Stockholms universitet

Kandidatuppsats

Statistiska institutionen

Bachelor thesis, Department of Statistics

Nr 2021:x

” The Effects of the Covid-19 Pandemic on the Swedish Housing Market”

“Effekter av Covid-19 Pandemin på den Svenska Bostadsmarknaden”

Emin Mahmutovic, Benjamin Sivac, Hampus Stridsman

Självständigt arbete 15 högskolepoäng inom Statistik III, HT2021

Handledare: Pär Stockhammar

1. Introduction

The spread of the Covid-19 virus was officially declared a pandemic by the ‘World Health Organisation’ (WHO) in March 2020. The implications of this pandemic have so far included nation-wide lockdowns, extensive quarantining procedures, and a multitude of other consequences. As a result, gradually more people have begun working from the comfort of their homes, leading to a change in consumer preferences as people are adapting to a more remote lifestyle. This includes an increase in demand of essential goods, but also mid-to-long term investments in housing. It is hypothesised that the quarantine lifestyle has led to an increase in the demand of houses and condominiums outside of the big cities where costs of living are lower, and it is more comfortable, peaceful, and enjoyable to work.

The goal of the study is to compare how the housing market has changed in different parts of the country. The hypothesis is that the development of housing prices have changed after March 2020, and that sparsely populated areas have seen a more rapid increase in prices compared to the big cities.

The expectation is that there has been an overall increase in housing prices throughout the country. One way to test for it is by constructing forecast models and comparing them to the actual price development since March 2020. If there is a clear break between the forecasts and the actual prices, it would indicate that there has been a change in structure. If there hasn’t been a change in price development, it is expected that the forecast model will capture the actual values.

2. Data

2.1. Description

When allowing a set of variables to move and change through time, the observed data is called time-series data, the variables are part of a time-series process. Measurements and studies on such data is called time-series analysis. Applications of time-series analysis include studying the past and predicting the future. The two most prominent methods of time-series analysis are linear regression estimation and ARIMA processes.

2.2. Study variables

The data surrounding the study variables of this report are provided by Hans Flink at 'Mäklarstatistik AB', containing monthly sales and prices of houses and condominiums in all 290 Swedish municipalities from the past 4 years (2017–2021). The data are split into two parts, condominiums, and villas respectively. The condominiums data focuses on the variable $\mathrm{Price/m^2}$ and is reported as the mean price in each municipality given in SEK. It also includes all sales in each municipality.[1]

The villas/houses data contains the variable 'purchase price coefficient', a quota between the purchase price of a house and its taxation value. Usually, the taxation value equals 75% of the current market value of the house. The coefficient can be used to tell if one has over- or underpaid for a house. An increase in the coefficient value indicates that prices are rising in relation to the taxation value.[2] In this study the variable purchase price coefficient will be shortened to *P/T*.

2.3. Dividing the datasets

In order to analyze the data, the datasets are divided into groups which are defined based on a report made by 'Swedish Agency for Growth Policy Analysis' from 2014. It proposes several good options for grouping. The best option groups based on geographic location, area size, population size, travel distance to a large district and the level of urbanization. This method generates three group categories: *Rural municipalities*, *dense municipalities*, and *metropolitan municipalities*, referred to as group 1, 2 and 3 respectively. With 290 municipalities in Sweden, it comes out to 130, 131 and 29 municipalities in each group respectively. [3]

Alternatively, using 'Statistics Sweden's' categorization of metropolitan areas, there would be 51 metropolitan municipalities and 239 municipalities which fall into the 'other' group. By using this method, the 'other' group becomes very heterogenous on several levels and generates much variance. Therefore, the first grouping method is more optimal.[4]

Furthermore, each dataset is divided into three different time periods. The first time period is the complete span of the time-series, 48 months. March 2020 is when Covid-19 was declared

---

[1] Svensk Mäklarstatistik, 2021. Prisutveckling bostadsrätter.
https://www.maklarstatistik.se/omrade/riket/#/bostadsratter/48m-prisutveckling
[2] Svensk Mäklarstatistik, 2021. Prisutveckling villor.
https://www.maklarstatistik.se/omrade/riket/#/villor/48m-prisutveckling
[3] Myndigheten för tillväxtpolitiska utvärderingar och analyser. 2014. Bättre statistik för bättre regional – och landsbygdspolitik. Östersund: Tillväxtanalys.
https://www.tillvaxtanalys.se/download/18.62dd45451715a00666f1f3a9/1586366195534/rapport_2014_04_rev1.pdf
[4] Statistikmyndigheten. Företeckning över storstadsområden med ingående kommuner I alfabetisk ordning.
https://www.scb.se/contentassets/c4b8142033a9440ca53725ca32321a74/storstadsomr.pdf

a global pandemic, the period before this, October 2017–March 2020 is defined as the period before covid. March 2020–September 2021 is defined as the period after covid.

2.4. Macroeconomic variables

Macroeconomics is the study of variables that affect the economics of a society at a grand scale, and the interplay between those variables. Such variables are vital to this analysis for possibly identifying non-trivial relationships and connections between other variables and for theoretical reasoning, which might amplify the complexity and depth of the study.

2.4.1. Inflation

Inflation is defined as the continuous increase in price. Consequently, this leads to a steady decrease in the purchasing power of money. Inflation is generally expressed in terms of inflation rate, the change in inflation from one moment to another given in per cent. The inflation rate is important because it influences unemployment rates and disposable income among other things, which are important to show how sales will develop with time. The inflation rate is measured using *CPI*, *consumer price index*, an index which measures changes prices of private domestic consumption. Though there are variants of this index that account for other types of consumption as well.

In this study, monthly CPI values from October 2017 to September 2021 are used, a total of 48 values entries. The base value of the index is in 1949, however, it is adjusted to 2017 for the sake of relevancy. [5] In this study, the variable CPI is used as inflation.

2.4.2. Unemployment

Unemployment is defined as the ratio between people that are currently not employed but are able to work and the number of people in total that are able to work. People that are not able to work, such as children, senior citizens, and sick people are disregarded. Unemployment has an intricate relationship with inflation, which is also dependent on interest rate. Levels of unemployment increase if overall price levels increase while disposable income remains the same. High levels of unemployment would decrease nominal wages, which would also decrease inflation rate. Labor force surveys conducted by Statistics Sweden measures unemployment between the ages of 15 to 74. [6]

Data on unemployment is supplied by 'Ekonomifakta'. It shows monthly unemployment in the last 4 years. According to 'Ekonomifakta', in January 2021, 'Statistics Sweden' changed their survey and therefore reports a timeseries break, noting that the data before and after the changes should not be compared. Inspecting the data there does not appear to be any large changes in unemployment for the period. Any changes seem to be small enough for us to continue to use the data.[7] The unemployment variable will be referred to as '*Unem*' in this study.

2.4.3. Economic tendency indicator

In Sweden, 'The National Institute of Economic Research' performs a monthly survey called the 'Economic Tendency Survey', a summary of the economy in the eyes of firms and consumers. This summary is presented as an indicator called the 'Economic Tendency Indicator'. There are several variants of this indicator, but from the consumers point of view,

[5] Statistiska Centralbyrån, 2021. Konsumentprisindex. https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__PR__PR0101__PR0101A/KPItotM/#

[6] Statistiska Centralbyrån, 2021. Arbetslösa 15–74 år (AKU) efter arbetslöshetstidens längd, kön och ålder. https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__AM__AM0401__AM0401L/NAKUArblosaTM/#

[7] Ekonomifakta, 2021. Arbetslöshet. https://www.ekonomifakta.se/fakta/arbetsmarknad/arbetsloshet/arbetsloshet/?graph=/25554/1,2/2017-/

it is the sum of the view of personal and national economic situation 12 months prior and future respectively and the addition of current consumption of capital goods. The purpose of the indicator is to be highly correlated with GDP (Gross domestic product) so that high values of the indicator suggest a fast-growing and healthy economy while lower values would suggest the opposite. It is given in index form, it is normally distributed with mean of 100 and standard deviation of 10, standardized and seasonally adjusted. It contains monthly data for the past 4 years. This study will be using the variable *'Household Confidence Indicator'* (*HCI*), which is a part of the 'Economic Tendency Indicator'.[8]

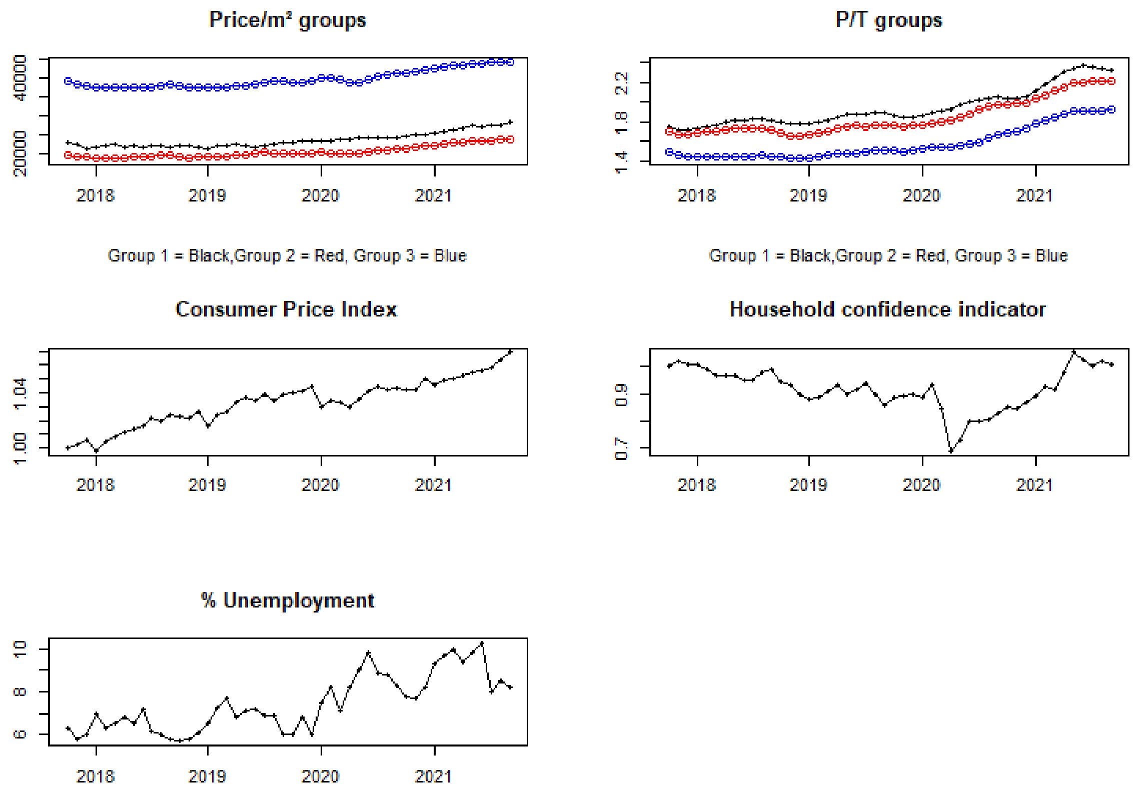


Figure 2.4.1 Respective research- and control variables used within this study

[8] Statistiska Centralbyrån, 2021. Konjunkturbarometer. https://www.scb.se/hitta-statistik/temaomraden/sveriges-ekonomi/konjunkturindikatorer/konjunkturbarometer/?showAllContentLinks=True#113147, Konjunkturinstitutet, 2021. Metodbok för Konjunkturbarometer. https://www.konj.se/metodbok, Konjunkturinstitutet, 2021. The Economic Tendency Survey. https://www.konj.se/english/about-nier/surveys.html

3. Models

3.1. Ordinary Least Squares (OLS)

For linear regression estimation, we use OLS (Ordinary least squares) estimation, which is based on minimizing the squared difference in distances between estimated model and sampled data. In this study, several multiple linear regressions will be estimated using a plethora of regressors at once to see which of them have the most influence and therefore relevance.

Though the OLS coefficient estimates are very important, dedicating space to explaining and deriving them is not very relevant or contributing to the understanding of this method, therefore, no equations are demonstrated.

3.1.1. Gauss–Markov theorems and associated tests

The OLS estimation of linear regression has the lowest sample variance and unbiased estimators if 5 conditions/assumptions are satisfied, which are collectively known as the Gauss–Markov theorem. Unbiasedness of an estimator is another way of saying that the expected value of the mean equals the actual mean after repeated sampling. Each assumption and the appropriate test of its validity is discussed below. [9]

3.1.1.1. Linear in parameters

The time-series process follows a linear model.

$$y_t = \beta_0 + \beta_1 x_{t1} + \cdots \beta_k x_{tk} + \varepsilon_t$$

Here, $t = 1,2, \ldots, n$ and n is the number of time periods/observations and u_t is the error in each of those periods.

One way to validate if the hypothesized regression model is linear is a 'regression specification error test', or 'RESET', which compares two functional forms of a regression and deems which is more appropriate for the data. Usually, the hypothesized functional form is compared to a similar model which differs by only a couple of polynomial terms. Whichever is statistically more significant is favored for analysis. [10]

$$y_t = \beta_0 + \beta_1 x_{t1} + \cdots \beta_k x_{tk} + \delta_1 \hat{y}^2 + \delta_2 \hat{y}^3 + \varepsilon_t$$

3.1.1.2. No perfect collinearity

The independent variables of the time-series process are neither constant nor consist of a perfect linear combination of other independent variables throughout time.

To test it, a 'variance inflation factor' (VIF) is used, whose size can approximately estimate the presence of multicollinearity. Rule of thumb says that if the statistic is larger than 10 (sometimes 5), there is serious multicollinearity. Another indicator of multicollinearity is the adjusted R-squared statistic that takes into consideration the number of regressors and their influence, besides already estimating explained variance in data. The VIF is calculated on each regressor

$$VIF_i = \frac{1}{1 - R_i^2} \; where \; i = 1,2 \ldots, k \; regressors.$$

In practice, we use the VIF to decide upon the degree of multicollinearity, but its square root has a different interpretation. Compared to if a regressor has 0 correlation with the other

[9] Wooldridge, 2016, pp. 339–344
[10] Wooldridge, 2016, pp. 277–278

regressors, the square root of the VIF is a scalar multiplied on the standard error, the higher the multicollinearity the higher the scalar and therefore higher standard error. [11]

3.1.1.3. Zero conditional mean

In each period t, the conditional expected value of an error equals 0 given all explanatory variables for all time periods. It means that across the whole time-series process, the mean of errors is equal to 0 and that there is no correlation between the errors and explanatory variables. The capital X represents the collection of all regressors.

$$E(\varepsilon_t|X) = 0 \; for \; t = 1,2,\dots,n$$

The easiest way to confirm this assumption is by plotting residuals against fitted values of a regression to see how errors develop through time. If the relationship seems to be fairly constant and around 0, the assumptions should be considered to be satisfied. Another indicator of this is to compare the regression model with itself without a constant and see how the parameter values and R-squared values change. If they don't change significantly, the assumption is most likely satisfied.

3.1.1.4. Homoskedasticity

In each period t, the conditional variance of errors given all explanatory variables is constant. In other words, the variance is constant throughout the time-series process for all explanatory variables. The capital X represents the collection of all regressors.

$$Var(\varepsilon_t|X) = Var(\varepsilon_t) = \sigma^2 \; for \; t = 1,2,\dots,n$$

The appropriate test of this assumption is the Breusch–Pagan test, which computes a chi-squared statistic that determines if the variance is constant or not. After the initial OLS regression estimation, the residuals are calculated, squared, and then fit against the original regressors as response variable to get another new regression. The null hypothesis is that the residuals are constant, i.e., homoscedastic, and the alternative assumes the opposite. The test statistic is a chi-squared statistic equal to $\chi_k^2 = nR^2$, where n is sample size and R^2 is the explained variance of the latter regression. The null hypothesis is rejected in favor of the alternative hypothesis if the statistic is greater than the critical value of the χ^2 distribution, or if the associated p-value of the test statistic is less than the significance level. [12]

Another relevant test is the ARCH test, short for "Autoregressive Conditional Heteroskedasticity", which is appropriate to use when the error variance in a time series follows an autoregressive model. If the error variance followed an ARMA model instead, a GARCH test would be more useful (generalized ARCH). ARCH is good for identifying increasingly volatile financial markets and adjusts forecasting models accordingly for better results.[13]

3.1.1.5. No serial correlation

Given all explanatory variables, there is no conditional correlation between any two error terms in any period t.

$$Corr(\varepsilon_t, \varepsilon_s|X) = 0 \; for \; all \; t \neq s$$

[11] Wooldridge, 2016, pp. 86

[12] Statology, 2020. The Breusch–Pagan Test: Definition & Example. https://www.statology.org/breusch-pagan-test/

[13] MathWorks, 2021. Engle's ARCH Test. https://www.mathworks.com/help/econ/engles-arch-test.html

The capital X represents the collection of all regressors, i.e., all lags of each variable, and the statement is valid for each combination of time periods that don't coincide.

This is also known as autocorrelation, and it means that two versions of a variable in time (lags) are linearly dependent on one another, they are correlated. An appropriate test for this is the 'Ljung–Box test' which computes a statistic based on sample autocorrelation and size for a group of lagged variables. It is chi-squared distributed. The difference between the second assumption (no perfect collinearity) and this assumption is that this assumption is correlation between a variable and its lagged version while the other assumption refers to correlation between a multitude of variables, regardless of lag.

$$Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k}.$$

n is the sample size, $\hat{\rho}_k^2$ is sample autocorrelation at lag k. The critical region of rejection is given by $Q > \chi^2_{1-\alpha,h}$ where $1-\alpha$ is the significance level, h is the degrees of freedom. [14]

Another appropriate test for serial correlation is the 'Cochrane–Orcutt' test, which tests if serial correlation exists and if they do, using estimation, models the errors so that they are minimized and maybe avoid potential bias. There are some slight losses of efficiency though (first difference lost).[15]

3.1.1.6. Normality

The error terms, ε_t, are independently, identically, normally distributed (i.i.d).

$$\varepsilon_t \sim i.i.d\ N(0,\sigma^2)$$

An informal but simple way of checking is to examine the 'Q-Q plot', which graphically places observed values against expected values and compares this distribution to a 45-degree line. The less the data deviates from the line, the higher the probability of the data being normally distributed. The 6th condition is optional and only necessary in some cases, though for the sake of entirety, it is included.

A formal test for normality is the Jarque–Bera test which checks the skewness and kurtosis of the error terms. If the errors are skewed it means that the curve of the normal distribution is uneven towards one side and that the slopes of the curve vary. Kurtosis is a method for explaining the shape of the curve. The kurtosis value for a normal distribution is 3, values above 3 produce a spikier distribution and values below 3 produce a flatter distribution. [16]

The null hypothesis of the Jarque–Bera test is that the skewedness is zero and that excess kurtosis is zero as well.

3.2. ARIMA

Another way to model time-series processes or stochastic processes is with ARIMA models, which stands for 'autoregressive integrated moving average', and is in fact a combination of possibly three components: an autoregressive process, and integrated process and a moving average process. They are three different descriptions of how a time-series process could be

[14] Montgomery, 2015, pp. 73–74
[15] Wooldridge, 2016, pp. 383–385
[16] Wikipedia, 2021. Jarque–Bera Test. https://en.wikipedia.org/wiki/Jarque%E2%80%93Bera_test

modeled. It is not obligatory to have one of each part, the number of parts depend on what best describes the process.

An autoregressive process of order p ($AR(p)$) is a time varying process in which the dependent variable is linearly related to its previous values and an error term.

$$AR(p): y_t = c + \sum_{i=1}^{p} \rho_i x_{t-i} + \varepsilon_t$$

In a moving average process of order q ($MA(q)$), the dependent variable is linearly related to its previous errors.

$$MA(q): y_t = \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}$$

Lastly, an integrated process of order d ($I(d)$) defines the dependent variable as the difference between the current value of the dependent variable and one of its previous values. Integration is only relevant if the process is not stationary.

$$I(d): \Delta y_t$$

Combined, they form the $ARIMA(p, d, q)$ model. The appearance of the ARIMA model will depend on the data, it is fitted similarly to a regression and the number of components from each part depends on the data. [17]

3.2.1. VAR Models

VAR stands for "Vector Auto Regression" and is a multivariate extension of the autoregressive AR(p) model. It is built upon a structure where each variable utilizes OLS to create linear functions of past lags of itself and other variables. The VAR model is useful for describing and analyzing multiple relationships within a system, for forecasting, and testing structural inferences. It is also useful for making predictions because of its multivariate nature.

An application of VAR models is the concept of 'Granger causality', which is a way of saying that a time-series process causes another time-series process. To say that one time-series granger causes the other, it means that the time-series look very similar, but one of them is lagged behind the other, it is shifted in time but not significantly different in any other aspect. One can explore such relationships with F or t-tests. For instance, using time-series data and macroeconomic variables. Below is an example of a VAR(1) model with k variables (the order denotes the lag). [18]

$$\begin{bmatrix} y_{1,t} \\ \vdots \\ y_{k,t} \end{bmatrix} = \begin{bmatrix} c_1 \\ \vdots \\ c_k \end{bmatrix} + \begin{bmatrix} a_{1,1} & a_{1,2} \\ \vdots & \vdots \\ a_{k,1} & a_{k,2} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ \vdots \\ y_{k,t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,t} \\ \vdots \\ \varepsilon_{k,t} \end{bmatrix}$$

3.2.1.1. Information criterion (Bayesian & Akaike)

One way to select appropriate models is with information criteria, two of the most prominent ones are the Bayesian one (BIC) and the Akaike one (AIC). The criteria are based on maximum likelihood estimates of the model parameters but are used slightly different. The

[17] Montgomery, 2015, pp. 334, 338, 363
[18] Montgomery, 2015, p. 496

information criteria measure relative likelihood by comparing how probable one model is to reduce information loss to the other. I.e., if we have three models, and one is significantly less likely to reduce information loss than the first model, the third model would be omitted because it is (comparatively) most likely irrelevant/useless.

The optimal AIC value should be as low as possible, either acquired through a small number of parameters or high log-likelihood (the latter would indicate that the data fits the model very well). The same goes for the BIC but there is a difference: the BIC also considers the number of samples used to estimate the parameters. The purpose of this variable is to penalize repeated sampling because by consistently considering more data, one can improve upon the precision of the parameter estimates by reducing the variance until the whole population has been sampled. Effectively, all data would be known, and the true population parameters would be acquired, which is not realistic. Therefore, if it is possible to get significant parameter estimates using few samples, it means that the samples are very representative of the population and good for general estimation. In essence, the lower the values of both BIC and AIC the better. [19]

$$BIC = k\,ln(n) - 2\,ln(\hat{L})$$

$$AIC = 2k - 2\,ln(\hat{L})$$

k is the number of parameters in each model, n is the number of samples taken and $\hat{L}$ is the maximum likelihood estimate of the parameters.

Both criteria are used to evaluate and compare various models and based on the joint results from the measurements, a final decision is made. No single criterion is better than the other, it depends on the situation. [20]

3.2.2. Autocorrelation function & Partial autocorrelation function

In time-series analysis, two helpful tools for identifying the functional form of ARIMA models are the autocorrelation function (ACF) and partial autocorrelation function (PACF). ACF is the autocorrelation between a variable and one of its lagged versions. The PACF does the same thing, but also accounts for partial correlation from other lags that might affect the regular autocorrelation and removes it. Therefore, PACF shows strictly variable-to-variable correlation while ACF is indirectly tainted by other lags.

The utility of these functions is to act as a decisive rule in model selection, specifically ARIMA models. When displaying the functions graphically, their values (correlations) are either going to consistently decrease in descending order (lags that are further back in time have less of an impact on future values due to uncertainty), or simply cut off after a number of lags (all lags afterwards aren’t statistically significant). This is useful for determining how many lags to use of each model. Importantly, ACF will cut off for $MA(q)$ models but trail off for $AR(p)$ models and PACF will cut off for $AR(p)$ models but trail off for $MA(q)$ models. Therefore, the ACF decides the order of the $MA(q)$ part of ARIMA and PACF decides the order of the $AR(p)$ part of ARIMA.

$$PACF\colon \tilde{y}_t = \phi_{11}\tilde{y}_{t-1} + \phi_{12}\tilde{y}_{t-2} + \dots + \phi_{1k}\tilde{y}_{t-k} + \varepsilon_t$$

$$ACF\colon \rho(k) = corr(y_{t+k}, y_t)$$

[19] Montgomery, 2015, pp. 76–77

[20] Montgomery, 2015, pp. 496–498

y_t is a time-series object with lag k, $\rho(k)$ is the regular Pearson correlation coefficient and $\tilde{y}_t$ is the mean-subtracted time-series object where ϕ_{kk} denotes the sample partial autocorrelation of one lag onto the other. In the equation of PACF above, the sample partial autocorrelations between lag 1 and all other variables/lags are shown. By regressing the PACF function, the partial autocorrelations are found.[21]

3.2.3. Stationarity

For an ARIMA model to be utilized, it must be stationary, which means that the probability distribution of the process must be constant throughout time. In other words, it implies that the mean value and variance need to remain constant through time, which is to say that the time-series process needs to fulfill the zero conditional mean assumption and homoskedasticity. [22]

To check for stationarity, a Dickey–Fuller test is performed. It explores the presence of unit roots. A unit root is a unique solution to the associated characteristic equation of every time-series process, if there are any unit roots present, the process is said to be non-stationary. The process can be made stationary by integrating it equal to the number of unit roots present. The Augmented Dickey–Fuller test also checks for autocorrelation and deals with it before dealing with possible unit roots. The test uses a regression equation which is based on a $AR(p)$ model and considers the parameter estimates of the regression to decide upon the state of unit roots. Below is an example of a test performed on a $AR(1)$ model:

$$x_t = \rho_1 x_{t-1} + \varepsilon_t$$

$$\Delta x_t = (\rho_1 - 1)x_{t-1} + \varepsilon_t = \delta_1 x_{t-1} + \varepsilon_t$$

If $\delta_1 = \rho_1 - 1 = 0$, the $AR(1)$ has a unit root and needs to be differenced to remove it. For the augmented version, the regression is different, and a test statistic is designed upon it.

$$\Delta x_t = \alpha + \beta t + \gamma x_{t-1} + \delta_1 \Delta x_t + \cdots + \delta_{p-1} \Delta x_{t-p+1} + \varepsilon_t$$

p is the lag of the $AR(p)$ process, α and β are constants. The test statistic is

$DF_\tau = \hat{\gamma}/SE(\hat{\gamma})$ and $\begin{cases} H_0: \gamma = 0 \\ H_a: \gamma < 0 \end{cases}$. If the null hypothesis is rejected, there is no unit root present, and the more negative the statistic the safer this conclusion is.[23]

3.2.4. Forecasting

Typically, after having examined a time-series process, the next step is to make a forecast into the future based on current and historical data. Forecasts are a subcategory of predictions, which are estimates of the present based on complementary data, whether that be historical data, current data or else. Forecasts are predictions of the future. Forecasts generally suffer from losses in precision the further into the future they estimate, and this error can be modeled once the future values actually come to pass. One can make forecasts with regression models, ARIMA processes and other methods.

Forecasting using VAR-models is a recursive procedure. It creates forecasts for each endogenous variable included in the system. Using the same VAR(1) formula found in 3.2.1., the step-by-step procedure for forecasting one-step-ahead forecasts is illustrated. By replacing the parameters with their estimates, one obtains:

---

[21] Montgomery, 2015, pp. 38, 348
[22] Montgomery, 2015, p. 329
[23] Wooldridge, 2016, pp. 575–577

$$\begin{bmatrix} \hat{y}_{1,t+1} \\ \vdots \\ \hat{y}_{k,t+1} \end{bmatrix} = \begin{bmatrix} \hat{c}_1 \\ \vdots \\ \hat{c}_k \end{bmatrix} + \begin{bmatrix} \hat{a}_{1,1} & \hat{a}_{1,2} \\ \vdots & \vdots \\ \hat{a}_{k,1} & \hat{a}_{k,2} \end{bmatrix} \begin{bmatrix} \hat{y}_{1,t} \\ \vdots \\ \hat{y}_{k,t} \end{bmatrix} + \begin{bmatrix} \hat{\varepsilon}_{1,t} \\ \vdots \\ \hat{\varepsilon}_{k,t} \end{bmatrix},$$

where t is all observations up to a certain time and ε for the error term has the value zero. Followed by a 2-step-ahead forecast, it would look like

$$\begin{bmatrix} \hat{y}_{1,t+2} \\ \vdots \\ \hat{y}_{k,t+2} \end{bmatrix} = \begin{bmatrix} \hat{c}_1 \\ \vdots \\ \hat{c}_k \end{bmatrix} + \begin{bmatrix} \hat{a}_{1,1} & \hat{a}_{1,2} \\ \vdots & \vdots \\ \hat{a}_{k,1} & \hat{a}_{k,2} \end{bmatrix} \begin{bmatrix} \hat{y}_{1,t+1} \\ \vdots \\ \hat{y}_{k,t+1} \end{bmatrix} + \begin{bmatrix} \hat{\varepsilon}_{1,t+1} \\ \vdots \\ \hat{\varepsilon}_{k,t+1} \end{bmatrix}.$$

This procedure is iterated upon to obtain forecasts farther into the future. However, with this in mind, keeping a low number of correlated variables and lags reduces the amount of estimation error entering the forecast, since each equation accounts for $1 + k$ lags for each included variable.[24]

3.3. Time-series decomposition

The best way to identify underlying patterns within a time-series process is to split it into components, which are characteristics of the time-series. The most common patterns include a trend-cycle component, a seasonal component, and a remainder component. There are mainly two ways to decompose a time-series process into these components: additive and multiplicative decomposition respectively.[25]

Using additive decomposition, the model has the following functional form

$$y_t = S_t + T_t + R_t,$$

where y_t is the time-series process, S_t is the seasonal component, T_t the trend component, and R_t the remainder component, at the time t. In multiplicative decomposition, the model has the following functional form instead:

$$y_t = S_t \times T_t \times R_t.$$

Both models are based on the same set of variables, though the way to combine them is different, as suggested by the name of the decompositions.

Additive decomposition is the most beneficial model when a time-series exhibits a seasonal component or a trend-cycle component with constant variation. Otherwise, a multiplicative decomposition is more appropriate when the variation is proportional to the level (average) of the time-series. An alternative to multiplicative decomposition is to log-transform the data until the variation remains stable over time, enabling the use of a log additive decomposition, which is better for identifying and representing the seasonal component

$$\log(X_t) = log(S_t) + \log(T_t) + \log(R_t).$$

3.3.1. Trend-cycle

Time-series data can sometimes exhibit trends, which are correlations between the existing data and unobserved factors which cause changes in mean value of the distribution. The trend can induce false correlation between a pair of time-series objects if they have similar trends because of the relation with the unobserved factors, which warrants its removal. The trend is

[24] Forecasting: Principles and Practice, 2021. Vector Autoregressions https://otexts.com/fpp2/VAR.html
[25] Montgomery, 2015, pp.55–56

estimated using regression and subtracted from the time-series. An example of a linear trend component is listed below. [26]

$$y_t = \alpha_0 + \alpha_1 t + \varepsilon_t$$

Time-series can also be detrended by differencing the time-series, similar to order of integration. A time-series object cannot be stationary if it has a trend because the mean value of the data changes with time, which is a major argument for removing it.

3.3.2. Seasonality

Consistent and recurring patterns throughout time-series are called seasons. A season has a length, and resets itself after it has passed, like the seasons of a year. Their cyclical nature makes them easy to predict, and this can be beneficial for some purposes. However, when analyzing time-series, the presence of seasonality often causes non-stationarity because the mean value and variance will change according to the season. Therefore, in order to estimate accurate models, data must first be seasonally adjusted to satisfy the stationarity assumption.

Most often, the presence of seasonality is much stronger than the time-series itself. A model estimated based on non-adjusted data would be a highly inaccurate description of the actual relation. I.e., any actual changes in variable values and their associated mean values and variances would be overshadowed by the seasonal pattern if it isn't removed.

The seasonal component can be modeled and removed in multiple ways. One way is to assign weights to data based on seasonal occurrence and divide these weights from the data accordingly to remove the impact of the season. Another is to add/subtract a fixed quantity from each value based on their occurrence in the season, it comes down to if the time-series model is described as a multiplicative or additive model. [27]

3.3.3. SEATS decomposition

"SEATS" stands for "Seasonal Extraction in ARIMA Time-series", a default method for seasonally adjusting data. It is a decomposition method that can extract seasonality from data with periods of 2, 4, 6 and 12, but falls short at other kinds of seasonality, such as daily-, hourly- or weekly data.

A time-series can be written in the form of:

$$x_t = \mu_t + \tilde{x}_t,$$

where $\mu_t$is a linearly deterministic component (outcome is guaranteed), and $\tilde{x}_t$ is a linearly inter-deterministic component (outcome is not guaranteed) with a white noise error that has zero mean and constant variance. In reality, these assumptions rarely hold true due to trend, seasonality, outliers, or calendar- and regression effects, which are treated as deterministic factors. To control for these, pre-adjustment is performed by SEATS, either by applying regular and seasonal differencing, or by transforming the data. Besides processing the deterministic effects, SEATS estimates the stochastic part of the time-series with ARIMA models, before removing the deterministic effects and proceeding with decomposition.

The decomposition follows the ARIMA-model-based (AMB), which estimates components through the use of centered moving averages, which are subtracted from the original time series accordingly. This process is repeated until desired results are achieved. To assess the

[26] Montgomery, 2015, pp. 48–50
[27] Montgomery, 2015, p. 52

quality of the seasonal adjustments, extensive testing, diagnostics, and other statistics are performed and computed. [28]

3.3.4. STL decomposition

STL is an acronym for "Seasonal and Trend decomposition using Loess", Loess is a local weighted regression used for fitting a smooth curve through points in a scatterplot. It works like this: for each value x of the independent variables, a function $f(x)$ is estimated using (sampled) neighboring values. These estimations and their smoothness greatly rely on the k values of the sample, or rather, how big of a "sliding window" is utilized. Each neighboring value is weighed according to their Euclidean distance to the x data point, which will later be used in the linear regression process to form Loess curves.[29] The Loess curves are instrumental part of the STL decomposition for being able to identify trends and cycles among other things.

The main idea of STL decomposition is to use two parameters to control how rapidly the trend-cycle and seasonal components change, or rather, how many observations are needed for estimating each cycle. That way, they act like sliding windows for dividing the data into different components, only leaving behind a remainder component which contains anomalies and irregularities. [30]

---

[28] JDemetra+. SEATS. https://sylwiagrudkowska.github.io/JDemetra-documentation/pages/theory/SA_SEATS.html
[29] Cleveland, William S, 1979, pp. 829–836.
[30] Forecasting: Principles and Practice, 2021. STL Decomposition. https://otexts.com/fpp2/stl.html

4. Chow test

A major part of the study is to compare sales and prices of houses before and after the Covid-19 outbreak, and the best way to do this is with a Chow test. Before performing the test, the data set of interest is split into two parts based on where a structural change is thought to exist, and a regression is fit to each part. The Chow test is an F-test which aims to compare a pair of regressions to deduce if they are statistically significantly different from each other. In this case, instead of only fitting a regression to the total data, two regressions are estimated, one regression based on pre-pandemic situation and one post-pandemic situation. Then they are compared to see if there are any significant statistical changes that have occurred in the structural break known a priori. To conduct the Chow test, the model errors are assumed to be from a i.i.d normal distribution with unknown variance. The Chow statistic is:

$$\frac{\frac{S_C - (S_1 + S_2)}{k}}{\frac{S_1 + S_2}{N_1 + N_2 - 2k}}$$

The hypothesis test is for a (e.g.) 2 regressor regression is:

$$\begin{cases} H_0: \alpha_0 = \alpha_1, \beta_0 = \beta_1, \delta_0 = \delta_1 \\ H_a: \alpha_0 \neq \alpha_1, \beta_0 \neq \beta_1, \delta_0 \neq \delta_1 \end{cases}$$

In this example, there are 2 regressors, 3 coefficients, S_C denotes the sum of squared residuals for the combined data, the others are the same measurement for each group with sample size N and number of parameters k (coefficients). The Chow statistic is F-distributed with k and $N_1 + N_2 - 2k$ degrees of freedom. [31]

[31] Wooldridge, 2016, p. 238

5. Imputation

The data at disposal should contain information on 290 municipalities on a monthly frequency for over 48 months, though some are missing, especially in smaller municipalities. The goal of imputation is to replace missing values of incomplete data sets by trying to estimate what those values might have been based on existing data. There are multiple ways of doing this. In the subsections below, there are 4 imputation methods listed. All methods will be tried to see which generates the best results, and differences between each will be discussed later on. [32]

5.1. Mean value imputation

Just like the name suggest, in mean value imputation, the mean value of the data is computed and imputed into the missing cases for each variable respectively. This method has the benefit of not changing the mean of the data, however, it increases the risk of running high correlation with the data (mean is generated from data). This is not an issue in univariate analysis, but because the focal point of this study is to use multivariate analysis, it becomes a problem. It is known as a static imputation model. [33]

5.2. (Partial) Listwise deletion/Complete case analysis

Listwise deletion, or complete case analysis, is not an imputation method per say, but it deals with missing data by removing any variable that has any. The problem with this method is that it can remove a lot of variables and sampled data because of a wide scattering of missing data. This would make any further use of the data highly inaccurate and possibly biased. There may be an argument for still consorting to that method if the data is missing completely at random (MCAR), in which case it would still provide fairly accurate results. For the purposes of this study, that method is too extreme and not trustworthy, so a partial listwise deletion is applied instead, where variables are removed only if they are missing 50% or more of their values. [34]

5.3. Regression imputation

In regression imputation, the variables of the data set are used to estimate a linear regression which is used to predict observed values and impute them into the incomplete cases. The main issue with this method is that there is no residual variance or uncertainty around the imputed estimates because they lie on the regression line, and because of this, they are most likely underestimating the true values. It overestimates the precision of the unknown data and most likely does not depict an accurate picture of the actual relationship, bias. A remedy is to use stochastic regression imputation, in which residual variance is added, though this is maybe not sufficient either, or perhaps beyond the scopes of this study. [35]

5.4. Cold/Hot deck imputation

Deck imputation in general revolves around picking similar values to the missing ones, either by random or some other algorithm and imputing them into the missing cases. In hot-deck imputation, values from the same data set are chosen. In cold-deck imputation, values from other sources are used, sources such as historic data or similar data. There are many variants

[32] Lohr, 2010, p. 346
[33] Lohr, 2010, pp. 347–348
[34] The Analysis Factor, 2014. When Listwise Deletion works for Missing Data. https://www.theanalysisfactor.com/when-listwise-deletion-works/
[35] Lohr, 2010, pp. 349–350

of each of these methods, so there is no definite method that grants the best results, it depends on the software of choice (not very practical to use by hand). Bias is a potential problem as always.[36]

5.5. Linear interpolation

Linear interpolation is a subcategory of polynomial interpolation, which is the concept of assuming a polynomial relationship to exist between two points, estimate the relationship and calculate any value in between the points. The same goes for linear interpolation, though the relationship is linear. The idea is to impute missing values by finding weighted averages using this method. It is a very easy thing to apply to data. There are some issues with this method, namely that all relationships aren't linear and that there still might exist outliers in data.

$$\frac{y_1 - y_0}{x_1 - x_0} = \frac{y - y_0}{x - x_0}$$

The geometric interpretation of the equation above is that the slopes at two parts of the same line are equal, and this is used to extract any value of a variable x or y, depending on which of them is already known, all other values are known. In the case of imputation, the other values are the known cases, and a value is imputed based on the assumption that the missing value lie perfectly across the line. [37]

5.6. Choice of imputation method

Considering that there is a trend with strong presence and weak seasonality in the data, at least in those areas where there are enough sales to even report any seasonality, by imputing missing values through static methods, any pattern/relationship would be disturbed and lost. E.g., using mean value imputation.

As for regression imputation, there were two issues that stood out: the data showed serious signs of correlation between variables, a consequence of prices on the housing market being related to each other. This generates a design matrix which is not invertible and makes it impossible to fit a regression model. This can be circumvented by imputing for each group, but it is not as promising as other methods, a conclusion based on diagrams and normalized RMSE.

As for cold- or hot deck imputation, methods that either carry the last observation forward or backwards to a similar missing value and replaces it, does produce acceptable results. However, it's not ideal for this data, as there is an issue with the presence of large consecutive gaps in data for specific variables, as a result of inconsistent documentation of data.

[36] Lohr, 2010, pp. 348–350
[37] Wikipedia, 2021. Linear interpolation. https://en.wikipedia.org/wiki/Linear_interpolation

6. Results

6.1. Imputation

After a partial-listwise deletion, or complete case analysis, has been applied, in which all observations that are missing 50% or more of their values are removed completely, the remaining data consists of $5 - 10\%$ missing values for $Price/m^2$ and P/T respectively.

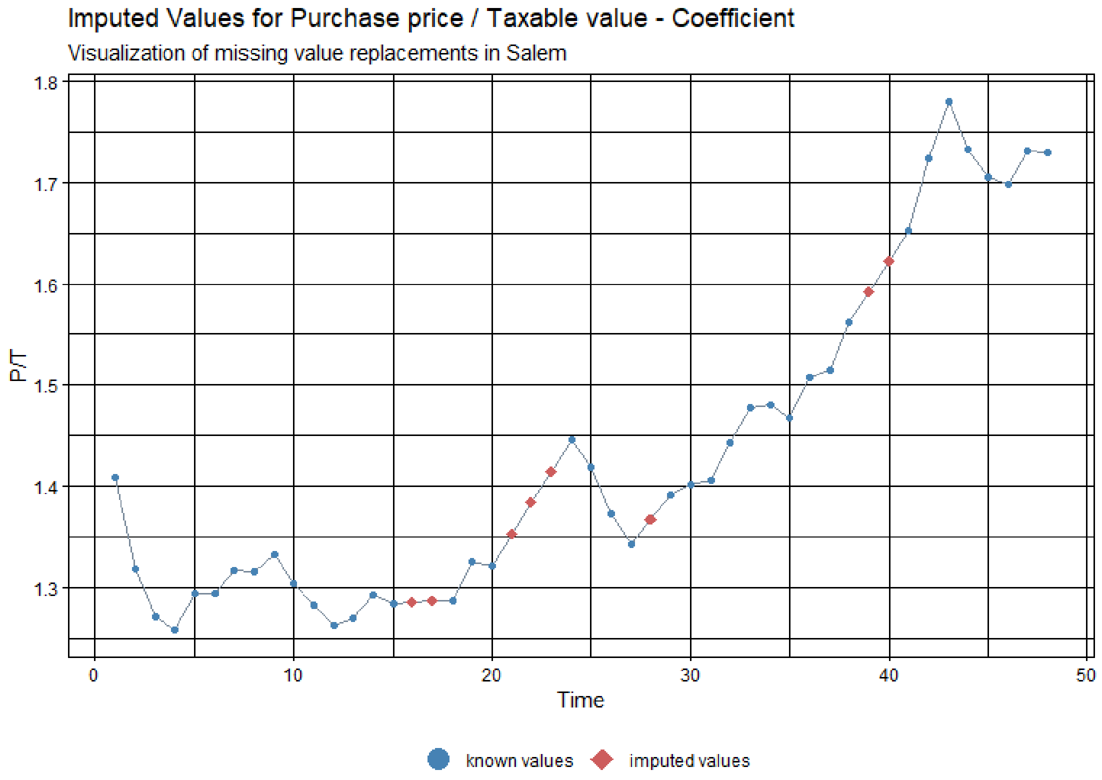


Figure 6.1.1 Imputed values for Salem

The most effective method of imputation was linear interpolation. After the missing data has been imputed, and the model has been identified, the trend- and seasonality component are brought back. The data is assumed to have a frequency of 12 because of its yearly nature, though the results showed that this produced inadequate models due to difficulties in identifying patterns or measurement error. Instead, an automatic built-in detection was used for seasonal patterns which produced the best results. Said results were identical regardless of imputing by group or separately on the data.

6.2. Decomposition

The SEATS decomposition plots, found in Appendix A, reveal that all three groups were similarly pre-adjusted before removing the deterministic parts for both variables $Price/m^2$ and P/T respectively. This proves that they were exposed to the same local patterns. In this, there appears to be no seasonal component for either group 1 or 2, as SEATS only fit group 3 with a Seasonal ARIMA-model (an off-shoot ARIMA model with a seasonal component) before splitting it up. However, their remainder component is relatively large and exhibits small seasonal patterns, but it is insignificant according to the ACF plots.

The ACF plots in appendix A also identifies potentially significant lags: judging by the plots, groups 1 and 2 seem to fit a VAR model of order 3, though group 2 seems to display high autocorrelation. As for group 3, both variables have very significant correlations at lag 1, suggesting that a VAR(1) model is adequate.

STL decomposition was also applied on the macroeconomic variables, as other methods weren't as thorough in removing the seasonal component from the time series. As such, the best results for smoothing were found to be a yearly and constant seasonal component, and the trend-cycle component to be approximately half of that of the seasonal component. Both variables experience less autocorrelation in recent lags and significant partial autocorrelations from older lags, as opposed to the study variables. They are expected to provide a complementary fit for the VAR-models of lower orders.

6.3. VAR models

A priority in choosing an adequate VAR model is to use models with fewer lags, good model fit and no correlation between residuals. This is to create the most significant model using the most amount of data. The AIC, BIC and Ljung–Box test are instrumental in choosing the optimal model. Considering that the data only spans over 48 months, no more than 6 lags are included to preserve the largest possible sample size. Other assessments of model relevance and forecasting potential include tests of normality, ARCH test, and granger causality test.

Summaries of all VAR-models can be found in Appendix B. The estimated VAR-models for each group of the variable $\mathrm{Price/m^2}$ utilizes both the HCI and CPI as significant and optimal endogenous variables. Consequently, unemployment was found to have no explanatory power within the predetermined number of lags. All unit roots were less than 1 in value, indicating stable and stationary models that are supported by high R-squared values and significant p-value of good model fit. For respective pre-covid models, they held similar results but with fewer lags and endogenous variables within the system. As opposed to $\mathrm{Price/m^2}$, all groups of $\mathrm{P/T}$ utilize the unemployment variable, while HCI is used to a lesser extent. The $\mathrm{P/T}$ models generally have a higher R-squared values than the $\mathrm{Price/m^2}$ models, but they are otherwise very similar to the $\mathrm{Price/m^2}$ models, passing the unit root tests and having high p-values.

The test results of the VAR models can be found in appendix B. The ARCH tests do not identify heteroskedasticity in any of the models and the results are highly significant. Looking at the residual plots for $\mathrm{P/T}$ group 2 and $\mathrm{Price/m^2}$ group 3 in figure B.4 and B.18, there is indication of heteroskedasticity in the residual plots. But since all models passed the ARCH test, the heteroskedastic pattern of the residuals is assumed to be within limits. Similarly, all models pass the tests for normality, but histogram for $\mathrm{P/T}$ group 2 and group 3 would suggest skewness in the residuals. Collectively, these findings prove that the Gauss–Markov assumptions of normal distribution and homoskedasticity are satisfied.

The test for Granger causality provides an insight to the relationship between the variables in the model. According to the tables found in Appendix B.13 and B26, the macroeconomic variables are the only ones proven to show Granger-causality, but less often in the VAR-models pertaining to P/T. An issue found with the Granger-causality test is that it is difficult to determine the direction of the causality for models with more than two endogenous variables.

6.4. Forecasts

Below are the estimated forecasts using a selection of VAR-models found in Appendix B. Decomposition was done by both SEATS and STL method, though it is impossible to simply inverse the trend- and seasonal adjustments after identifying the time series. Instead, one takes each of the trend and seasonal components, estimates the same VAR-models as their respective remainder component, performs the same 18- respective 12 months ahead forecasts, and adds them on to each other (additive components).

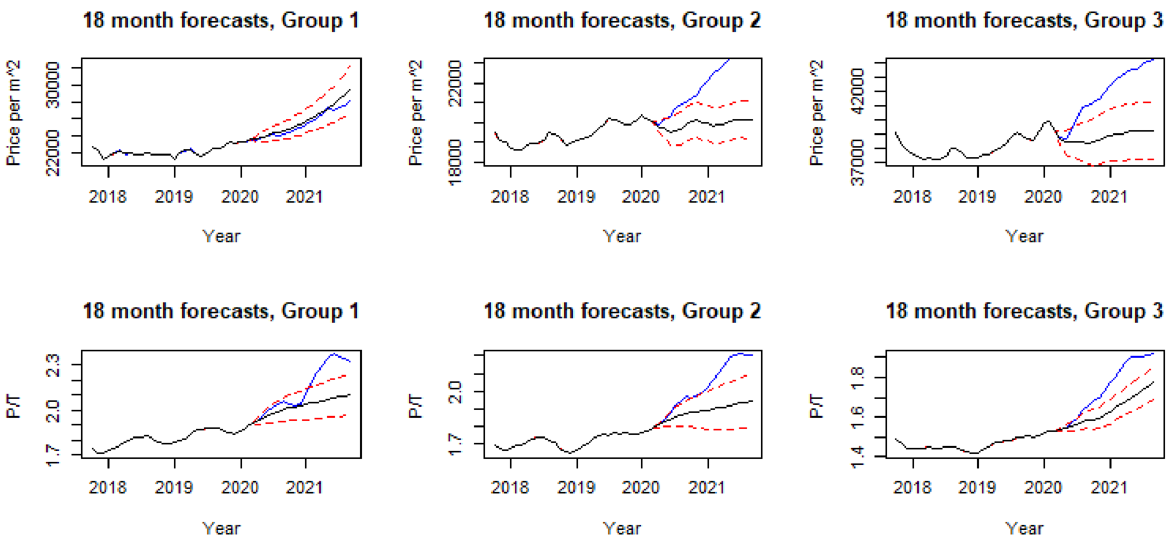


Figure 6.4.1 18 months ahead forecasts for all groups of $Price/m^2$ and P/T, before the covid outbreak, with blue lines for actual change.

The 18-month forecasts have narrow upper and lower bounds. For groups 2 and 3, the actual change after covid outbreak falls outside of the forecasted confidence intervals. All except two models forecast and increase in the study variable. Variable $Price/m^2$ group 2 and group 3 does not forecast an increase in the study variables value, rather they forecast values that varies around the mean. Regarding the implications of the differences between the forecasted values and actual values, that will later be discussed along with Chow tests.

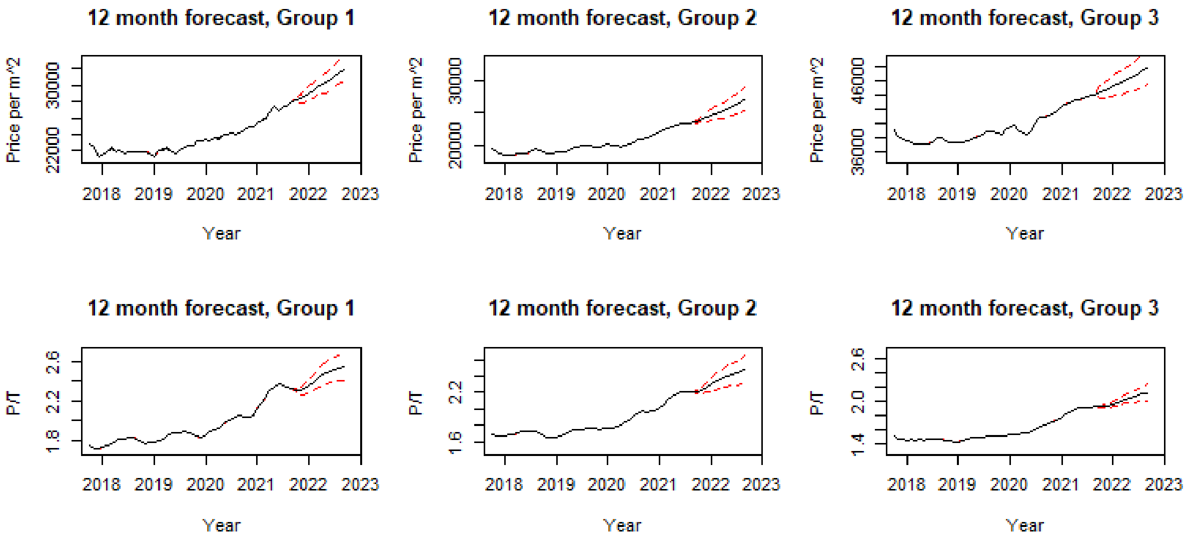


Figure 6.4.2 12 months ahead future forecasts for all groups of $Price/m^2$ and P/T.

For the 12 month's forecast, all of the models predict an increase in prices for houses and condominiums, the forecasts all have narrow upper and lower bounds. Only P/T group 1 and group 3 having predictions that seem to flatten out a bit during the last months of the forecast.

6.5. Chow tests

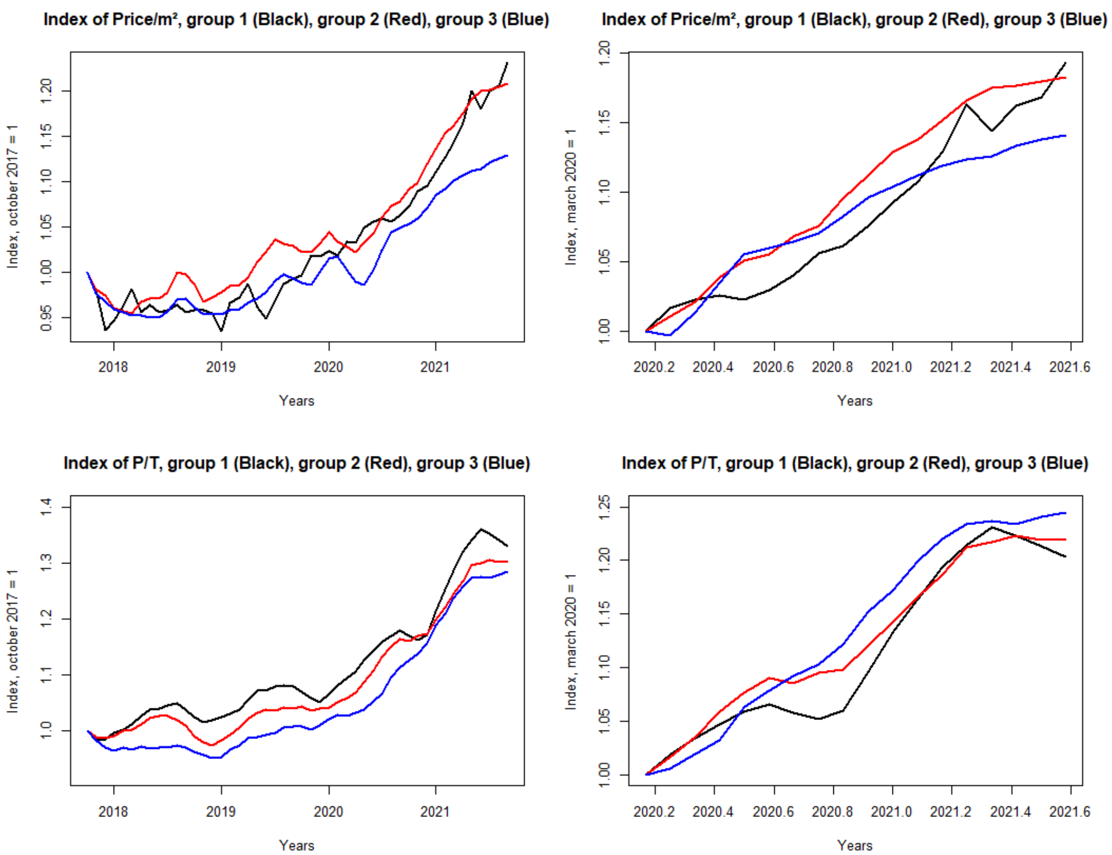
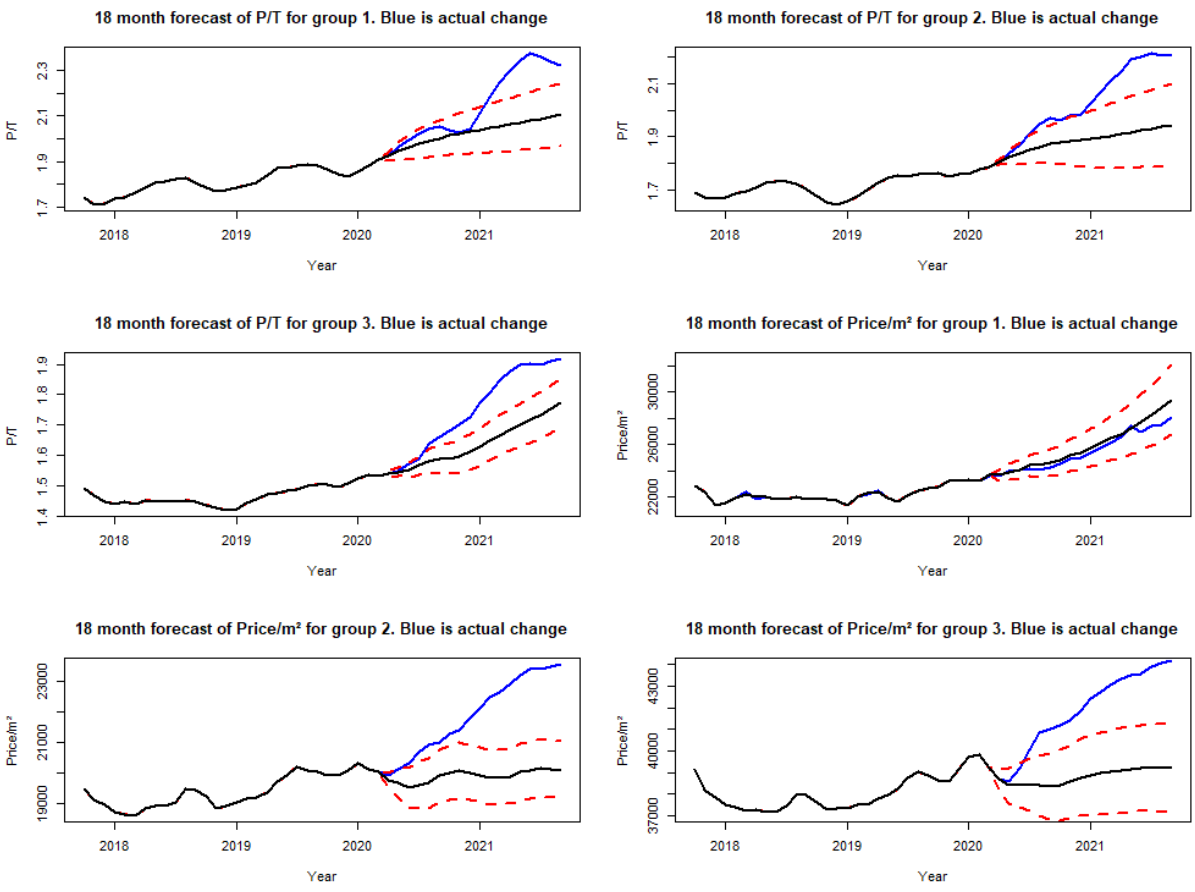


Figure 6.5.1 Index of Price / m^2 and P/T for 48 months and after covid.

All of the chow tests are found in appendix C. As can be seen in table C.1, the Chow test for the time series of $Price/m^2$ does not follow a clear pattern. It can be seen that only two tests for the variable spanning all 48 months have structural breaks while all of the after covid tests do have structural breaks. The graphs show that all of the time series have increased through the years with group 1 and group 2 increasing the most. In the after covid graph, group 1 and group 2 have had a very similar increase while group 3 has had the lowest increase of the time series.

In table C.2 we see that the P/T time series for all 48 months indicate a lot of structural breaks. While the variable before covid is not that interesting, it is worth noting that almost of the tests indicate a structural change. For the variable after covid, only group 1 and group 2 have had a structural break. When these results are compared to the graphs, the variables in after covid have different shapes, though they all end similarly. In table C.3 and C.4, the tests for sales for variable $Price/m^2$ and P/T both have no structural breaks except ones, for $Price/m^2$ group 1. The time-series follow a clear seasonal pattern.

All Chow tests are done with the mean of the forecast as one time-series and the actual values as the other time-series. It is not possible to perform a Chow test with the forecasted time-series mean value and upper- and lower boundaries.



Figur 6.5.2 18 months forecasts for variables P/T and $Price - m^2$, blue time series is actual values.

In table C.5 and C.6 we see that the chow test identifies structural breaks in all forecasted models and the results are highly significant, except for the time series $Price/m^2$ group 1.

The tests only compare the mean of the forecasted values with the actual data. The graphs show that for all models except $Price/m^2$ group 1, the actual increase in the variable is higher than what has been modelled. For $Price/m^2$ group 1, the actual values always stay within the confidence interval, and closely follow the mean. For P/T group 1, the actual values stay within the forecasted confidence interval for the first nine months before exceeding the forecasted values. In all models except for $Price/m^2$ group 1, the forecasts underestimate the true values.

6.6. Changes in study variables, before and after Covid outbreak

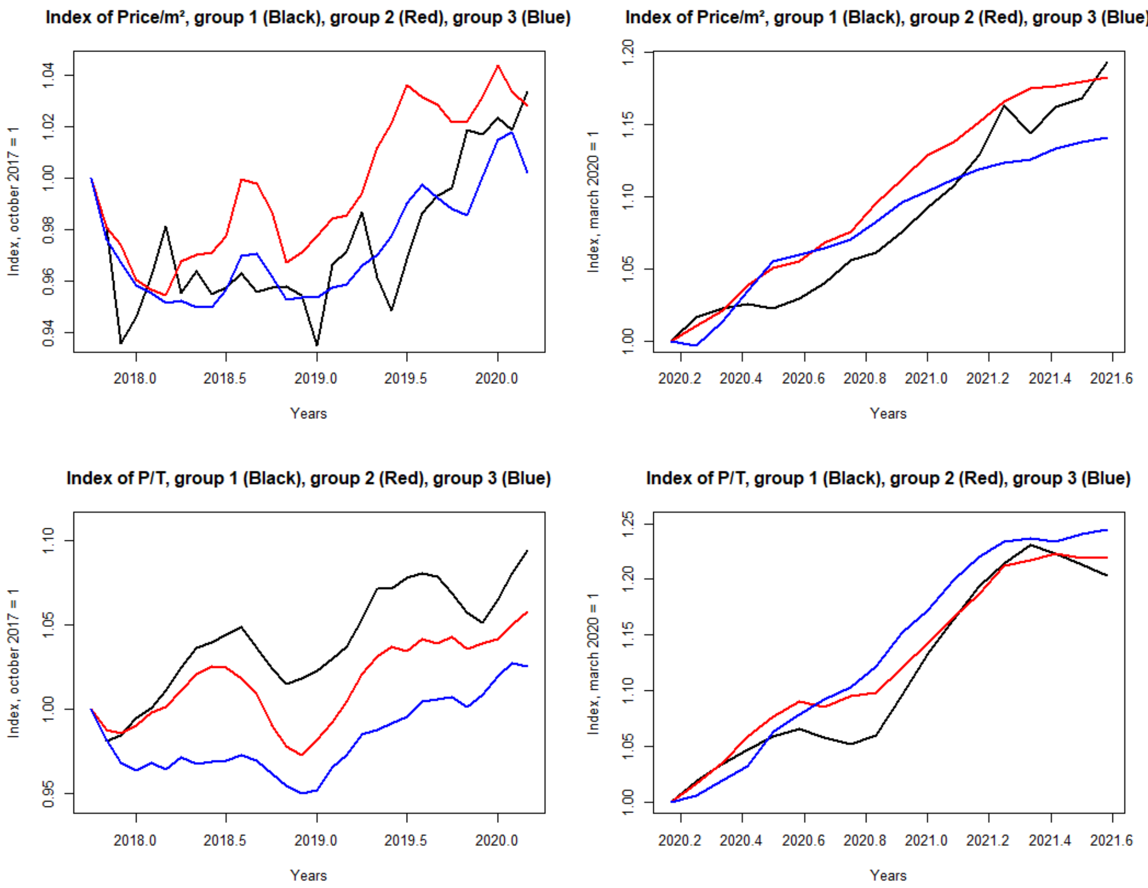


Figure 6.6.1 Index of $Price/m^2$ and P/T for time periods before and after covid.

The easiest way to see if the variables $Price/m^2$ and P/T have increased more in one group then another is to plot their indexes for the relevant time periods. After March 2020, $Price/m^2$ group 1 has had the largest increase, closely followed by group 2. $Price/m^2$ has increased by more than 19% for group 1, group 2 is only a little behind at 18%. Group 3 has had the lowest increase at 14%. Looking at the data from October 2017 to March 2020, there was a decline in prices for many years for all the groups. Group 3 ended up 0.2% above what it started at in October 2017, group 1 and group 2 grew 3% and 2% respectively.

For P/T, in between October 2017 and March 2020, all three groups had an increase in the P/T variable, with group 1 increasing the most. It increased by about 9%, group 2 just north of 5% and group 3 approximately 2.5%. From March 2020 to September 2021, Group 3 has had the largest increase, with group 2 and 1 following behind. Group 3 increase by more than 24%, group 2 by 21.95% and group 1 by 20%.

7. Discussion

7.1. Chow tests

The Chow test is an important test to formally compare two time series and detect any structural breaks, but it is severely lacking when it comes to providing in-depth information of the nature of those discoveries. In several situations, a pair of time-series had structural breaks, but the starting and end positions were very similar. This is because the test only measures if the structure of the models differs. To actually understand the relationship between a pair of time-series, one needs to compare their graphs. This is not to say that the test is worthless, but there may be other tests better for the task.

When looking at the sales for the variables $Price/m^2$ and P/T, in all but one case, there are no structural breaks. This should not be surprising since it is expected that a person moving from one group to another would cause a sale in each group. Overall, the datasets for the sales are quite uninteresting and not worth investigating any further, though it was worth it to see if the datasets would indicate any change.

Comparisons of the actual data lead to inconclusive results. For $Price/m^2$ after covid, all of the tests resulted in structural breaks. When those results are compared to the graph for the same datasets, while it is true that group 1 and 2 follows different paths, and therefore have different structures, they end similarly. This indicates that in the last 18 months since March 2020, the prices for condominiums have increased similarly for the two groups, with group 2 increasing more in the early months, slowly flattening out, while group 1 has had a slower start, it has steadily gone up as of September 2021.

The Chow test for the forecasted models is somewhat limited, in that by decomposing the data, to then recompose it, it ended up with a vector for the mean, and the confidence bounds. When performing the Chow test, the VAR models were unavailable in the 'sctest' function in R. Instead, only the mean was used. Tests on the confidence bounds seemed to be pointless because they almost always resulted in structural breaks, as expected. When comparing the tests with the graph, it is clear that the tests are unreliable when it comes to providing useful information of the two time-series. Note that the actual time series breaks from the forecasted models quite quickly. In only $Price/m^2$ of group 1 does the forecasted model match the actual data very well. For P/T group 1, the actual data lies within the prediction interval for the first 9 months but then it outgrows the model. All the other models underestimate the actual data in the first few months.

It is evident that the Chow test is lacking when it comes to determine if there exist differences between two time-series. It is also clearly insufficient on its own. It is clear from the graphs which of the time-series values lie within the prediction interval of the forecasts. It should also be noted that the forecasts for $Price/m^2$ group 2 and 3 both have much flatter forecasts then the other models, so that they fail to capture the true values should not be surprising. It is however unlikely that models that forecasted a bigger slope would be able to capture the true values either, the increase in prices have simply been too high for any of the models to capture the change. Collectively, with the exception of one graph, they point towards a higher increase than expected.

7.2. Changes in *P/T* and $Price/m^2$, before and after covid outbreak

Looking at the data, it is evident that the initial hypothesis of the study cannot be confirmed. As of March 2020, the prices of condominiums and houses have increased rapidly across the country. The prior has increased the most in small municipalities, followed by those in group 2. There has also been a rapid increase in prices for group 3.

When it comes to houses, it is group 3 that sees the greatest increase, followed by group 2, and then group 1. Therefore, it is not safe to say that smaller municipalities have had a higher price increase then larger ones.

Appendix A. Decomposition plots

Figure A.1 *Plots of extracted patterns for P/T, group 1*

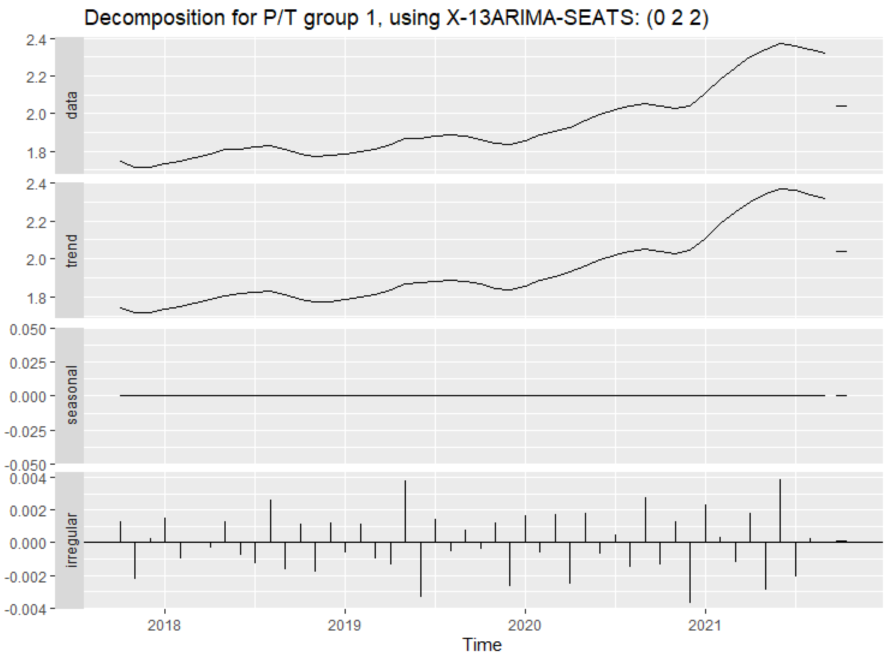


Figure A.2 *ACF/PACF plots for the remainder component of P/T, group 1*

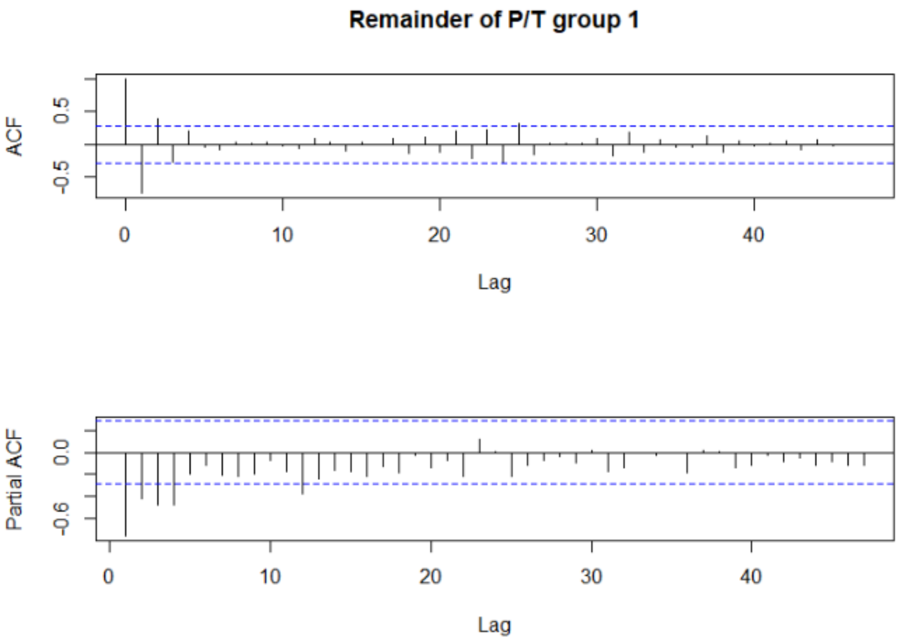


Figure A.3 *Plots of extracted patterns for P/T, group 2*

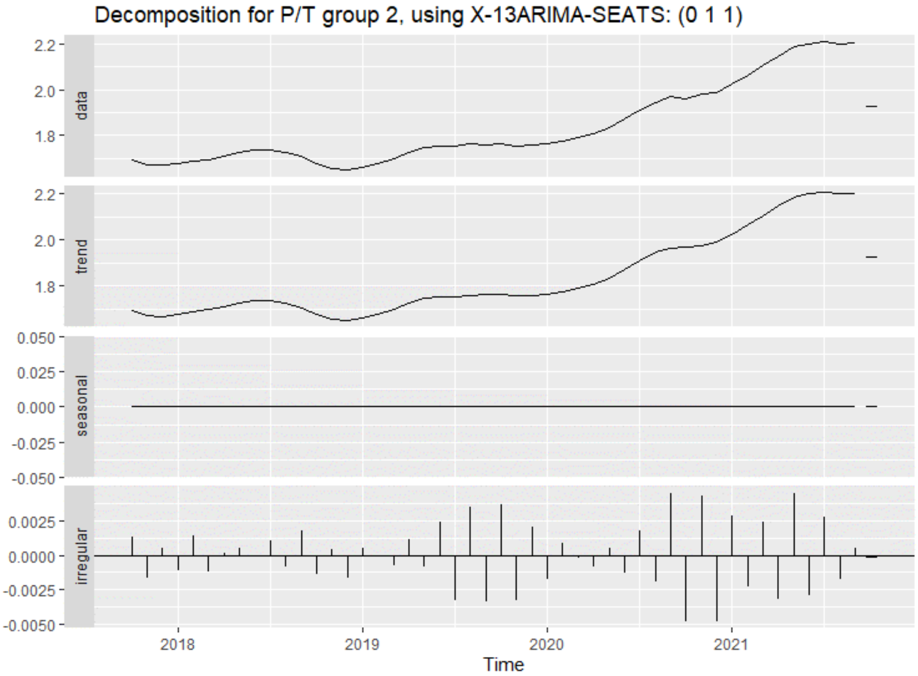


Figure A.4 *ACF/PACF plots for the remainder component of P/T, group 2*

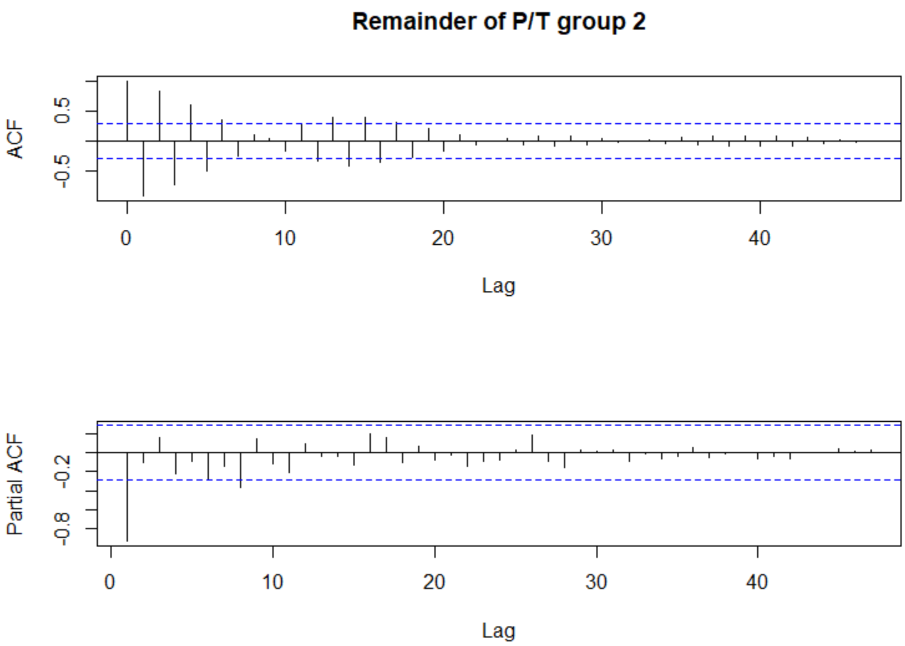


Figure A.5 *Plots of extracted patterns for P/T, group 3*

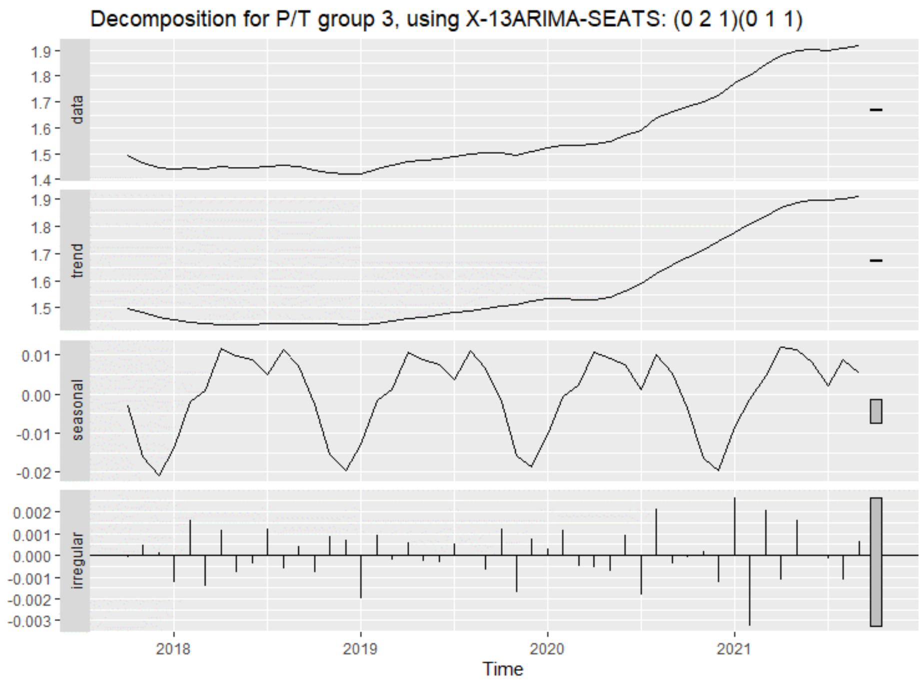


Figure A.6 *ACF/PACF plots for the remainder component of P/T, group 3*

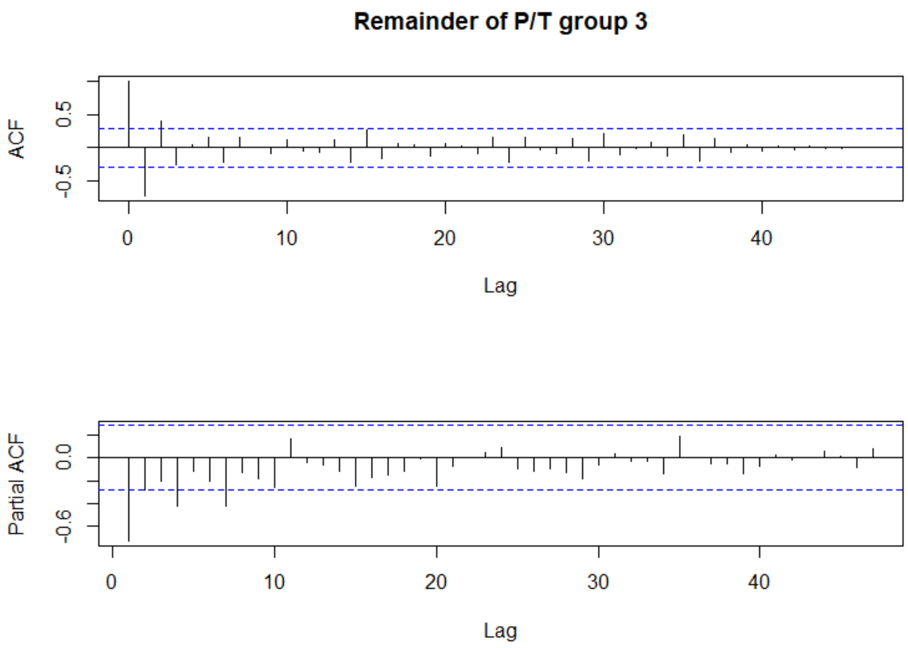


Figure A.7 *Plots of extracted patterns for $Price/m^2$, group 1*

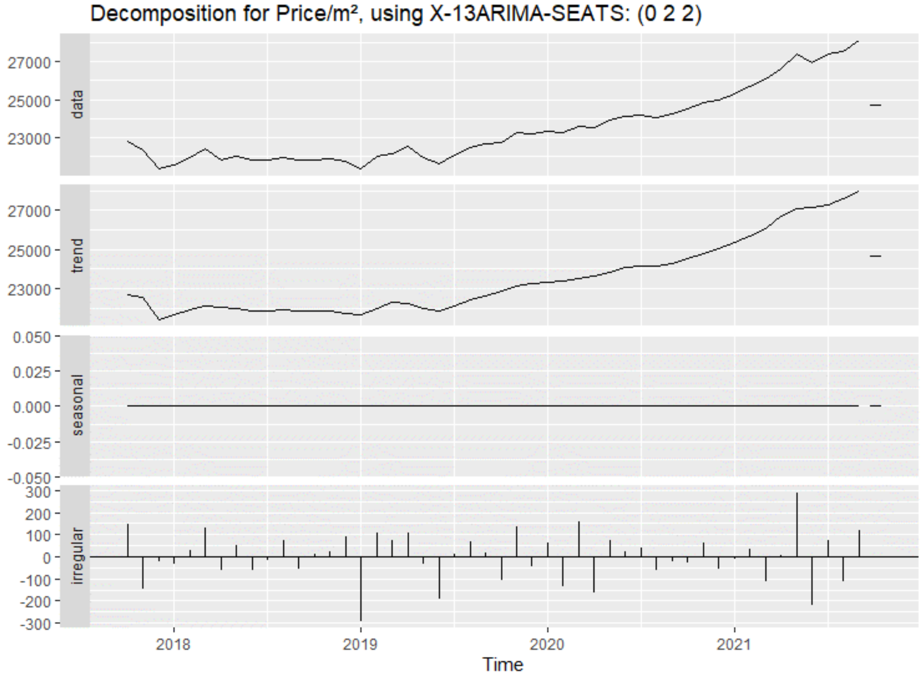


Figure A.8 *ACF/PACF plots for the remainder component of $Price/m^2$, group 1*

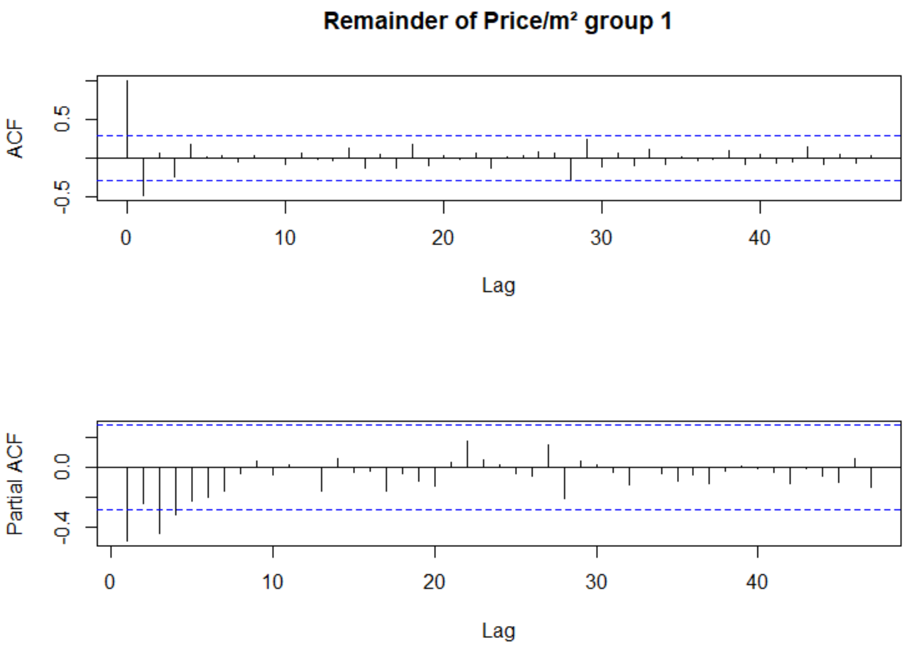


Figure A.9 *Plots of extracted patterns for $Price/m^2$, group 2*

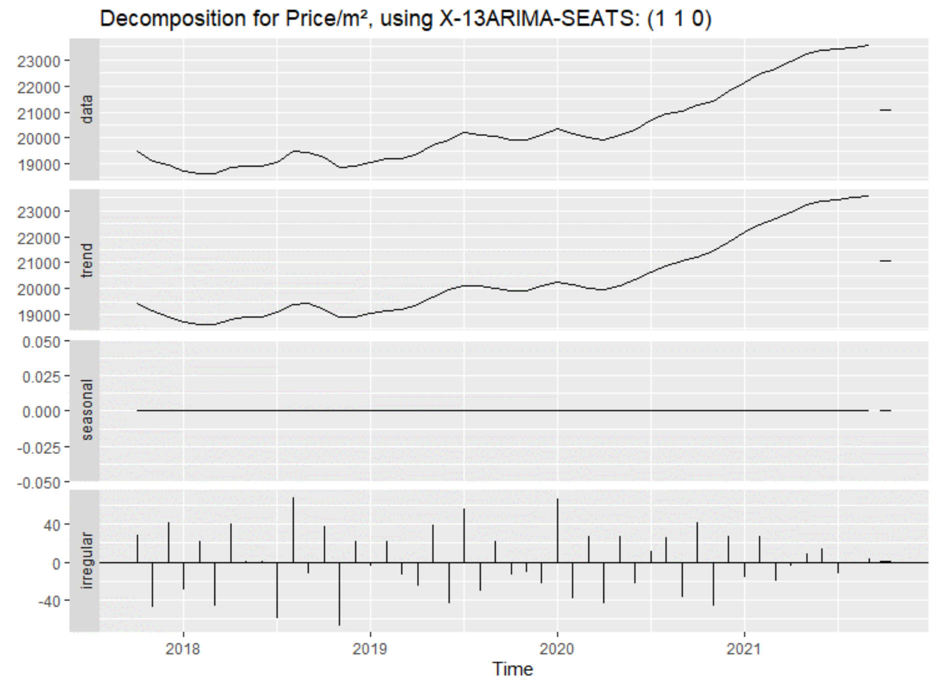


Figure A.10 *ACF/PACF plots for the remainder component of $Price/m^2$, group 2*

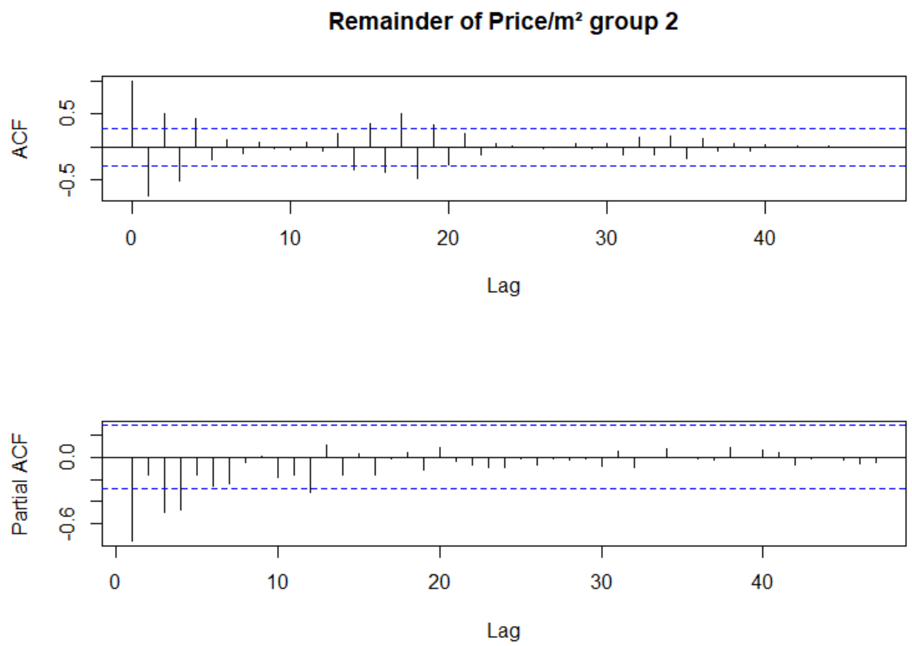


Figure A.11 *Plots of extracted patterns for $Price/m^2$, group 3*

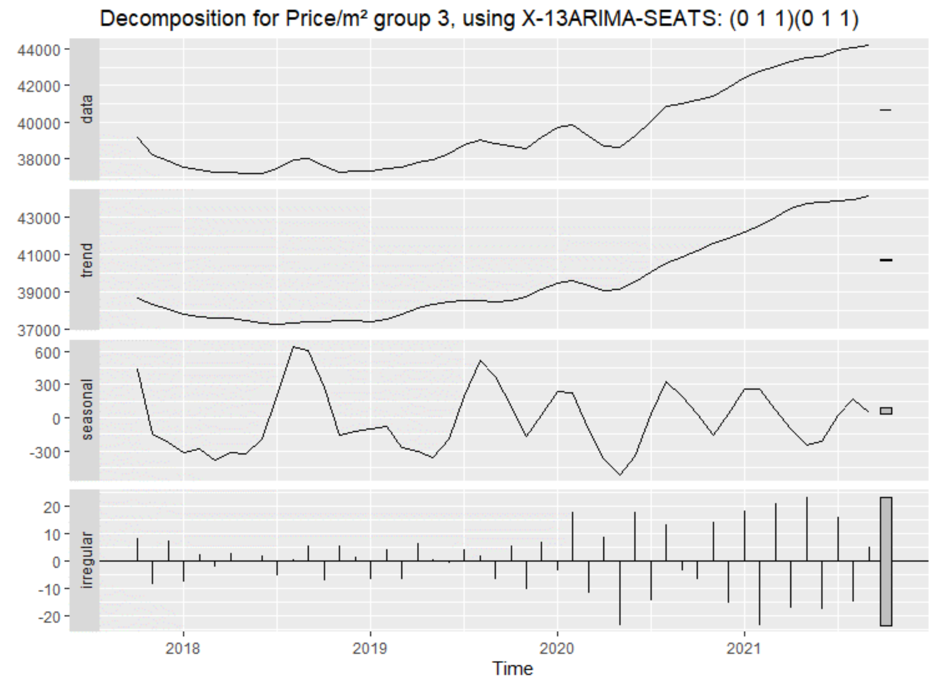


Figure A.12 *ACF/PACF plots for the remainder component of $Price/m^2$, group 3*

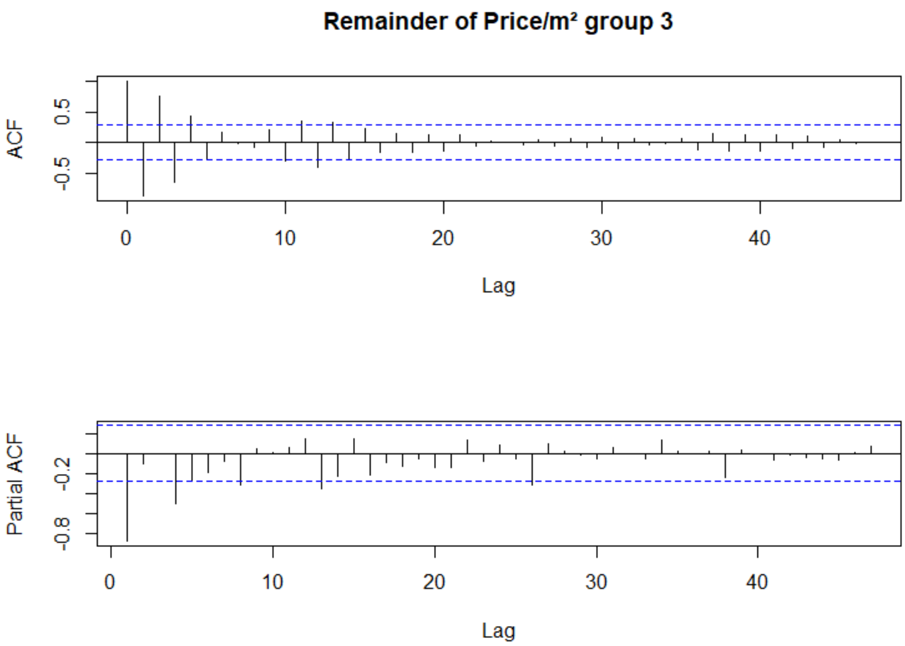


Figure A.13 *Plots of extracted patterns for Consumer Price Index*

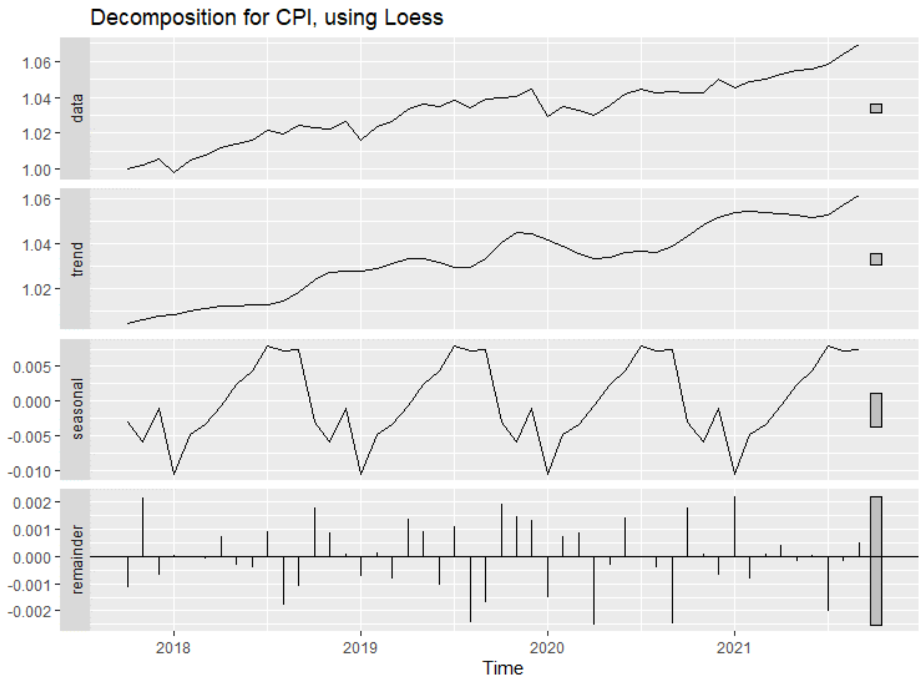


Figure A.14 *ACF/PACF plots for the remainder component of Consumer Price Index*

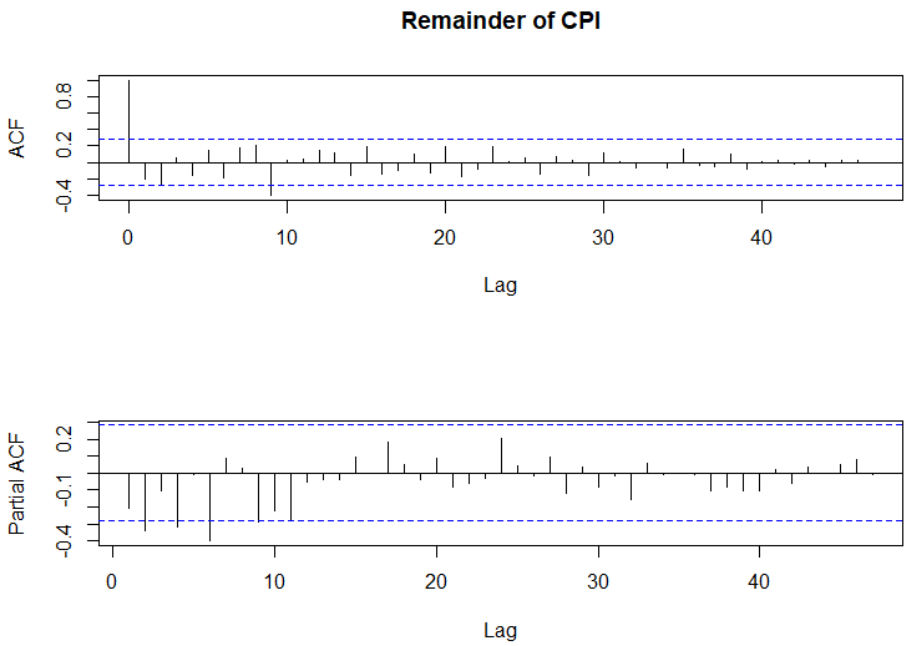


Figure A.15 *Plots of extracted patterns for Household Confidence Indicator*

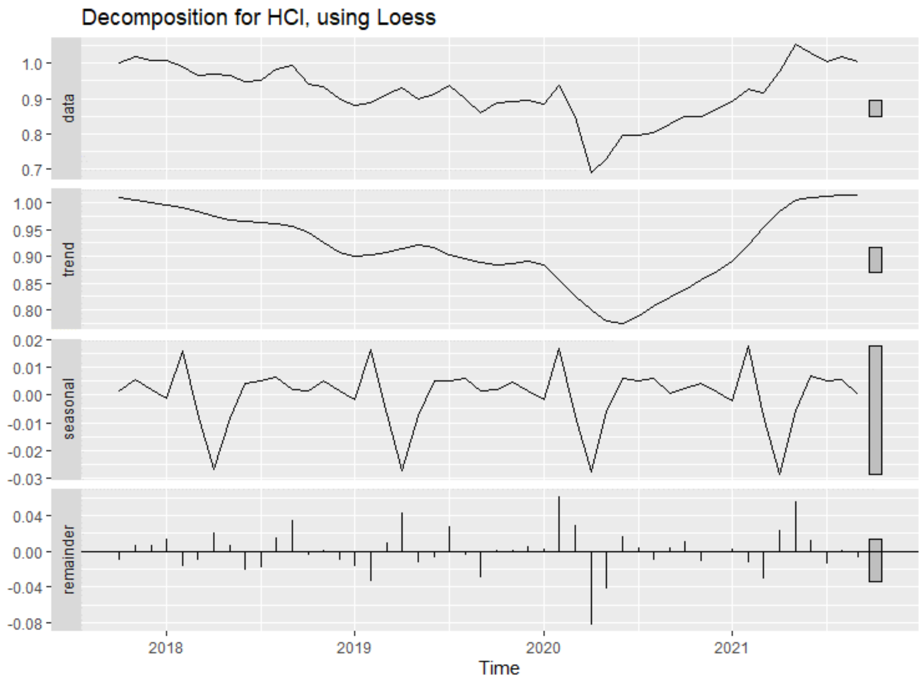


Figure A.16 *ACF/PACF plots for the remainder component of Household Confidence Indicator*

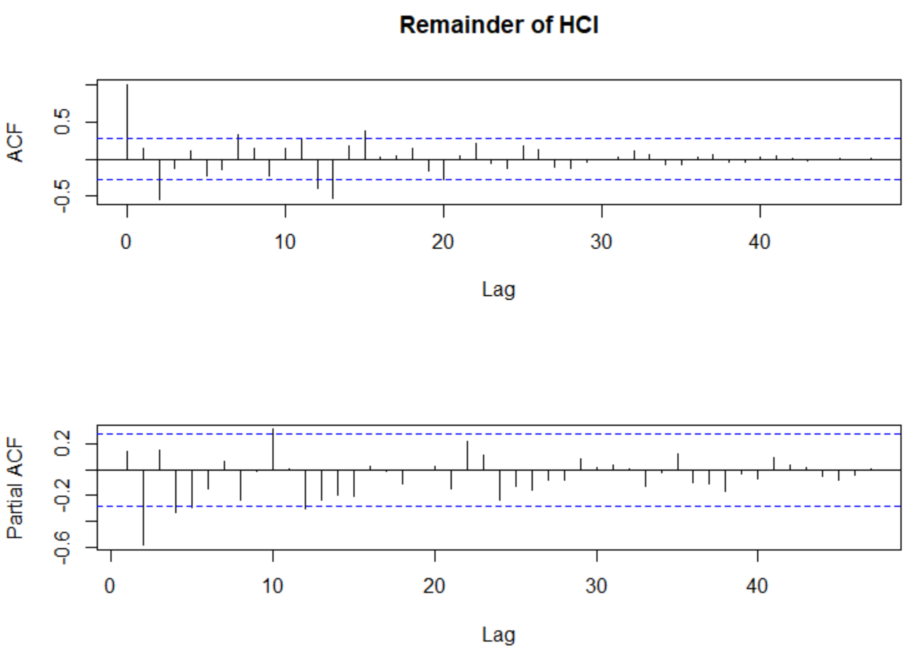


Figure A.17 *Plots of extracted patterns for Unemployment*

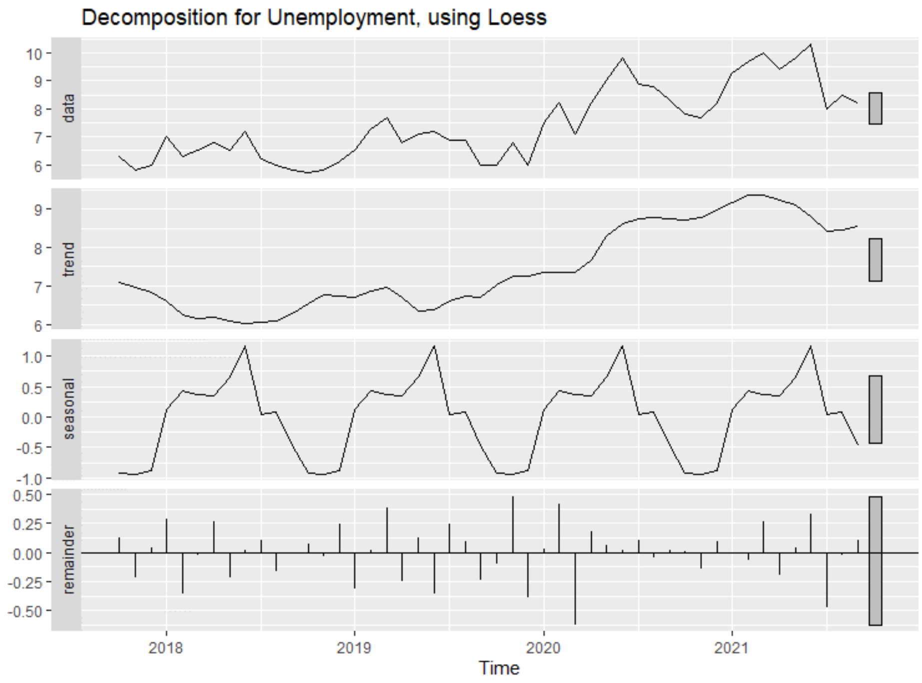
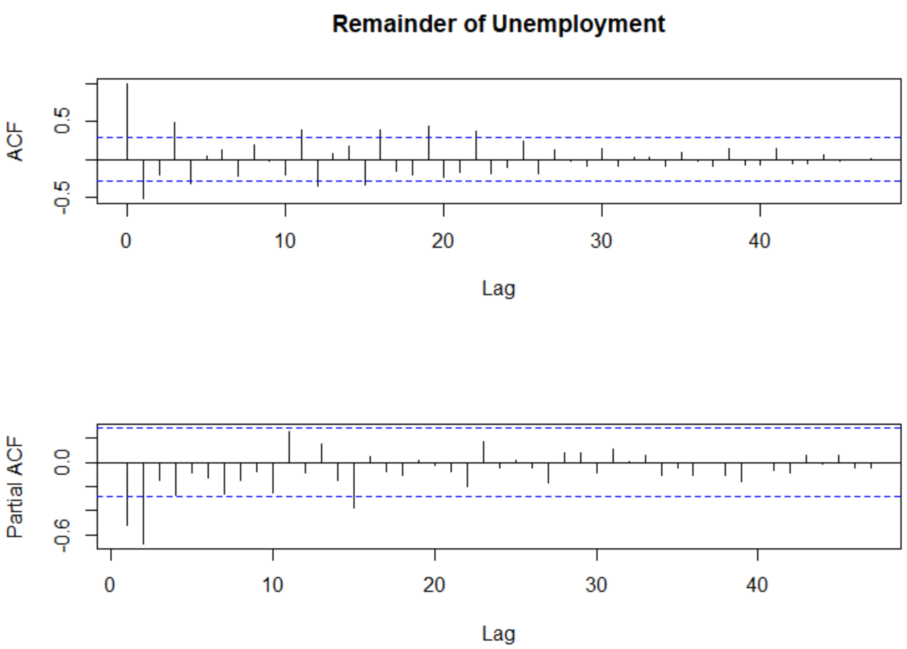


Figure A.18 *ACF/PACF plots for the remainder component of Unemployment*



Appendix B. Var-models

Table B.1 *Summary output for each VAR-model of P/T groups, using 48 months.*

	VAR(3)		VAR(3)		VAR(3)	
Predictors	*Estimates*	*p*	*Estimates*	*p*	*Estimates*	*p*
P/T.lag1	-1.26	**<0.001**	-0.98	**<0.001**	-1.39	**<0.001**
CPI.lag1	0.05	0.698				
P/T.lag2	-0.96	**<0.001**	0.18	0.397	-1.08	**<0.001**
CPI.lag2	-0.35	**0.003**				
P/T.lag3	-0.57	**<0.001**	0.26	0.086	-0.50	**0.004**
CPI.lag3	0.11	0.377				
Unem.lag1			0.00	0.096	0.00	**0.005**
Unem.lag2			0.00	**0.013**	0.00	**0.018**
Unem.lag3			0.00	**0.030**	0.00	**0.007**
CIH.lag1					0.03	**<0.001**
CIH.lag2					-0.01	0.146
CIH.lag3					0.02	**<0.001**
Observations	45		45		45	
R^2 / R^2 adjusted	0.811 / 0.782		0.889 / 0.872		0.777 / 0.721	

Table B.2 *Summary output for each VAR-model of P/T groups, using 30 months.*

	VAR(4)		VAR(3)		VAR(3)	
Predictors	*Estimates*	*p*	*Estimates*	*p*	*Estimates*	*p*
P/T.lag1	-1.81	**<0.001**	-0.77	**0.002**	-1.35	**<0.001**
CPI.lag1	0.12	0.282				
P/T.lag2	-1.94	**<0.001**	0.43	0.143	-1.10	**<0.001**
CPI.lag2	-0.36	**0.005**				
P/T.lag3	-1.64	**<0.001**	0.41	0.064	-0.57	**0.009**
CPI.lag3	-0.24	0.081				
P/T.lag4	-0.69	**0.001**				
CPI.lag4	0.09	0.407				
Unem.lag1			0.00	0.101	0.00	**0.024**
Unem.lag2			0.00	**0.047**	0.00	**0.028**
Unem.lag3			0.00	0.065	0.00	**0.011**
CIH.lag1					0.03	**0.001**
CIH.lag2					-0.01	**0.034**
CIH.lag3					0.03	**<0.001**
Observations	26		27		27	
R^2 / R^2 adjusted	0.915 / 0.877		0.873 / 0.836		0.822 / 0.733	

Figure B.3 *Residuals for VAR(3)-model of P/T group 1, using 48 months.*

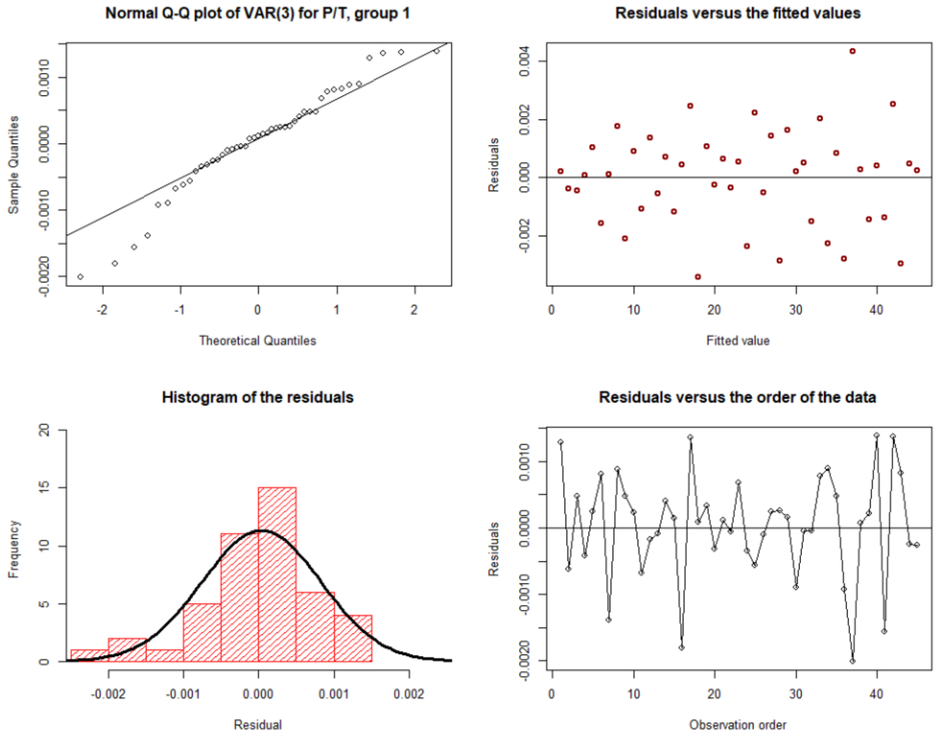


Figure B.4 *Residuals for VAR(3)-model, P/T group 2, using 48 months.*

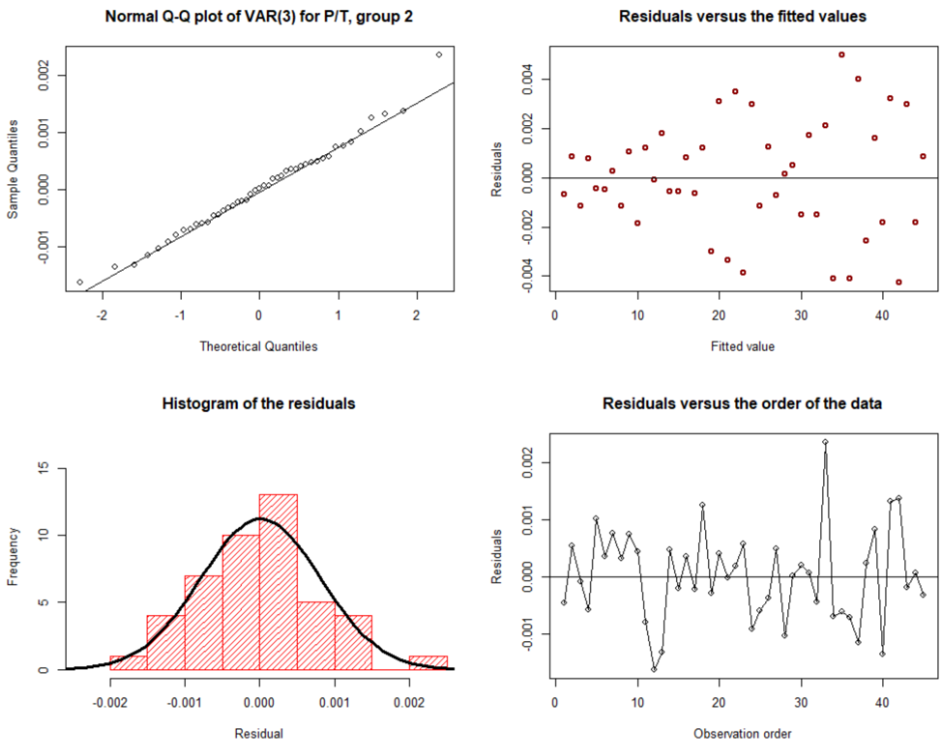


Figure B.5 *Residuals for VAR(3)-model, P/T group 3, using 48 months.*

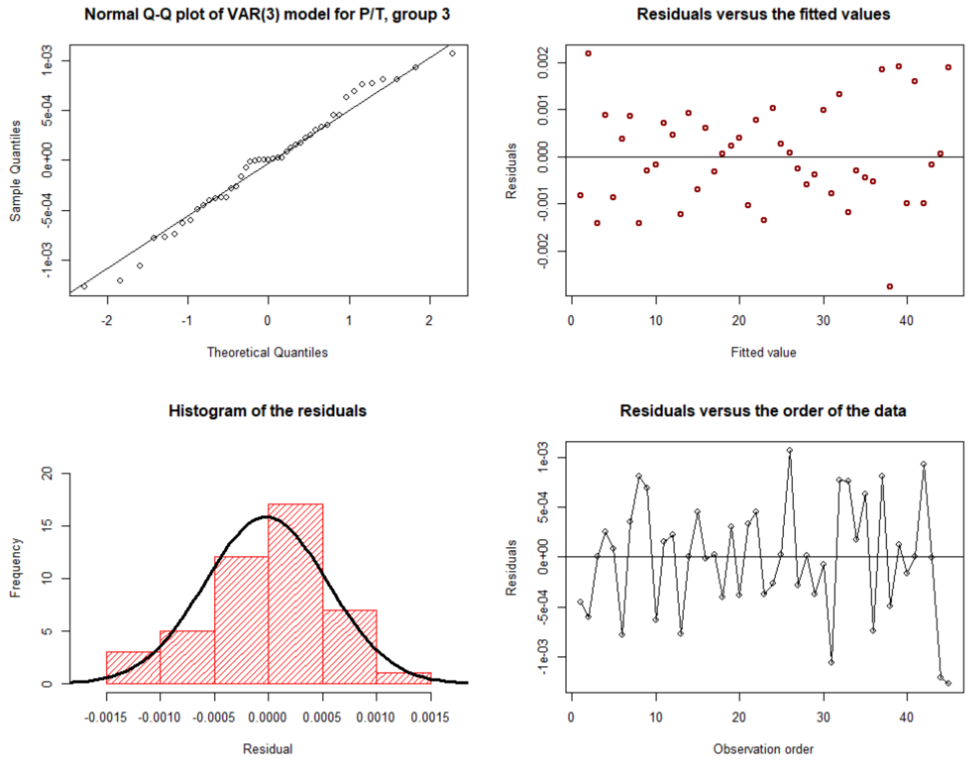


Figure B.6 *Residuals for VAR(4)-model of P/T group 1, using 30 months.*

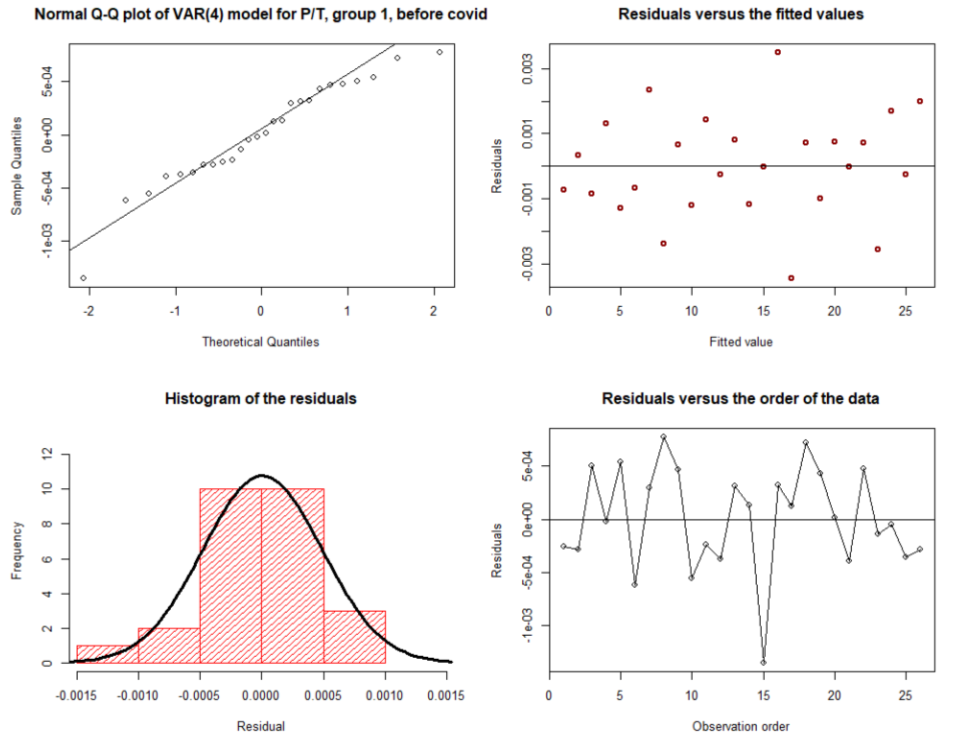


Figure B.7 *Residuals for VAR(3)-model of P/T group 2, using 30 months.*

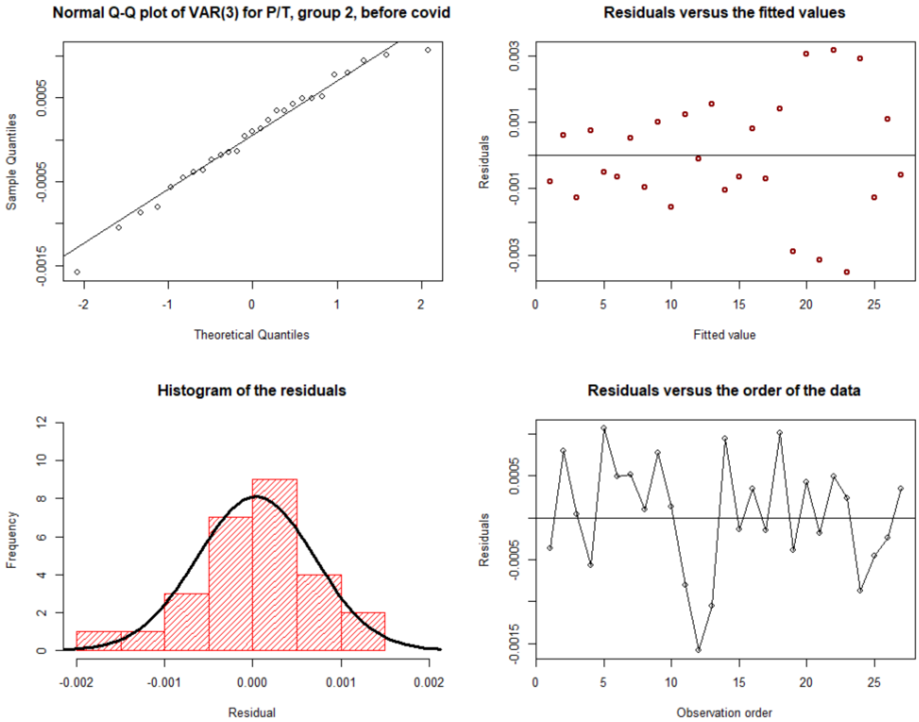


Figure B.8 *Residuals for VAR(3)-model of P/T group 3, using 30 months.*

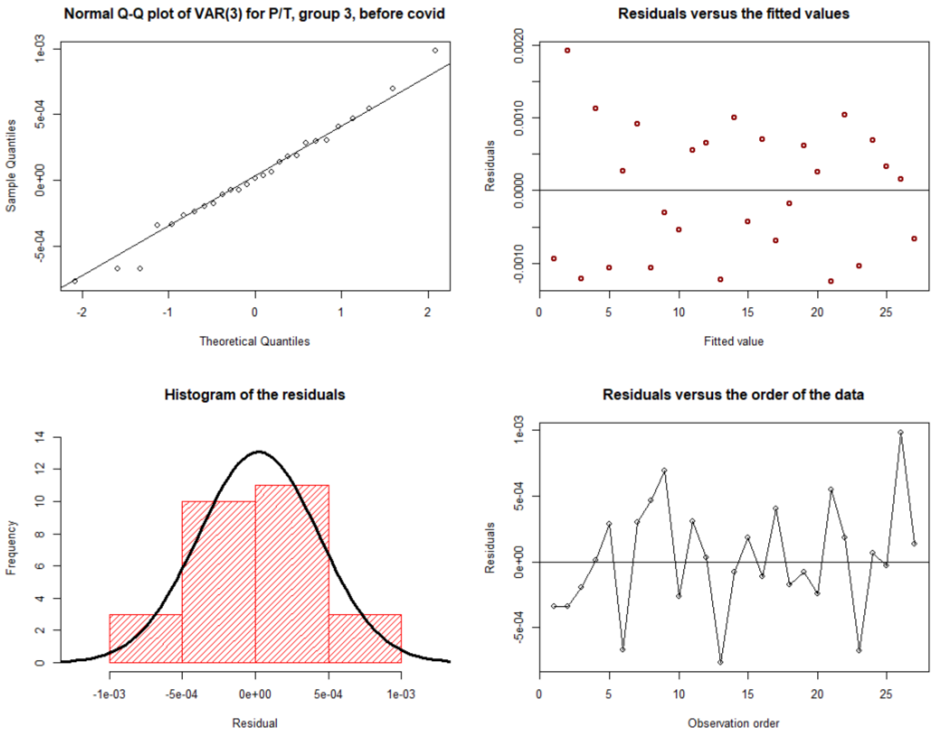


Table B.9 *ARCH-test results for P/T VAR-models.*

VAR-models	Group	Months	Chi-squared	P-value
VAR(3)	1	48	29.437	0.3401
VAR(3)	2	48	19.74	0.8415
VAR(3)	3	48	126.34	0.1097
VAR(4)	1	30	24.473	0.9275
VAR(3)	2	30	31.475	0.2521
VAR(3)	3	30	111.37	0.3928

Table B.10 *JB-test results for P/T VAR-models.*

VAR-models	Group	Months	Chi-squared	P-value
VAR(3)	1	48	3.9934	0.4069
VAR(3)	2	48	1.8678	0.7601
VAR(3)	3	48	1.2499	0.9743
VAR(4)	1	30	2.6406	0.6196
VAR(3)	2	30	1.3597	0.8512
VAR(3)	3	30	2.8712	0.8248

Table B.11 *Test results for Skewness, P/T VAR-models.*

VAR-models	Group	Months	Chi-squared	P-value
VAR(3)	1	48	3.7713	0.1517
VAR(3)	2	48	1.4659	0.4805
VAR(3)	3	48	0.33903	0.9525
VAR(4)	1	30	2.3213	0.3133
VAR(3)	2	30	0.91371	0.6333
VAR(3)	3	30	1.823	0.61

Table B.12 *Test results for kurtosis, P/T VAR-models.*

VAR-models	Group	Months	Chi-squared	P-value
VAR(3)	1	48	0.22214	0.8949
VAR(3)	2	48	0.40193	0.8179
VAR(3)	3	48	0.91085	0.8228
VAR(4)	1	30	0.31931	0.8524
VAR(3)	2	30	0.44604	0.8001
VAR(3)	3	30	1.0482	0.7896

Table B.13 *Test results for Granger Causality, P/T VAR-models.*

Cause	Var-models	Group	Months	F-statistic	P-value	Direction
P/T	VAR(3)	1	48	0.88042	0.455	
CPI	VAR(3)	1	48	4.7065	0.004511	CPI → P/T
P/T	VAR(3)	2	48	2.3229	0.08148	
Unem	VAR(3)	2	48	2.3951	0.0746	
P/T	VAR(3)	3	48	1.4503	0.2023	
HCI	VAR(3)	3	48	6.8894	3.258e-06	→ P/T & Unem
Unem	VAR(3)	3	48	2.1527	0.05318	

P/T	VAR(4)	1	30	0.90888	0.4692	
CPI	VAR(4)	1	30	5.4864	0.001486	P/T→CPI
P/T	VAR(3)	2	30	0.83471	0.4824	
Unem	VAR(3)	2	30	1.5446	0.2171	
P/T	VAR(3)	3	30	2.2569	0.05138	
Unem	VAR(3)	3	30	2.9937	0.01344	→P/T & HCI
HCI	VAR(3)	3	30	5.2559	0.000252	→P/T & Unem

Table B.14 *Summary output for each VAR-model of* $Price/m^2$*, using 48 months.*

	VAR(4)		**VAR(3)**		**VAR(1)**	
Predictors	*Estimates*	*p*	*Estimates*	*p*	*Estimates*	*p*
Price/m² lag1	-0.86	**<0.001**	-1.08	**<0.001**	-0.89	**<0.001**
CPI.lag1	19847.87	**0.041**	-5651.26	**0.034**	765.04	0.231
HCI.lag1	1357.85	**0.013**	287.95	**0.040**		
Price/m² lag2	-0.59	**0.001**	-0.69	**0.001**		
CPI.lag2	-2689.85	0.780	-1340.07	0.609		
HCI.lag2	-2068.44	**<0.001**	51.78	0.689		
Price/m² lag3	-0.71	**<0.001**	-0.46	**0.001**		
CPI.lag3	-26627.50	**0.008**	5447.05	**0.033**		
HCI.lag3	1644.36	**0.010**	-188.59	0.219		
Price/m² lag4	-0.38	**0.008**				
CPI.lag4	12021.32	0.230				
HCI.lag4	-1297.62	**0.026**				
Observations	44		45		47	
R^2 / R^2 adjusted	0.767 / 0.679		0.797 / 0.747		0.790 / 0.781	

Table B.15 *Summary output for each VAR-model of $Price/m^2$, using 30 months*

	VAR(3)		VAR(3)		VAR(2)	
Predictors	*Estimates*	*p*	*Estimates*	*p*	*Estimates*	*p*
Price/m².lag1	-0.65	**<0.001**	-0.95	**<0.001**	-0.76	**0.001**
CPI.lag1	20698.16	0.118	-1380.32	0.716	-219.24	0.769
Price/m².lag2	-0.43	**0.015**	-0.72	**0.003**	-0.03	0.908
CPI.lag2	-18292.20	0.129	-5017.21	0.183	789.96	0.290
Price/m².lag3	-0.59	**0.001**	-0.64	**0.001**		
CPI.lag3	-25635.87	0.050	7396.01	**0.040**		
Observations	27		27		28	
R^2 / R^2 adjusted	0.629 / 0.523		0.795 / 0.737		0.555 / 0.481	

Figure B.16 *Residuals for VAR(4)-model of $Price/m^2$ group 1, using 48 months*

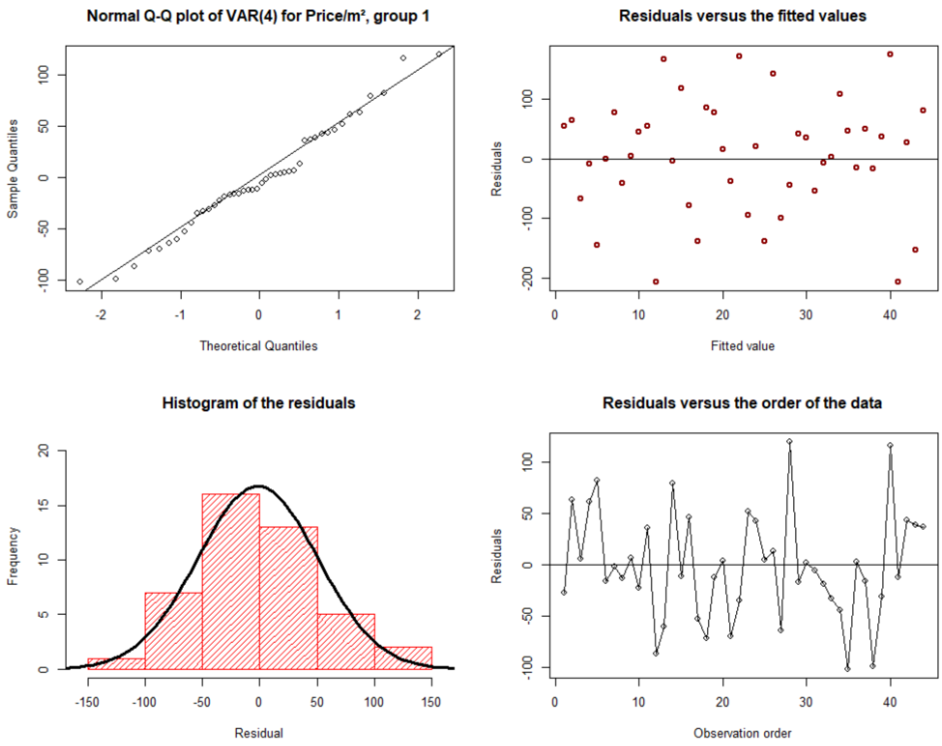


Figure B.17 *Residuals for VAR(3)-model of* $Price/m^2$ *group 2, using 48 months*

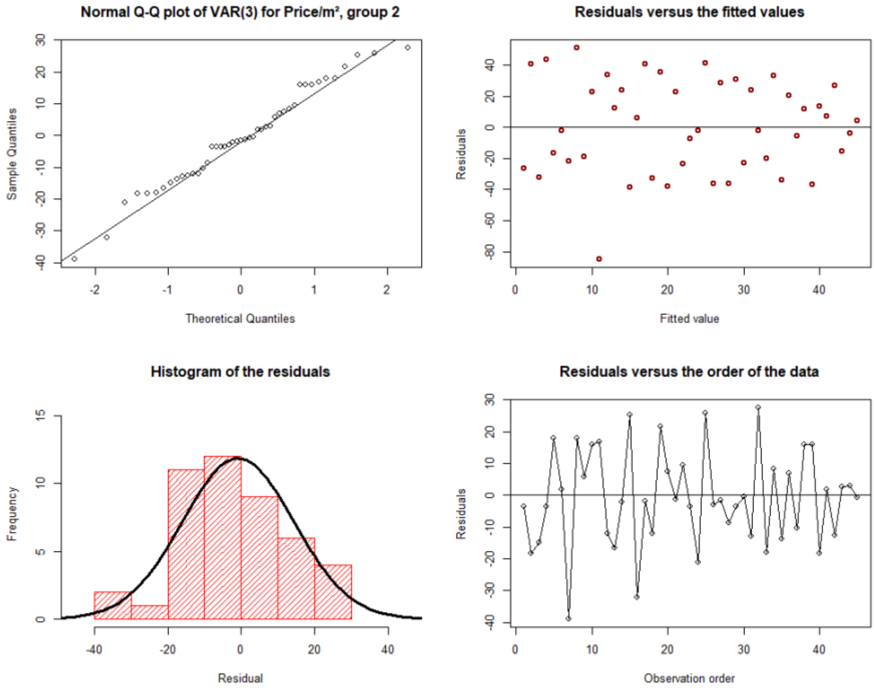


Figure B.18 *Residuals for VAR(1)-model of* $Price/m^2$ *group 3, using 48 months*

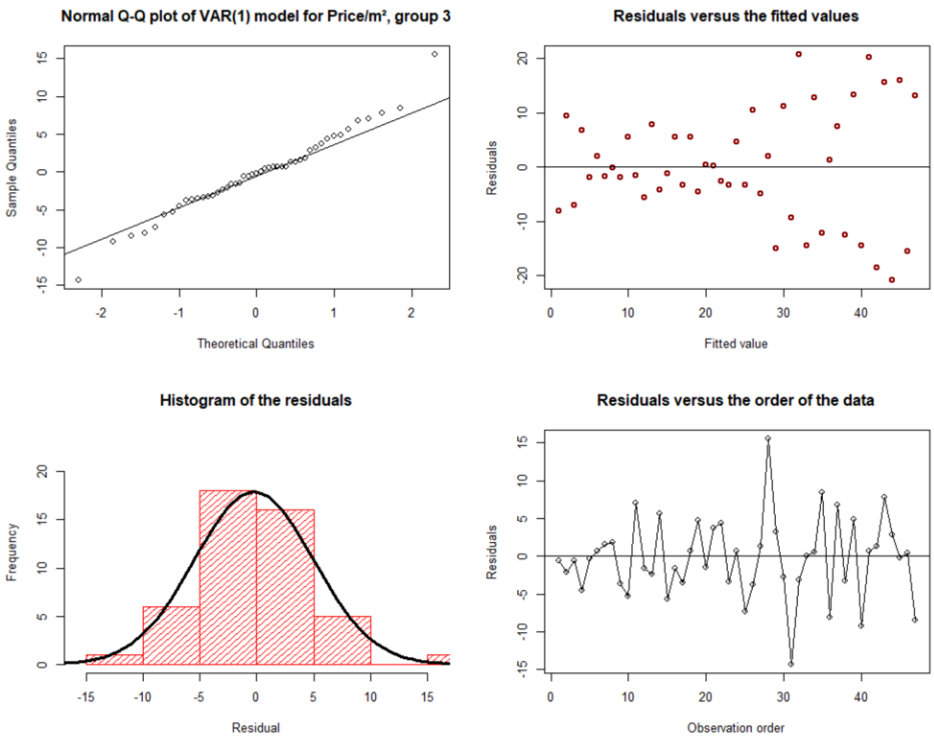


Figure B.19 *Residuals for VAR(3)-model of* $Price/m^2$ *group 1, using 30 months*

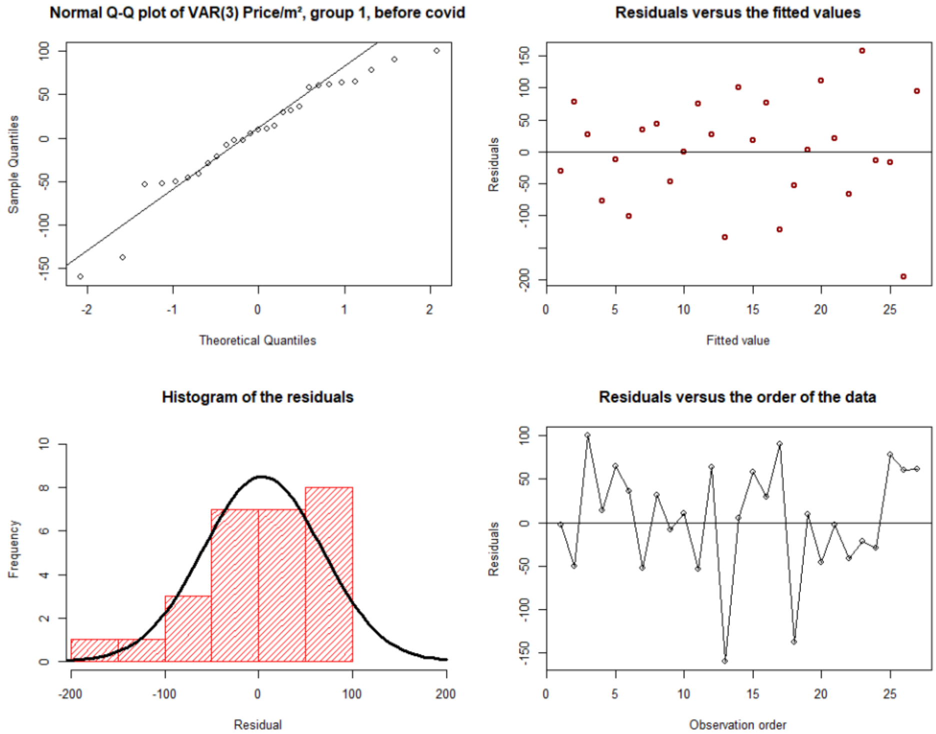


Figure B.20 *Residuals for VAR(3)-model of* $Price/m^2$ *group 2, using 30 months*

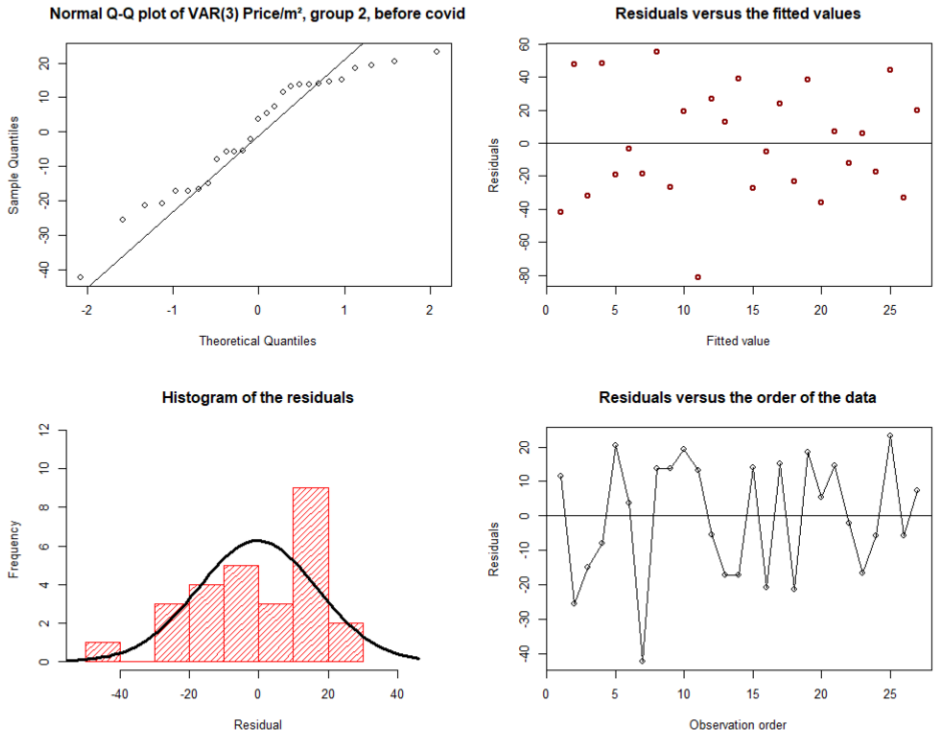


Figure B.21 *Residuals for VAR(2)-model of $Price/m^2$ group 3, using 30 months*

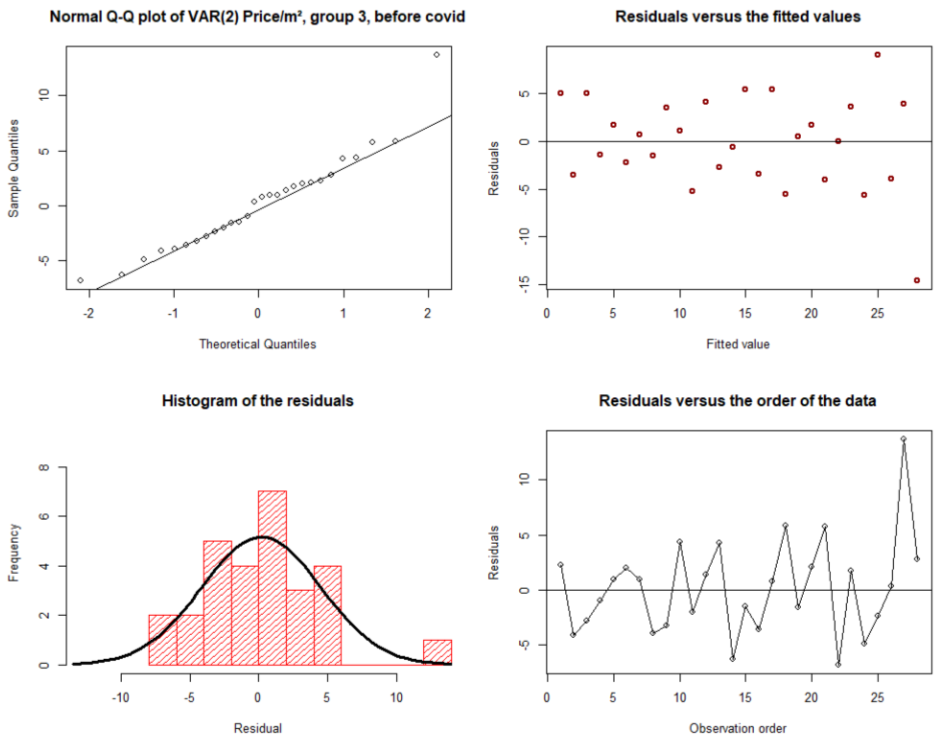


Table B.22 *ARCH-test results for $Price/m^2$ VAR-models.*

VAR-models	Group	Observations	Chi-squared	P-value
VAR(4)	1	48	139.79	0.5836
VAR(3)	2	48	83.445	0.9617
VAR(1)	3	48	6.4321	0.696
VAR(3)	1	30	21.443	0.7651
VAR(3)	2	30	26.418	0.4955
VAR(2)	3	30	17.421	0.4944

Table B.23 *JB-test results for $Price/m^2$ VAR-models.*

VAR-models	Group	Observations	Chi-squared	P-value
VAR(4)	1	48	2.4451	0.8746
VAR(3)	2	48	3.6292	0.7267
VAR(3)	3	48	4.3135	0.3652
VAR(3)	1	30	3.3102	0.5073
VAR(2)	2	30	3.5175	0.4752
VAR(2)	3	30	8.1876	0.08494

Table B.24. *Test results for Skewness, , $Price/m^2$ VAR-models.*

VAR-models	Group	Observations	Chi-squared	P-value
VAR(4)	1	48	2.2619	0.5199
VAR(3)	2	48	1.4659	0.4805
VAR(3)	3	48	1.2409	0.5377

VAR(3)	1	30	2.8332	0.2425
VAR(2)	2	30	3.0773	0.2147
VAR(2)	3	30	5.404	0.06707

Table B.25. *Test results for Kurtosis, , $Price/m^2$ VAR-models.*

VAR-models	Group	Observations	Chi-squared	P-value
VAR(4)	1	48	0.1832	0.9803
VAR(3)	2	48	0.32999	0.9543
VAR(3)	3	48	3.0726	0.2152
VAR(3)	4	30	0.47699	0.7878
VAR(2)	5	30	0.44018	0.8024
VAR(2)	6	30	2.7837	0.2486

Table B.26. *Test results for Granger Causality, $Price/m^2$ VAR-models.*

Cause	Var-models	Group	Observations	F-statistic	P-value	Direction
$Price/m^2$	VAR(4)	1	48	0.52497	0.8351	
CPI	VAR(4)	1	48	3.4085	0.001727	→$Price/m^2$ & HCI
HCI	VAR(4)	1	48	3.3905	0.001807	→$Price/m^2$ & HCI
$Price/m^2$	VAR(3)	2	48	1.1254	0.3525	
CPI	VAR(3)	2	48	2.8691	0.01241	→$Price/m^2$ & HCI
HCI	VAR(3)	2	48	3.5529	0.002996	→$Price/m^2$ & CPI
$Price/m^2$	VAR(1)	3	48	0.4176	0.5198	
CPI	VAR(1)	3	48	1.4761	0.2276	
$Price/m^2$	VAR(3)	1	30	1.3085	0.2843	
CPI	VAR(3)	1	30	4.1197	0.01195	→$Price/m^2$
$Price/m^2$	VAR(3)	2	30	0.84974	0.4746	
CPI	VAR(3)	2	30	3.286	0.02986	→$Price/m^2$
$Price/m^2$	VAR(2)	3	30	0.19165	0.8262	
CPI	VAR(2)	3	30	0.70928	0.4971	

Appendix C. Chow-tests

Table C.1 Chow test for Price/m^2

Timeseries 1	Timeseries 2	P-value	Result
Price/m^2 group 1	Price / m^2 group 2	0.003738	Break
Price/m^2 group 2	Price /m^2 group 1	0.1927	No Break
Price/m^2 group 1	Price / m^2 group 3	0.0327	Break
Price/m^2 group 3	Price / m^2 group 1	0.8424	No Break
Price/m^2 group 2	Price / m^2 group 3	0.4249	No Break
Price/m^2 group 3	Price / m^2 group 2	0.379	No Break
Price/m^2 group 1 before covid	Price / m^2 group 2 before covid	0.3469	No Break
Price/m^2 group 2 before covid	Price / m^2 group 1 before covid	0.5326	No Break
Price/m^2 group 1 before covid	Price / m^2 group 3 before covid	0.2879	No Break
Price/m^2 group 3 before covid	Price / m^2 group 1 before covid	0.3102	No Break
Price/m^2 group 2 before covid	Price / m^2 group 3 before covid	0.002859	Break
Price/m^2 group 3 before covid	Price / m^2 group 2 before covid	0.001983	Break
Price/m^2 group 1 after covid	Price / m^2 group 2 after covid	0.01548	Break
Price/m^2 group 2 after covid	Price / m^2 group 1 after covid	0.002541	Break
Price/m^2 group 1 after covid	Price / m^2 group 3 after covid	0.001187	Break

Price/m^2 group 3 after covid	Price / m^2 group 1 after covid	0.00003432	Break
Price/m^2 group 2 after covid	Price / m^2 group 3 after covid	0.0004843	Break
Price/m^2 group 3 after covid	Price / m^2 group 2 after covid	0.0001267	Break

Table C.2 Chow test for P/T

Timeseries 1	Timeseries 2	P-value	Result
P/T group 1	P/T group 2	0.3418	No Break
P/T group 2	P/T group 1	0.001004	Break
P/T group 1	P/T group 3	0.5074	No Break
P/T group 3	P/T group 1	0.0008072	Break
P/T group 2	P/T group 3	0.006146	Break
P/T group 3	P/T group 2	0.001509	Break
P/T group 1 before covid	P/T group 2 before covid	0.00356	Break
P/T group 2 before covid	P/T group 1 before covid	0.1023	No Break
P/T group 1 before covid	P/T group 3 before covid	0.000001883	Break
P/T group 3 before covid	P/T group 1 before covid	0.0002811	Break
P/T group 2 before covid	P/T group 3 before covid	0.0002739	Break

P/T group 3 before covid	P/T group 2 before covid	0.001182	Break
P/T group 1 after covid	P/T group 1 after covid	0.006355	Break
P/T group 2 after covid	P/T group 1 after covid	0.001362	Break
P/T group 1 after covid	P/T group 3 after covid	0.2464	No Break
P/T group 3 after covid	P/T group 1 after covid	0.06168	No Break
P/T group 2 after covid	P/T group 3 after covid	0.4812	No Break
P/T group 3 after covid	P/T group 2 after covid	0.5596	No Break

Table C.3 Chow test for sales of P/T

Timeseries 1	Timeseries 2	P-value	Result
Sale for P/T group 1	Sale for P/T group 2 after covid	0.3389	No Break
Sale for P/T group 2 after covid	Sale for P/T group 1	0.6959	No Break
Sale for P/T group 1 after covid	Sale for P/T group 3 after covid	0.6922	No Break
Sale for P/T group 3 after covid	Sale for P/T group 1 after covid	0.6913	No Break

Sale for P/T group 2 after covid	Sale for P/T group 3 after covid	0.7883	No Break
Sale for P/T group 3 after covid	Sale for P/T group 2 after covid	0.3284	No Break
Sale for P/T group 1 before covid	Sale for P/T group 2 before covid	0.8095	No Break
Sale for P/T group 2 before covid	Sale for P/T group 1 before covid	0.6537	No Break
Sale for P/T group 1 before covid	Sale for P/T group 3 before covid	0.8148	No Break
Sale for P/T group 3 before covid	Sale for P/T group 1 before covid	0.5529	No Break
Sale for P/T group 2 before covid	Sale for P/T group 3 before covid	0.8954	No Break
Sale for P/T group 3 before covid	Sale for P/T group 2 before covid	0.7551	No Break
Sale for P/T group 1 after covid	Sale for P/T group 2 after covid	0.8165	No Break
Sale for P/T group 2 after covid	Sale for P/T group 1 after covid	0.5945	No Break

Sale for P/T group 1 after covid	Sale for P/T group 3 after covid	0.7973	No Break
Sale for P/T group 3 after covid	Sale for P/T group 1 after covid	0.9568	No Break
Sale for P/T group 2 after covid	Sale for P/T group 3 after covid	0.2294	No Break
Sale for P/T group 3 after covid	Sale for P/T group 2 after covid	0.3547	No Break

Table C.4 Chow test for sales of Price/m^2

Timeseries 1	Timeseries 2	P-value	Results
Sale for Price/m^2 group 1 after covid	Sale for price / m^2 group 2 after covid	0.002201	Break
Sale for Price/m^2 group 2 after covid	Sale for price / m^2 group 1 after covid	0.3509	No Break
Sale for Price/m^2 group 1 after covid	Sale for price / m^2 group 3 after covid	0.001996	Break
Sale for Price/m^2 group 3 after covid	Sale for price / m^2 group 1 after covid	0.3564	No Break
Sale for Price/m^2 group 2 after covid	Sale for price / m^2 group 3 after covid	0.282	No Break
Sale for Price/m^2 group 3 after covid	Sale for price / m^2 group 2 after covid	0.2254	No Break
Sale for Price/m^2 group 1 before covid	Sale for price / m^2 group 2 before covid	0.9059	No Break

Sale for Price/m^2 group 2 before covid	Sale for price / m^2 group 1 before covid	0.5564	No Break
Sale for Price/m^2 group 1 before covid	Sale for price / m^2 group 3 before covid	0.7606	No Break
Sale for Price/m^2 group 3 before covid	Sale for price / m^2 group 1 before covid	0.8305	No Break
Sale for Price/m^2 group 2 before covid	Sale for price / m^2 group 3 before covid	0.3775	No Break
Sale for Price/m^2 group 3 before covid	Sale for price / m^2 group 2 before covid	0.6701	No Break
Sale for Price/m^2 group 1 after covid	Sale for price / m^2 group 2 after covid	0.3337	No Break
Sale for Price/m^2 group 2 after covid	Sale for price / m^2 group 1 after covid	0.5914	No Break
Sale for Price/m^2 group 1 after covid	Sale for price / m^2 group 3 after covid	0.5501	No Break
Sale for Price/m^2 group 3 after covid	Sale for price / m^2 group 1 after covid	0.6598	No Break
Sale for Price/m^2 group 2 after covid	Sale for price / m^2 group 3 after covid	0.6545	No Break
Sale for Price/m^2 group 3 after covid	Sale for price / m^2 group 2 after covid	0.5264	No Break

Table C.5 Chow test for P/T forecasts

Timeseries 1	Timeseries 2	P-Value	Result
P/T group 1 forecast	P/T group 1 actual change	2.2 * 10^-16	Break
P/T group 1 actual values	P/T group 1 forecast	1.123*10^-9	Break
P/T group 2 forecast	P/T group 2 actual values	2.2 * 10^-16	Break
P/T group 2 actual values	P/T group 2 forecast	2.2 * 10^-16	Break
P/T group 3 forecast	P/T group 3 actual values	1.555*10^-7	Break
P/T type 3 actual values	P/T group 3 forecast	1.444 * 10^- 5	Break

Table C.6 Chow test for $\mathrm{Price/m^2}$ forecasts

Timeseries 1	Timeseries 2	P-value	Result
$\mathrm{Price/m^2}$ group 1 forecast	$\mathrm{Price/m^2}$ group 1 actual values	0.04891	Break
$\mathrm{Price/m^2}$ group 1 actual values	$\mathrm{Price/m^2}$ group 1 forecast	0.01004	Break
$\mathrm{Price/m^2}$ group 2 forecast	$\mathrm{Price/m^2}$ group 2 actual values	2.2*10^-16	Break
$\mathrm{Price/m^2}$ group 2 actual values	$\mathrm{Price/m^2}$ group 2 forecast	5.551*10^-15	Break
$\mathrm{Price/m^2}$ group 3 forecast	$\mathrm{Price/m^2}$ group 3 actual values	2.2*10^-16	Break
$\mathrm{Price/m^2}$ group 3 actual values	$\mathrm{Price/m^2}$ group 3 forecast	2.2*10^-16	Break

Figure C.7 Index of variables Price/m^2 and P/T for 48 months and before and after covid.

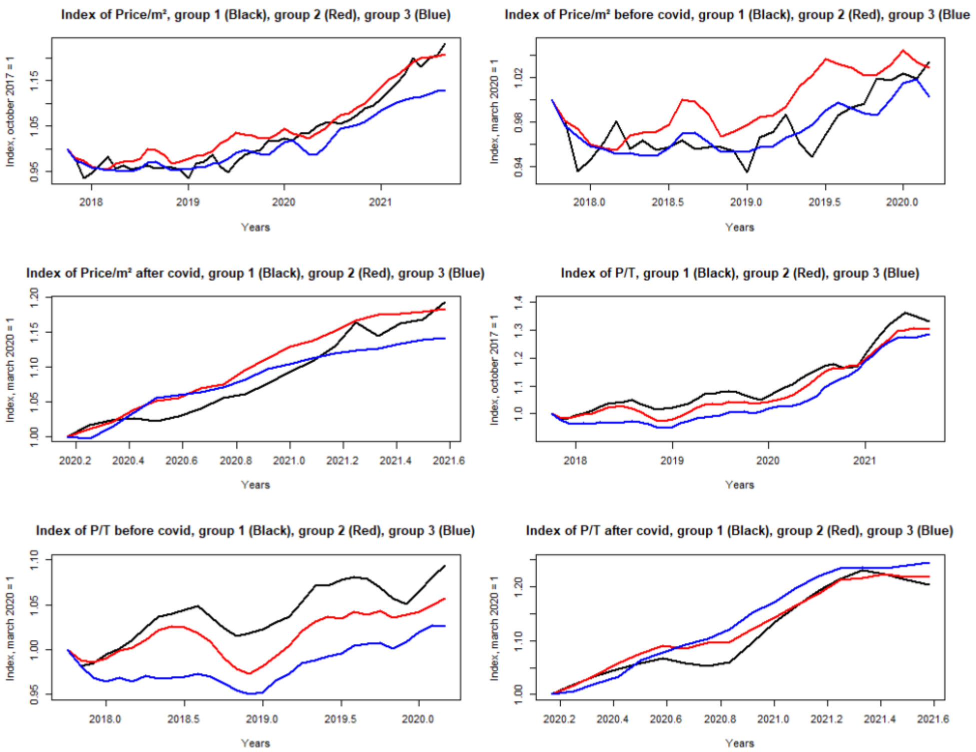
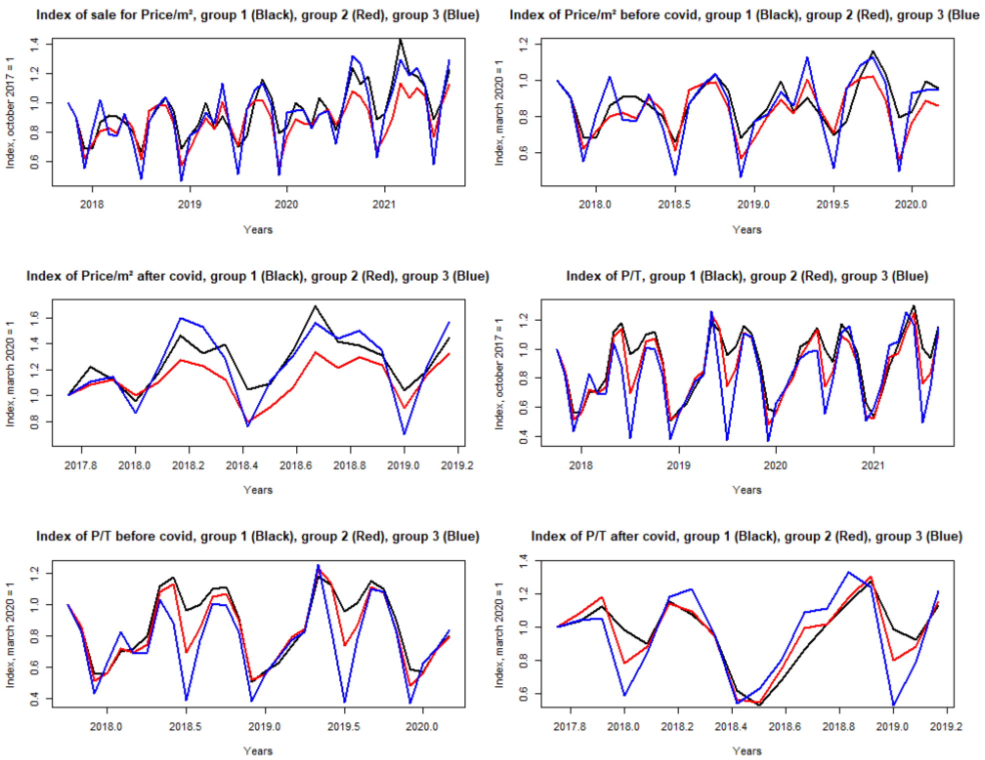


Figure C.8 Index of sales for variables Price/m^2 and P/T for 48 months and before and after covid



References

Sharon L, L. 2010. Sampling: Design and Analysis. 2nd ed. Boston: Brooks/Cole.

Jeffrey M, W. 2016. Introductory Econometrics: A Modern Approach. 6th ed. Boston: Cengage Learning.

Douglas C, M., Cheryl L, J. and Murat, K. 2015. Introduction to Time-series Analysis and Forecasting. 2nd ed. New Jersey: John Wiley & Sons, Inc.

Nyquist, H. 2017. Statistikens Grunder. 1st ed. Studentlitteratur

Myndigheten för tillväxtpolitiska utvärderingar och analyser. 2014. Bättre statistik för bättre regional – och landsbygdspolitik. Österstund: Tillväxtanalys. [online] (2/1 - 22) Available at: https://www.tillvaxtanalys.se/download/18.62dd45451715a00666f1f3a9/1586366195534/rapport_2014_04_rev1.pdf

Statistikmyndigheten. Företeckning över storstadsområden med ingående kommuner I alfabetisk ordning. [online] (2/1 - 22) Available at: https://www.scb.se/contentassets/c4b8142033a9440ca53725ca32321a74/storstadsomr.pdf

Svensk Fastighetsförmedling, 2021. Bostadspriser. [online] Available at: https://www.svenskfast.se/guider/bostadspriser/

Statistiska Centralbyrån, 2021. Konsumentprisindex. [online] (14/12-21) Available at: https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__PR__PR0101__PR0101A/KPItotM/#

Riksbanken, 2021. Reporänta. [online] (25/11-21) Available at: https://www.riksbank.se/sv/statistik/sok-rantor--valutakurser/reporanta-in--och-utlaningsranta/

Statistiska Centralbyrån, 2021. Arbetslösa 15–74 år (AKU) efter arbetslöshetstidens längd, kön och ålder. [online] (16/12-21) Available at: https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__AM__AM0401__AM0401L/NAKUArblosaTM/#

Statistiska Centralbyrån, 2021. Disponibel inkomst för hushåll efter region, hushållstyp och ålder. [online] (27/1-21) Available at: https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__HE__HE0110__HE0110G/Tab4bDispInkN/#

Statistiska Centralbyrån, 2021. Konjunkturbarometer. [online] (2/2–21) Available at: https://www.scb.se/hitta-statistik/temaomraden/sveriges-ekonomi/konjunkturindikatorer/konjunkturbarometer/?showAllContentLinks=True#113147

Konjunkturinstitutet, 2021. Metodbok för Konjunkturbarometer. [PDF] Available at: https://www.konj.se/metodbok

Konjunkturinstitutet, 2021. The Economic Tendency Survey. [online] Available at: https://www.konj.se/english/about-nier/surveys.html

Wikipedia, 2021. Linear interpolation. [online] (14/9-21) Available at: https://en.wikipedia.org/wiki/Linear_interpolation

Anomaly, 2019. Extracting Seasonality and Trend from Data: Decomposition Using R. [online] Available at: https://anomaly.io/seasonal-trend-decomposition-in-r/index.html

Javlacalle, 2016. ACF and PACF Formula. Stackexchange.com Programming blog, [blog] 16 december. Available at: https://stats.stackexchange.com/questions/129052/acf-and-pacf-formula

The Analysis Factor, 2014. When Listwise Deletion works for Missing Data. [online] Available at: https://www.theanalysisfactor.com/when-listwise-deletion-works/

Statology, 2020. The Breusch–Pagan Test: Definition & Example. [online] Available at: https://www.statology.org/breusch-pagan-test/

Statistiska Centralbyrån, 2021. Konsumentprisindex (1949=100). [online] (14/12-21) Available at: https://www.scb.se/hitta-statistik/statistik-efter-amne/priser-och-konsumtion/konsumentprisindex/konsumentprisindex-kpi/pong/tabell-och-diagram/konsumentprisindex-kpi/kpi-index-1949100/

Ekonomifakta, 2021. Arbetslöshet. [online] (27/12-21) Available at: https://www.ekonomifakta.se/fakta/arbetsmarknad/arbetsloshet/arbetsloshet/?graph=/25554/1,2/2017-/

Statistics Sweden, 2021. Labor Force Surveys (LFS). [online] (16/12-21) Available at: https://www.scb.se/en/finding-statistics/statistics-by-subject-area/labour-market/labour-force-surveys/labour-force-surveys-lfs/

Forecasting: Principles and Practice, 2018. Time series components. [online] Available at: https://otexts.com/fpp2/components.html

Cleveland, William S. 1979. "Robust Locally Weighted Regression and Smoothing Scatterplots". Journal of the American Statistical Association. 74 (368).

JDemetra+. Seasonal adjustment methods. [online] Available at: https://sylwiagrudkowska.github.io/JDemetra-documentation/pages/theory/

JDemetra+. SEATS. [online] Available at: https://sylwiagrudkowska.github.io/JDemetra-documentation/pages/theory/SA_SEATS.html

Forecasting: Principles and Practice, 2018. STL Decomposition. [online] Available at: https://otexts.com/fpp2/stl.html

Wikipedia, 2021. Local Regression. [online] (24/12-21) Available at: https://en.wikipedia.org/wiki/Local_regression

MathWorks, 2021. Engle's ARCH Test. [online] Available at: https://www.mathworks.com/help/econ/engles-arch-test.html

Wikipedia, 2021. Jarque–Bera Test. [online] Available at: https://en.wikipedia.org/wiki/Jarque%E2%80%93Bera_test

Wikipedia, 2021. Kurtosis. [online] Available at:

https://en.wikipedia.org/wiki/Kurtosis

Wikipedia, 2021. X-13ARIMA-SEATS. [online] Available at:
https://en.wikipedia.org/wiki/X-13ARIMA-SEATS

Forecasting: Principles and Practice, 2021. Vector Autoregressions. [online] Available at:
https://otexts.com/fpp2/VAR.html

Svensk Mäklarstatistik, 2021. Prisutveckling bostadsrätter. [online] Available at:
https://www.maklarstatistik.se/omrade/riket/#/bostadsratter/48m-prisutveckling

Svensk Mäklarstatistik, 2021. Prisutveckling villor. [online] Available at:
https://www.maklarstatistik.se/omrade/riket/#/villor/48m-prisutveckling