

MAE5870-Avaliação do Comportamento da oferta de residências USA (1959-2019)

Luis Correia No. USP 745724

25 de maio de 2019

1 Descrição do Estudo

O estudo em questão tem como objetivo identificar comportamentos periódicos, sazonalidade e características da demanda por residências nos Estados Unidos. Para isso, analisaremos a oferta de imóveis novos nos últimos 51 (cinquenta e um) anos, isto é, desde Jan/1968 até Abr/2019.

A fonte de dados é o *United States Census Bureau*, órgão especializado em pesquisas do governo americano que, a partir do governo Obama passou a disponibilizar informações detalhadas sobre o desempenho da economia americana.

Selecionamos das bases de dados as residências consideradas finalizadas, isto é, com status **COMPLETIONS** e de abrangência nacional, **US**.

```
US_HCompl <- which( (HData$cat_code=="COMPLETIONS") &
                    (HData$dt_code=="TOTAL") & (HData$et_code=="E_NONE") &
                    (HData$geo_code=="US"))
USHousing <- zoo(HData[US_HCompl,7], as.yearmon(HData[US_HCompl,1]))
tsUSHousing <- as.ts(USHousing)

autoplot.zoo(tsUSHousing)
```

Através da simples avaliação a série de dados, observa-se ciclos de crescimento de aproximadamente 5-7 anos no período entre 1968 a 1990, e a partir deste período um intenso crescimento da oferta com duração de cerca de 15 anos, quando houve a grande expansão do mercado de *Real State* americano devido à grande oferta de crédito.

Este período de expansão sofre então em 2008 o grande revés relacionado à crise dos sub-primes (*Subprime Mortgage Crisis*) onde os créditos destinados ao mercado imobiliário simplesmente evaporaram deixando uma grande parte dos mutuários sem poder financeiro para quitar suas dívidas.

A partir de 2011, o mercado imobiliário americano estabilizou-se passando novamente a fases de crescimento contínuo, já a partir do 2o. mandato do Presidente Barack Obama e em continuidade ao Governo Trump.

2 Análise Estatística

Vamos inicialmente verificar o grau de correlação em que a quantidade de interesse **X**: *Número de residências finalizadas* por período de tempo

```
par(mfrow = c(1,2)) # set up the graphics

acf(tsUSHousing, type="correlation")
pacf(tsUSHousing)
```

As funções de Auto-Correlação e Auto-Correção Parcial demonstram que as observações têm uma alta correlação com as medidas do período anterior sugerindo que o número de imóveis disponibilizados em cada período é ajustado à demanda apresentada no período anterior.

Decompondo-se a série de interesse, obtivemos os seguintes resultados:

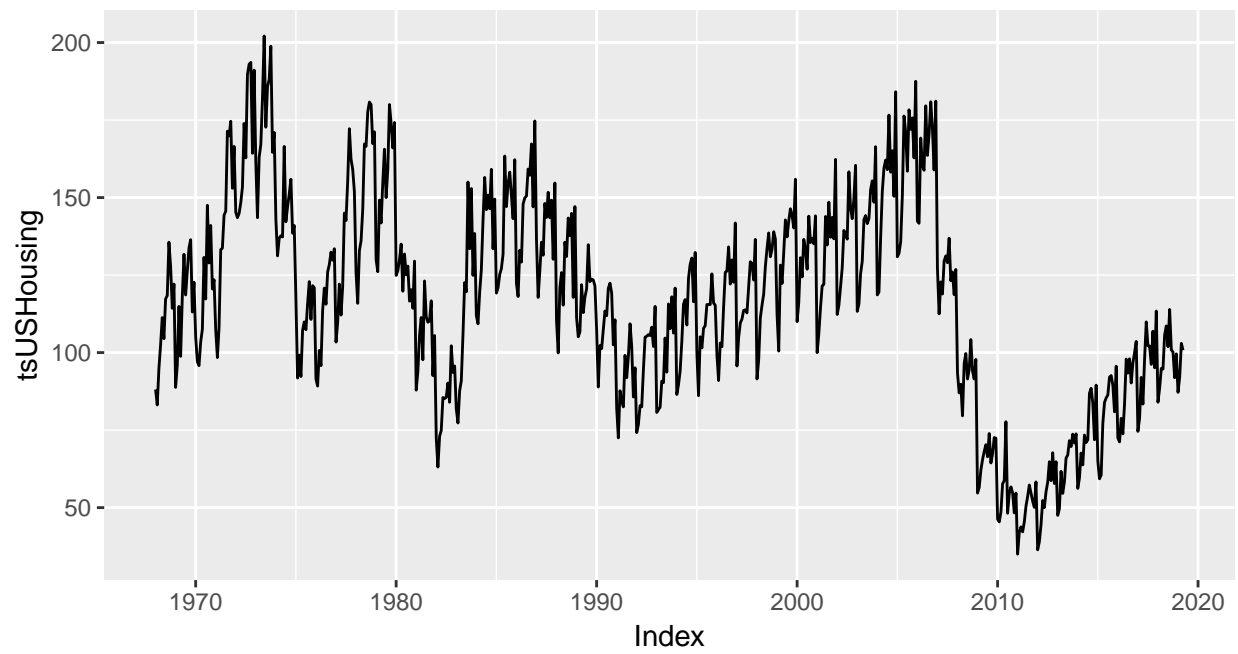


Figure 1: Oferta de Imóveis USA (1968-2019)

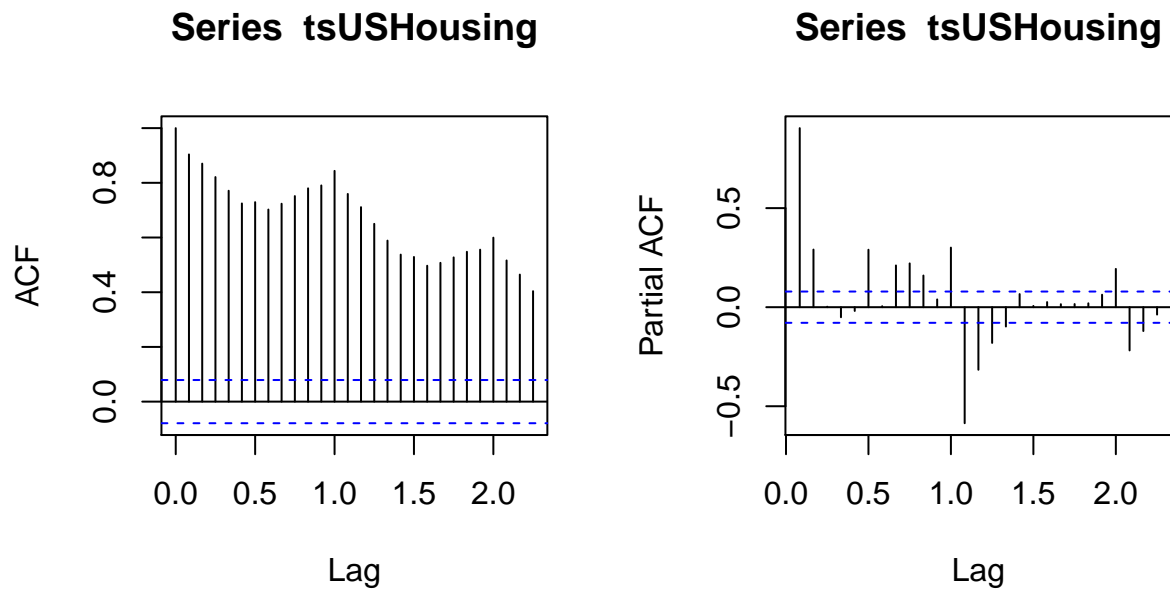


Figure 2: Auto-Correlações Geral e Parcial x_t

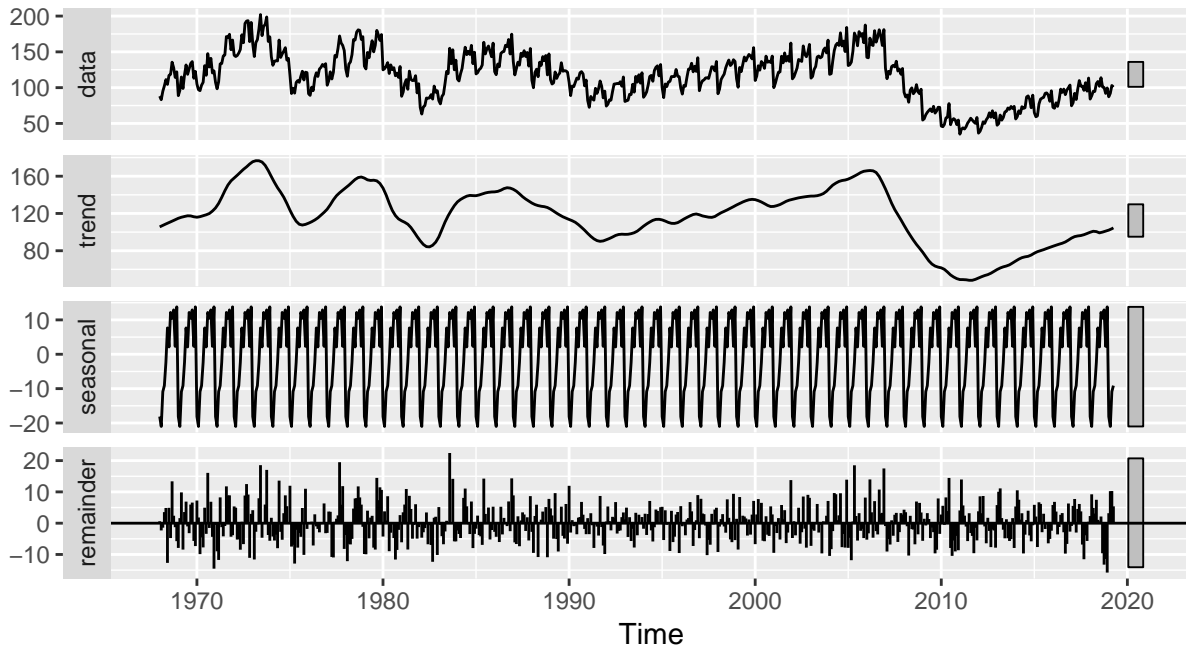


Figure 3: Decomposição da Série - Housing-US

```
tsUSHousing.comp <- stl(tsUSHousing,"periodic")
autoplot(tsUSHousing.comp)
```

Aqui pode ser observada uma forte componente sazonal de 12 meses, os ciclos de 05-07 anos entre 1968 e 1992, bem como a grande expansão do mercado imobiliário americano a partir de 1993 até a crise de 2008.

3 Análise Exploratória

3.1 Avaliação das diferenças com lags variáveis

Calculamos as diferenças da série z_1 com $\text{lag}=1$ para remover a forte tendência apresentada na série em estudo e também como forma de avaliar se a variância pode ser considerada constante para a modelagem desejada.

As séries transformadas consistem em aplicar as diferenças de $\text{lag}=1$ ($z_t = x_t - x_{t-1}$) e, aplicando sobre a série z_t a diferença de $\text{lag}=12$ para eliminar a sazonalidade da série, ou seja, fazendo $w_t = z_t - z_{t-12}$.

Os gráficos das séries resultantes demonstram que a série w_t possui uma menor variabilidade ao longo do tempo, sendo mais estável que a série z_t .

```
autoplot(z1_12, ylab="w_t")
```

A verificação da sazonalidade fica clara através da plotagem da série z_t por período (meses) indicando que há uma um comportamento periódico em função do lag anual, conforme mostra o gráfico abaixo, com diferentes níveis dependendo do mês, maiores concentrações nos meses de finais de trimestre (Q1, Q2 e Q4).

```
monthplot(z1)
```

Para isso, vamos aplicar sobre a série $z_t = x_t - x_{t-1}$ a fazendo $w_t = z_t - z_{t-12}$ de forma a remover a sazonalidade e, sobre esta série prosseguir com o ajuste do modelo.

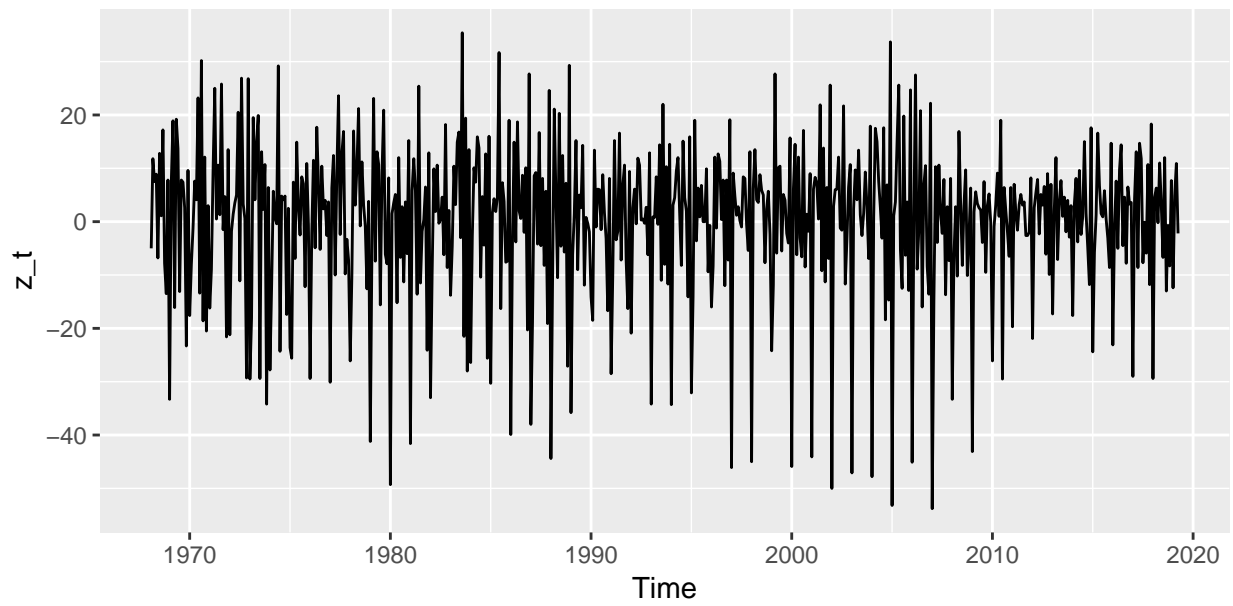


Figure 4: Autoplot da série z_t

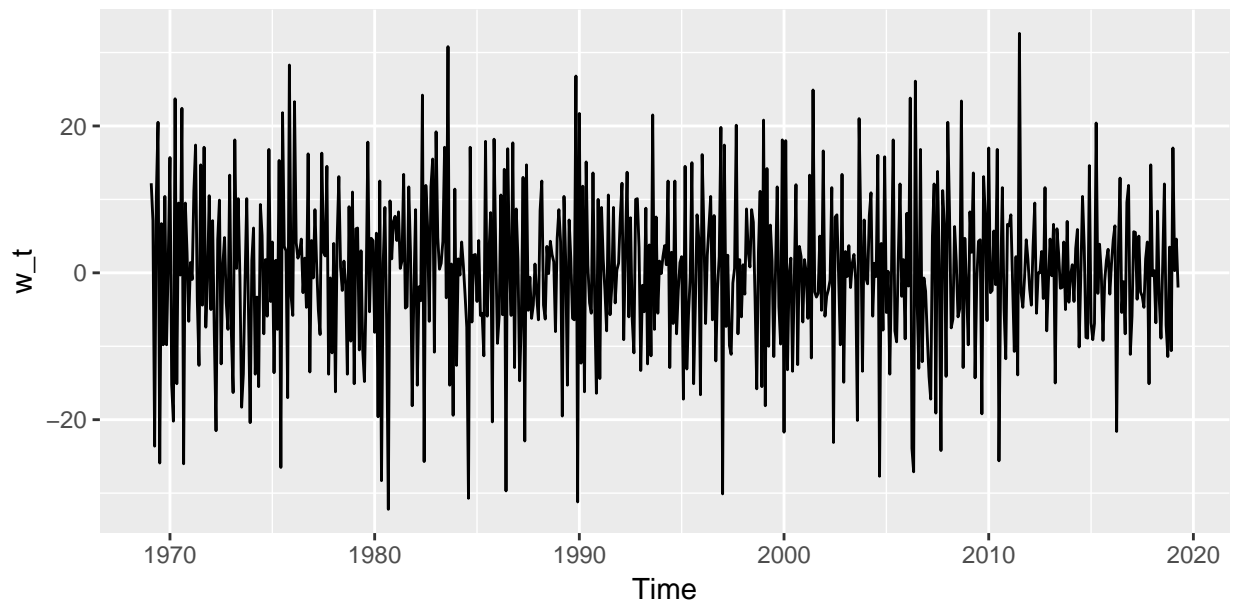


Figure 5: Autoplot da série w_t

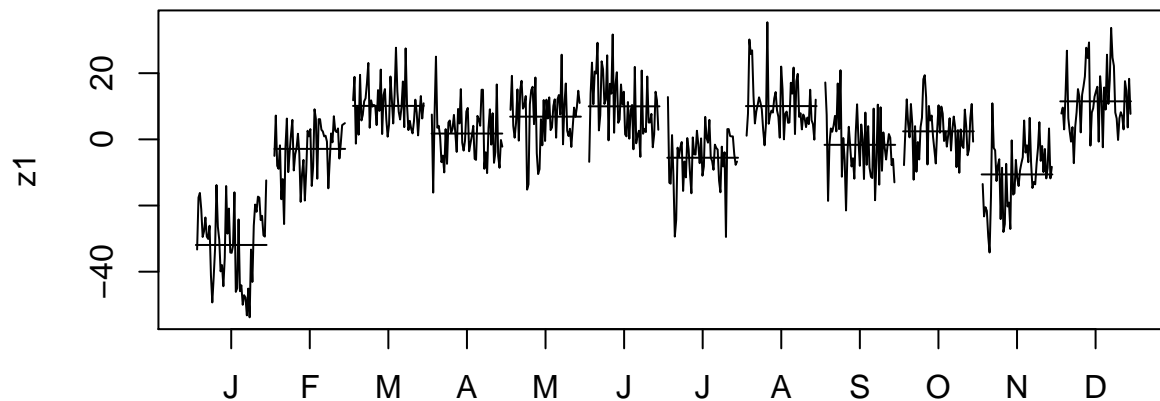


Figure 6: Monthplot da série z_t

```
monthplot(z1_12,ylab=expression(w_t))
```

Fazendo a transformação e utilizando a série w_t reduzimos a variabilidade da série e iremos prosseguir com ela para a modelagem. Sua decomposição (abaixo) demonstra que a tendência foi eliminada, restando as características próprias da série e que podem suscitar os comentários que auxiliem na interpretação do modelo ajustado.

Notemos também que a sazonalidade fica agora padronizada e denota o comportamento da oferta de imóveis durante o ano, com maiores concentrações de oferta no 2o. trimestre (Q2) e depois no último trimestre do ano (Q4), período em que normalmente os anos fiscais das empresas é encerrado, e onde normalmente há a distribuição de bônus e gratificações aos funcionários, favorecendo a comercialização desses imóveis.

Verifica-se também que o resto resultante da decomposição da série fica relativamente estável e com média 0 (zero), sem variações expressivas que possam sugerir heterocedasticidade da série.

```
z1_12.comp <- stl(z1_12,"periodic")
autoplot(z1_12.comp)
```

3.2 Avaliação das Funções de Correlação

As funções de Auto-Correlação e Auto-Correção Parcial da série w_t demonstra demonstram que as observações têm uma alta correlação com as medidas do período anterior sugerindo que o numero de imóveis disponibilizados em cada período é ajustado à demanda apresentada no periodo anterior.

```
par(mfrow = c(1,2)) # set up the graphics

acf(z1_12, type="correlation")
pacf(z1_12)
```

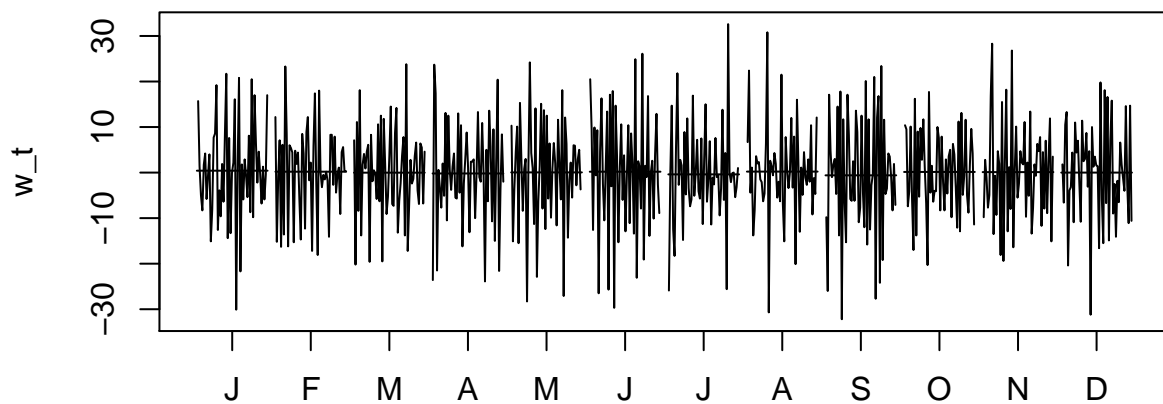


Figure 7: Monthplot da série w_t

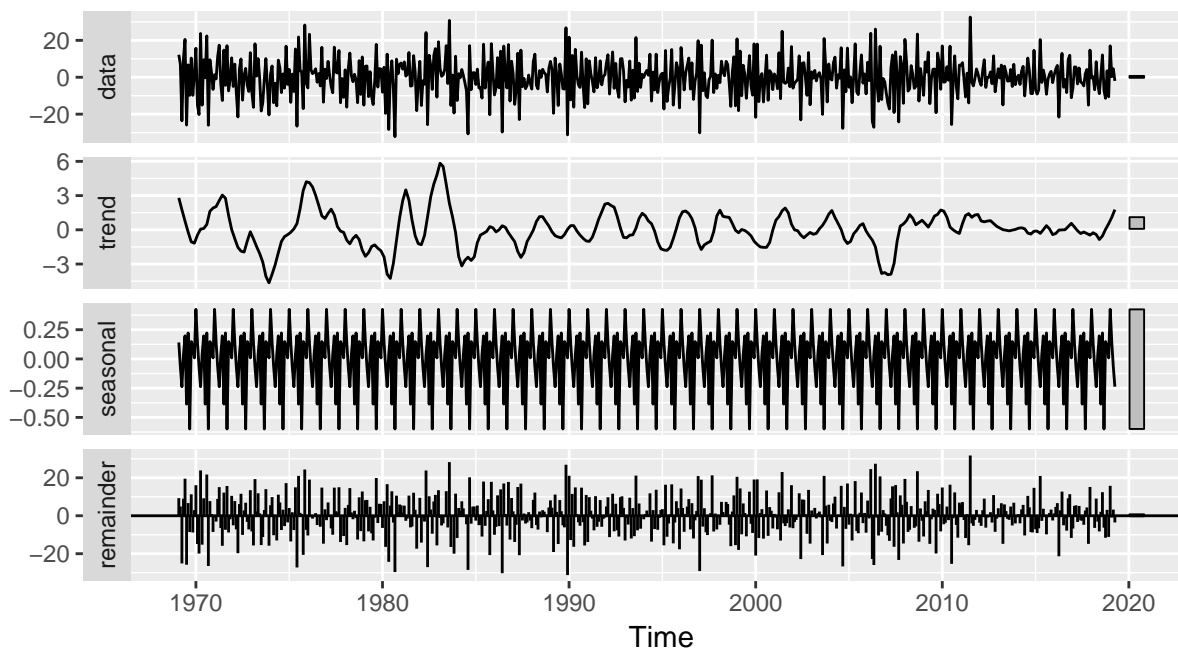


Figure 8: Decomposição da Série w_t

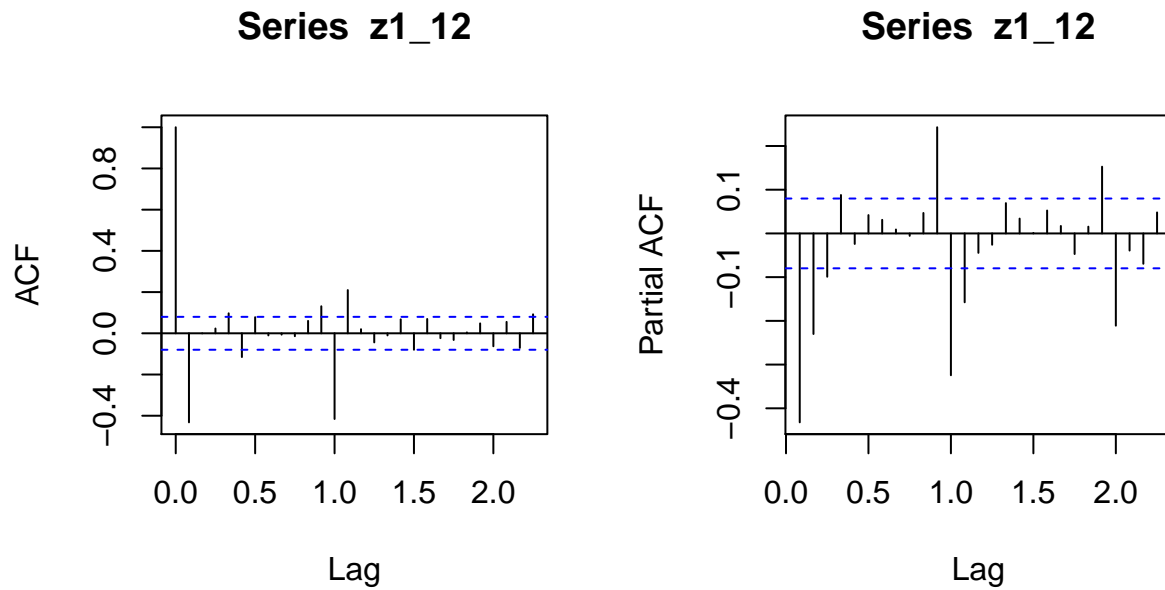


Figure 9: Auto-Correlações Geral e Parcial de w_t

4 Ajuste do Modelo

Para a série w_t utilizaremos a função `auto.arima` do R para ajustar o modelo

O modelo é **SARIMA(4,0,2)(0,0,2)[12]** com média $\mu = 0$, bem como a visão geral de sua análise de resíduos, e é dado por:

```
summary(USHousing.fit1_12)
```

```
## Series: z1_12
## ARIMA(4,0,2)(0,0,2)[12] with zero mean
##
## Coefficients:
##      ar1      ar2      ar3      ar4      ma1      ma2      sma1      sma2
##     -0.222  0.4246  0.3009  0.2475 -0.2890 -0.4926 -0.7566 -0.1140
## s.e.   0.180  0.1051  0.0711  0.0435  0.1843  0.1697  0.0410  0.0415
##
## sigma^2 estimated as 59.17: log likelihood=-2089.8
## AIC=4197.6  AICc=4197.9  BIC=4237.22
##
## Training set error measures:
##              ME      RMSE      MAE MPE MAPE      MASE      ACF1
## Training set -0.04590013 7.640825 6.062775 NaN  Inf  0.4145187 -0.001907732
```

```
checkresiduals(USHousing.fit1_12)
```

```
##
## Ljung-Box test
##
## data: Residuals from ARIMA(4,0,2)(0,0,2)[12] with zero mean
## Q* = 9.4271, df = 16, p-value = 0.8948
```

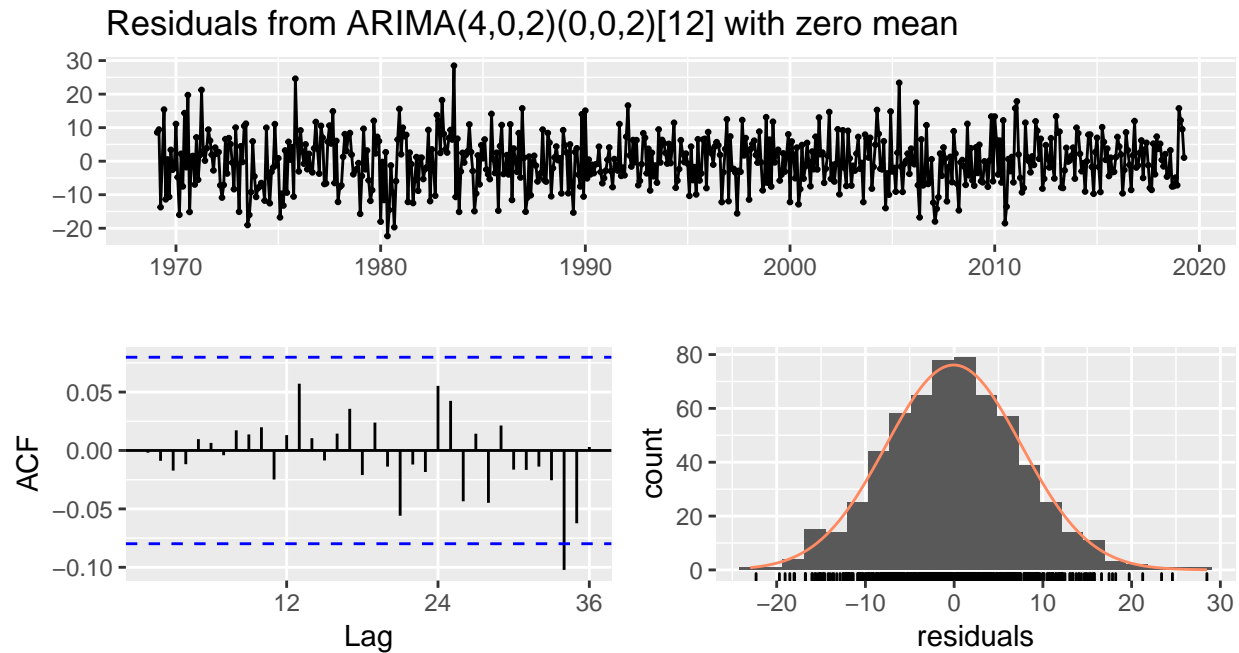


Figure 10: Análise de Resíduos do Modelo Ajustado

```
##
## Model df: 8.    Total lags used: 24
```

O teste de McLeod-Li nos mostra que não há nenhum lag crítico em que a suposição de não-autocorrelação possa não estar atendida.

```
TSA::McLeod.Li.test(USHousing.fit1_12)
```

De maneira análoga, a verificação de normalidade dos resíduos parece ser atendida. Podemos aferir isto através do gráfico Q-Q Plot.

```
par(mfrow = c(1,2)) # set up the graphics

plot(USHousing.fit1_12$residuals,ylab="Residuals w_t")

qqnorm(USHousing.fit1_12$residuals)
qqline(USHousing.fit1_12$residuals)
```

Neste sentido, para o modelo ajustado, podemos considerar que os lags são não correlacionados, confirmando as condições de regularidade exigidas para o ajuste do modelo estão satisfeitas.

Efetuamos o reajuste do modelo utilizando o método iterativo da função `sarima()` do R e os resultados foram os mesmos obtidos com a função `auto.arima()`

```
USHousing.fitSARIMA <- sarima(z1_12, 4, 0, 2, P=0, D=0, Q=2, S=12,no.constant=FALSE)
```

```
summary(USHousing.fitSARIMA$fit)
```

```
##
## Call:
## stats::arima(x = xdata, order = c(p, d, q), seasonal = list(order = c(P, D,
##     Q), period = S), xreg = xmean, include.mean = FALSE, transform.pars = trans,
##     fixed = fixed, optim.control = list(trace = trc, REPORT = 1, reltol = tol))
```

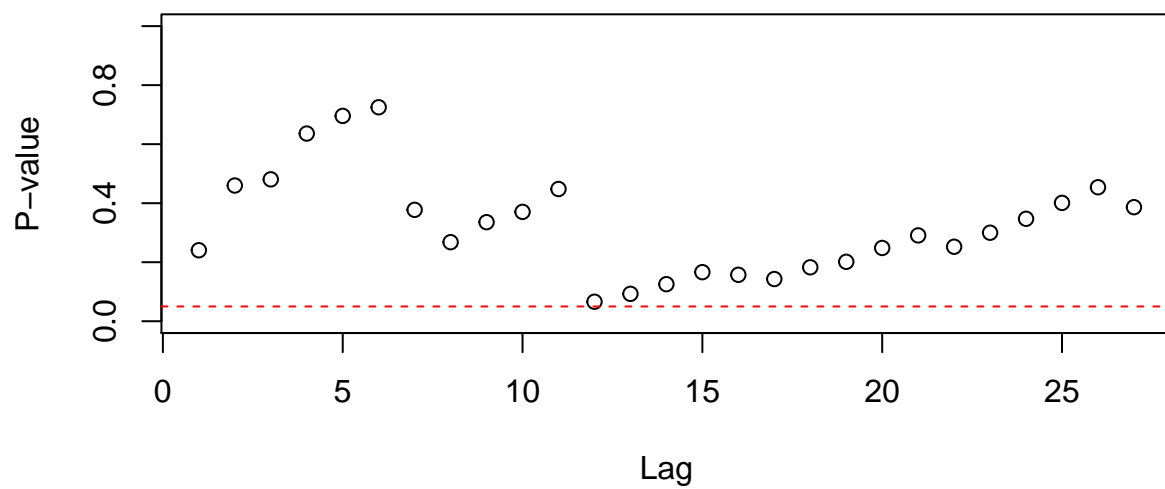



Figure 11: McLeod-Li Analysis para o Modelo Ajustado

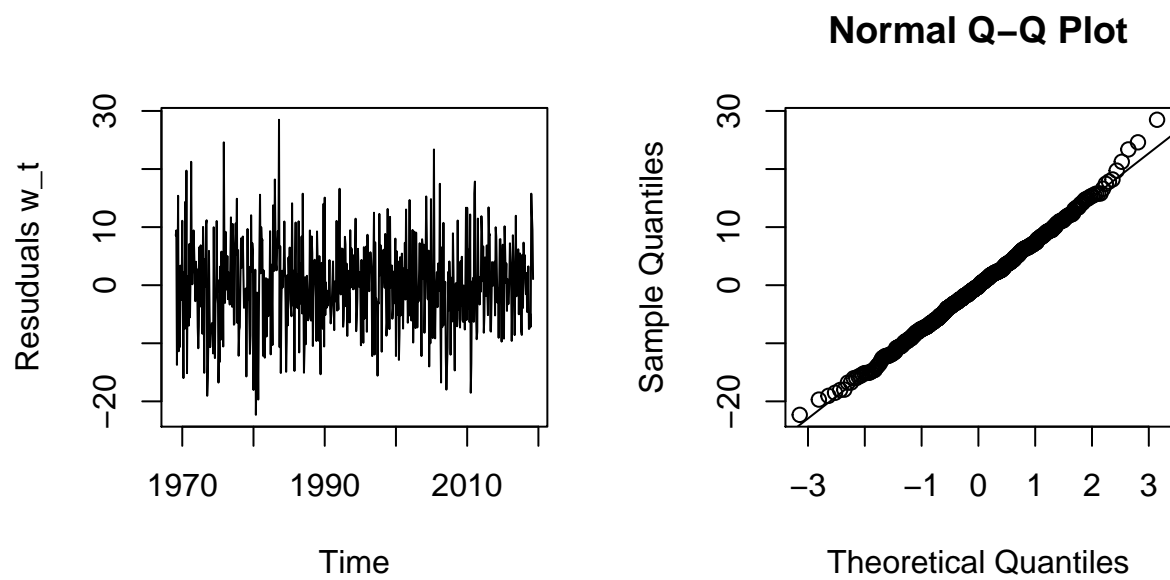


Figure 12: Análise de Resíduos do Modelo Ajustado

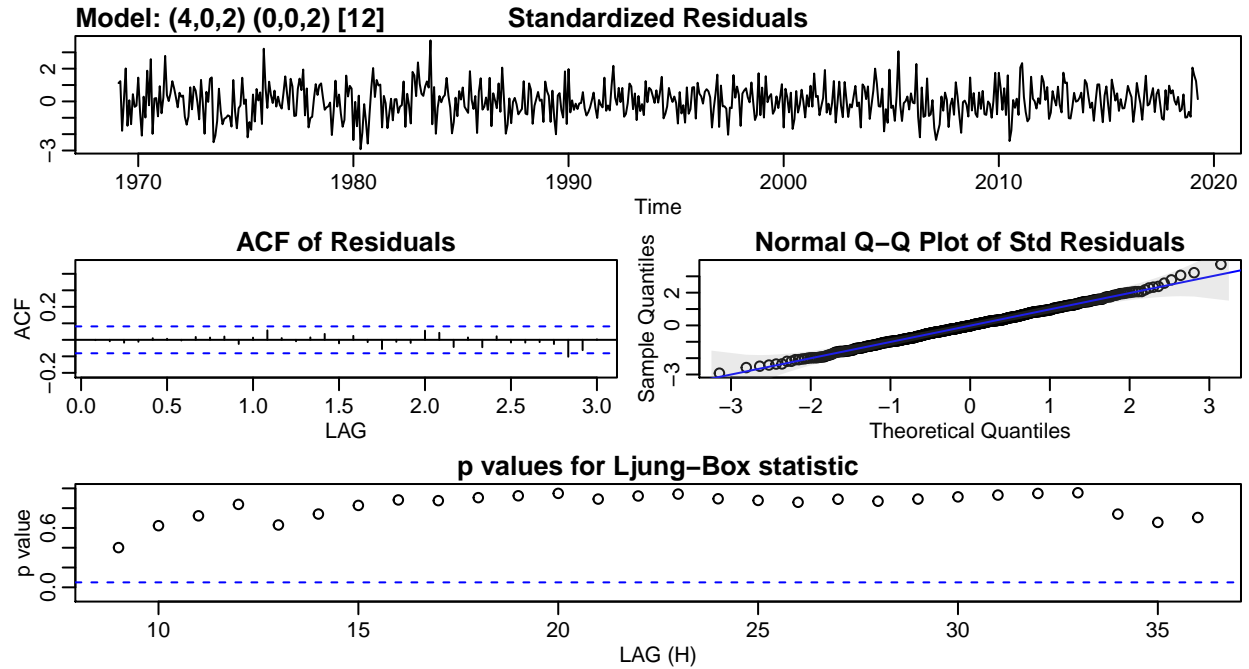


Figure 13: Modelo SARIMA(4,0,2)(0,0,2)[12]

```
##
## Coefficients:
##      ar1      ar2      ar3      ar4      ma1      ma2      sma1      sma2
##     -0.222  0.4246  0.3009  0.2475  -0.2890  -0.4926  -0.7566  -0.1140
## s.e.   0.180  0.1051  0.0711  0.0435   0.1843   0.1697   0.0410   0.0415
##      xmean
##     -0.0013
## s.e.   0.0418
##
## sigma^2 estimated as 58.38:  log likelihood = -2089.8,  aic = 4199.6
##
## Training set error measures:
## Warning in trainingaccuracy(f, test, d, D): test elements must be within
## sample
##
##      ME RMSE MAE MPE MAPE
## Training set NaN  NaN NaN NaN  NaN
```

5 Análise de Resíduos

Analisando os resíduos da série ajustada, podemos observar que são distribuídos normalmente com média $\mu = 0$ e variância σ^2 constante. Desta forma a suposição de $a_t \sim N(0, \sigma^2)$ é plausível.

Para garantir isto, avançaremos na análise de resíduos plotando seu gráfico de Auto-Correlações e executando o Teste de Box-Ljung para heterocedasticidade.

```
autoplot(USHousing.errors, ylab="a_t")
```

A avaliação dos Resíduos confirmam a suposição de homocedasticidade e não auto-correlação.

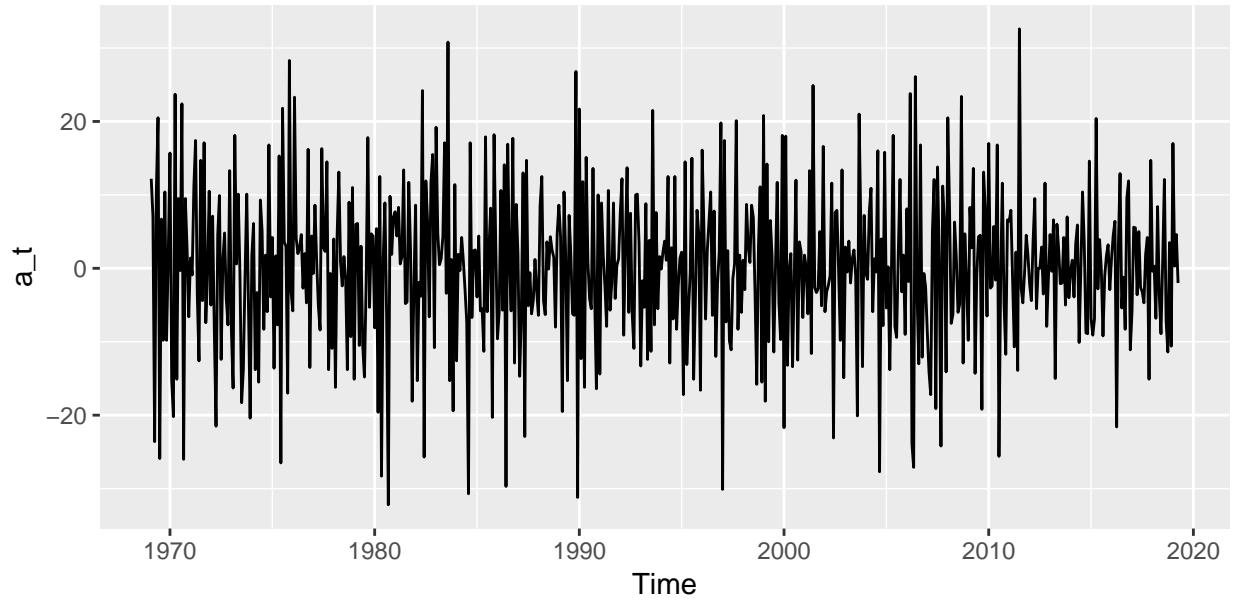


Figure 14: Análise de a_t para o Modelo Ajustado

```
tsdiag(USHousing.fit1_12)
```

Além disso, o teste de Box-Ljung apresentou p -value não significativa (com $p = 0.8916$) ou seja, aceitamos a hipótese nula de que a variância dos erros é constante.

```
Box.test(USHousing.fit1_12$residuals, type="Ljung", lag=3, fitdf=1)
```

```
##
## Box-Ljung test
##
## data: USHousing.fit1_12$residuals
## X-squared = 0.22946, df = 2, p-value = 0.8916
```

6 O Modelo Final Ajustado

O modelo ajustado pode ser representado da seguinte forma:

$$(1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - \phi_4 B^4)w_t = (1 + \Theta_1 B^{12} + \Theta_2 B^{24})(1 + \theta_1 B + \theta_2 B^2)a_t \quad (1)$$

Substituindo-se os valores obtidos no ajuste e fazendo as simplificações necessárias, o modelo final pode ser escrito da seguinte forma:

$$(1 - 0.222B - 0.4246B^2 - 0.3009B^3 - 0.2475B^4)w_t = (1 - 0.7566B^{12} + 0.1140B^{24})(1 + 0.2890B + 0.4926B^2)a_t \quad (2)$$

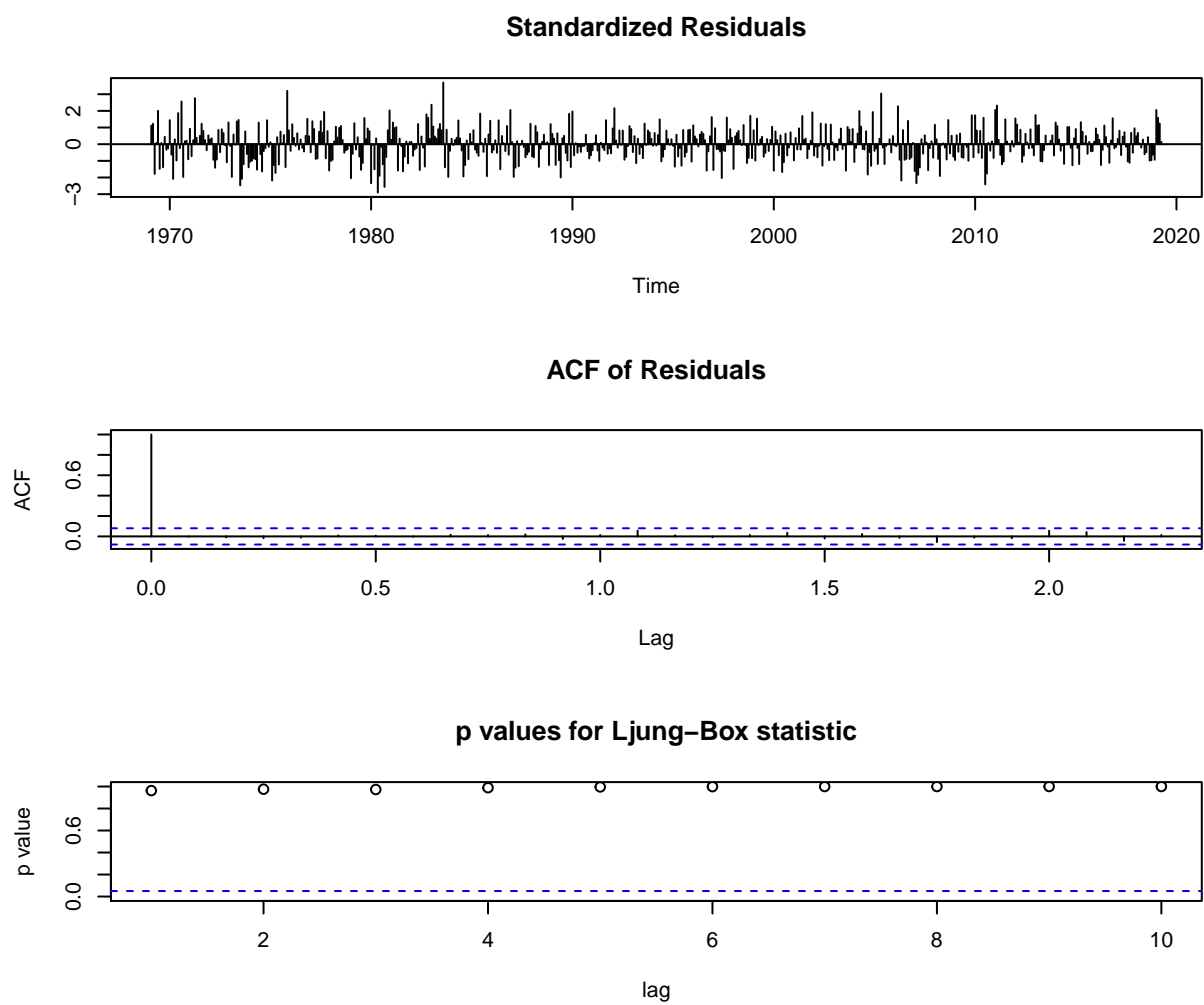


Figure 15: Diagnóstico do Modelo Ajustado

7 Conclusões

O modelo ajustado sugere que as quantidades de oferta de imóveis nos Estados Unidos pode ser representado por um modelo Sazonal Misto Auto-Regressivo e Médias Moveis (SARIMA) que depende das quantidades dos lags 1 a 4 e efeito sazonal de 12 (doze) meses. Importante notar que neste modelo esta dependência é positiva no lag=1 e negativa nos lags=2 a 4.

A parte aleatória sugere uma combinação linear complexa que depende dos lags=24 e 48, ou seja, o ciclo de crescimento/decrescimento é repetido a cada 04 (quatro) anos aproximadamente, consoante com os períodos eleitorais americanos, onde incentivos e políticas economicas específicas podem influenciar diretamente o mercado imobiliário.

O mercado imobiliário mundial, em especial o americano, foi fortemente afetado pela crise dos sub-primes em 2007-2008, portanto o ajuste deste modelo pode ser mais complexo de interpretar pois há 02 períodos onde o cenário economico comportava-se de forma bastante distinta.

Estudos adicionais pode incluir algumas ações não cobertas por este trabalho mas que podem suscitar novas interpretações adicionais:

- Analisar separadamente os períodos P1:1968-1991, P2:1992-2007, P3:2008-2011 e P4:2012-2019;
- Ajustar modelos mais simples e interpretáveis para cada um desses períodos;
- Analisar o paralelo entre as políticas econômicas desses períodos em relação ao impacto no mercado imobiliário;
- Sugerir ações que poderiam atenuar/reduzir tendências de declínio da oferta de imóveis como mitigadores de riscos ao investimento no setor.