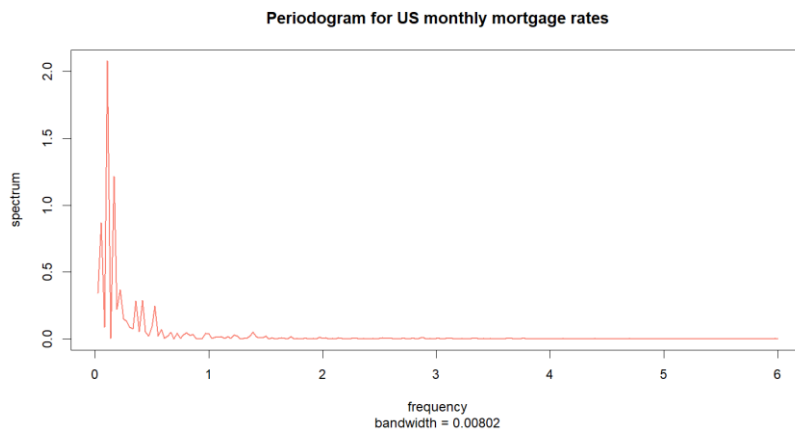


Рисунок 5: Периодограмма для *mortgage rates*

Из периодограммы видно, что доминирующий пик находится вблизи нулевой частоты. График показывает несколько высоких волн вокруг пика, что указывает на отсутствие четкой тенденции. Нечеткая тенденция также отражает флуктуацию временного ряда из общего графика. Кроме того, линия становится почти полностью плоской после интенсивных волн вблизи 0 с чрезвычайно малыми волнами.

4. Augmented Dickey-Fuller (ADF) тест

H_0 : Временной ряд нестационарен.

H₁: Временной ряд является стационарным.

Результат:

Dickey-Fuller статистика: -3.5918

p-значение: 0.03379

Таким образом, мы отвергаем нулевую гипотезу, что означает, что временной ряд ставок по ипотеке является стационарным.

Стационарность временного ряда ставок по ипотеке означает, что средняя ставка по ипотеке, а также разброс ставок по ипотеке не меняются со временем. Однако, поскольку мы все еще можем четко наблюдать тенденцию к снижению, мы считаем, что это связано с тем, что ставки по ипотеке имеют небольшие значения от 1% до 13%, и поэтому тест ADF считает изменения ставок по ипотеке незначительными.

5. Заключение

Ставки по ипотеке имеют высокую автокорреляцию по месяцам, что означает, что ставка по ипотеке месяца зависит от ставки(ей) по ипотеке предыдущего месяца(ов).

В целом, ставки по ипотеке в США имеют тенденцию к снижению от начала до конца периода. Поскольку цены на жилье имеют тенденцию расти с течением времени, ставки по ипотеке предлагались ниже, чтобы побудить людей покупать дома. В сложные экономические времена ставки по ипотеке, как правило, снижаются еще больше, чтобы побудить людей брать кредиты и тратить больше. Однако сразу после того, как экономика начала расти, например, в период после пандемии, ставки по ипотеке снова показали признаки роста.

Часть 2.2: Временной ряд цен по жилью

1. Общий график

В целом, из первого наблюдения видно, что временной ряд имеет тенденцию к росту.

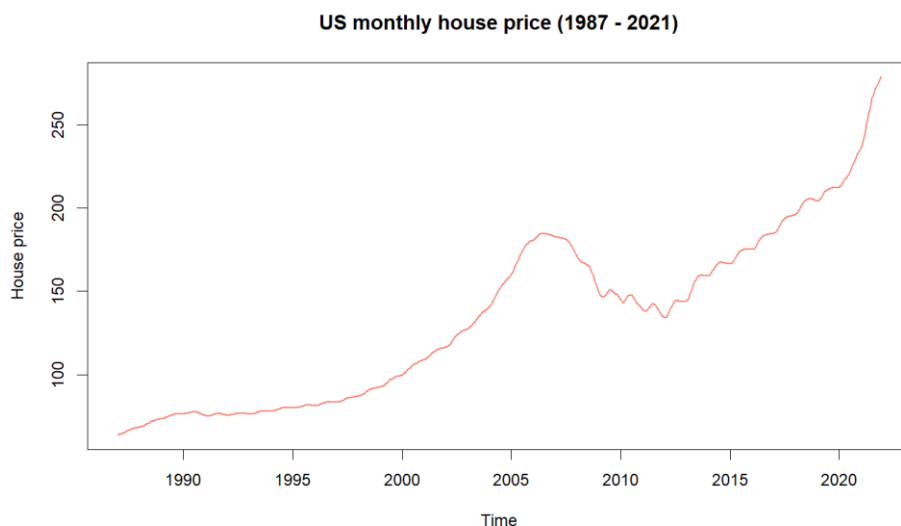


Рисунок 6: График ежемесячных изменений *house price* в США (1987–2021 гг.)

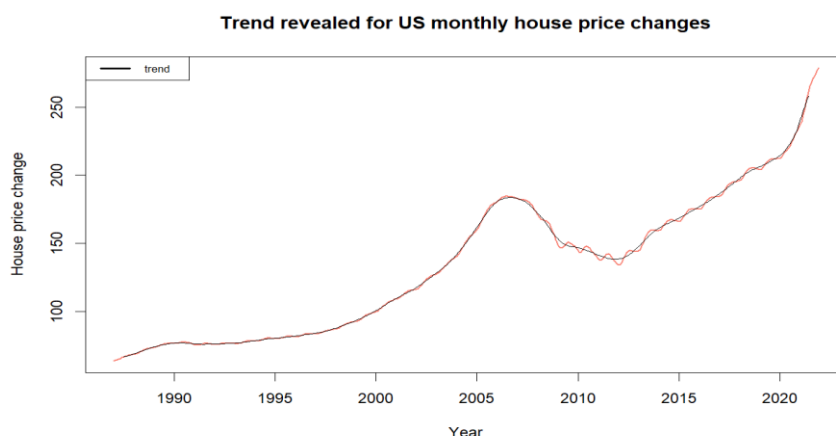


Рисунок 7: График тенденции ежемесячных изменений *house price* в США (1987–2021 гг.)

С 1987 по 2000 год цены на жилье постепенно росли. С 2000 года они продолжали расти, но более быстрыми темпами, пока не достигли пика в 2006 году. Из-за мирового финансового кризиса 2007–2009 годов цены на жилье начали снижаться в результате рецессии и вернулись к предыдущей положительной тенденции только примерно через 4 года после кризиса. С 2009 по 2020 год восходящая тенденция цен на жилье в США была не такой «плавной» (с большими колебаниями), как раньше. Также можно увидеть, что с 2020 года цены на жилье росли быстрее своего колеблющегося тренда, что можно объяснить инфляцией, вызванной пандемией COVID-19.

На этом графике мы не видим никаких признаков сезонности или циклического движения, что является для нас признаком того, что следует отвергнуть нашу первую гипотезу о потенциальной сезонности цен на жилье.

В то же время видно, что средняя цена на жилье и дисперсия цен на жилье не были постоянными, а менялись с течением времени. Поэтому мы можем временно предположить, что временной ряд не является стационарным.

Более того, нетрудно заметить, что поведение цен на жилье различается в течение двух признанных периодов рецессии. В то время как цены на жилье упали в рецессию 2007–2009 годов, в рецессию COVID-19 они фактически выросли, а не упали, как мы ожидали при формулировании нашей 3-й гипотезы. Поэтому у нас есть основания отвергнуть 3-ю гипотезу.

Гипотеза 3: В течение двух признанных периодов экономических рецессий цены на жилье имели тенденцию к снижению.

В поисках ответа на наши ложные предположения нам удалось найти три основные причины:

- Во-первых, ставки по ипотеке в период рецессии COVID-19 были ниже, чем в период Великой рецессии (3% по сравнению с 6%), что помогло еще больше увеличить спрос на рынке недвижимости.
- Во-вторых, хотя снижение ставок по ипотеке в 2008 году также способствовало увеличению спроса, оно также увеличило количество новых домов, построенных для продажи. Поскольку новых списков домов становилось все больше (предложение превышало реальный спрос), цены на жилье начали быстро падать в 2008 году. Напротив, в период COVID-19 наблюдался высокий спрос на жилье и более низкое предложение на рынке, что привело к росту цен.
- Наконец, из-за опыта пузыря на рынке недвижимости в 2008 году рынок недвижимости стал более жестко регулироваться правительством США, чтобы избежать очередного краха⁶.

Гипотеза 3: Отвергается.

2. Автокорреляция

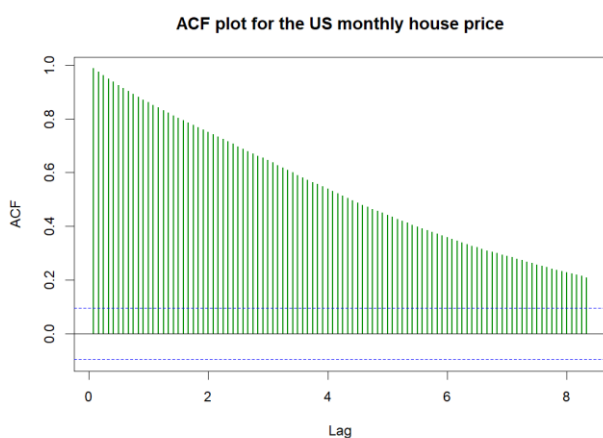


Рисунок 8: График ACF для *house price*

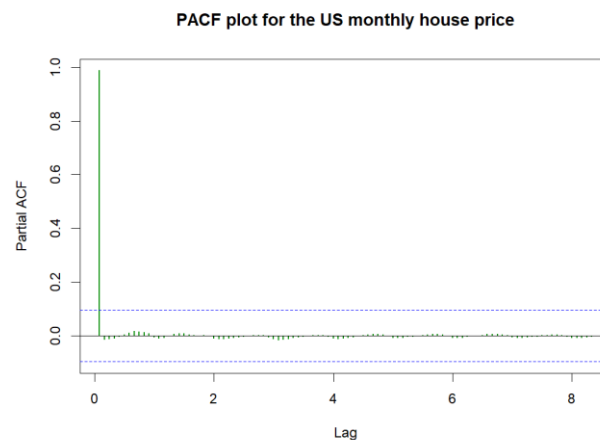


Рисунок 9: График PACF для *house price*

⁶ Zumwalt, W. (November 15, 2021). "The Difference in the 2008 and 2021 Housing Bubble". <https://storymaps.arcgis.com/stories/ca5fe58da85240a19c4cfc144cd28552>

График автокорреляции постепенно уменьшается со всеми пиками выше нулевой границы значимости, что показывает, что цены на жилье имеют высокую автокорреляцию с предыдущим периодом и, следовательно, это не белый шум. График частичной автокорреляции показывает, что мы должны учитывать только один лаг, пик которого выше линии значимости.

В целом, существует тенденция, поскольку график ACF постепенно уменьшается, и нет никаких признаков сезонности, поскольку нет видимых колебаний в любой период времени.

3. Периодограмм

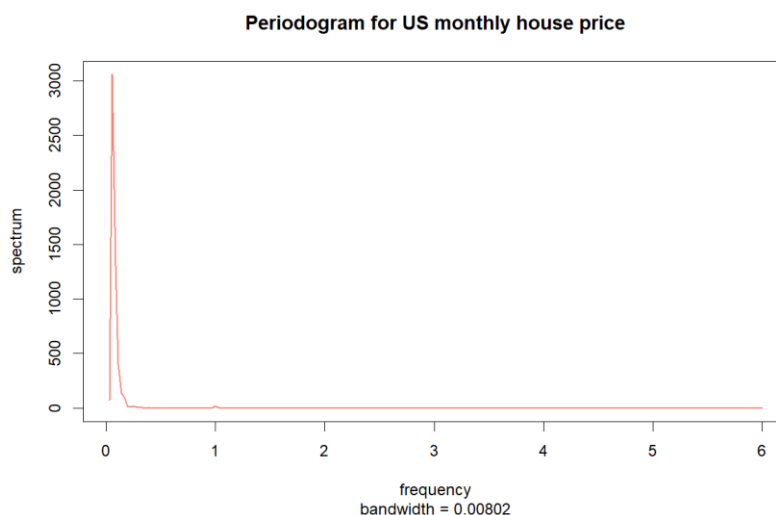


Рисунок 10: Периодограмма для *house price*

Из периодограммы мы видим, что есть только один пик вблизи нулевой частоты. Это говорит о том, что есть существенное изменение на одной волне, следовательно, есть общая тенденция. В то же время можно предположить нестационарную характеристику нашего временного ряда. Поскольку линия становится почти полностью плоской после очень низкого пика, когда частота равна 1, график показывает, что нет никаких мелких волн (почти плоских).

4. Augmented Dickey-Fuller (ADF) тест

H_0 : Временной ряд нестационарен.

H_1 : Временной ряд является стационарным.

Результат:

Dickey-Fuller статистика: -0.94345

p-значение: 0.9473

Мы не можем отвергнуть нулевую гипотезу, что означает, что временной ряд цен на жилье нестационарен.

Нестационарность временного ряда цен на жилье означает, что средняя цена на жилье, а также разброс цен на жилье меняются со временем. Возможным объяснением этого является то, что по мере роста экономики США с течением времени ее граждане стали более

способны покупать дома. Как правило, растущий спрос на рынке недвижимости приводил к росту цен на жилье.

5. Заключение

Цены на жилье сильно автокоррелированы по месяцам, что означает, что цена на жилье месяца зависит от цены(цен) на жилье предыдущего месяца(ов).

В целом, цены на жилье в США имеют тенденцию к росту с начала до конца периода. Однако мировой финансовый кризис 2007–2009 годов стал определяющим фактором, который впервые обозначил быстрое падение по сравнению с устойчивым ростом предыдущего периода. Более того, после этой глобальной рецессии закономерности роста цен на жилье не такие гладкие, как раньше, и, по сути, мировая рецессия полностью изменила динамику цен на жилье в США в следующем десятилетии.

Часть 3: Разложение временного ряда

Часть 3.1: Временной ряд ставок по ипотеке

Тенденция и сезонность

Как показано в части 2, наш временной ряд ставок по ипотеке не имеет сезонной характеристики. Поэтому мы могли бы пропустить процесс десезонализации (используя гармоническую регрессию) и сразу выполнить полиномиальную регрессию. Свойства регрессии показаны в таблице ниже:

Residuals:					
Min	1Q	Median	3Q	Max	
-1.77046	-0.37629	-0.03477	0.39445	1.51521	
Coefficients:					
			Estimate	Std. Error	t value Pr(> t)
(Intercept)			9.955e+03	1.311e+03	7.592 2.09e-13 ***
poly(index(Y_m), 2, raw = TRUE)1			-9.719e+00	1.308e+00	-7.427 6.31e-13 ***
poly(index(Y_m), 2, raw = TRUE)2			2.372e-03	3.264e-04	7.268 1.81e-12 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1					
Residual standard error: 0.6107 on 417 degrees of freedom					
Multiple R-squared: 0.9238, Adjusted R-squared: 0.9234					
F-statistic: 2526 on 2 and 417 DF, p-value: < 2.2e-16					

Таблица 1: Полиномиальная регрессия

Уравнение регрессии:

$$y = 9.955e+03 + 2.372e-03 * \text{индекс} + -9.9719e+00 * \text{индекс}^2$$

p-значение регрессии составляет $2,2e-16 < 0,05$, можно сделать вывод, что регрессия значима. p-значение коэффициентов составляет $2,09e-13$, $6,31e-13$, $1,81e-12 < 0,05$, можно сделать вывод, что все коэффициенты значимы.

Затем мы извлекли остатки из полиномиальной регрессии и получили следующий рисунок.

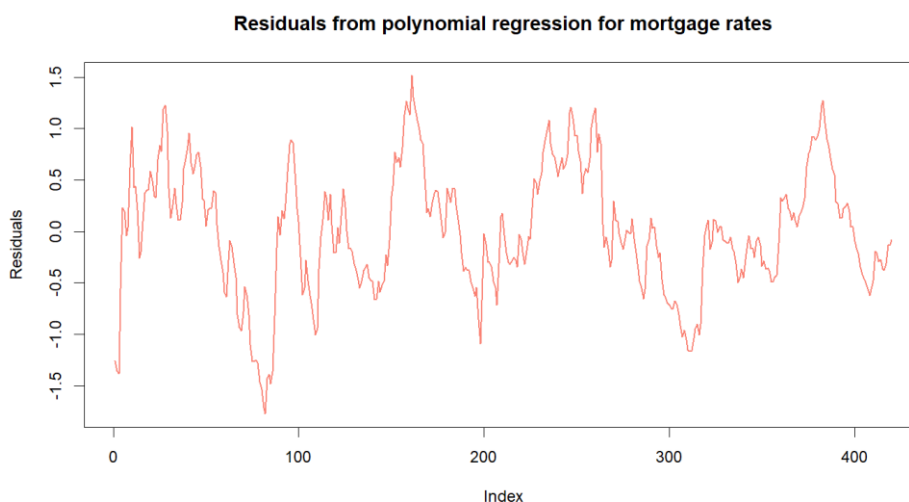


Рисунок 11: Остатки полиномиальной регрессии для *mortgage rates*

По сравнению с рисунками 1 и 2, этот рисунок не показывает никакой тенденции к снижению во временном ряду.

Augmented Dickey-Fuller (ADF) тест

H_0 : Временной ряд нестационарен.

H_1 : Временной ряд является стационарным.

Результат:

Dickey-Fuller статистика: -3.9163

p-значение: 0.01335

Мы провели тест ADF для остатков полиномиальной регрессии, и полученный результат подтверждает стационарность.

Наконец, мы сохранили данные для остатков полиномиальной регрессии, чтобы перейти к следующему этапу.

Часть 3.2: Временной ряд цен по жилью

1. Тренд

Чтобы удалить ранее наблюдавшийся восходящий тренд, мы использовали технику, называемую дифференциацией, для стабилизации среднего значения.

Однако после **первой дифференциации** наш график выявил наличие сезонности, которую нельзя было наблюдать в предыдущем анализе.

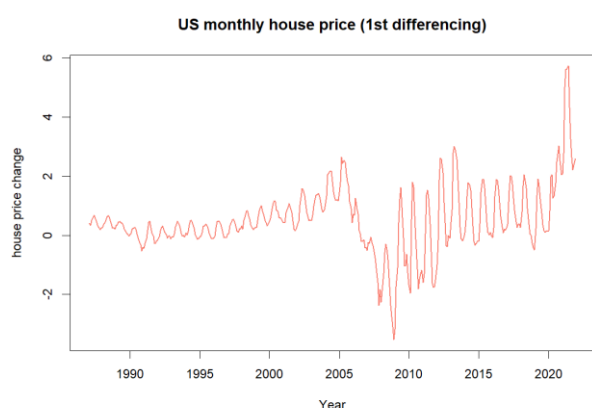


Рисунок 12: График ежемесячной *house price* в США (1987–2021 гг.) после первой дифференциации

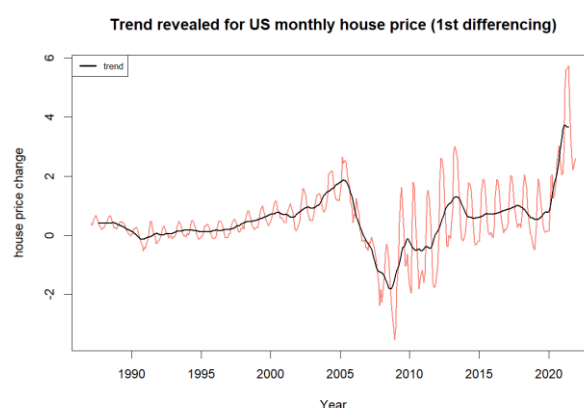


Рисунок 13: График тренда ежемесячной *house price* в США (1987–2021 гг.) после первой дифференциации

Как можно видеть, значение цены на жилье теперь ближе к 0. Подобно нашему предыдущему анализу, мировой финансовый кризис 2007–2009 гг. ознаменовал существенное изменение способа колебания цен на жилье. В течение этого периода цены на жилье колебались не по какой-либо повторяющейся схеме, но после кризиса цены на жилье начали колебаться более широко, чем до 2006 г.

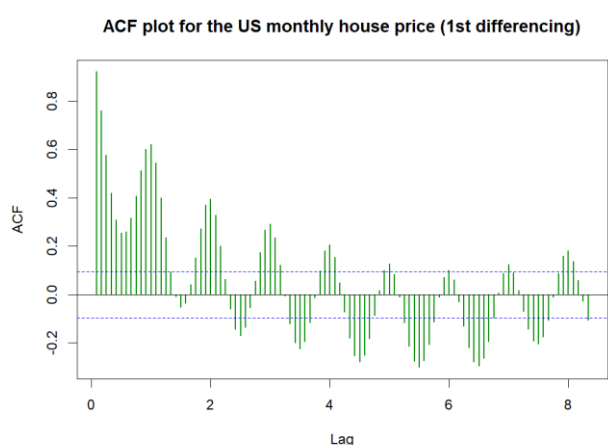


Рисунок 14: График ACF для *house price* после первой дифференциации

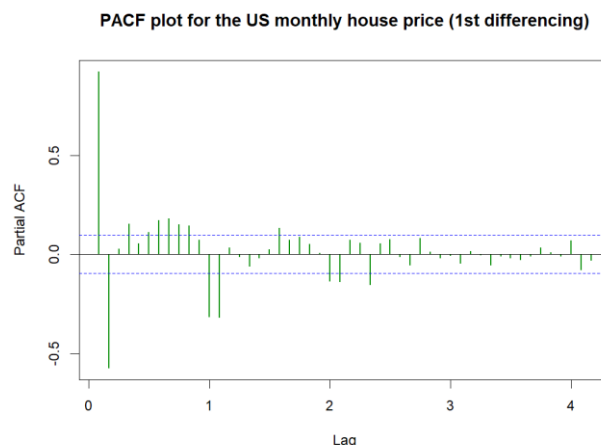


Рисунок 15: График PACF для *house price* после первой дифференциации

График ACF слева показывает, что волны следуют повторяющемуся шаблону после каждых 12 лагов, что означает, что существует сезонное изменение в течение года (12 месяцев). График PACF справа также показывает несколько значительных лагов перед резким обрывом после лага 12, что также доказывает годовую сезонность.

В предварительном анализе мы отвергли 1-ю гипотезу о наличии сезонности в ценах на жилье. Однако, благодаря тому, что показал график ACF, у нас есть основания все-таки подтвердить эту гипотезу.

Мы продолжили применять **дифференциацию во второй раз**. На этот раз мы можем быть уверены, что успешно удалили все оставшиеся признаки тренда, и то, что осталось, — это более четкая визуализация сезонного временного ряда.

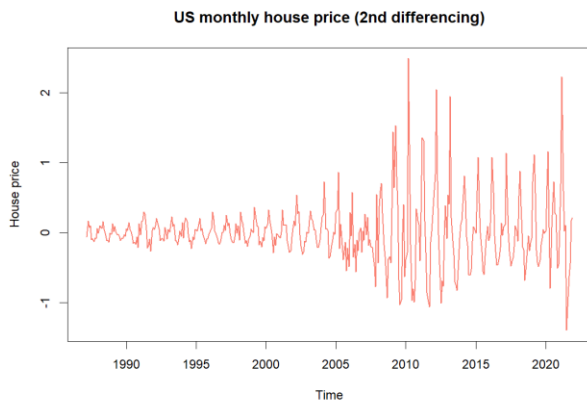


Рисунок 16: График ежемесячной *house price* в США (1987–2021 гг.) после второй дифференциации

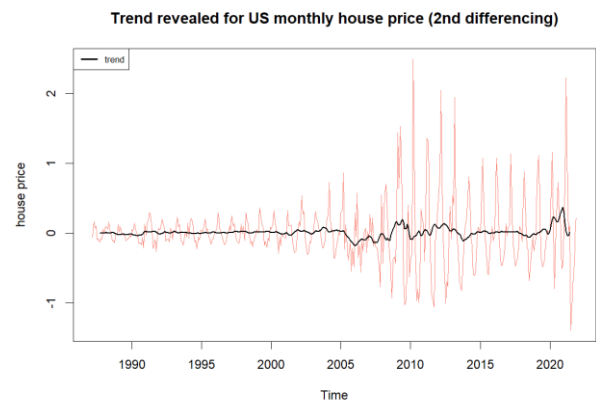


Рисунок 17: График тренда ежемесячной *house price* в США (1987–2021 гг.) после второй дифференциации

Самые большие волны приходятся на период с 2009 по 2013 год, поскольку рынок был затронут рецессией 2007–2009 годов и мерами правительства по стабилизации экономики в целом. Разница в диапазоне колебаний между периодом после рецессии и периодом до рецессии совершенно очевидна. Помимо того, что тренд теперь удален визуально, мы также отметили, что наш временной ряд имеет аддитивную сезонность, поскольку величина колебаний не увеличивается со временем, а резко увеличивается из-за мирового финансового кризиса 2007–2009 годов. Начиная с этого периода, диапазон колебаний становится уже, но все еще шире, чем в период до мирового финансового кризиса. В трех периодах (до мирового финансового кризиса, во время мирового финансового кризиса, после мирового финансового кризиса) величина колебаний стабильна и не усиливается со временем. Таким образом, мы подтвердили утверждение, сделанное в анализе части 2, о том, что глобальная рецессия 2007–2009 годов навсегда изменила рынок недвижимости.

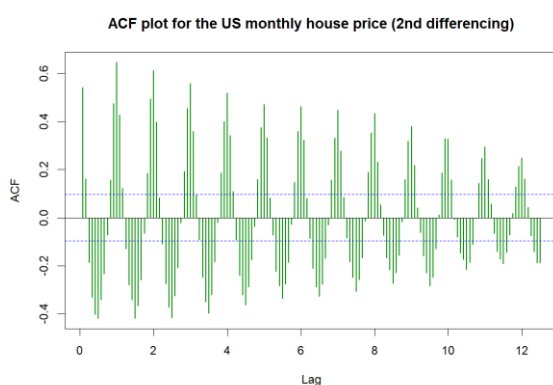


Рисунок 18: График ACF для *house price* после второй дифференциации

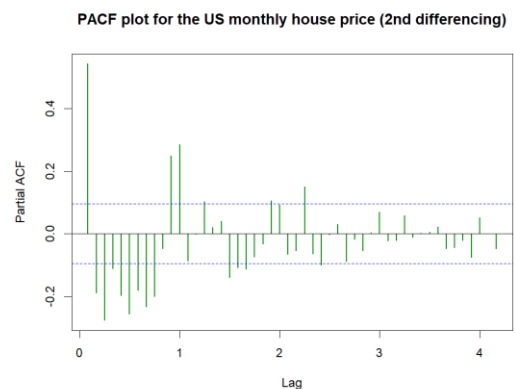


Рисунок 19: График PACF для *house price* после второй дифференциации

Поскольку сезонность проявляется более четко, можно заметить, что цены на жилье меняют свою модель каждые 6 месяцев, поскольку пики на графике ACF достигают пиков после каждых 12 лагов и возвращаются примерно к тому же уровню после каждых 12 лагов. Кроме того, ACF медленно затухает, что означает, что будущие значения цен на жилье сильно коррелируют с их прошлыми значениями. На графике PACF почти все первые 12 лагов значимы, а за лагом 12 следует падение значений PACF и приближение к нулю к концу.

Гипотеза 1: В течение года наблюдается сезонная динамика цен на жилье: они имеют тенденцию снижаться в зимние месяцы и расти в летние месяцы.

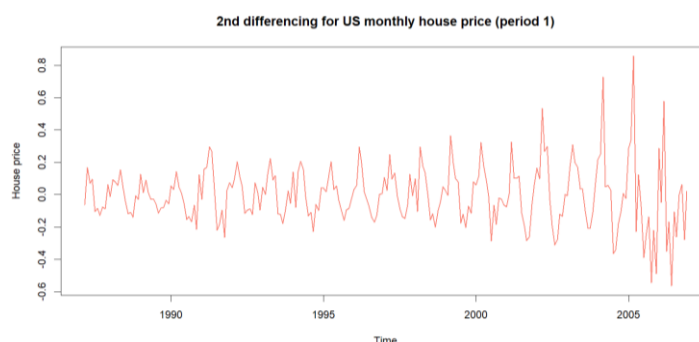


Рисунок 20: Общий график *house price* в период до мирового финансового кризиса (после второй дифференциации)

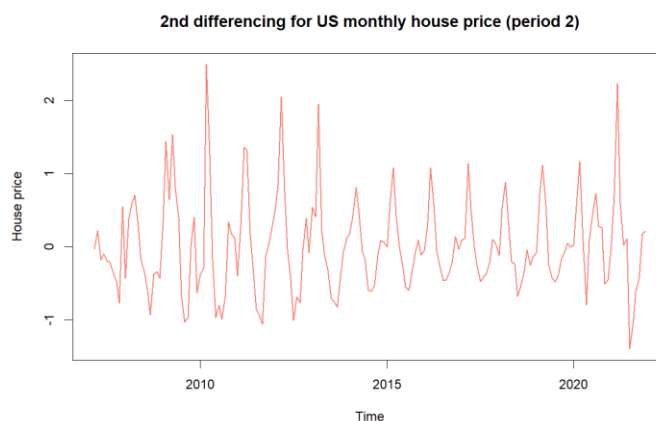


Рисунок 21: Общий график *house price* в период после мирового финансового кризиса (после второй дифференциации)

Действительно, мы разделили данные на два периода, чтобы наблюдать более внимательно. Период 1 (до глобальной финансовой рецессии) длится с 1987 по 2006 год, а период 2 (включая глобальную финансовую рецессию и рецессию COVID-19) длится с 2007 по 2021 год. Для обоих периодов мы заметили, что, хотя сезонность и существует, нет никаких доказательств, подтверждающих, что цены на жилье растут в летние месяцы и падают в зимние месяцы. Фактически, существует много периодов, когда цены на жилье достигают пика зимой, затем постепенно снижаются, пока не достигают дна летом, и снова начинают расти до зимы. Поэтому мы отвергли первую гипотезу.

Гипотеза 1: Отвергается.

2. Сезонность

Начиная с этого шага, мы решили применить два разных подхода к нашему исследованию цен на жилье. Для первого подхода мы провели спектральный анализ с гармонической регрессией, чтобы проверить наши предположения о сезонности. Для второго подхода мы применили 3-й раунд дифференциации, чтобы полностью исключить любые признаки сезонности.

Подход 1: Спектральный анализ

Теперь, когда восходящий тренд удален из нашего ряда, мы начинаем процесс устранения сезонности, проводя спектральный анализ, в котором применяем гармоническую регрессию.

Примечание: Причина, по которой мы не проводили регрессию по циклам для исходных данных, а только для данных, которые были дифференцированы дважды, заключается в том, что без дифференцирования наши данные не выявили признаков сезонности ни на общем графике, ни на графиках ACF.

Периодограмма

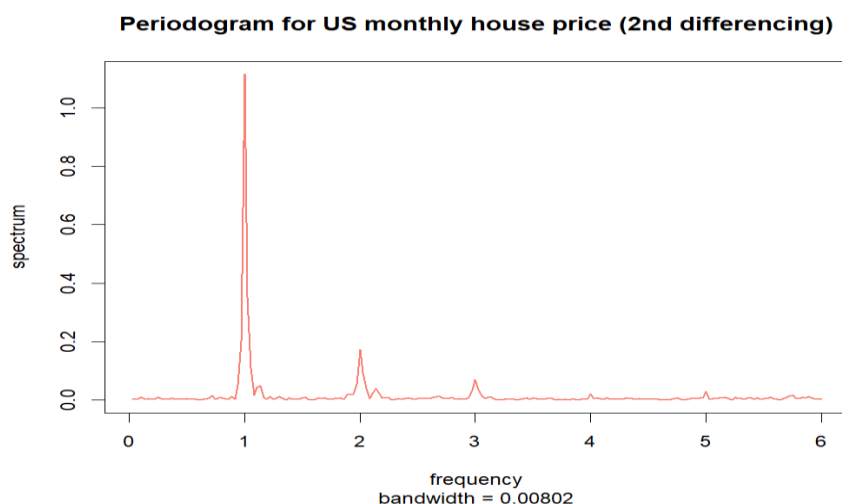


Рисунок 22: Периодограмма для *house price* (после второй дифференциации)

На графике выше мы видим очевидные пики спектрограмм, которые указывают на потенциальную сезонность в периоды 12 месяцев, 6 месяцев или 4 месяца. После этих пиков волны чрезвычайно малы с мини-пиками на частотах 4 и 5, поэтому мы их не учитывали.

Гармоническая регрессия

После определения гармоник мы построили для них регрессионную модель, используя метод наименьших квадратов.

```

Residuals:
    Min       1Q   Median       3Q      Max
-1.12625 -0.20550  0.02141  0.15762  1.91137

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.006100   0.016515   0.369   0.712
omega1c1     0.346100   0.023399  14.791 < 2e-16 ***
omega1s1     0.121977   0.023314   5.232 2.68e-07 ***
omega2c2    -0.010351   0.023370  -0.443   0.658
omega2s2     0.144289   0.023342   6.182 1.53e-09 ***
omega3c3     0.007573   0.023356   0.324   0.746
omega3s3     0.092030   0.023356   3.940 9.56e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3376 on 411 degrees of freedom
Multiple R-squared:  0.423,    Adjusted R-squared:  0.4146
F-statistic: 50.21 on 6 and 411 DF,  p-value: < 2.2e-16

```

Таблица 2: Результат гармонической регрессии для *house price* после второй дифференциации

Таблица выше показывает, что частота одного года является наиболее значимой, при этом р-значения как синуса, так и косинуса находятся ниже значимого уровня. Другие частоты 6 месяцев и 4 месяца также значимы, и хотя не всегда, это также доказывает, что эти гармоники присутствуют. Теперь мы проверяем остатки, чтобы увидеть, помогла ли нам эта модель улучшить существующие пики.

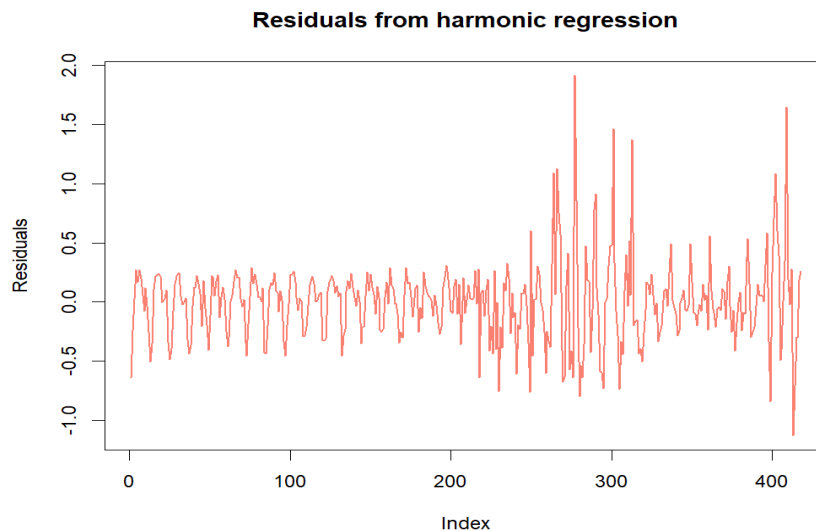


Рисунок 23: Остатки гармонической регрессии для *house price* после второй дифференциации

По сравнению с пиками на рис. 16, пики теперь ближе к 0, что является хорошим знаком.

Мы также попытались построить регрессию на полиномах для остатков. Однако результат оказался незначимым при р-значении = 0,48 ($> 0,05$). В то же время коэффициенты также незначимы.

```

Residuals:
    Min       1Q   Median       3Q      Max
-1.14572 -0.20796  0.02343  0.16121  1.90491

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -2.004e-02  3.287e-02  -0.610   0.542
poly(index(res_h2), 1, raw = TRUE)  9.564e-05  1.360e-04   0.703   0.482

Residual standard error: 0.3354 on 416 degrees of freedom
Multiple R-squared:  0.001188, Adjusted R-squared:  -0.001213
F-statistic: 0.4949 on 1 and 416 DF, p-value: 0.4822

```

Таблица 3: Полиномиальная регрессия для остатков гармонической регрессии *house price*

Поэтому для дальнейшего анализа мы решили использовать не остатки полиномиальной регрессии, а предыдущие остатки гармонической регрессии.

Augmented Dickey-Fuller (ADF) тест

Мы провели ADF-тест для остатков гармонической регрессии, и полученный результат подтверждает стационарность.

H_0 : Временной ряд нестационарен.

H_1 : Временной ряд является стационарным.

Результат:

Dickey-Fuller статистика: -12.005

p-значение: 0.01

Наконец, мы сохраняем данные для остатков регрессии (после удаления тренда путем дифференцирования и обнаружения сезонности), чтобы перейти к следующему этапу.

Подход 2: Третья дифференцирования

Мы применили **дифференциацию в третий раз**. На этот раз нам удалось очистить сезонные характеристики нашего временного ряда.

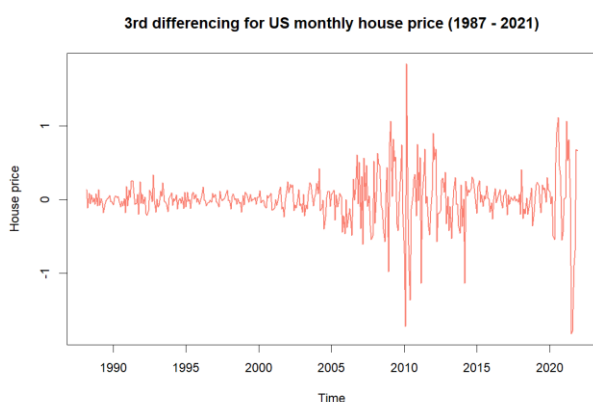


Рисунок 24: График ежемесячной *house price* в США (1987–2021 гг.) после третьей дифференциации

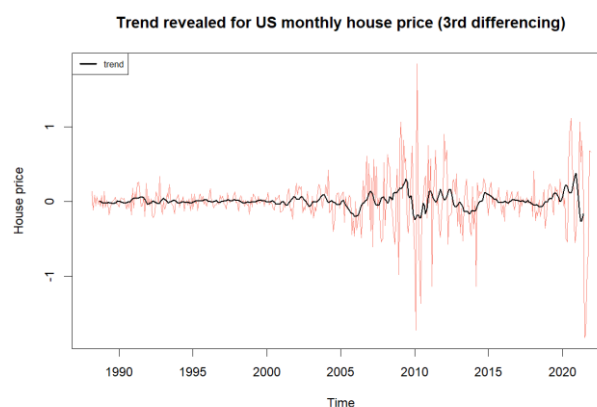


Рисунок 257: График тренда ежемесячной *house price* в США (1987–2021 гг.) после третьей дифференциации

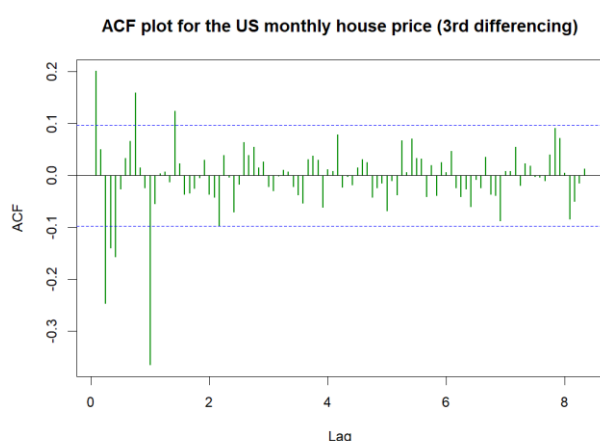


Рисунок 26: График ACF для *house price* после третьей дифференциации

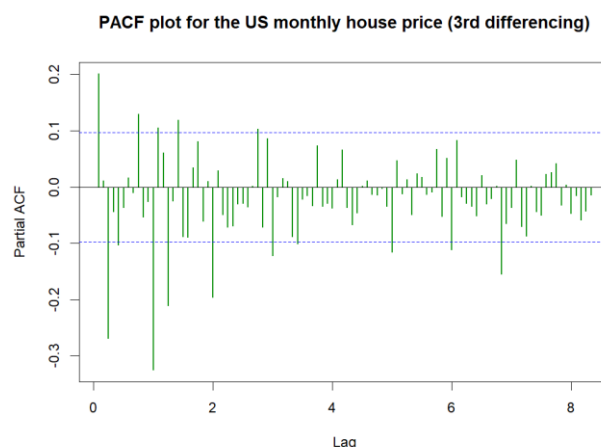


Рисунок 27: График PACF для *house price* после третьей дифференциации

Как видно из графика ACF слева, нет повторяющегося шаблона определенного количества лагов, а график справа показывает детрендированный и десезонированный график цен на жилье. Неудивительно, что во время глобальной рецессии и несколько лет спустя наблюдаются значительные турбулентности. Наконец, график PACF информирует нас о значительных лагах, пики которых превышают пороговое значение. Пиковые значения чередуются между положительными и отрицательными корреляциями, что указывает на более высокий порядок скользящего среднего. Мы перейдем к более подробной информации, когда будем использовать этот график для выбора подходящей модели прогнозирования.

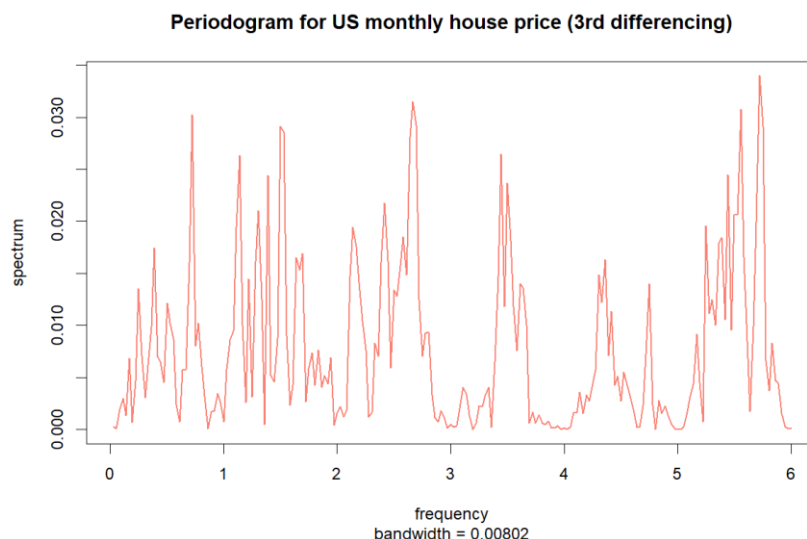


Рисунок 28: Периодограмма ежемесячной *house price* в США после третьей дифференциации

Периодограмма цен на жилье на уровне 3-й разницы не имеет доминирующих пиков или четких волн, указывающих на сезонность или тенденцию.

Augmented Dickey-Fuller (ADF) тест

H_0 : Временной ряд нестационарен.

H_1 : Временной ряд является стационарным.

Результат:

Dickey-Fuller статистики: -7.6923

p-значение: 0.01

Мы провели тест ADF для остатков гармонической регрессии, и полученный результат подтверждает стационарность.

Наконец, мы сохранили трехкратно разностные данные о *цене на жилье* для дальнейшего анализа.

3. Заключение

Наши данные о ценах на жилье, по-видимому, имели тенденцию к росту, но не показывали никаких признаков сезонности, пока мы не применили дифференциацию во второй раз.

При первом подходе, после проведения регрессии для гармоник, мы вывели остатки регрессии. График ACF выявил годовую сезонность (и, что удивительно, стационарность), из-за чего мы пытаемся использовать **сезонную модель ARIMA** для этих остатков в следующей части исследования.

При втором подходе мы применили 3-й раунд дифференциации и получили ряд, очищенный от тренда и сезонности, следовательно, стационарности. Мы сохранили ряд с тройной разницей и пытаемся использовать для него **модель ARIMA**.

Часть 4: Автономные динамические модели временных рядов

Часть 4.1: Динамическая модель ставок по ипотеке

Мы разделили наши данные на две части: обучаемая часть включает значения с января 1987 года по декабрь 2020 года, а тестируемая часть включает значения с января 2021 года по декабрь 2021 года, что также означает, что мы попытаемся спрогнозировать ставки по ипотеке в США на 12 последовательных месяцев.

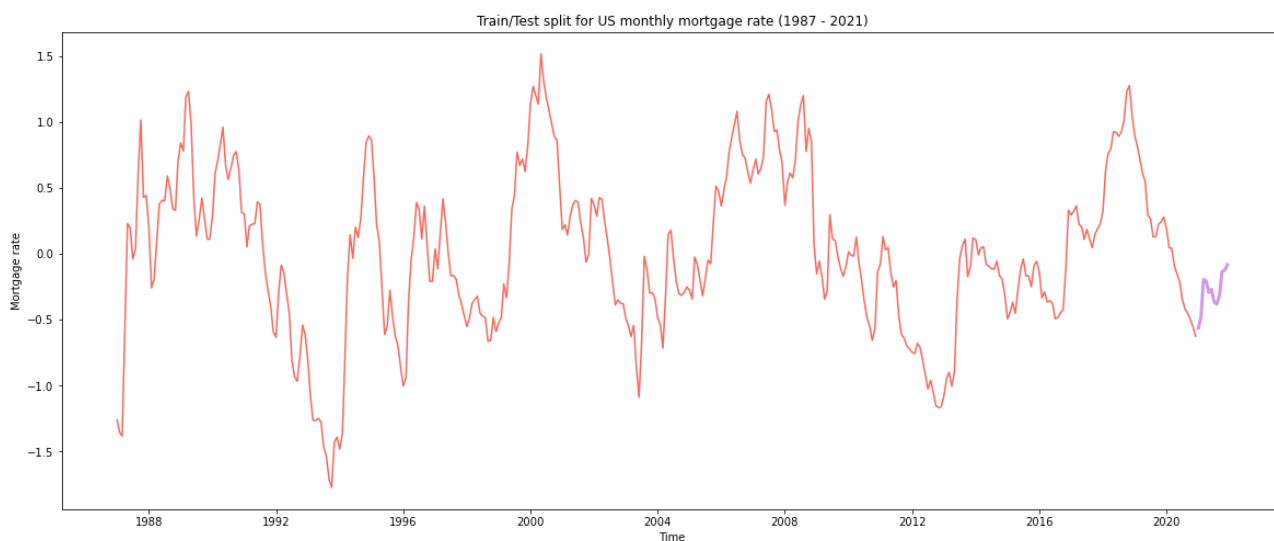


Рисунок 29: Разделение обучающего и тестового наборов (12 месяцев) для *mortgage rates*

Следующие шаги основаны на методологии Бокса-Дженкинса.

Шаг 1: Идентификация: стационарность и определение параметров модели.

Два теста ADF из предыдущей части показали, что наши данные, а также остатки гармонической регрессии, являются стационарными.

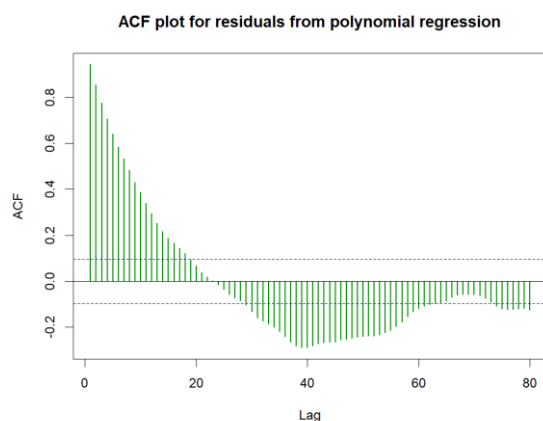


Рисунок 30: График ACF для остатков полиномиальной регрессии для *mortgage rates*

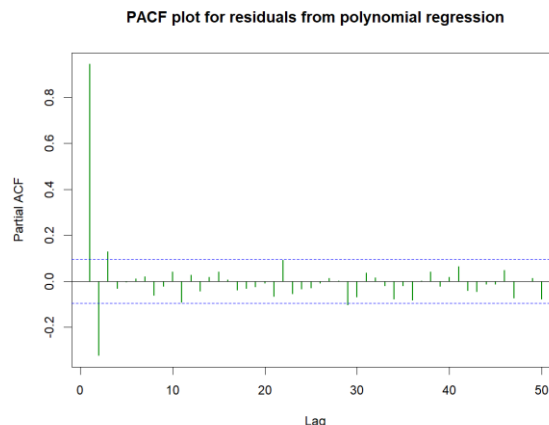


Рисунок 31: График ACF для остатков полиномиальной регрессии для *mortgage rate*

Однако из графика PACF мы видим, что частичная автокорреляция затухает после лага 3. Мы также учли лаг 2, поскольку он гораздо более значим, чем лаг 3. График ACF спадает с положительного на отрицательный, что указывает на то, что в наших данных все еще существует тенденция.

Поэтому мы решили включить первую дифференциацию для остатков из полиномиальной регрессии для ставок по ипотеке.

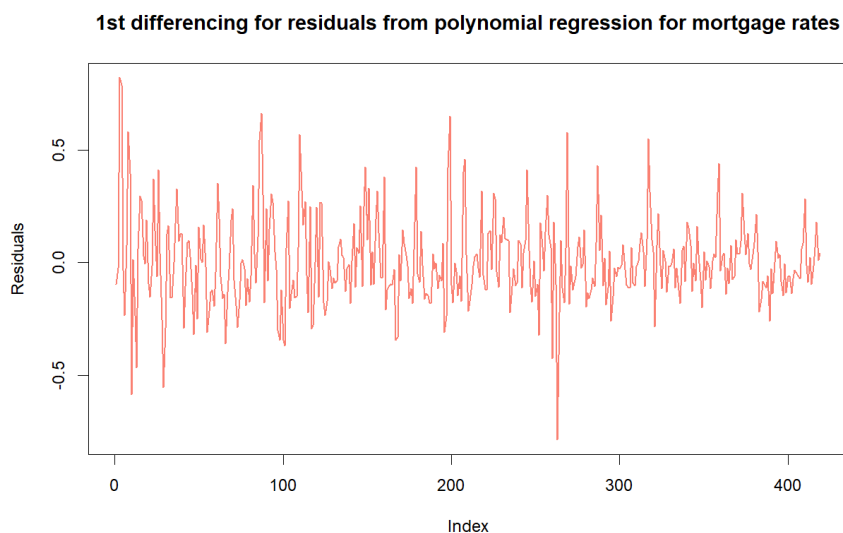


Рисунок 32: Общий график остатков полиномиальной регрессии для *mortgage rates* (после первой дифференциации)

После применения метода дифференцирования наши данные потеряли 1 значение, что равно значению ставки по ипотеке, зафиксированному в декабре 2021 года. Мы считаем это справедливой потерей в обмен на набор данных с исключенным трендом.

Поэтому мы разделили наши данные на две части: обучаемая часть включает значения с января 1987 года по ноябрь 2021 года, а тестируемая часть включает значения с декабря 2021 года по ноябрь 2021 года, что также означает, что мы попытаемся спрогнозировать ставки по ипотеке в США на 12 последовательных месяцев.

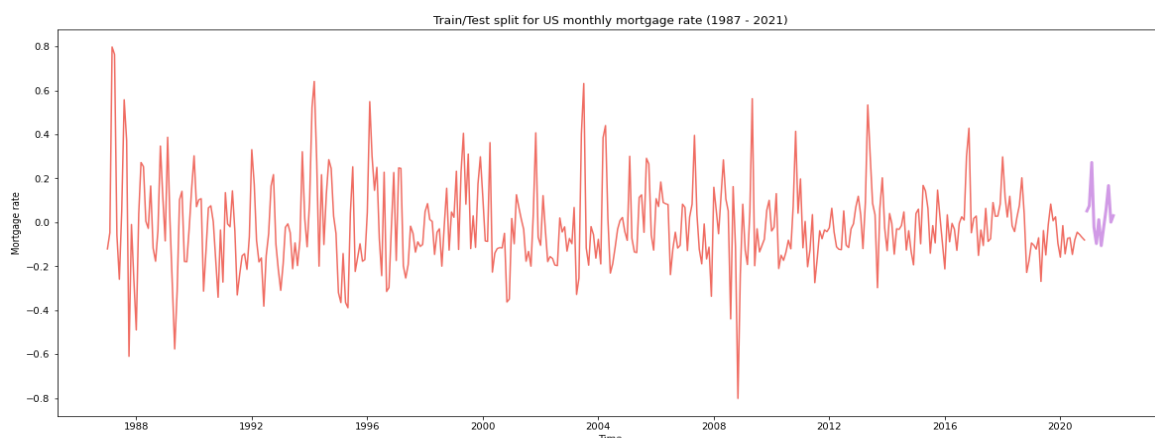


Рисунок 33: Разделение обучающего и тестового наборов (12 месяцев) для *mortgage rates* (после первой дифференциации)

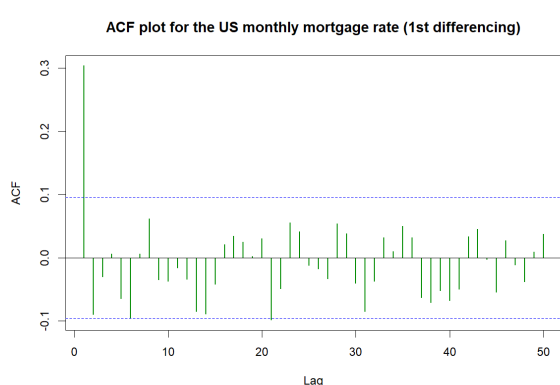


Рисунок 34: График ACF для *mortgage rates* (после первой дифференциации)

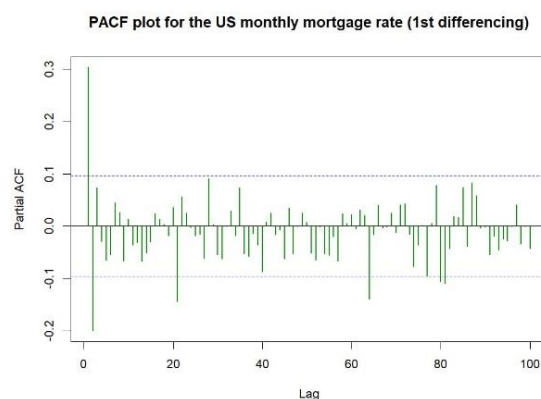


Рисунок 25: График PACF для *mortgage rates* (после первой дифференциации)

Графики ACF и PACF показывают убывающую характеристику. Из графика ACF мы видим, что лаги 1 и 21 значимы. Из графика PACF мы должны учитывать лаги 1, 2, 21, 64, 80 и 81, которые имеют пики за пределами значимого уровня.

Поэтому из этих графиков мы решили попробовать модель ARMA (p,q) с $p = 1$, $p = 2$, $p = 21$, $p = 64$, $p = 80$ или $p = 81$ и $q = 1$ или $q = 21$.

Шаг 02: Оценка

Помимо наших собственных предположений, полученных из графиков ACF и PACF, мы использовали функцию `auto.arima` в программе R. Она предложила модель MA (1), которая включает компонент MA первого порядка.

Чтобы оптимизировать предложенную модель, мы удалили остатки и проверили графики ACF и PACF, чтобы скорректировать несколько параметров, и в конце предоставили сравнение между потенциальными моделями.

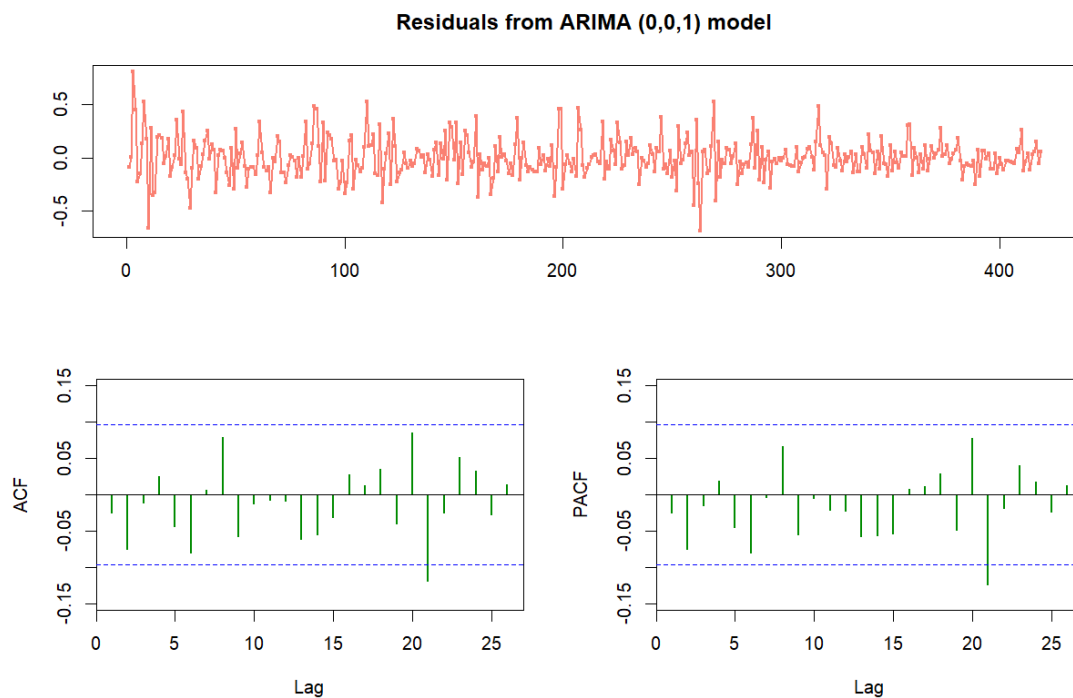


Рисунок 36: Остатки из модели MA (1) для *mortgage rates*

Графики ACF и PACF показывают резкий срез после лага 21. На графике ACF резкий срез после значимого всплеска лага 21. На графике PACF значимым является только лаг 21. Однако мы уже учли эти лаги в предыдущем наблюдении. Поэтому мы скорректировали параметры, запустили модели на Python и получили следующие результаты:

Model	AIC	BIC
ARIMA (0, 0, 1)	-114.57	302.35
ARIMA (1, 0, 1)	-115.07	292.67
ARIMA (1, 0, 21)	-121.01	281.89
ARIMA (2, 0, 1)	-141.26	191.47
ARIMA (2, 0, 21)	-133.20	215.57
ARIMA (21, 0, 1)	-154.65	113.94
ARIMA (21, 0, 21)	-181.60	-5.21
ARIMA (64, 0, 1)	-198.52	-102.31
ARIMA (64, 0, 21)	-201.28	-105.07
ARIMA (80, 0, 1)	-215.29	-203.27
ARIMA (80, 0, 21)	-214.67	-198.64
ARIMA (81, 0, 1)	-214.19	-194.15
ARIMA (81, 0, 21)	-207.00	-106.78

Таблица 4: Сравнение моделей *mortgage rate* на основе критериев AIC и BIC

Мы учли оба критерия: AIC и BIC, поэтому решили выбрать модель ARMA (80, 1) в качестве оптимальной для дальнейшего анализа. Эта модель характеризуется восьмидесятым порядком AR и первым порядком MA.

	coef	std err	z	P> z	[0.025	0.975]
const	-0.0168	0.004	-4.210	0.000	-0.025	-0.009
ar.L1	0.6306	0.419	1.506	0.132	-0.190	1.452
ar.L2	-0.3234	0.157	-2.066	0.039	-0.630	-0.017
ar.L3	0.1090	0.119	0.917	0.359	-0.124	0.342
ar.L4	-0.0300	0.078	-0.382	0.702	-0.184	0.124
ar.L5	-0.0332	0.081	-0.410	0.682	-0.192	0.126
ar.L6	-0.0928	0.079	-1.178	0.239	-0.247	0.062
ar.L7	0.0258	0.084	0.307	0.759	-0.139	0.191
ar.L8	0.0778	0.074	1.046	0.296	-0.068	0.223
ar.L9	-0.1345	0.079	-1.703	0.089	-0.289	0.020
ar.L10	0.0021	0.089	0.024	0.981	-0.172	0.176
ar.L11	-0.0171	0.082	-0.208	0.835	-0.178	0.144
ar.L12	-0.0323	0.088	-0.368	0.713	-0.204	0.140
ar.L13	-0.0736	0.099	-0.747	0.455	-0.267	0.120
ar.L14	0.0287	0.086	0.334	0.739	-0.140	0.198
ar.L15	-0.0667	0.091	-0.732	0.464	-0.246	0.112
ar.L16	-0.0144	0.080	-0.180	0.858	-0.172	0.143
ar.L17	0.0092	0.073	0.126	0.899	-0.133	0.152
ar.L18	0.0252	0.072	0.348	0.728	-0.117	0.167
ar.L19	-0.1036	0.074	-1.391	0.164	-0.250	0.042
ar.L20	0.1227	0.089	1.371	0.170	-0.053	0.298
ar.L21	-0.2142	0.100	-2.153	0.031	-0.409	-0.019
ar.L22	0.0652	0.110	0.590	0.555	-0.151	0.282
ar.L23	0.0229	0.084	0.273	0.785	-0.142	0.188
ar.L24	-0.0186	0.077	-0.241	0.810	-0.170	0.133
ar.L25	-0.0788	0.080	-0.986	0.324	-0.235	0.078
ar.L26	0.0261	0.086	0.302	0.763	-0.143	0.196
ar.L27	-0.1140	0.078	-1.471	0.141	-0.266	0.038
ar.L28	0.0940	0.092	1.024	0.306	-0.086	0.274
ar.L29	0.0036	0.084	0.043	0.966	-0.161	0.168
ar.L30	-0.0786	0.084	-0.939	0.348	-0.243	0.085
ar.L31	-0.0456	0.089	-0.514	0.607	-0.220	0.128
ar.L32	-0.0143	0.082	-0.174	0.862	-0.176	0.147
ar.L33	0.0223	0.089	0.251	0.801	-0.152	0.196
ar.L34	-0.0677	0.090	-0.756	0.450	-0.243	0.108
ar.L35	0.1056	0.083	1.278	0.201	-0.056	0.268
ar.L36	-0.0833	0.090	-0.923	0.356	-0.260	0.094
ar.L37	-0.0935	0.084	-1.114	0.265	-0.258	0.071
ar.L38	-0.0214	0.089	-0.241	0.809	-0.195	0.152
ar.L39	0.0143	0.070	0.203	0.839	-0.124	0.152
ar.L40	-0.1190	0.077	-1.544	0.123	-0.270	0.032
ar.L41	0.0031	0.092	0.034	0.973	-0.178	0.184
ar.L42	0.0383	0.086	0.445	0.656	-0.130	0.207
ar.L43	-0.1100	0.093	-1.183	0.237	-0.292	0.072
ar.L44	0.0297	0.097	0.306	0.760	-0.161	0.220
ar.L45	-0.0789	0.078	-1.006	0.314	-0.232	0.075
ar.L46	0.0344	0.089	0.385	0.700	-0.141	0.209
ar.L47	-0.0499	0.084	-0.596	0.551	-0.214	0.114
ar.L48	0.0023	0.082	0.028	0.977	-0.158	0.163
ar.L49	-0.0003	0.085	-0.003	0.998	-0.166	0.166
ar.L50	-0.0149	0.092	-0.163	0.871	-0.195	0.165
ar.L51	-0.0345	0.084	-0.410	0.682	-0.199	0.130
ar.L52	-0.0635	0.081	-0.779	0.436	-0.223	0.096
ar.L53	-0.0122	0.091	-0.133	0.894	-0.191	0.167
ar.L54	-0.0174	0.085	-0.205	0.838	-0.184	0.149
ar.L55	-0.0857	0.077	-1.108	0.268	-0.237	0.066
ar.L56	0.0072	0.082	0.088	0.930	-0.153	0.167
ar.L57	-0.0479	0.075	-0.641	0.521	-0.194	0.099
ar.L58	0.0292	0.077	0.379	0.705	-0.122	0.180
ar.L59	-0.0594	0.086	-0.691	0.489	-0.228	0.109
ar.L60	0.0332	0.081	0.409	0.683	-0.126	0.193
ar.L61	-0.0136	0.083	-0.165	0.869	-0.176	0.149
ar.L62	-0.0091	0.083	-0.111	0.912	-0.171	0.153
ar.L63	0.0710	0.074	0.956	0.339	-0.075	0.217
ar.L64	-0.1701	0.079	-2.144	0.032	-0.326	-0.015
ar.L65	-0.0100	0.110	-0.091	0.928	-0.226	0.206
ar.L66	0.0547	0.075	0.728	0.466	-0.093	0.202
ar.L67	-0.0148	0.082	-0.180	0.857	-0.176	0.146
ar.L68	-0.0354	0.088	-0.404	0.686	-0.207	0.136
ar.L69	0.0176	0.081	0.218	0.827	-0.140	0.176
ar.L70	-0.0168	0.079	-0.213	0.832	-0.172	0.138
ar.L71	0.0177	0.086	0.205	0.838	-0.152	0.187
ar.L72	0.0359	0.080	0.451	0.652	-0.120	0.192
ar.L73	0.0093	0.083	0.112	0.911	-0.153	0.172
ar.L74	-0.0800	0.085	-0.940	0.347	-0.247	0.087
ar.L75	-0.0758	0.096	-0.786	0.432	-0.265	0.113
ar.L76	0.0518	0.086	0.602	0.547	-0.117	0.220
ar.L77	-0.0934	0.085	-1.099	0.272	-0.260	0.073
ar.L78	-0.0285	0.100	-0.285	0.775	-0.224	0.167
ar.L79	0.1683	0.079	2.122	0.034	0.013	0.324
ar.L80	-0.1696	0.073	-2.323	0.020	-0.313	-0.026
ma.L1	-0.2635	0.431	-0.611	0.541	-1.109	0.581
sigma2	0.0266	0.002	12.917	0.000	0.023	0.031
=====						
Ljung-Box (L1) (Q):		0.07	Jarque-Bera (JB):		20.78	
Prob(Q):		0.80	Prob(JB):		0.00	
Heteroskedasticity (H):		0.46	Skew:		0.27	
Prob(H) (two-sided):		0.00	Kurtosis:		3.97	
=====						

Таблица 5: Результат модели ARMA (80, 1) для *mortgage rates*

В таблице выше показан результат модели ARMA (80, 1), в которой мы сосредоточились на коэффициентах и обнаружили, что значимыми являются только следующие компоненты: авторегрессионные члены с лагами 2, 21, 64, 79, 80..

Шаг 3: Диагностическая проверка: остаточные ошибки

Мы проверили остатки выбранной модели и провели тест Льюнга-Бокса.

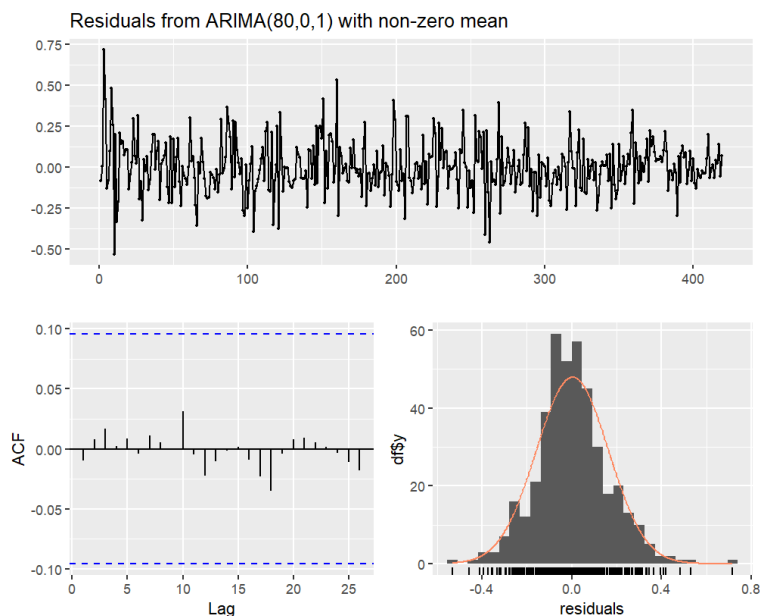


Figure 37: Residuals from the ARMA (80, 1) model

Все пики графика ACF теперь находятся в пределах значимости, что свидетельствует об отсутствии автокорреляций.

Ljung-Box тест

H_0 : Остатки распределены независимо.

H_1 : Остатки не распределены независимо и демонстрируют последовательную корреляцию.

Результат:

Ljung-Box статистика: 27.718847

p-значение: 1.0

Нам не удалось отвергнуть H_0 . Тест Льюнга-Бокса показывает, что остатки не имеют остаточных автокорреляций. Поэтому мы можем подтвердить, что остатки модели ARMA (80, 1) являются белым шумом.

Таким образом, теперь у нас есть сезонная модель ARIMA, которая соответствует требованиям и готова к прогнозированию.

Процесс прогнозирования

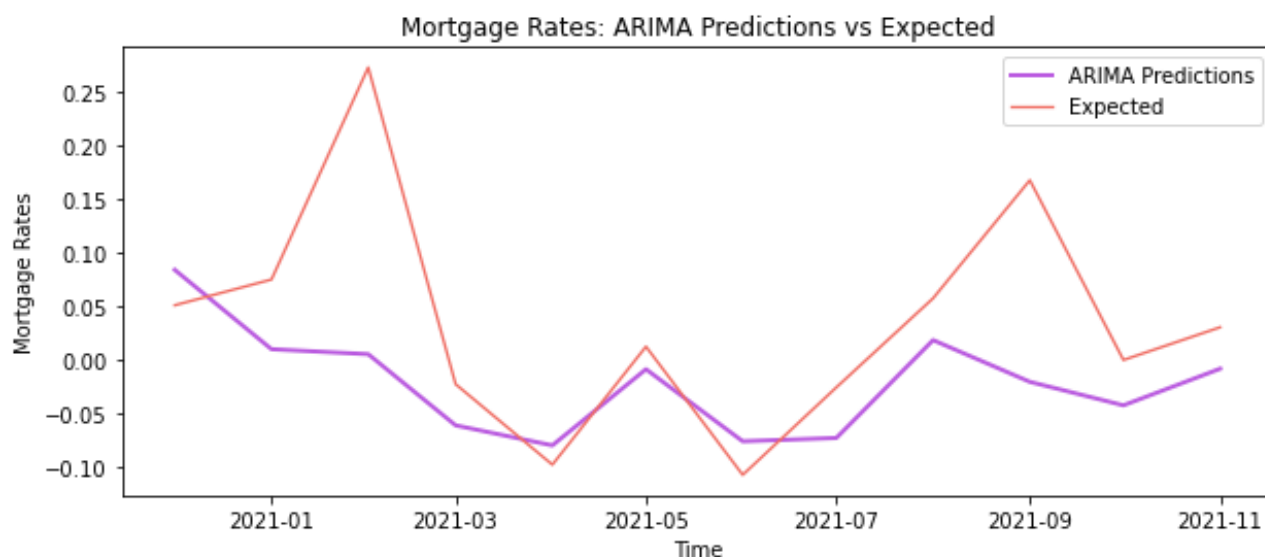


Рисунок 38: Сравнительный график ожидаемых значений и значений, предсказанных ARIMA

На сравнительном графике мы видим, что наша модель довольно хорошо следует закономерностям тестового набора. Прогнозы довольно точны для периодов с марта 2021 года по июль 2021 года и с октября 2021 года по ноябрь 2021 года. Самый большой разрыв наблюдается в период с января 2021 года по февраль 2021 года и с августа 2021 года по октябрь 2021 года. В 2020 году и в первые месяцы 2021 года ставки по ипотеке резко упали в ответ на пандемию COVID-19. Эти рекордно низкие ставки в значительной степени зависели от политики Федеральной резервной системы в эпоху COVID. Однако сразу после восстановления экономики в середине 2021 года ставки по ипотеке начали расти. Наша модель не знает ни мер правительства, ни того, насколько быстро экономика восстановилась, что объясняет разрыв между прогнозами и ожидаемыми значениями.

Часть 4.2: Динамическая модель цен на жилье

В этой части мы построим потенциальные модели на основе результатов предыдущей части. Для каждого подхода мы выберем наиболее подходящую модель и попытаемся сравнить, какой подход лучше в прогнозировании, посмотрев на сравнительные графики.

Подход 01: Оптимальная модель для остатков гармонической регрессии

После применения метода дифференциации для двух раз для удаления тренда наши данные потеряли 2 значения, которые равны 2 значениям цен на жилье, зарегистрированным в ноябре 2021 года и декабре 2021 года. Таким образом, наш временной ряд цен на жилье сократился до 418 значений.

Мы разделили наши данные на две части: обучаемая часть включает значения с января 1987 года по октябрь 2019 года, а тестируемая часть включает значения с ноября 2019 года по октябрь 2021 года, что также означает, что мы попытаемся спрогнозировать цены на жилье в США на 24 последовательных месяца.

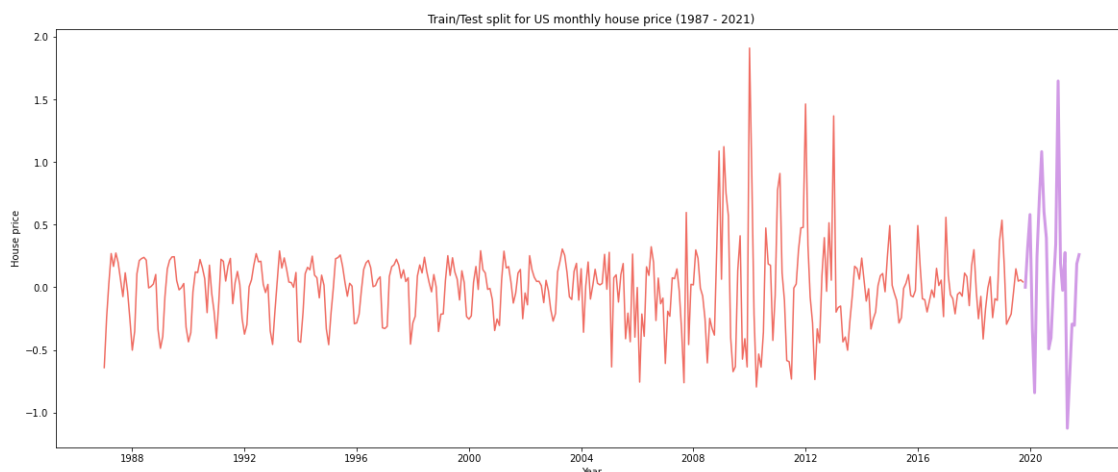


Рисунок 39: Разделение обучающего и тестового наборов (24 месяца) для *house price*

Следующие шаги основаны на методологии Бокса-Дженкинса.

Шаг 1: Идентификация: стационарность и определение параметров модели.

В части 3 мы доказали, что наши окончательные данные свободны от стационарности, используя тест ADF.

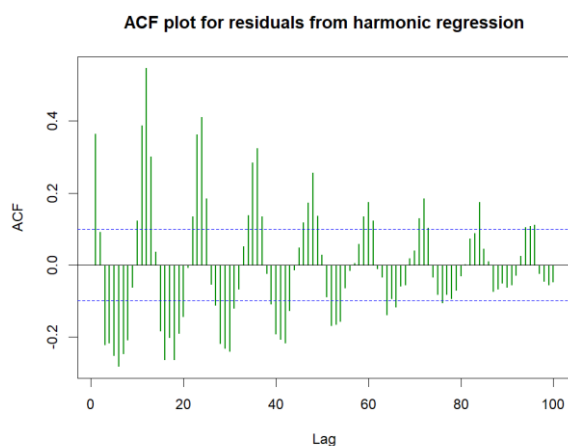


Рисунок 40: График ACF для остатков гармонической регрессии

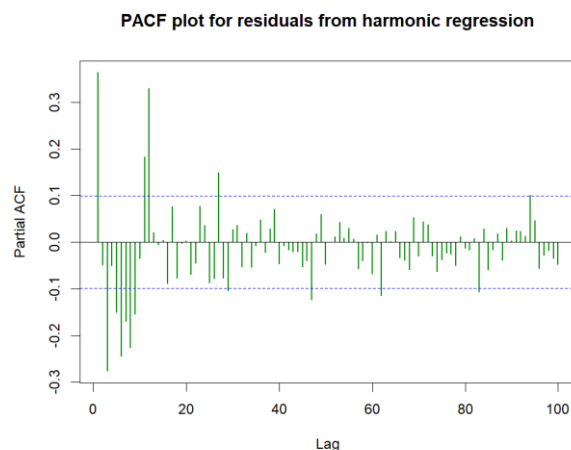


Рисунок 41: График цены на жилье (после второй дифференциации) от гармонической регрессии

Графики ACF и PACF показывают затухающую характеристику, то есть автокорреляция и частичная автокорреляция постепенно становятся менее значимыми. На графике ACF есть явные доказательства годовой сезонности, поскольку пики достигают пика после каждых 12 лагов (синусоидальный рисунок). На графике PACF большинство первых 12 лагов значимы, пока не наступает внезапный срез после лага 12, после которого пики начинают сужаться между положительной и отрицательной зонами.

С постепенно геометрически уменьшающейся ACF и PACF, которая значима только для нескольких лагов, временной ряд может быть сезонным AR (12) процессом.

Шаг 02: Оценка

Помимо наших собственных предположений, полученных из графиков ACF и PACF, мы использовали функцию `auto.arima` в программе R. Она предложила модель SARIMA с параметрами $(3, 0, 2)(0, 1, 2)[12]$, которая указывает на несезонный компонент MA второго порядка с несезонным компонентом AR третьего порядка и сезонную дифференциацию с сезонным вторым порядком MA и временным интервалом годового сезонного паттерна.

Чтобы оптимизировать предложенную модель, мы изъяли остатки и проверили графики ACF и PACF, чтобы скорректировать несколько параметров и предоставить сравнение между потенциальными моделями в конце.

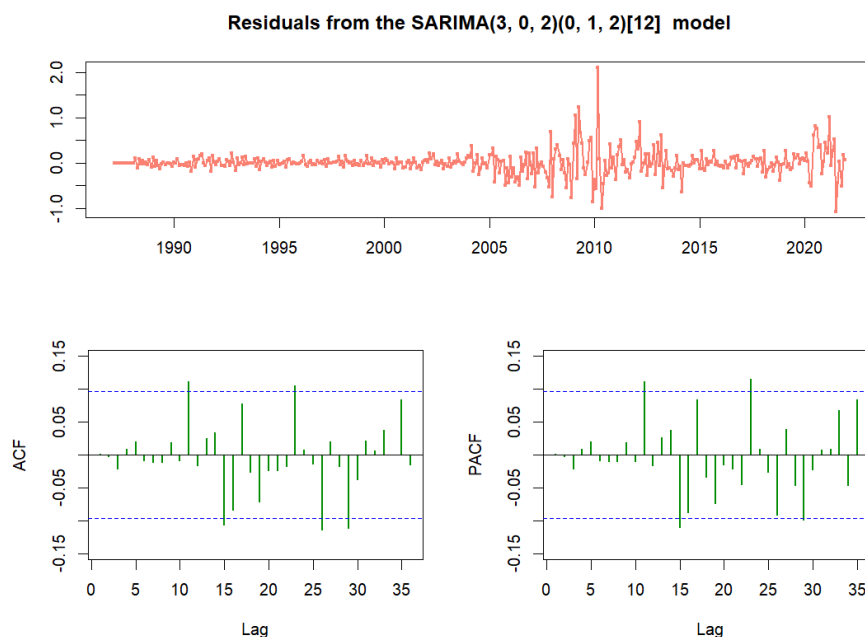


Рисунок 42: Общий график, графики ACF и PACF для остатков из модели SARIMA (3, 0, 2)(0, 1, 2)[12] для *house price*

Графики показывают, что есть значительные пики, соответствующие скользящему среднему и авторегрессионным компонентам. Нет никакого шаблона сужения, и лаги значимы в случайном порядке, поэтому для модели SARIMA мы попробовали дополнительные члены AR, соответствующие $p = 11$, $p = 15$, $p = 23$, $p = 29$ и $p = 12$ (из шага 1).

Поэтому мы скорректировали параметры, запустили модели на Python и получили следующие результаты:

Model	AIC	BIC
SARIMA (3, 0, 2) (0, 1, 2) [12]	33.29	64.85
SARIMA (11, 0, 2) (0, 1, 2) [12]	12.12	79.25
SARIMA (12, 0, 2) (0, 1, 2) [12]	12.02	79.09
SARIMA (15, 0, 2) (0, 1, 2) [12]	9.55	88.46
SARIMA (23, 0, 2) (0, 1, 2) [12]	12.71	123.18
SARIMA (29, 0, 2) (0,1,2) [12]	17.59	151.74

Таблица 6: Сравнение моделей цен на жилье на основе критериев AIC и BIC (1-ый подход)

Мы учли оба критерия: AIC и BIC, поэтому решили выбрать модель SARIMA (12, 0, 2) (0, 1, 2) [12] в качестве оптимальной модели для дальнейшего анализа. Эта модель характеризуется несезонным вторым порядком MA с двенадцатым порядком AR, сезонным дифференцированием с сезонным вторым порядком MA и временным интервалом годового сезонного паттерна.

	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.1585	0.107	1.488	0.137	-0.050	0.367
ar.L2	0.0488	0.133	0.366	0.714	-0.213	0.310
ar.L3	-0.3049	0.036	-8.370	0.000	-0.376	-0.234
ar.L4	0.0202	0.056	0.360	0.719	-0.090	0.131
ar.L5	-0.0258	0.067	-0.383	0.702	-0.158	0.106
ar.L6	-0.1261	0.046	-2.762	0.006	-0.216	-0.037
ar.L7	-0.0641	0.040	-1.613	0.107	-0.142	0.014
ar.L8	-0.0495	0.039	-1.262	0.207	-0.126	0.027
ar.L9	-0.0315	0.041	-0.765	0.444	-0.112	0.049
ar.L10	-0.0411	0.038	-1.084	0.278	-0.115	0.033
ar.L11	0.1609	0.037	4.380	0.000	0.089	0.233
ar.L12	0.3493	0.082	4.238	0.000	0.188	0.511
ma.L1	-0.0096	0.115	-0.084	0.933	-0.234	0.215
ma.L2	-0.0278	0.127	-0.219	0.827	-0.276	0.221
ma.S.L12	-1.0598	0.089	-11.950	0.000	-1.234	-0.886
ma.S.L24	0.1603	0.067	2.386	0.017	0.029	0.292
sigma2	0.0533	0.002	21.822	0.000	0.049	0.058
=====						
Ljung-Box (L1) (Q):			0.00	Jarque-Bera (JB):		1070.58
Prob(Q):			1.00	Prob(JB):		0.00
Heteroskedasticity (H):			14.56	Skew:		0.95
Prob(H) (two-sided):			0.00	Kurtosis:		10.98
=====						

Таблица 7: Результат модели SARIMA (12, 0, 2) (0, 1, 2) [12] для *house price* (1-ый подход)

В таблице выше показан результат модели SARIMA (12, 0, 2) (0, 1, 2) [12], в которой мы сосредоточились на коэффициентах и обнаружили, что значимыми являются только следующие компоненты: несезонные авторегрессионные члены с лагами 3, 6, 11, 12 и сезонные скользящие средние члены с лагами 12, 24..

Шаг 3: Диагностическая проверка: остаточные ошибки

Мы проверили остатки выбранной модели и провели тест Льюнга-Бокса.

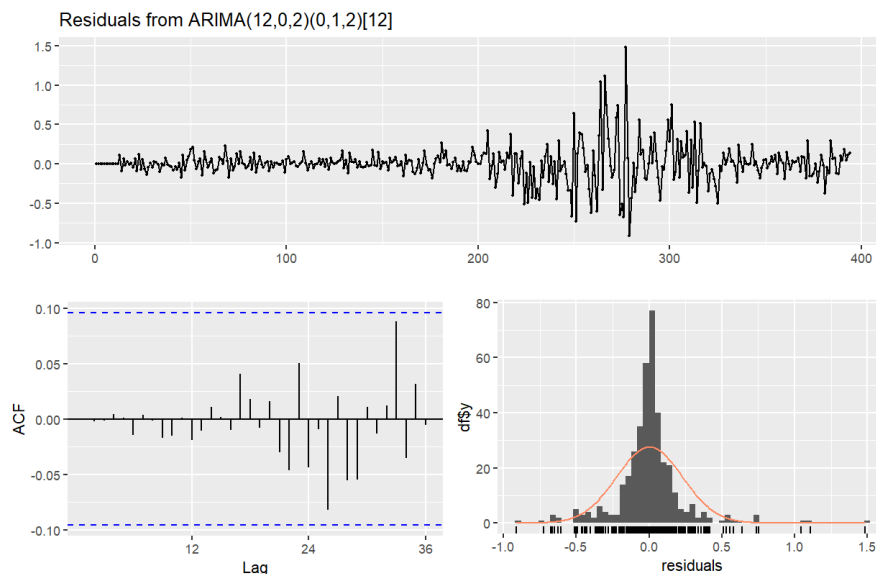


Рисунок 43: Остатки из модели SARIMA (12, 0, 2)(0, 1, 2)[12]

Все пики графика ACF теперь находятся в пределах значимости, что свидетельствует об отсутствии автокорреляций.

Ljung-Box тест

H_0 : Остатки распределены независимо.

H_1 : Остатки не распределены независимо и демонстрируют последовательную корреляцию.

Результаты:

Ljung-Box статистика: 0.360938

p-значение: 0.999999

Нам не удалось отвергнуть H_0 . Тест Льюнга-Бокса показывает, что остатки не имеют остаточных автокорреляций. Поэтому мы можем подтвердить, что остатки модели SARIMA (12, 0, 2)(0, 1, 2)[12] являются белым шумом.

Таким образом, теперь у нас есть сезонная модель ARIMA, которая соответствует требованиям и готова к прогнозированию.

Forecasting process

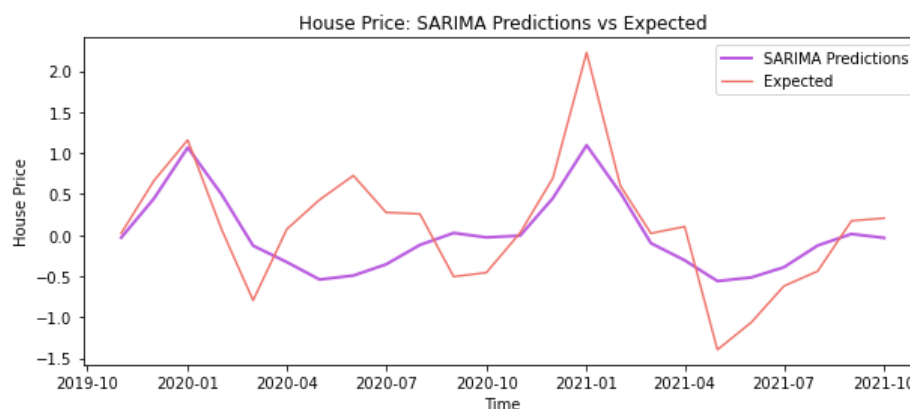


Рисунок 44: Сравнительный график ожидаемых значений и значений, предсказанных SARIMA

На сравнительном графике мы видим, что наша модель довольно хорошо следует закономерностям тестового набора. Прогнозы довольно точны для периодов ноябрь 2019 г. — март 2020 г., ноябрь 2020 г. — март 2021 г. и июль 2021 г. — октябрь 2021 г. Самый большой разрыв наблюдается в период с апреля 2020 г. по август 2020 г., что можно объяснить тем, что это первые месяцы COVID-19, и для таких исключительных времен моделям, основанным на наборе поездок в обычное время, очень сложно предсказать такое сильное падение.

Подход 02: Оптимальная модель для временного ряда без сезонных колебаний (после трехкратного дифференциации)

После применения метода дифференциации в течение трех раз наши данные потеряли в общей сложности 14 значений, что равно 14 значениям цен на жилье, регистрируемых ежемесячно с ноября 2020 года по декабрь 2021 года. Мы считаем это справедливой потерей в обмен на стационарный, максимально детрендированный и десезонированный набор данных, а также потому, что пандемия COVID-19 является внешним фактором, который, вероятно, подорвет точность наших прогнозов.

Поэтому мы разделили наши данные на две части: обучаемая часть включает значения с января 1987 года по октябрь 2019 года, а тестируемая часть включает значения с ноября 2019 года по октябрь 2020 года, что также означает, что мы попытаемся спрогнозировать цены на жилье в США на 12 последовательных месяцев.

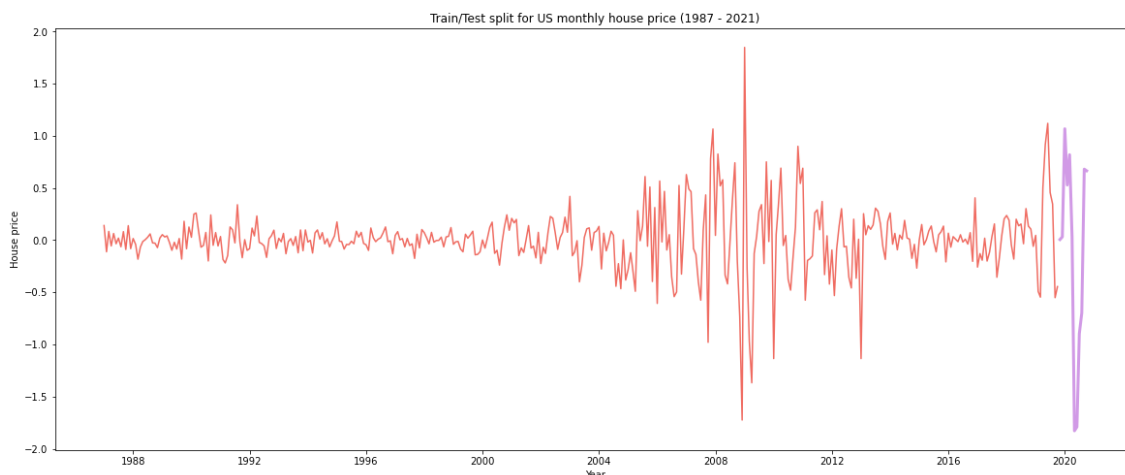


Рисунок 45: Разделение обучающего и тестового наборов (12 месяцев) для *house price*

Следующие шаги основаны на методологии Бокса-Дженкинса.

Шаг 1: Идентификация: стационарность и определение параметров модели.

В третьей части мы доказали, что наши окончательные данные свободны от стационарности с помощью теста ADF.

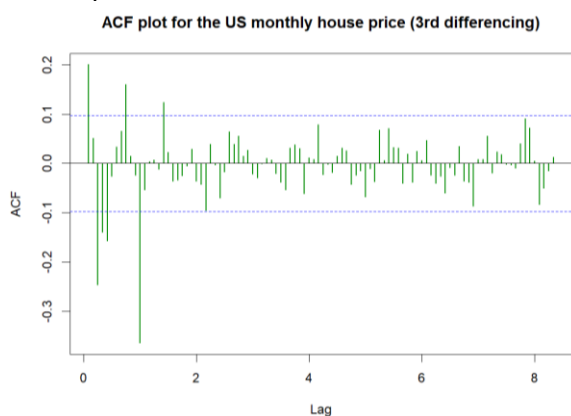


Рисунок 46: График ACF для *house price* (после третьей дифференциации)

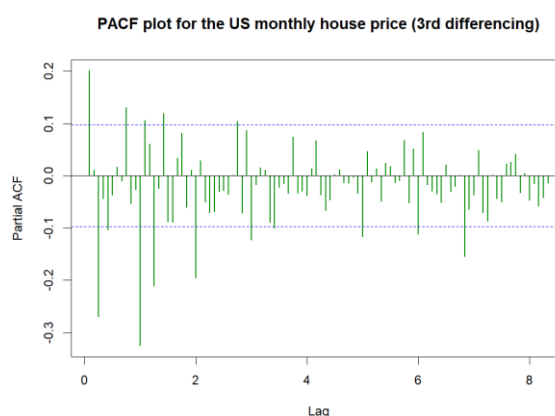


Рисунок 47: График ACF для *house price* (после третьей дифференциации)

Из графика ACF видно, что после лага 12 есть отсечка, после чего лаги незначительны, за исключением лага 17. Из графика PACF, похоже, есть затухающая характеристика, но нет существенных лагов, после которых PACF затухает. Следовательно, из этих графиков мы можем вывести только компонент МА (12) или МА (17). Дальнейшие предположения будут сделаны на следующем этапе.

Шаг 02: Оценка

Помимо наших собственных предположений, полученных из графиков ACF и PACF, мы использовали функцию `auto.arima` в программе R, и она предложила модель ARIMA с параметрами (0,0,4), что указывает на компонент МА четвертого порядка.

Чтобы оптимизировать предложенную модель, мы удалили остатки и проверили графики ACF и PACF, чтобы скорректировать несколько параметров, затем мы представили сравнение между потенциальными моделями в конце

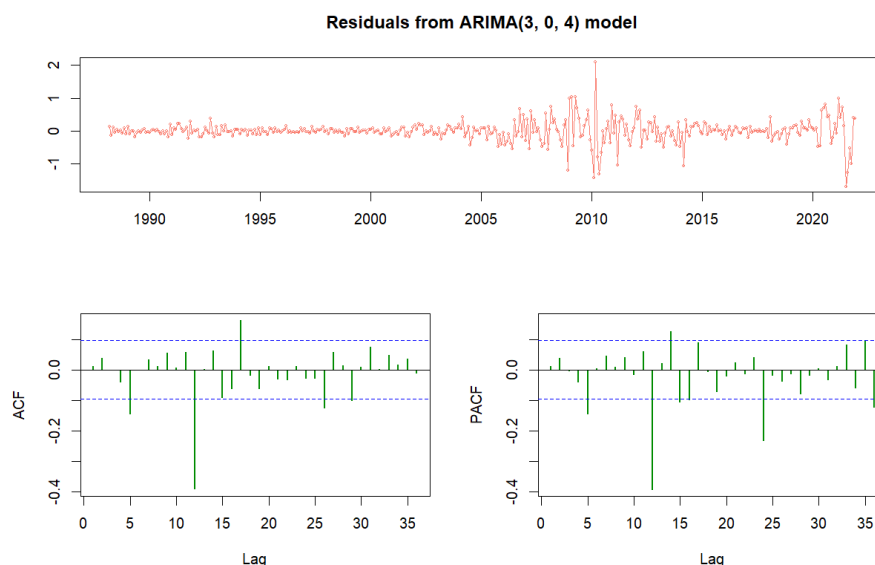


Рисунок 48: Общий график, графики ACF и PACF для остатков из модели ARIMA (3, 0, 4) (3, 0, 4)

После вывода остатков предлагаемой модели мы видим, что оба графика спадают до 0. Мы выбрали следующие параметры для неоцененной модели ARMA (p, q) из предлагаемой: p = 5, p = 12, p = 14, p = 24, q = 5, q = 12, q = 17.

Мы скорректировали несезонный порядок AR и запустили модели на Python и получили следующие результаты:

Model	AIC	BIC
ARIMA (3, 0, 4)	167.55	191.41
ARIMA (5, 0, 4)	145.16	188.90
ARIMA (12, 0, 4)	90.94	162.51
ARIMA (14, 0, 4)	90.17	169.69
ARIMA (24, 0, 4)	83.46	202.75
ARIMA (5, 0, 5)	121.15	168.86
ARIMA (5, 0, 12)	50.07	125.62
ARIMA (5, 0, 17)	60.96	156.23
ARIMA (12, 0, 5)	86.94	162.49
ARIMA (12, 0, 12)	65.26	168.65
ARIMA (12, 0, 14)	65.53	189.87
ARIMA (14, 0, 5)	89.92	185.35
ARIMA (14, 0, 12)	51.05	174.32
ARIMA (14, 0, 17)	60.77	231.75
ARIMA (24, 0, 5)	86.28	209.55
ARIMA (24, 0, 12)	59.55	210.65
ARIMA (24, 0, 17)	69.63	240.61

Таблица 8: Сравнение моделей цен на жилье на основе критериев AIC и BIC (2-ой подход)

Мы учли оба критерия: AIC и BIC, поэтому решили выбрать модель ARIMA (5, 0, 12) в качестве оптимальной для дальнейшего анализа. Эта модель характеризуется двенадцатым порядком MA и пятым порядком AR.

	coef	std err	z	P> z	[0.025	0.975]
const	0.0010	0.002	0.595	0.552	-0.002	0.004
ar.L1	0.1603	0.057	2.791	0.005	0.048	0.273
ar.L2	0.0672	0.054	1.241	0.214	-0.039	0.173
ar.L3	-0.1861	0.061	-3.033	0.002	-0.306	-0.066
ar.L4	0.0954	0.056	1.714	0.086	-0.014	0.205
ar.L5	-0.0190	0.071	-0.270	0.787	-0.157	0.119
ma.L1	0.0870	0.064	1.370	0.171	-0.037	0.211
ma.L2	-0.0564	0.074	-0.767	0.443	-0.200	0.088
ma.L3	-0.1464	0.062	-2.343	0.019	-0.269	-0.024
ma.L4	-0.1431	0.075	-1.910	0.056	-0.290	0.004
ma.L5	-0.0576	0.076	-0.757	0.449	-0.207	0.092
ma.L6	-0.0623	0.058	-1.067	0.286	-0.177	0.052
ma.L7	-0.0476	0.050	-0.956	0.339	-0.145	0.050
ma.L8	-0.0212	0.057	-0.373	0.710	-0.133	0.090
ma.L9	-0.0175	0.051	-0.344	0.731	-0.117	0.082
ma.L10	0.0227	0.052	0.434	0.664	-0.080	0.125
ma.L11	0.1942	0.046	4.253	0.000	0.105	0.284
ma.L12	-0.7079	0.054	-13.054	0.000	-0.814	-0.602
sigma2	0.0583	0.004	15.053	0.000	0.051	0.066
=====						
Ljung-Box (L1) (Q):			0.08	Jarque-Bera (JB):		935.42
Prob(Q):			0.78	Prob(JB):		0.00
Heteroskedasticity (H):			15.68	Skew:		0.85
Prob(H) (two-sided):			0.00	Kurtosis:		10.36
=====						

Таблица 9: Результат модели ARIMA (5, 0, 12) для *house price* (2-ой подход)

В таблице выше показан результат модели ARIMA (5, 0, 12), в которой мы сосредоточились на коэффициентах и обнаружили, что значимыми являются только следующие компоненты: авторегрессионные члены с лагами 1, 3, 4 и скользящие средние члены с лагами 3, 4, 11, 12.

Шаг 3: Диагностическая проверка: остаточные ошибки

Мы проверили остатки выбранной модели и провели тест Льюнга-Бокса.

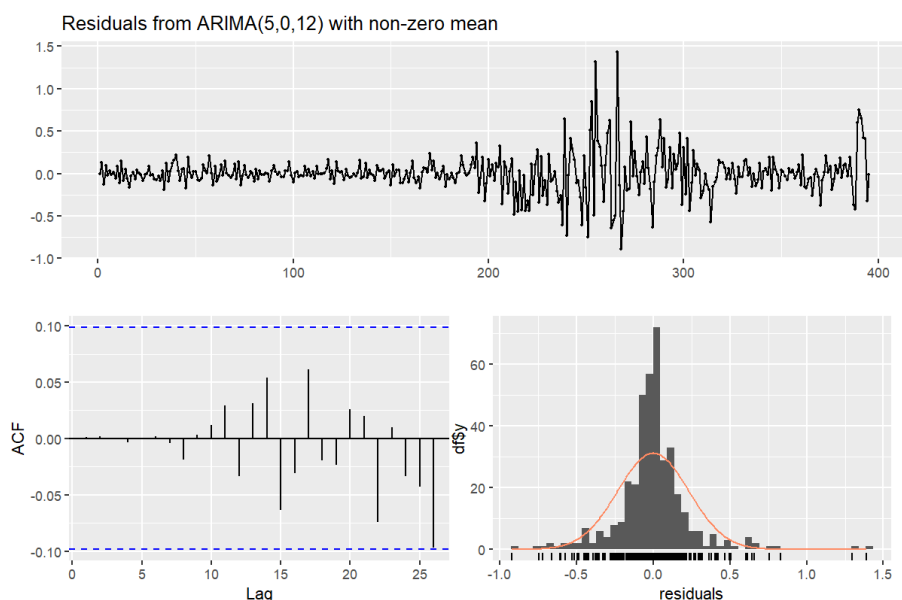


Рисунок 49: Остатки модели ARIMA (5, 0, 12)

Все пики графика ACF теперь находятся в пределах значимости, что свидетельствует об отсутствии автокорреляций.

Лjung-Box тест

H_0 : Остатки распределены независимо.

H_1 : Остатки не распределены независимо и демонстрируют последовательную корреляцию.

Результат:

Ljung-Box статистика: 2.159917

p-значение: 0.994961

Нам не удалось отвергнуть H_0 . Тест Льюнга-Бокса показывает, что остатки не имеют остаточных автокорреляций. Поэтому мы можем подтвердить, что остатки модели SARIMA (12, 0, 2)(0, 1, 2)[12] являются белым шумом.

Таким образом, теперь у нас есть сезонная модель ARIMA, которая соответствует требованиям и готова к прогнозированию.

Процесс прогнозирования

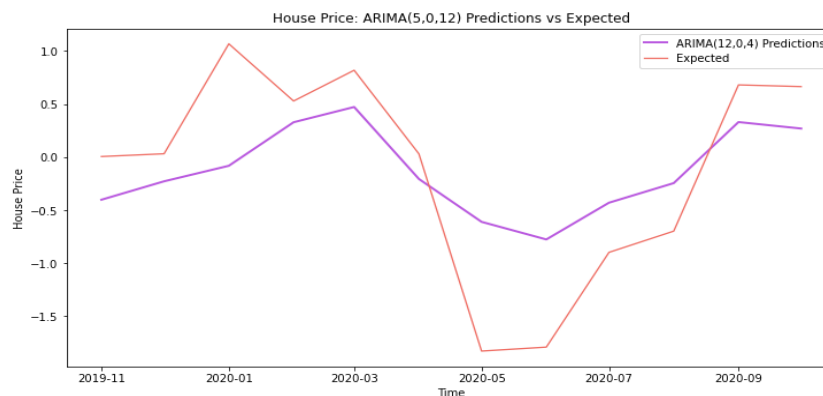


Рисунок 50: Сравнительный график ожидаемых значений и значений, предсказанных ARIMA

На сравнительном графике мы видим, что наша модель довольно хорошо следует тренду тестового набора. Прогнозы довольно точны для периодов март 2020 г. – апрель 2020 г. и август 2020 г. – сентябрь 2020 г., где есть пересечения. Самый большой разрыв наблюдается в период с апреля 2020 г. по август 2020 г., что можно объяснить тем, что это первые месяцы COVID-19, и для таких исключительных времен моделям, основанным на наборе поездов в обычное время, очень сложно предсказать такое сильное падение. (Подобный большой разрыв также наблюдался в модели нашего 1-го эксперимента, что говорит о том, что это нестабильный период, который трудно предсказать).

Закдючение

Подводя итог, в нашем исследовании нам удалось изучить поведение цен на жилье и ставок по ипотеке в США с 1987 по 2021 год, в течение которых было два периода рецессии.

Что касается ставки по ипотеке, мы применили полиномиальную регрессию, вывели из нее остатки и применили к ним дифференциацию первого порядка, чтобы очистить тренд. Нам потребовалось довольно много времени, чтобы получить результат соответствующей модели для ставок по ипотеке, поскольку из-за высокого порядка AR модель показала себя относительно хорошо.

Что касается цены на жилье, после удаления обнаруженного тренда двумя раундами дифференциации мы применили два подхода для обработки сезонности и построения моделей. Для подхода 01 мы построили гармоническую регрессию, изъяли остатки для прогнозирования значений на следующие 2 года с использованием сезонной модели ARIMA. Для подхода 02 мы дифференцировали данные в третий раз и использовали десезонализированные данные для прогнозирования значений на следующий год с использованием модели ARIMA. Сравнив эффективность двух моделей с реальными значениями, мы пришли к выводу, что сезонная модель ARIMA для остатков из гармонической регрессии не только лучше прогнозирует, но и может хорошо прогнозировать по крайней мере на 2 года, тогда как модель ARIMA для данных с разностью третьего порядка может достаточно хорошо прогнозировать только на один год.

Нам также удалось проверить три гипотезы. Первая гипотеза была отклонена, потому что, несмотря на базовую сезонность, цены на жилье фактически выросли больше в зимние месяцы и упали в летне-осенние месяцы. Вторая гипотеза была подтверждена, потому что в течение обоих периодов рецессии ставки по ипотеке всегда снижались. Третья гипотеза была отклонена, потому что, хотя цены на жилье упали во время мирового финансового кризиса в 2007–2009 годах, они выросли во время пандемии COVID-19 и в последующий период.

Наши окончательные выводы из нашего исследования таковы:

- летние и осенние месяцы являются наиболее подходящим временем года для потенциальных покупателей и владельцев домов из-за увеличения как спроса, так и предложения;
- ставки по ипотеке имеют тенденцию снижаться во время рецессий, но это не следует воспринимать слишком серьезно, поскольку в целом снижение происходит с течением времени;
- не всегда цены на жилье будут падать во время рецессий, поскольку ФРС может проводить политику разницы, поэтому покупателям следует обращать внимание на обновленную политику центрального банка.