UNIVERSITY OF TARTU
FACULTY OF SCIENCE AND TECHNOLOGY
INSTITUTE OF MATHEMATICS AND STATISTICS

Enelin Haviko

Modelling volume of savings deposits

Mathematics and Statistics Curriculum

(Mathematical Statistics)

Master's thesis (30 ECTS)

Supervisors:

Prof. Kalev Pärna

Jonathan Diamant, MSc

TARTU 2022

Modelling volume of savings deposits

Master’s thesis

Enelin Haviko

Abstract. The objective of this master’s thesis is to find an appropriate time series model to forecast the volume of private customers’ savings deposits of one undisclosed Swedish financial institution. Models such as Holt, Holt-Winters, ARIMA and ARIMAX are fitted to the data under analysis. The best performing model is ARIMA showing approximately 20% lower error measures during the out-of-sample test period compared to the best ARIMAX models.

CERCS research specialisation: P160 Statistics, operations research, programming, financial and actuarial mathematics

Keywords: Time series analysis, ARIMA processes, forecasting

Kogumishoiuste mahu modelleerimine

Magistritöö

Enelin Haviko

Lühikokkuvõte. Käesoleva magistritöö eesmärk on leida sobiv aegridade mudel prognoosimaks ühe avalikustamata Rootsi finantsasutuse eraisikute kogumishoiuste mahtu. Andmetele rakendatavateks mudeliteks on Holt, Holt-Winters, ARIMA ja ARIMAX. Parima prognoosivõimega mudeliks osutus ARIMA, mille keskmised prognoosivead olid võrreldes parimate ARIMAX mudelitega ligikaudu 20% madalamad.

CERCS teaduseriala: P160 Statistika, operatsioonianalüüs, programmeerimine, finants- ja kindlustusmatemaatika

Märksõnad: Aegridade analüüs, ARIMA protsessid, prognoosimine

Contents

Introduction

The financial crisis in the 2008 showed the importance of stable funding and sufficient liquidity in the financial sector. Regulators have since put liquidity and stable funding in the financial sector to their top priorities. One of the focuses has been on non-maturity deposits (NMDs) as they are an important part of commercial bank's funding and are characterized by two options. First, customers can withdraw their money any time without penalty as there is no contractual maturity and second, banks are allowed to change the deposit rate at any point in time as deposit rate is administratively set and reprice overnight.

In April 2016, the Basel Committee on Banking Supervision (BCBS) announced the standard on capital framework for interest rate risk in the banking book (IRRBB). IRRBB refers to the current or prospective risk to the bank's capital and earnings arising from adverse movements in interest rates that affect the bank's banking book positions. Changes in interest rates have an effect in the present value of future cash flows and the bank's earnings by altering interest rate-sensitive income and expenses, i.e. net interest income (NII). Excessive interest rate risk can pose a significant threat to a bank's current capital base and future earnings if not managed appropriately. One of the findings by BCBS was that banks should document, monitor, and regularly update primal assumptions regarding NMDs behaviour and balances. Although, depositors are free to withdraw their money at any time, NMD balances have historically proved to be relatively stable. Therefore, banks should analyse its depositor base to be able to identify the proportion of core deposits, i.e. NMDs which are unlikely to be redrawn or reprice even under significant changes in interest rate environment. (Basel Committee on Banking Supervision, 2016)

The aim of the study is to find the best model to forecast the volume of private customers' savings deposits of one undisclosed financial institution. The financial institution has previously implemented a deposit model where it assesses both core volumes and their duration. Next steps include building a more sophisticated model to forecast volumes for the NII forecast and the funding plan. This is especially needed in the condition where NII sensitivity has increased significantly due to excess liquidity. With rates being negative, until recently, and the financial institution's limited will and capacity to discourage depositors, deposits have been seen as an expensive source of funding the banks operations. By being able to forecast their volumes in a comprehensive manner given different scenarios the financial institution will be better equipped to manage its funding and costs.

In Section 1 an overview of previous research on the topic of deposit volume forecasting is presented. The focus is on determinants such as deposit rate, consumer confidence indicator, purchasing managers' index, gross domestic product, house price index and quantitative easing. Section 2 provides the reader the theory of time series analysis, including Holt, Holt-Winters, ARIMA and ARIMAX models, and its application in R software. Finally, in Section 3 the data under analysis is described and different time series models are fitted. Furthermore, the models are compared, validated and the best model is chosen. The evaluation of the models is based on three principles – goodness of fit, accuracy and simplicity.

1 Literature review

There are numerous analyses done to investigate determinants of deposit volumes. The next subchapters give an overview of the different factors that can possibly affect NMD balances. The choice of variables is based on previous studies in the area of deposit volumes and time series forecasting.

1.1 Deposit rate

The deposit rate has for some time been considered to have high explanatory power on deposit volumes as it should be one of the key factors considered by customers when allocating assets. For example, Masson et al. (1998) found a positive effect of interest rates on private savings in industrial countries. However, as the deposit rates have recently been decreasing and close to zero, but deposit volumes on the other hand have grown, there have been alternative research outcomes. Bank of Japan (2014) showed that NMD balances decrease in a high interest rate environment and vice versa.

1.2 Consumer confidence indicator

Consumer confidence indicator (CCI) is produced by the Directorate General for Economic and Financial Affairs (DG ECFIN) of the European Commission to indicate economic perceptions and expectations. It is calculated as the arithmetic average of the balances (in percentage points) of the answers to the questions on the financial situation of households, the general economic situation, unemployment expectations and savings (Directorate General for Economic and Financial Affairs, 2016). The questionnaire covers respondent's view on the economic situation over the previous and upcoming 12 months. Two example questions are shown in Figure 1.

How do you expect the number of people unemployed in this country to change over the next 12 months? The number will...

+ +	increase sharply
+	increase slightly
=	remain the same
−	fall slightly
− −	fall sharply
N	don't know.

How do you think that consumer prices have developed over the last 12 months? They have…

+ +	risen a lot
+	risen moderately
=	risen slightly
−	stayed about the same
− −	fallen
N	don't know.

Figure 1: Sample questions from the CCI questionnaire. Source: Directorate General for Economic and Financial Affairs (2016, p 36).

The indicator takes values between -100 and 100. A positive value indicates that consumers are rather optimistic about the conditions. However, a negative value refers to a more pessimistic view.

The hypothesis that increased uncertainty about the future economic situation causes greater savings rates has been tested numerous times. For example, an article written by Kłopocka in 2016 proved that consumer confidence indicator has predictive power for Polish future household saving rates. The multiple linear regression analysis (OLS technique) of time-series showed that four lags of changes in CCI explain 23% of the variation of changes in total household saving rate. There are more examples about the influence of economic uncertainty on savings from Carroll et al. (2012), Mody et al. (2012), Carroll and Samwick (1998).

1.3 Purchasing managers' index

The purchasing managers' index (PMI) is a business cycle indicator for the economy, reported by Swedbank in cooperation with Sveriges Inköps- och Logistikförbund (Silf). It measures the activity level of purchasing managers both in the manufacturing and the services sector. On monthly basis, purchasing managers are surveyed and the two indexes – PMI manufacturing and PMI services – are calculated. A level above 50 indicates expansion, while an index level of below 50 signals a contraction. The aim is to quickly measure the current state of the economy as purchasing managers generally have early access to their company's performance, which can be a leading indicator of overall economic performance. (Swedbank, 2022)

1.4 Gross domestic product

Gross domestic product (GDP) is one of the fundamental economic growth measures. One of the main theories is that a general long-term increase in GDP creates a surplus in the economy which is transferred to account balances and vice versa. Chaturvedi et al. confirmed in 2009 that saving rate is positively affected by the real GDP growth rate.

1.5 House price index

The house price index (HPI) measures inflation in the residential real estate market. The index calculation includes all kinds of new and existing residential property like flats, detached houses and terraced houses, purchased by households. (Eurostat, 2022a)

There are two hypotheses regarding HPI and savings deposits. Firstly, selling real estate at high market price is increasing wealth and account balances. Secondly, in order to buy a property, one needs to save money and thus savings deposits are increasing. The positive relationship between Sweden's HPI and household savings was proved by Kennedy and Andersen (1994). An alternative result was provided by Klerck and Salame (2017) where northern Europe countries were analysed. The paper concluded that one percent increase in house prices will decrease savings 0.113 percent.

1.6 Quantitative easing

Following the financial crises in the 2008, central banks have started using a new tool for monetary policy, such as large-scale asset purchases (LSAPs), also known as quantitative easing (QE). Although, it was first used by Bank of Japan already in 2001, it has become widely used over the last decade with the aim of stimulating demand to increase inflation (Grimaldi et al. 2021).

The concept behind it is that a central bank starts purchasing government bonds on the secondary market. As a result, yields for those safe bonds fall and it becomes more attractive for investors to turn to alternative assets. In this way, the lower yields for government bonds spread onwards to other parts of the financial markets. Lower interest rates help the banks to decrease rates for loans and deposits, which in turn increases companies' willingness to invest and households' incentive to consume. As a result, asset prices rise and so does the wealth of both households and companies. (Sveriges Riksbank, 2021a)

2 Time Series Analysis

This section introduces the theory behind modelling and forecasting time series and is based on the book by Chatfield (2004: pp. 1, 5, 10-14, 19-20, 23-24, 34-35, 37-38, 48, 62, 66, 73-89, 245-246), unless noted otherwise.

2.1 Fundamentals

Time series is defined as a collection of observations x_t made sequentially through time and is referred to as discrete when observations are taken only at specific times, usually at equal time intervals. Observations x_t are realizations of a random variable X_t, $t \in Z$, and a family of random variables $(X_t)_{t \in Z}$ is called a stochastic, or random, process.

One of the main goals in analysing time series is to predict the future values as accurately as possible, given the information available about the past. The estimate of x_{T+h} based on the values $x_1, \dots, x_T$ is denoted $\hat{x}_{T+h|T}$.

There are four main components in time series forecasting:

1) preliminary analysis,
2) choice of suitable model,
3) model calibration and goodness of fit,
4) estimating future values and calculating confidence intervals.

2.1.1 Decomposition

The first part of preliminary analysis is plotting the data. It enables to see patterns, unusual fluctuations, changes over time, and relationships between variables. Traditional methods of time series analysis decompose the variation in a time series into trend T_t, seasonal component S_t and irregular component ε_t. In case of additive decomposition, the formula is given as

$$x_t = T_t + S_t + \varepsilon_t,$$

and in case of multiplicative decomposition as

$$x_t = T_t S_t \varepsilon_t.$$

The additive decomposition is applicable when the magnitude of the seasonal component is roughly constant through the series, while the multiplicative method is preferred when the seasonal effect is directly proportional to the mean. A logarithmic transformation can be used to make the seasonal effect from multiplicative to additive.

2.1.2 Stationarity

The next element of interest is stationarity. A random process $(X_t)_{t\in Z}$ is said to be (strictly) stationary if random vectors $(X_1, \dots, X_m)$ and $(X_{1+q}, \dots, X_{m+q})$ have the same distribution, m and q are integers. In other words – its statistical properties, like mean and variation, do not depend on the time at which the series is observed. If random vectors $(X_1, \dots, X_m)$ and $(X_{1+q}, \dots, X_{m+q})$ have all up to l-th moments equal, then the time series is called l-th order (weakly) stationary. Thus, for second order weakly stationary stochastic process $(X_t)_{t\in Z}$ by definition

$$E(X_j) = E(X_{j+q}),$$

$$E({X^2}_j) = E({X^2}_{j+q}),$$

$$E(X_j X_i) = E(X_{j+q} X_{i+q}),$$

where $1 \leq j \leq i \leq m, q \in Z$. This means that mean and variation are constant. In this thesis, mainly the second order weakly stationary processes are considered. In addition, it is reckoned that a time series is stationary if the underlying stochastic process is stationary. Stationarity is essential because it gives more accurate and stable parameter estimates compared to a series whose statistical properties keep changing over time.

From stationarity definition it can be concluded that a time series with changing variance or mean is not stationary. Transformations like logarithms are useful for stabilising the variance of a time series. Trend can be eliminated by computing the differences between consecutive observations, i.e. differencing the series,

$$\nabla x_t = x_t - x_{t-1} = (1 - B)x_t,$$

where B is backward shift operator, $Bx_t = x_{t-1}$. Similarly, $B(Bx_t) = B^2 x_t = x_{t-2}$.

For non-seasonal data, first-order differencing is usually enough to attain apparent stationarity. Occasionally the differenced data does not look stationary, and it may be necessary to difference the data a second time to obtain a stationary series:

$$\nabla^2 x_t = \nabla x_t - \nabla x_{t-1} = x_t - 2x_{t-1} + x_{t-2} = (1 - 2B + B^2)x_t = (1 - B)^2 x_t.$$

Seasonal component can be removed by using seasonal difference. For this, the difference between observation and the previous observation from the same season is calculated

$$\nabla_s x_t = x_t - x_{t-s} = (1 - B^s)x_t,$$

where s is the number of seasons. Seasonal difference followed by first-order difference is calculated as

$$\nabla \nabla_s x_t = (1 - B)(1 - B^s)x_t = x_t - x_{t-1} - x_{t-s} + x_{t-s-1}.$$

There are number of so-called unit root tests available for determining whether differencing is required to attain stationarity, for example, Phillips-Perron test (Phillips and Perron, 1988) and augmented Dickey-Fuller (ADF) test. Both test the null hypothesis that a unit root is present (non-stationarity).

2.1.3 Autocorrelation

The third element of interest in preliminary analysis is autocorrelation. It measures the linear relationship between observations at different distances apart. The autocorrelation function (ACF) of a second order weakly process $(X_t)_{t \in Z}$ at lag k is

$$\rho_X(k) = cor(X_t, X_{t+k}) = \frac{\gamma(k)}{\gamma(0)} = \frac{\gamma(k)}{\sigma^2}, k \in Z,$$

where $\gamma(k) = cov(X_t, X_{t+k}) = E[(X_t - E(X_t))(X_{t+k} - E(X_{t+k}))], k \in Z,$ is the auto-covariance function of the process (X_t).

The theoretical autocorrelation function is estimated by the sample autocorrelation function which is calculated as

$$r_k = \hat{\rho}_x(k) = cor(x_t, x_{t+k}) = \frac{\sum_{t=1}^{T-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{T}(x_t - \bar{x})^2},$$

where k is the lag and T is the length of the time series. Thus, r_1 measures the relation between x_t and x_{t+1}, r_2 measures the relation between x_t and x_{t+2}, and so forth.

Another useful tool for analysing relation between observations at different distances apart is partial autocorrelation function (PACF). It measures the relationship between X_t and X_{t+k} after removing the effect of lags $1, \ldots, k-1$, and is noted as $\pi_X(k)$. For instance, the sample partial autocorrelation of order two $\hat{\pi}_x(2)$ measures the excess correlation between x_t and x_{t+2} not accounted by r_1.

A discrete process that consists of mutually independent and identically distributed (iid) random variables is called white noise. Further on, it is assumed that the random variables are normally distributed with zero mean and variance δ^2, indicated by the notation $(X_t) \sim WN(0, \delta^2)$. Hence, with large T, approximately 95% of the sample autocorrelations r_k and sample partial autocorrelations $\hat{\pi}_k$ are expected to fall between $\pm 1.96/\sqrt{T}$ (Brockwell and Davis, 2002 pp. 16-20). In addition, white noise is a second order weakly stationary process, meaning that a time series generated from uncorrelated random variables is second order weakly stationary.

2.2 Models

In the following subsections different time series models are described. As deposits have historically proven to grow in time, then the focus is on methods that allow trend.

2.2.1 Holt and Holt-Winters

One of the popular methods for forecasting of time series with a trend is Holt's linear trend method. Exponential smoothing assigns more weight to recent observations and less weight to observations further in the past. The forecasting equation of Holt's method is

$$\hat{x}_{t+h|t} = l_t + b_t h,$$

where l_t denotes an estimate of the level of the series at time t

$$l_t = \alpha x_t + (1-\alpha)\hat{x}_t = \alpha x_t + (1-\alpha)(l_{t-1} + b_{t-1}),$$

and b_t denotes an estimate of the trend of the series at time t

$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1}.$$

α is the smoothing parameter for the level, $0 < \alpha < 1$, and β is the smoothing parameter for the trend, $0 < \beta < 1$.

Another similar and widely used method is Holt-Winters' seasonal method. It is an extension of Holt's method to capture seasonality. Depending on the nature of the seasonal component, either multiplicative or additive forecasting equation is used. The formula for the multiplicative method is

$$\hat{x}_{t+h|t} = (l_t + b_t h)S_{t+h-s},$$

where $h = 1, \dots, s$, and

$$l_t = \alpha \frac{x_t}{S_{t-s}} + (1-\alpha)(l_{t-1} + b_{t-1}),$$

$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1},$$

$$S_t = \gamma \frac{x_t}{l_t} + (1-\gamma)S_{t-s}.$$

The formula for the additive method is

$$\hat{x}_{t+h|t} = (l_t + b_t h) + S_{t+h-s},$$

where $h = 1, \dots, s$, and

$$l_t = \alpha(x_t - S_{t-s}) + (1-\alpha)(l_{t-1} + b_{t-1}),$$

$$b_t = \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1},$$

$$S_t = \gamma(x_t - l_t) + (1-\gamma)S_{t-s}.$$

The estimation of model parameters can be done minimising a forecast error measure. In R software, the sum of square errors is used:

$$\sum_{t=1}^{T} (\hat{x}_t - x_t)^2.$$

2.2.2 ARIMA model

Another widely used class of time series models is ARIMA. While exponential smoothing models are built based on the trend and seasonality in the data, the objective of ARIMA models is to describe the autocorrelation. ARIMA(p, d, q) is an acronym for autoregressive integrated moving-average and is formed by combining autoregressive process of order p, AR(p), d degree of differencing and moving-average process of order q, MA(q).

Autoregressive process

Let $E(X_t) = \mu$ be constant. A second order weakly process $(X_t)_{t \in Z}$ is called an autoregressive process of order p and denoted as AR(p) if

$$\tilde{X}_t = Z_t + \alpha_1 \tilde{X}_{t-1} + \cdots + \alpha_p \tilde{X}_{t-p},$$

where $\tilde{X}_t = X_t - \mu$, $Z_t \sim \mathrm{WN}(0, \sigma^2)$ is white noise and α_i, $i = 1, \ldots, p$, are parameters. The model specifies variable's dependency on previous p values. In this thesis only such AR(p) processes are considered for which $cov(X_t, Z_i) = 0$ for all $i > t$. These processes are called causal (Kangro, 2016). Using backward shift operator B, a compact form of the AR(p) process can be written as

$$\varphi(B)\tilde{X}_t = Z_t$$

where $\varphi(x)$ is polynomial of order p such that $\varphi(x) = 1 - \sum_{i=1}^{p} \alpha_i x^i$.

AR(p) process is second order weakly stationary if and only if the modules of the roots of the polynomial $\varphi(x)$ lie outside the unit circle, $|x_i| > 1, i = 1, \ldots, p$.

Moving-average process

A process $(X_t)_{t \in Z}$ is called a moving-average process of order q and denoted as MA(q) if

$$\tilde{X}_t = Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q},$$

where $\tilde{X}_t = X_t - \mu$, $Z_t \sim \mathrm{WN}(0, \sigma^2)$ and β_i, $i = 1, \ldots, q$, are parameters. The model describes regression error's dependency on previous q forecast errors.

A compact form of the MA(q) process can be written as

$$\tilde{X}_t = \theta(B)Z_t,$$

where $\theta(x)$ is polynomial of order q such that $\theta(x) = 1 + \sum_{i=1}^{q} \beta_i x^i$.

MA(q) process is invertible if it can be rewritten in the form of an autoregressive process, possibly of infinite order, whose coefficients form a convergent sum. The same applies if the modules of the roots of the polynomial $\theta(x)$ all lie outside the unit circle, $|x_i| > 1, i = 1, \dots, p$.

ARMA model

Combination of AR(p) and MA(q) processes is called ARMA(p, q) process and expressed as

$$\tilde{X}_t = \alpha_1 \tilde{X}_{t-1} + \cdots + \alpha_p \tilde{X}_{t-p} + Z_t + \beta_1 Z_{t-1} + \cdots + \beta_q Z_{t-q},$$

where $\tilde{X}_t = X_t - \mu$, $(X_t)_{t \in Z}$ is second order weakly stationary and $Z_t \sim \mathrm{WN}(0, \sigma^2)$. A compact form of the formula can be written as

$$\varphi(B)\tilde{X}_t = \theta(B)Z_t,$$

where φ and θ are respectively polynomials of order p and q. ARMA(p, q) model can be presented as

- moving-average process $\tilde{X}_t = \frac{\theta(B)}{\varphi(B)} Z_t = \psi(B)Z_t$, if the modules of the roots of the polynomial $\varphi(x)$ all lie outside the unit circle, $|x_i| > 1, i = 1, \dots, p$,
- autoregressive process $\frac{\varphi(B)}{\theta(B)} \tilde{X}_t = \pi(B)\tilde{X}_t = Z_t$, if the modules of the roots of the polynomial $\theta(x)$ all lie outside the unit circle, $|x_i| > 1, i = 1, \dots, p$.

In addition, the sample values of PACF of an ARMA(p, q) model are not significantly different from zero after lag p and the sample values of ACF cut off after lag q (see Table 2).

Table 2: Properties of ACF and PACF

Model	**ACF**	**PACF**
AR(p)	Decay towards zero	Decay to zero after lag p
MA(q)	Decay to zero after lag q	Decay towards zero
ARMA(p, q)	Decay to zero after lag q	Decay to zero after lag p

ARIMA model

Previous models assumed stationarity. In practice most time series are non-stationary. ARIMA(p, d, q) model can be used for forecasting non-stationary time series. This is a generalization of ARMA(p, q) processes to incorporate differenced time series. Let d be a non-negative integer and $W_t = (1 - B)^d X_t$, then $(X_t)_{t \in Z}$ is an ARIMA(p, d, q) process if W_t is a causal ARMA(p, q) process:

$$\varphi(B) W_t = \varphi(B)(1 - B)^d X_t = \theta(B) Z_t,$$

where $Z_t \sim \mathrm{WN}(0, \sigma^2)$, φ and θ are polynomials of order p and q.

Let d and D be non-negative integers and $Y_t = (1 - B)^d (1 - B^s)^D X_t$. A process $(X_t)_{t \in Z}$ is called a seasonal ARIMA$(p, d, q) \times (P, D, Q)$ process with period s if Y_t is a casual ARMA (p, q) process:

$$\varphi_p(B) \Phi_P(B^s) Y_t = \varphi_p(B) \Phi_P(B^s)(1 - B)^d (1 - B^s)^D X_t = \theta_q(B) \Theta_Q(B^s) Z_t,$$

where $Z_t \sim \mathrm{WN}(0, \sigma^2)$ and φ_p, Φ_P, θ_q, Θ_Q are polynomials of order p, P, q, Q.

Once the values p, d and q have been identified, the estimation of model parameters can be done. The following description of methods is based on course material by Kangro (2016). A common technique is to use the maximum likelihood estimation (MLE), which finds the values of the parameters so that the probability of obtaining the data observed is maximised. Let $w = (W_1, \dots, W_T)$ be a multivariate normal random variable. When maximizing the log-likelihood function by δ^2, the estimate of variance is $\hat{\delta}^2 = \frac{1}{T} w \Omega^{-1}(\alpha, \beta) w^T$, where $\alpha =$

$(\alpha_1, \dots, \alpha_p)$, $\beta = (\beta_1, \dots, \beta_q)$ and Ω is the correlation matrix. The parameters $\alpha_1, \dots, \alpha_p \beta_1, \dots, \beta_q$ can be found by maximizing the expression

$$-\frac{T}{2}\ln(w\Omega^{-1}(\alpha,\beta)w^T) - \frac{1}{2}\ln\left(|\Omega(\alpha,\beta)|\right)$$

by $\alpha = (\alpha_1, \dots, \alpha_p)$ and $\beta = (\beta_1, \dots, \beta_q)$, where $|\Omega(\alpha,\beta)|$ is the determinant of the correlation matrix.

Another technique for estimating the model parameters is conditional least squares method. Let z_t denote the residuals from ARIMA model, $t = 1, \dots, T$. Conditional least squares method finds the parameters $\alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q$ by minimising the sum of squares of residuals

$$\sum_{t=1}^{T} {z_t}^2,$$

where

$$z_t = w_t - \sum_{i=1}^{p} \alpha_i w_{t-i} - \sum_{j=1}^{q} \beta_j z_{t-j}.$$

The default technique in R software is using conditional least square method to find starting values, and then applying maximum likelihood method.

2.2.3 ARIMAX model

This paragraph is based on the course material by Kangro (2016).

The time series models introduced in the previous sections are only based on past observation of the series and are not considering other variables. In this subchapter, an extension of ARIMA model is introduced to include explanatory variable in the model.

In the following let's assume that time series X_t and V_t are stationary with zero mean or differenced to achieve stationarity. If X_t is dependent on V_t, then ARIMAX model can be defined as linear regression with ARIMA errors:

$$X_t = \gamma V_t + \eta_t,$$

where η_t follows an ARIMA process. In comparison of standard linear regression, the errors η_t are not presumed to be iid. The ARIMAX model with m regressors $(V_{1,t}, V_{2,t}, \dots, V_{m,t})$ is defined as

$$X_t = \sum_{i=1}^{m} \gamma_i V_{i,t} + \eta_t,$$

where η_t follows an ARIMA process.

There are three main steps in estimation of ARIMAX model parameters:

1) estimate the model $X_t = \gamma V_t + \eta_t$,
2) examine the sample residuals from the model and find suitable ARIMA model for η_t,
3) estimate the parameters $\delta^2, \gamma, \alpha_1, \dots, \alpha_p, \beta_1, \dots, \beta_q$ using, for example, maximum likelihood method.

2.3 Goodness of fit

ACF plot, or alternatively correlogram, is the first indicator of goodness of fit of a statistical model. If not more than two or three of the first 30-40 sample autocorrelations of the residuals fall outside $\pm 1.96/\sqrt{T}$, then the fitted model is suitable (Brockwell and David, 2002: p. 36).

In addition to treating each r_k separately while observing ACF plot, there are statistical tests called portmanteau tests that consider a whole set of r_k values as a group. One such test is the Box-Pierce test, which is based on the statistic

$$Q = T \sum_{k=1}^{K} {r_k}^2,$$

where T is the number of observations and K is the number of chosen lags in scope (chosen somewhat arbitrarily, typically in the range 15 to 30). Another similar and often used test is the Ljung-Box test, based on

$$Q_* = T(T+2) \sum_{k=1}^{K} \frac{{r_k}^2}{T-k}.$$

Both statistics are from χ^2 distribution with $(K - p)$ degrees of freedom, where p is the number of parameters in the model. If the test is applied to the autocorrelations from the original series, not from the model, the then degrees of freedom is just K. The null hypothesis of a portmanteau test is that the autocorrelations for the chosen lags in the population from which the sample is taken are all zero. If the p-value of the test is very small, then there is strong evidence that the autocorrelations are not from an iid sequence.

Another way of examining goodness-of-fit is comparing AIC, AICc or BIC of competing models:

- $AIC = 2p - 2\ln(\hat{L})$,
- $AICc = AIC + \frac{2p^2+2p}{T-p-1}$,
- $BIC = p\ln(T) - 2\ln(\hat{L})$,

where p is the number of parameters in the model, $\hat{L}$ is the maximum value of the likelihood function of the model and T is the sample size. The preferred model is the one with the lowest AIC, AICc or BIC value. (Brockwell and David, 2002: pp. 173-174)

2.4 Forecasting

When a suitable model has been found, the estimation of future values can be done. A wide variety of different forecasting procedures is available. It is important to acknowledge that each of them is based on specific assumptions and no single method is universally applicable. Subjective judgement can be combined with statistical approach when, for example, choosing an appropriate model and adjusting the resulting forecasts, especially in case the forecasting horizon is long-term.

In case of univariate prediction methods, the model forecasts are based on the past observations and the fitted residuals. Assuming that the ARIMA model equation is known, then the point forecasts can be calculated by

1) expanding the equation so that X_t is on the left side and all other inputs are on the right,
2) replacing t with $T + h$ and random variables with the observations,

3) replacing past values of the series with the observed values, past errors with the corresponding residuals, future values of the series with their forecasts and future errors with zero.

Let's consider a time series $x_t, t = 1, \dots, T$, which corresponds to the $\text{ARIMA}(1,0,0)(0,1,1)_{12}$ model

$$\varphi_1(B)(1 - B^{12})X_t = \Theta_1(B^{12})Z_t.$$

This can be presented in a form

$$X_t = \alpha_1(X_{t-1} - X_{t-13}) + X_{t-12} + Z_t + \beta_1 Z_{t-12}.$$

The goal is to find the forecasts $\hat{x}_{T+h|T}$ and their confidence intervals. When following the steps 2) and 3), the forecasts are

$$\hat{x}_{T+1|T} = \hat{\alpha}_1(x_T - x_{T-12}) + x_{T-11} + \hat{\beta}_1 Z_{T-11}$$

$$\hat{x}_{T+2|T} = \hat{\alpha}_1(\hat{x}_{T+1|T} - x_{T-11}) + x_{T-10} + \hat{\beta}_1 Z_{T-10}$$

$$\dots$$

$$\hat{x}_{T+h|T} = \hat{\alpha}_1(\hat{x}_{T+h-1|T} - x_{T+h-13}) + x_{T+h-12} + \hat{\beta}_1 Z_{T+h-12}$$

The 95% prediction interval is given by

$$\hat{x}_{T+h|T} \pm 1.96\sqrt{{\sigma_h}^2},$$

where ${\sigma_h}^2$ is the variance of the forecasting error made at time T when forecasting h steps ahead

$${\sigma_h}^2 = var(e_{T+h}) = var(X_{T+h} - \hat{x}_{T+h|T}) = \sigma^2\left(1 + \sum_{j=1}^{h-1} {\psi^2}_j\right).$$

In case of ARIMAX model, the estimation of future values is done by combining the forecasts of the regression part of the model and ARIMA part of the model. For the regression model, the forecasts of the predictor variable need to be obtained first.

2.5 Application in R

This subsection gives an overview of different functions in R useful for time series analysis and is based on the online textbook by Hyndman and Athanasopoulos (2018).

The function *ts()* is used to produce a time series object. In order to compute and plot the estimates of autocorrelation and partial autocorrelation, functions *acf()* and *pacf()* can be applied to the series. Through `lag.max` one can specify the maximum number of lags at which to calculate the ACF and PACF. Different tests like *pp.test()* and *adf.test*() are available in R to check the stationarity of a time series.

Arima() function can be used to fit an ARIMA model to a time series. Through `xreg` argument it allows to specify independent variables. The regressors must have the same number of rows as dependent variable `x`. In case the series are non-stationary, it is possible to apply differencing within the *Arima()* function. For example, the R command

```
Arima(X, xreg=V, order=c(1,1,0))
```

will fit the model $\nabla X_t = \gamma \nabla V_t + \nabla \eta_t$, where $\nabla \eta_t$ is an ARIMA$(1,0,0)$ error and $Z_t \sim WN(0, \delta^2)$. This is equivalent to fitting the model $X_t = \gamma V_t + \eta_t$, where η_t is an ARIMA$(1,1,0)$ error and $Z_t \sim WN(0, \delta^2)$.

The *auto.arima()* function can be used to automatically find a suitable ARIMA model. The function uses a combination of unit root tests and minimization of the AIC, AICc or BIC to find a suitable ARIMA model. Unit root tests are applied for determining the number of differences, while goodness of fit statistics are used for choosing the values of p and q. In this thesis, *auto.arima()* is used for comparison purposes. The main analysis for finding suitable ARIMA models is based on manual analysis and usage of *Arima()* function.

Methods for checking residuals are conveniently integrated into one function *checkresiduals()*, which will produce a time plot, ACF plot and histogram of the residuals, and do a Ljung-Box test with the correct degrees of freedom.

Once a suitable time series model is found, *forecast()* function can be used for generating forecasts and their confidence intervals.

3 Empirical Study

In this section different time series analysis methods were applied on the financial institution's savings deposits volume. The models with exogenous variables included Stibor 1-month, consumer confidence indicator (CCI), purchasing managers' index (PMI), gross domestic product (GDP), housing price index (HPI) and quantitative easing (QE).

In paragraph 3.1 the data under analysis is described. In the next two paragraphs, different time series models are fitted and validated. In paragraph 3.4, the summary of the results is presented.

3.1 Data

The financial institution offers a wide range of deposit products based on different characteristics like maturity and interest. Overview of the characteristics is brought out in Table 3.

Table 3: Characteristics of the deposits offered by the financial institution

Deposit type	**Characteristics**
Transaction account	0% interest rate, no maturity, option to add and withdraw money without a fee, physical card linked to the account
Savings deposit	Administratively set interest rate, no maturity, option to add and withdraw money without a fee, no physical card linked to the account
Term deposit	Fixed interest rate, defined maturity, no option to add or withdraw money without a fee, no physical card linked to the account

Different deposit types and customer segments in the financial institution are assumed to behave differently, and therefore, separate deposit volume models are needed to be implemented. This thesis is focusing on savings deposits from private customers.

To evaluate forecasting accuracy, the datasets were divided into training and test set. The training set covers period from 2011 until 2020 and the test set period of 2021.

Savings deposits' volume

As a first step, the volumes of savings deposits were analysed. The data was retrieved from the financial institution's data warehouse and consisted of end of month volumes from January 2011 until December 2021. For business purposes, the volumes were indexed. From Figure 4 it can be seen that there is a positive trend, no cyclicity and no apparent seasonality.

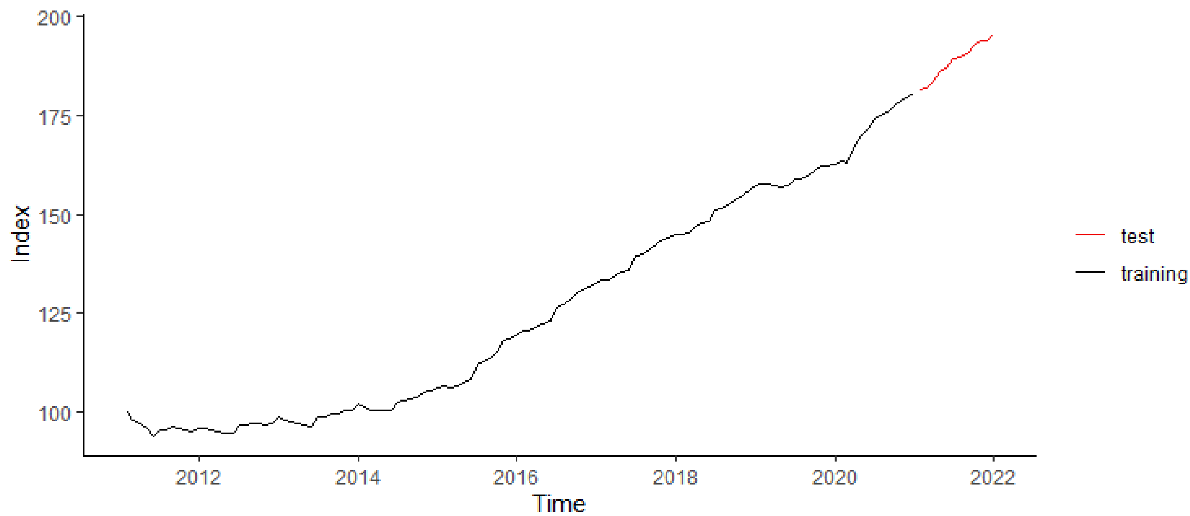

Figure 4: Savings deposits' volumes from January 2011 until December 2021. January 2011 = 100.

Explanatory variables

In the financial institution, the deposit rate for savings deposits is mainly administratively set and linked to market rate. Therefore, the first exogenous variable included in the analysis was Stibor 1-month which was used as proxy for deposit rate. The data was retrieved from Sveriges Riksbank (2022b) and Swedish Financial Benchmark Facility (2022) websites and consisted of monthly average values from January 2011 until December 2021 (see Figure 5). One can see that there is a negative trend during 2012 and 2016, but no apparent seasonality.

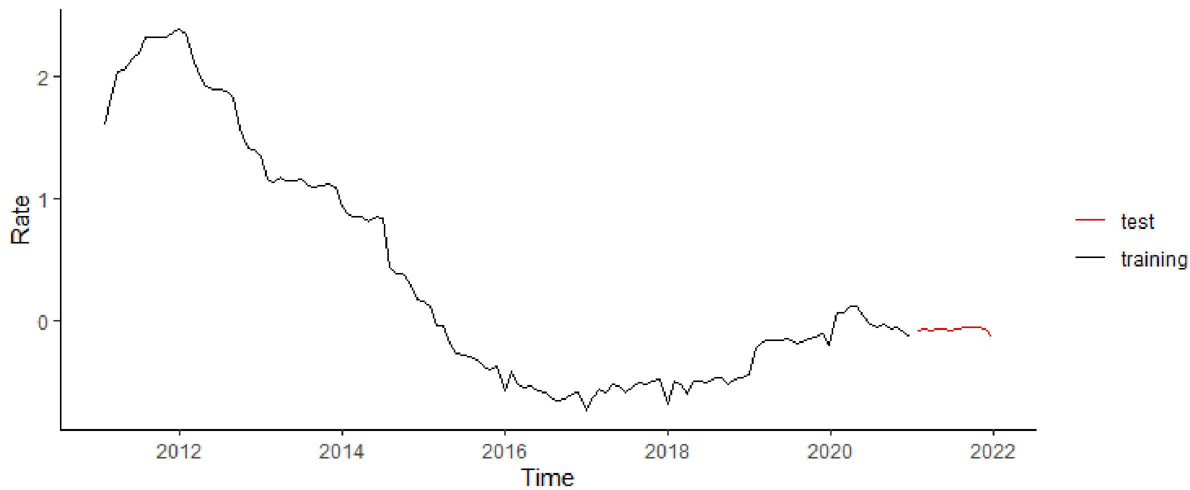

Figure 5: Stibor 1-month values from January 2011 until December 2021

The second explanatory variable was Sweden's consumer confidence indicator. The data was taken from Eurostat database (2022b) and included monthly values from January 2011 until December 2021. By looking at Figure 6, it can be seen that except from the period in the beginning of 2020, the series is stationary – no overall trend, cyclicity nor seasonality.

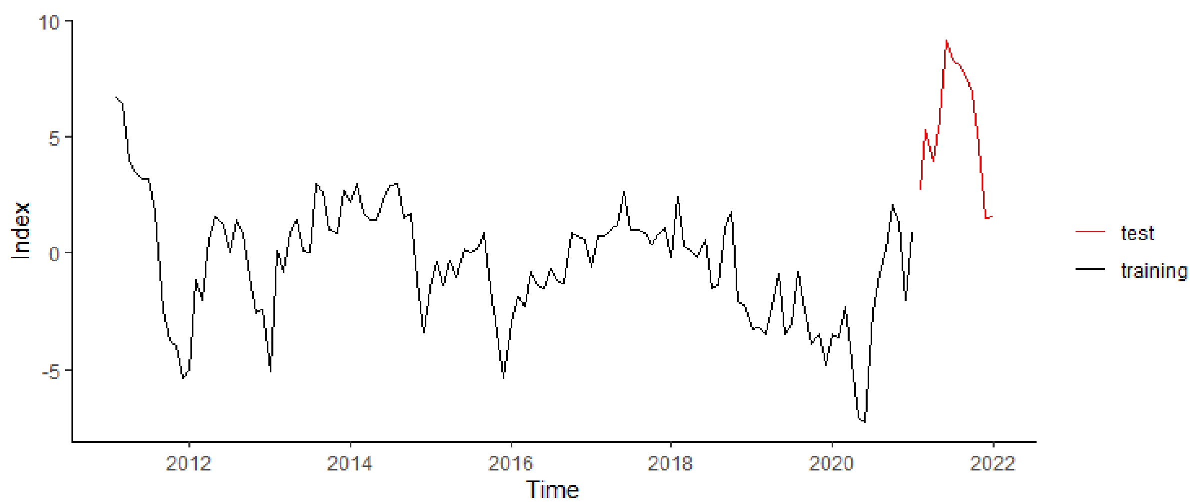

Figure 6: Sweden's consumer confidence indicator from January 2011 until December 2021

The next independent variable was Sweden's purchasing managers' index. The observations were retrieved from Swedbank database (2022) and contained monthly values of PMI manufacturing and PMI services from January 2011 until December 2021 (see Figure 7). Except from the period in the beginning of 2020, the series are stationary.

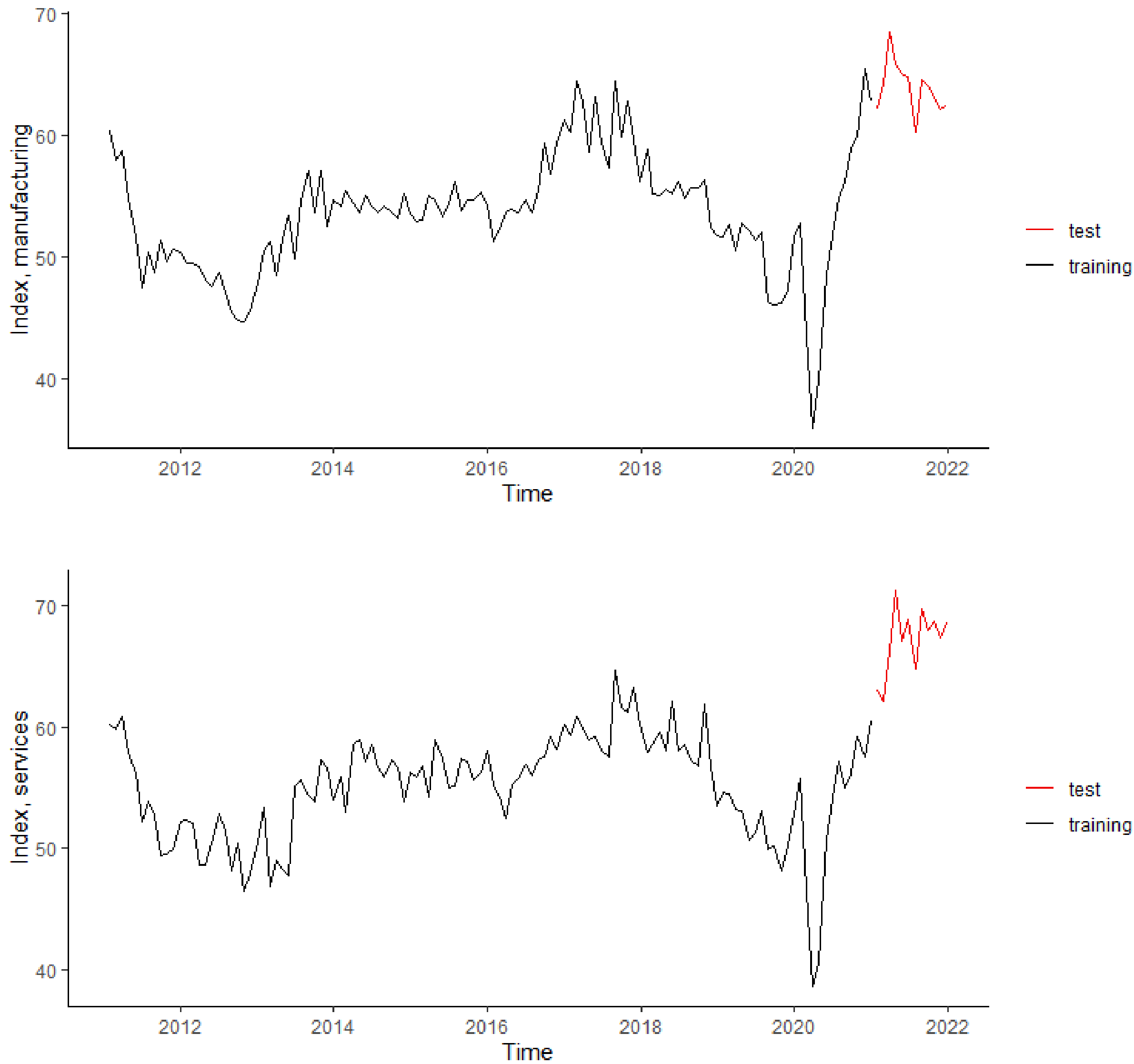

Figure 7: Sweden's purchasing managers' indexes from January 2011 until December 2021

The fourth independent variable in interest was Sweden's gross domestic product in constant prices. The observations were retrieved from the database of Statistics Sweden (2022) and consisted of quarterly values from Q1 2011 until Q4 2021. The data was transformed to monthly by multiplying the quarterly value by share of the month. For example, if the number of days in Q1 2011 was 90, then GDPs for January, February and March was calculated as

$$GDP_{Jan\ 2011} = GDP_{Q1\ 2011} * \frac{31}{90},$$

$$GDP_{Feb\ 2011} = GDP_{Q1\ 2011} * \frac{28}{90},$$

$$GDP_{Mar\ 2011} = GDP_{Q1\ 2011} * \frac{31}{90}.$$

Figure 8 shows that there is a positive trend and an evident seasonality.

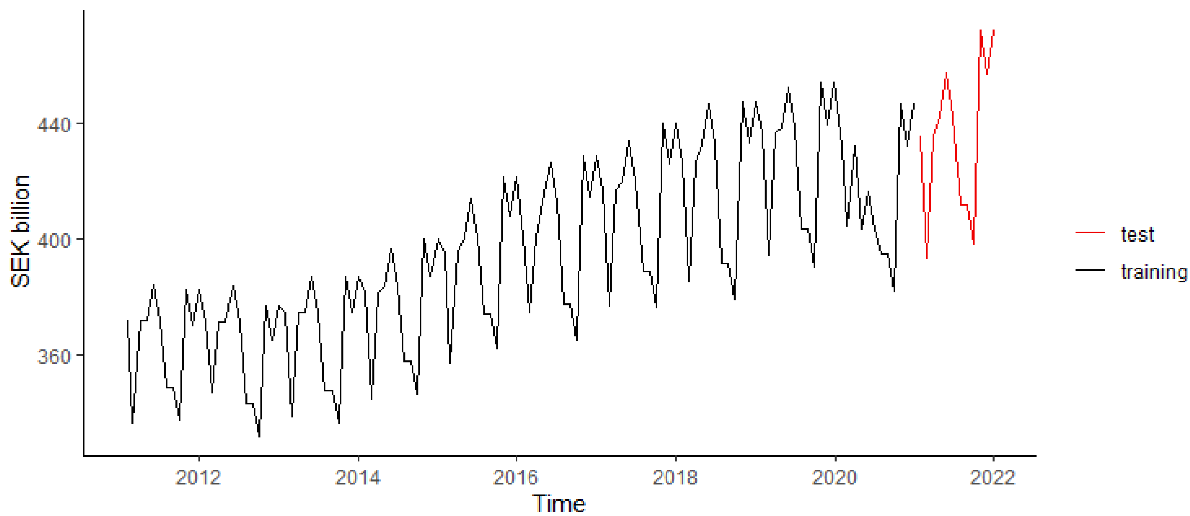

Figure 8: Sweden's gross domestic product from January 2011 until December 2021

The next explanatory variable included in the thesis was Sweden's house price index. The data was retrieved from Eurostat (2022a) database and included quarterly values from Q1 2011 until Q4 2021. The quarterly data was interpolated to monthly. From Figure 9 one can see that there is an overall positive trend with no seasonal component.

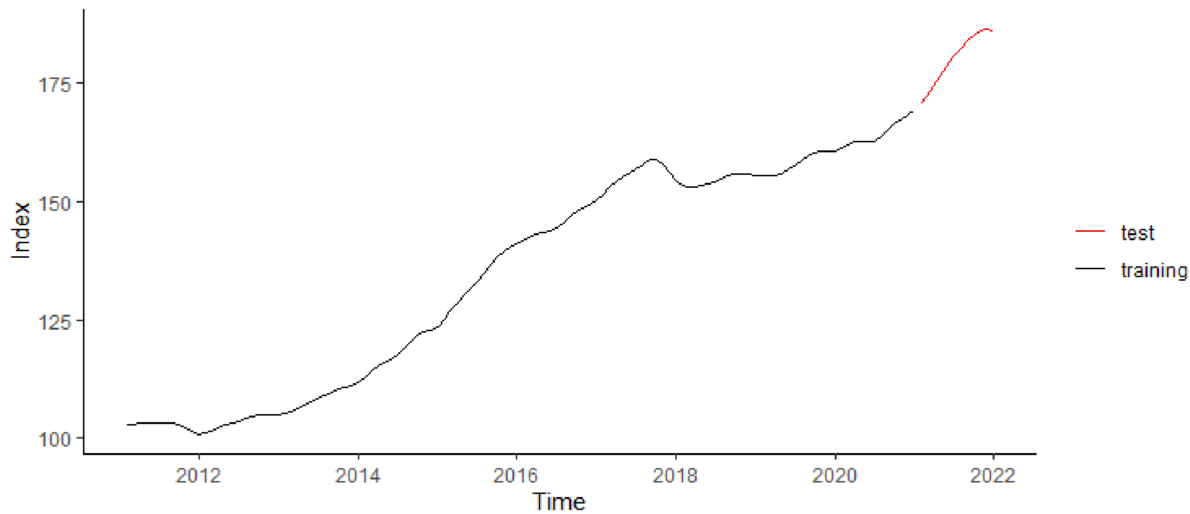

Figure 9: Sweden's house price index from January 2011 until December 2021. Year 2010 = 100.

Last hypothesis was that quantitative easing has impact on savings volumes. Sveriges Riksbank had been purchasing large-scale asset from the end of 2014. The volumes of the holdings of government bonds were retrieved from Sveriges Riksbank's (2022a) website and consisted of quarterly values from Q4 2014 until Q4 2021. The quarterly data was transformed to monthly similarly to GDP - multiplying the quarterly values by share of the month. Figure 10 shows that during the years 2015 and 2017 quantitative easing increased rapidly and after that slowed down.

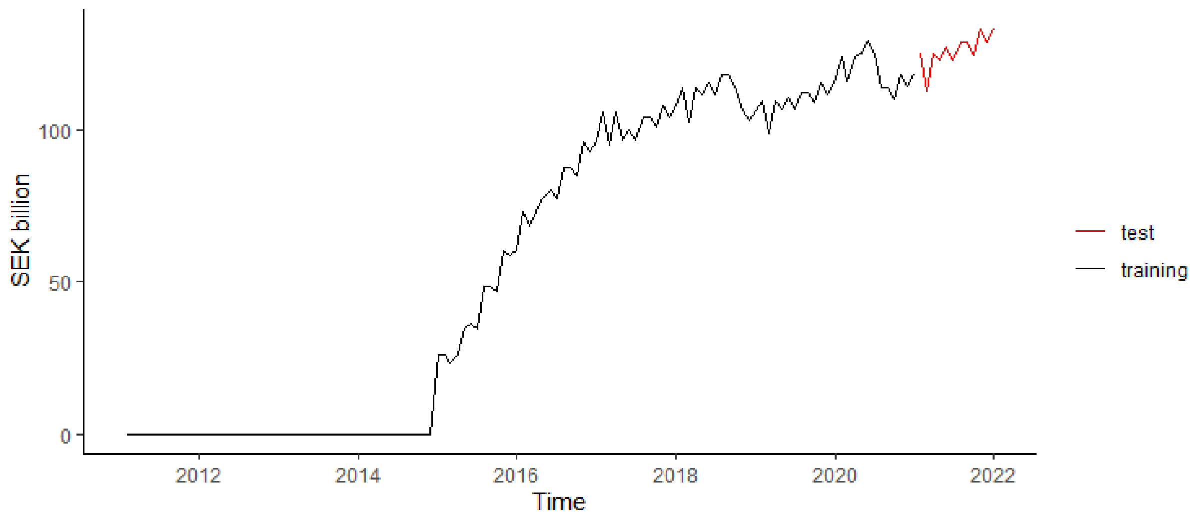

Figure 10: The Riksbank's total holdings of government bonds from January 2011 until December 2021

3.2 Modelling

In this section time series models are tried out and compared, with the purpose of finding the most appropriate model for forecasting savings deposits. The input information, e.g. ACF and PACF plots, for selecting the models are pointed out in appendices. The best models are chosen based on the Ljung-Box test and goodness of fit statistics (AIC, AICc, BIC).

3.2.1 Holt and Holt-Winters

The first methods investigated in the thesis were Holt's linear trend method and Holt-Winters' seasonal method, which are simple time series techniques appropriate for data with a trend. From Figure 4 it was seen that savings deposits have a positive trend, but no evident seasonality. Therefore, Holt's linear trend method was applied. The forecasting equation is

$$\hat{x}_{t+h|t} = l_t + b_t h,$$

where

$$l_t = 0.9947x_t + 0.0053(l_{t-1} + b_{t-1}),$$

$$b_t = 0.1048(l_t - l_{t-1}) + 0.8952b_{t-1}.$$

The AIC, AICc and BIC values were accordingly 569.7120, 570.2383 and 583.6494.

3.2.2 ARIMA model

Figure 4 showed that the time series of savings volumes is not stationary and differencing is needed. After first order differencing, the series had evident seasonal component. Due to that, an additional seasonal difference was taken after which the series is stationary. The results are shown in Figure 11.

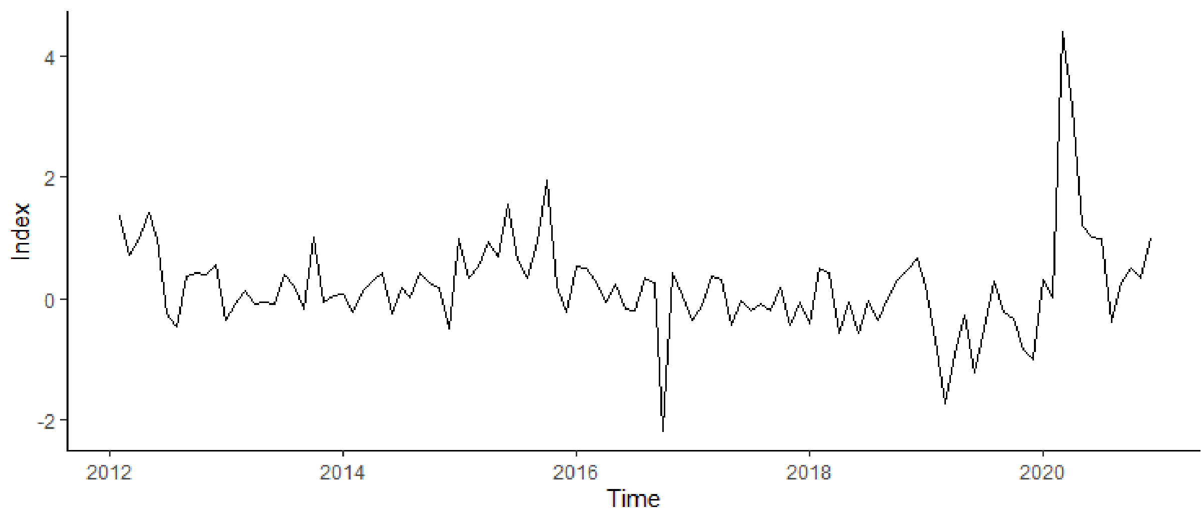

Figure 11: First order and seasonally differenced savings deposits

The next step was to find appropriate ARIMA models based on the sample ACF and PACF. The significant spikes at lag 1 and 2 in the ACF suggested $\text{ARIMA}(0,1,2)(0,1,0)_{12}$. Although, the p-value of Ljung-Box test was over 0.05, the 12th lag of ACF of the fitted model was outside the confidence interval, indicating the usefulness of a seasonal term. Therefore, $\text{ARIMA}(0,1,2)(0,1,1)_{12}$ was tried out as well.

The next group of models in interest were ARIMA models with AR component. The PACF showed significant spikes at lag 1 and 12 and therefore both $\text{ARIMA}(1,1,0)(0,1,0)_{12}$ and $\text{ARIMA}(1,1,0)(1,1,0)_{12}$ were fitted.

Finally, the models with combination of MA and AR components were tried out. The comparison of the models based on Ljung-Box test and goodness of fit statistics is summarized in Table 12.

Table 12: Comparison of fitted models for savings deposits

Model	p-value of L-B test	AIC	AICc	BIC
$\text{ARIMA}(0,1,2)(0,1,0)_{12}$	0.0511	240.38	240.61	248.4
$\text{ARIMA}(0,1,2)(0,1,1)_{12}$	0.4384	227.81	228.20	238.5
$\text{ARIMA}(1,1,0)(0,1,0)_{12}$	0.1535	235.34	235.45	240.68
$\text{ARIMA}(1,1,0)(1,1,0)_{12}$	0.2391	223.02	223.25	231.04
$\text{ARIMA}(1,1,0)(0,1,1)_{12}$	0.5157	221.46	221.69	229.48
$\text{ARIMA}(1,1,1)(0,1,1)_{12}$	0.5877	221.84	222.23	232.53
$\text{ARIMA}(1,1,2)(0,1,1)_{12}$	0.4913	221.29	221.88	234.65
$\text{ARIMA}(0,1,1)(1,1,0)_{12}$	0.1486	228.94	229.18	236.96
$\text{ARIMA}(1,1,1)(1,1,0)_{12}$	0.3308	223.29	223.69	233.99
$\text{ARIMA}(1,1,1)(1,1,1)_{12}$	0.4772	223.51	224.1	236.87
$\text{ARIMA}(1,1,2)(1,1,1)_{12}$	0.3868	222.97	223.81	239.01

The best model with lowest goodness of fit statistics and high p-value of Ljung-Box test was $\text{ARIMA}(1,1,0)(0,1,1)_{12}$:

$$
\begin{aligned}
X_t &= (1+\alpha_1)(X_{t-1}-X_{t-13}) - \alpha_1(X_{t-2}-X_{t-14}) + X_{t-12} + Z_t + {\beta_1}^{*} Z_{t-12} \\
&= 1.5227(X_{t-1}-X_{t-13}) - 0.5227(X_{t-2}-X_{t-14}) + X_{t-12} + Z_t \\
&\quad - 0.5023 Z_{t-12},
\end{aligned}
$$

where X_t is savings deposits' volume at time t and $Z_t \sim \text{WN}(0, 0.6576^2)$.

3.2.3 ARIMAX model

Stibor 1-month

In order to fit ARIMAX model to savings deposits, the first step was to estimate linear regression. In Figure 5 it was seen that Stibor 1-month is not stationary. According to the unit root tests and review of ACF and PACF plot, the series was stationary after first difference (see Figure 13).

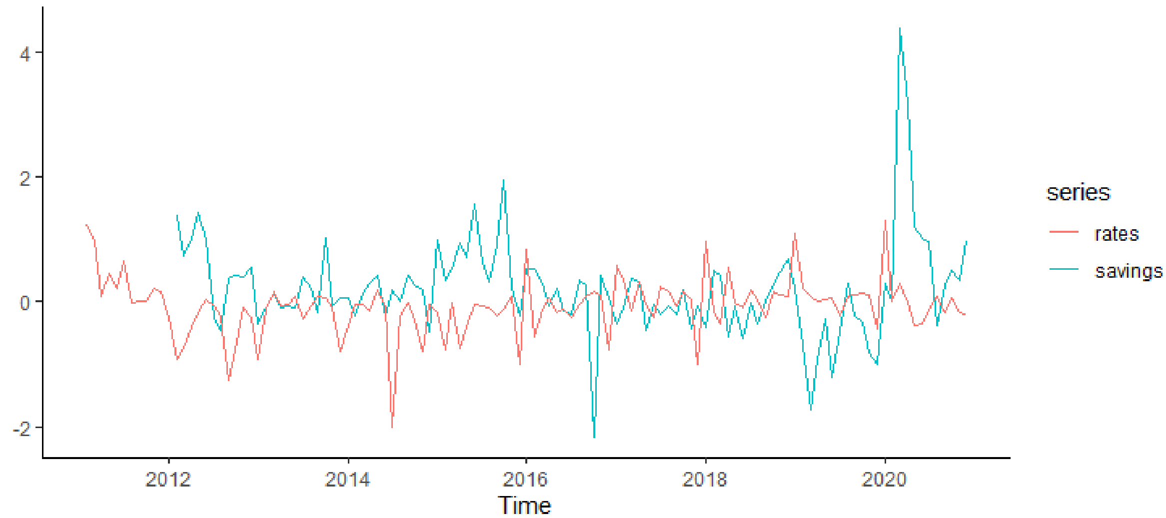

Figure 13: First order and seasonally differenced savings deposits compared to differenced values of Stibor 1-month

The initial estimated linear regression model with stationary variables was

$$\nabla\nabla_{12}X_t = -1.3339\nabla V_t + \eta_t,$$

where $\nabla\nabla_{12}X_t$ is first order and seasonally differenced savings deposits' volume at time t, ∇V_t is first order differenced Stibor 1-month value at time t and η_t is model residual at time t. The residuals from linear regression model were stationary and based on the sample ACF and PACF, several ARIMA models were fitted. The results are summarized in Table 14.

Table 14: Comparison of fitted models for residuals from linear regression model with Stibor 1-month

Model	**p-value of L-B test**	**AIC**	**AICc**	**BIC**
$ARIMA(0,0,2)(0,0,0)_{12}$	0.1408	242.09	242.49	252.79
$ARIMA(0,0,2)(0,0,1)_{12}$	0.4414	229.61	230.2	242.97
$ARIMA(1,0,0)(0,0,0)_{12}$	0.1722	237.25	237.48	245.27
$ARIMA(1,0,0)(1,0,0)_{12}$	0.2921	224.96	225.35	235.65
$ARIMA(1,0,0)(0,0,1)_{12}$	0.5117	223.45	223.85	234.15
$ARIMA(1,0,1)(0,0,1)_{12}$	0.5297	223.77	224.36	237.13

$\text{ARIMA}(1,0,2)(0,0,1)_{12}$	0.5284	222.24	223.08	238.27
$\text{ARIMA}(0,0,1)(1,0,0)_{12}$	0.1703	230.76	231.15	241.45
$\text{ARIMA}(1,0,1)(1,0,0)_{12}$	0.3467	225.27	225.86	238.63

One of the suitable models was linear regression with $\text{ARIMA}(1,0,2)(0,0,1)_{12}$ errors

$$\nabla\nabla_{12}X_t = 0.4574\nabla V_t + \eta_t,$$

where $\eta_t = 0.9797\eta_{t-1} + Z_t - 0.5213Z_{t-1} - 0.2686Z_{t-2} - 0.6456Z_{t-12} + 0.3366Z_{t-13} + 0.1734Z_{t-14}$ and $Z_t \sim \text{WN}(0, 0.6425^2)$.

Consumer confidence index

The CCI graph in paragraph 3.1 showed that the indicator is stationary. From Figure 15 it can be seen that the differenced savings volumes tend to have negative relation to CCI.

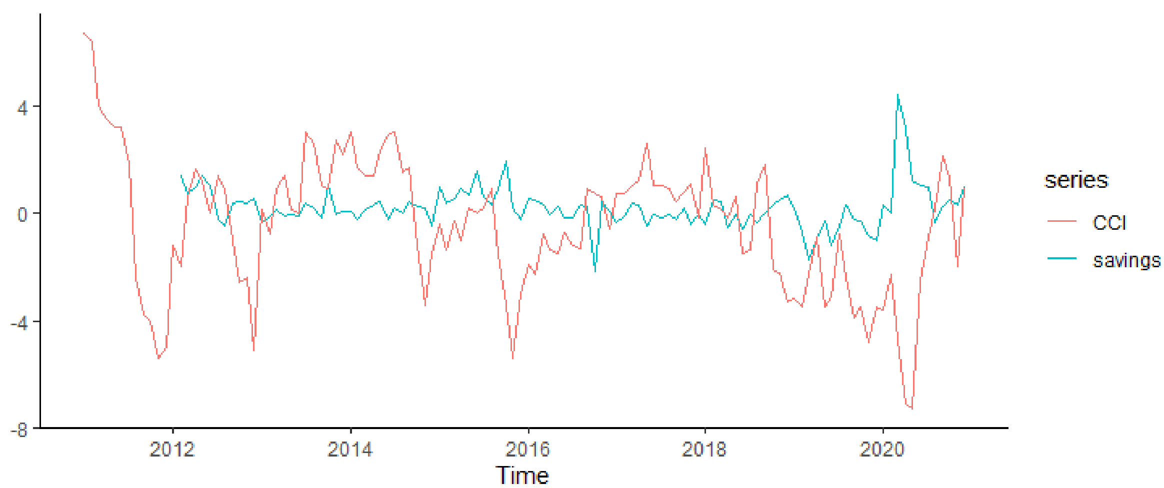

Figure 15: First order and seasonally differenced savings deposits compared to consumer confidence index

The initially fitted linear regression model was

$$\nabla\nabla_{12}X_t = -0.0993V_t + \eta_t,$$

where $\nabla\nabla_{12}X_t$ is first order and seasonally differenced savings deposit volume at time t, V_t is consumer confidence index at time t and η_t is model residual at time t. The residuals η_t were stationary. Based on the ACF and PACF plot, different ARIMA models were fitted to the residuals from linear regression model. The summary of the results is shown in Table 16.

Table 16: Comparison of fitted models for residuals from linear regression model with CCI

Model	p-value of L-B test	AIC	AICc	BIC
$\text{ARIMA}(0,0,3)(0,0,0)_{12}$	0.0988	234.89	235.48	248.25
$\text{ARIMA}(0,0,3)(0,0,1)_{12}$	0.5965	224.52	225.36	240.56
$\text{ARIMA}(1,0,0)(0,0,0)_{12}$	0.1191	232.65	232.89	240.67
$\text{ARIMA}(1,0,0)(1,0,0)_{12}$	0.2380	221.35	221.74	232.04
$\text{ARIMA}(1,0,0)(0,0,1)_{12}$	0.4712	221.58	221.98	232.28
$\text{ARIMA}(1,0,1)(0,0,1)_{12}$	0.5449	221.67	222.26	235.03
$\text{ARIMA}(0,0,1)(1,0,0)_{12}$	0.1190	226.34	226.73	237.03
$\text{ARIMA}(1,0,1)(1,0,0)_{12}$	0.3220	221.29	221.89	234.66

Linear regression with $\text{ARIMA}(1,0,0)(1,0,0)_{12}$ errors was showing the best results. The formula is

$$\nabla\nabla_{12}X_t = -0.0688V_t + \eta_t,$$

where $\eta_t = 0.4712\eta_{t-1} - 0.4265\eta_{t-12} + 0.2010\eta_{t-13} + Z_t$ and $Z_t \sim \text{WN}(0, 0.6569^2)$.

Purchasing managers' index

Figure 7 showed that PMI manufacturing and PMI services are both stationary and have had quite similar values during the time scope of this thesis. Further on, only PMI services is used as the linear regression was showing slightly better results. The comparison of differenced savings volumes and PMI services is shown in Figure 17.

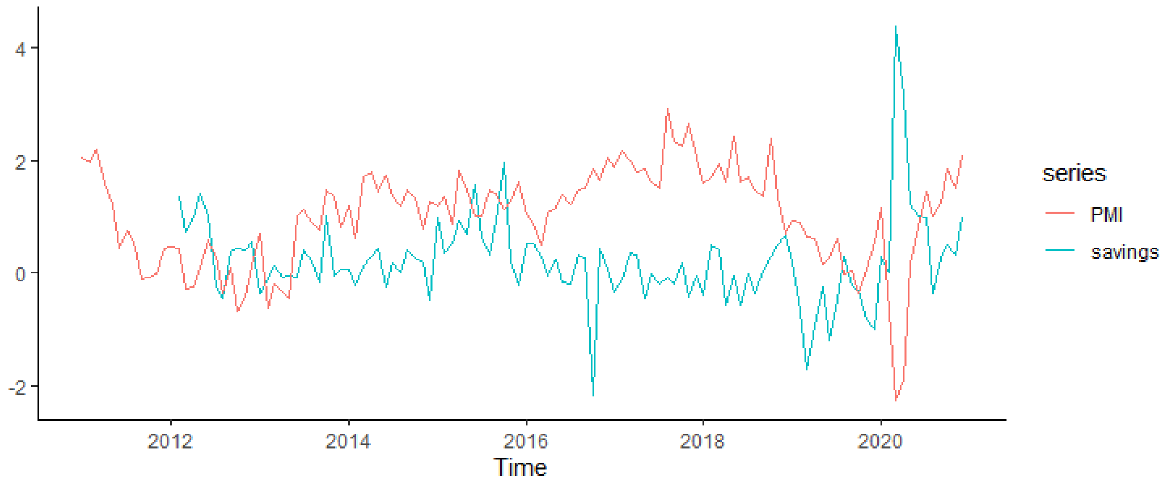

Figure 17: First order and seasonally differenced savings deposits compared to purchasing managers index (services)

The initial linear regression model was

$$\nabla\nabla_{12}X_t = 0.0033V_t + \eta_t,$$

where $\nabla\nabla_{12}X_t$ is first order and seasonally differenced savings deposit volume at time t, V_t is purchasing managers index (services) at time t and η_t is model residual at time t. The residuals from the model were stationary. Based on the sample ACF and PACF, several models were fitted to the residuals from linear regression model. The results are summarized in Table 18.

Table 18: Comparison of fitted models for residuals from linear regression model with PMI

Model	p-value of L-B test	AIC	AICc	BIC
$\text{ARIMA}(0,0,1)(0,0,0)_{12}$	0.0777	239.38	239.61	247.39
$\text{ARIMA}(0,0,1)(0,0,1)_{12}$	0.3281	221.96	222.35	232.65
$\text{ARIMA}(1,0,0)(0,0,0)_{12}$	0.2061	235.27	235.51	243.29
$\text{ARIMA}(1,0,0)(1,0,0)_{12}$	0.3856	221.71	222.10	232.40
$\text{ARIMA}(1,0,0)(0,0,1)_{12}$	0.7279	217.7	218.09	228.39
$\text{ARIMA}(1,0,1)(0,0,1)_{12}$	0.7270	219.40	220.00	232.77
$\text{ARIMA}(0,0,1)(1,0,0)_{12}$	0.1546	225.43	225.82	236.12

$\text{ARIMA}(1,0,1)(1,0,0)_{12}$	0.3220	222.99	223.59	236.36

One of the best models was linear regression with $\text{ARIMA}(1,0,0)(0,0,1)_{12}$ errors

$$\nabla\nabla_{12}X_t = 0.0027V_t + \eta_t,$$

where $\eta_t = 0.4517\eta_{t-1} + Z_t - 0.5935Z_{t-12}$ and $Z_t \sim \text{WN}(0, 0.6375^2)$.

Gross domestic product

In paragraph 3.1 it was seen that GDP is not stationary time series. Results of unit root tests and review of the sample ACF and PACF suggested taking both regular and seasonal difference. The differenced volumes of GDP can be seen in Figure 19.

Figure 19: First order and seasonally differenced savings deposits compared to first order and seasonally differenced volumes of GDP

As both savings deposits and GDP are stationary after first order and seasonal difference, the linear regression can be assigned to original values and differences can be taken from the model residuals. Thus, the estimated linear regression model was

$$X_t = 0.3204V_t + \eta_t,$$

where X_t is savings deposit volume at time t, V_t is GDP at time t and η_t is model residual at time t. The residuals from linear regression were stationary after first order and seasonal difference.

As a first step, $\text{ARIMA}(0,1,1)(0,1,0)_{12}$ was fitted to the model residuals. The p-value of Ljung-Box test was below 0.05 and there were significant spikes at lags 2 and 12 in the ACF plot of the residuals from the fitted model. Therefore, $\text{ARIMA}(0,1,3)(0,1,1)_{12}$ was tried out as well.

Next, $\text{ARIMA}(1,1,0)(0,1,0)_{12}$ was fitted. There was a significant spike at lag 12 in the PACF plot of the residuals and due to that $\text{ARIMA}(1,1,0)(1,1,0)_{12}$ was tried.

Finally, the models with combination of MA and AR components were tried out. The comparison of the models based on Ljung-Box test and goodness of fit statistics is summarized in Table 20.

Table 20: Comparison of fitted models for residuals from linear regression model with GDP

Model	p-value of L-B test	AIC	AICc	BIC
$\text{ARIMA}(0,1,1)(0,1,0)_{12}$	0.0240	240.19	240.42	248.21
$\text{ARIMA}(0,1,3)(0,1,1)_{12}$	0.5857	225.43	226.27	241.46
$\text{ARIMA}(1,1,0)(0,1,0)_{12}$	0.0541	234.49	234.72	242.50
$\text{ARIMA}(1,1,0)(1,1,0)_{12}$	0.3297	221.61	222.00	232.30
$\text{ARIMA}(1,1,0)(0,1,1)_{12}$	0.4914	221.01	221.40	231.70
$\text{ARIMA}(1,1,1)(0,1,1)_{12}$	0.4484	220.72	221.32	234.09
$\text{ARIMA}(2,1,1)(0,1,1)_{12}$	0.6337	219.89	220.73	235.93
$\text{ARIMA}(1,1,1)(1,1,0)_{12}$	0.3515	220.95	221.54	234.31

Linear regression with $\text{ARIMA}(2,1,1)(0,1,1)_{12}$ errors was having the lowest goodness of fit statistics and highest p-value of Ljung-Box test. The formula is

$$X_t = -0.0147V_t + \eta_t,$$

where

$$\eta_t = 2.3210(\eta_{t-1} - \eta_{t-13}) - 1.6542(\eta_{t-2} - \eta_{t-14}) + 0.3332(\eta_{t-3} - \eta_{t-15}) + \eta_{t-12} + Z_t - 0.8926Z_{t-1} - 0.5812Z_{t-12} + 0.5188Z_{t-13}$$

and $Z_t \sim \mathrm{WN}(0, 0.6409^2)$.

House price index

Figure 9 showed that house price index is not stationary time series. Unit root tests and review of ACF and PACF plot implied taking first order and seasonal difference. The differenced volumes of HPI can be seen in Figure 21.

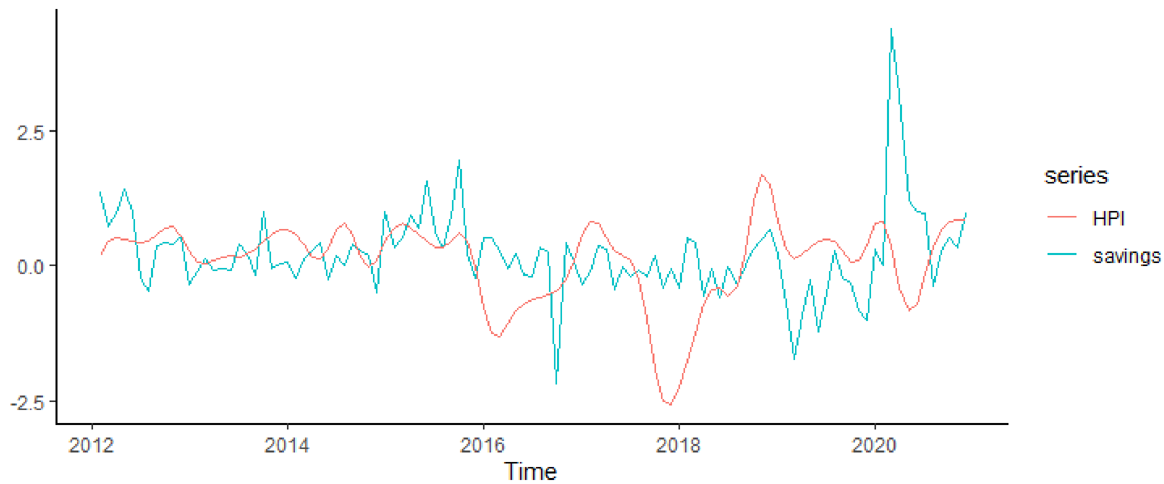

Figure 21: First order and seasonally differenced savings deposits compared to first order and seasonally differenced values of house price index

Similarly to GDP, the first step was to fit linear regression with non-differenced values

$$X_t = 0.9380V_t + \eta_t,$$

where X_t is savings deposit volume at time t, V_t is house price index at time t and η_t is model residual at time t. After that, first order and seasonal difference was taken from the model residuals η_t. Several models were fitted, and the comparison of the models based on Ljung-Box test and goodness of fit statistics is summarized in Table 22.

Table 22: Comparison of fitted models for residuals from linear regression model with house price index

Model	p-value of L-B test	AIC	AICc	BIC
ARIMA$(0,1,1)(0,1,0)_{12}$	0.0226	241.29	241.52	249.31
ARIMA$(0,1,2)(0,1,0)_{12}$	0.0696	240.86	241.25	251.55
ARIMA$(0,1,3)(0,1,0)_{12}$	0.0873	239.11	239.7	252.47
ARIMA$(0,1,1)(0,1,1)_{12}$	0.2843	227.99	228.38	238.68
ARIMA$(0,1,2)(0,1,1)_{12}$	0.4314	227.16	227.75	240.52
ARIMA$(0,1,3)(0,1,1)_{12}$	0.6526	225.25	226.09	241.29
ARIMA$(1,1,0)(0,1,0)_{12}$	0.1197	235.98	236.21	244
ARIMA$(1,1,0)(1,1,0)_{12}$	0.2350	223.22	223.61	233.91
ARIMA$(1,1,1)(0,1,0)_{12}$	0.1262	237.47	237.86	248.16
ARIMA$(1,1,0)(0,1,1)_{12}$	0.5357	221.67	222.06	232.36
ARIMA$(1,1,1)(0,1,1)_{12}$	0.6118	222.68	223.28	236.05
ARIMA$(0,1,1)(1,1,0)_{12}$	0.1508	227.64	228.03	238.33
ARIMA$(1,1,1)(1,1,0)_{12}$	0.3076	224.26	224.86	237.63

One of the best models was linear regression with ARIMA$(1,1,0)(0,1,1)_{12}$ errors

$$X_t = 0.2141V_t + \eta_t,$$

where $\eta_t = 1.4995(\eta_{t-1} - \eta_{t-13}) - 0.4995(\eta_{t-2} - \eta_{t-14}) + \eta_{t-12} + Z_t - 0.5048Z_{t-12}$ and $Z_t \sim \mathrm{WN}(0, 0.6549^2)$.

Quantitative easing

In Figure 10 it was seen that quantitative easing is non-stationary time series. The series was stationary after first order and seasonal difference (see Figure 23).

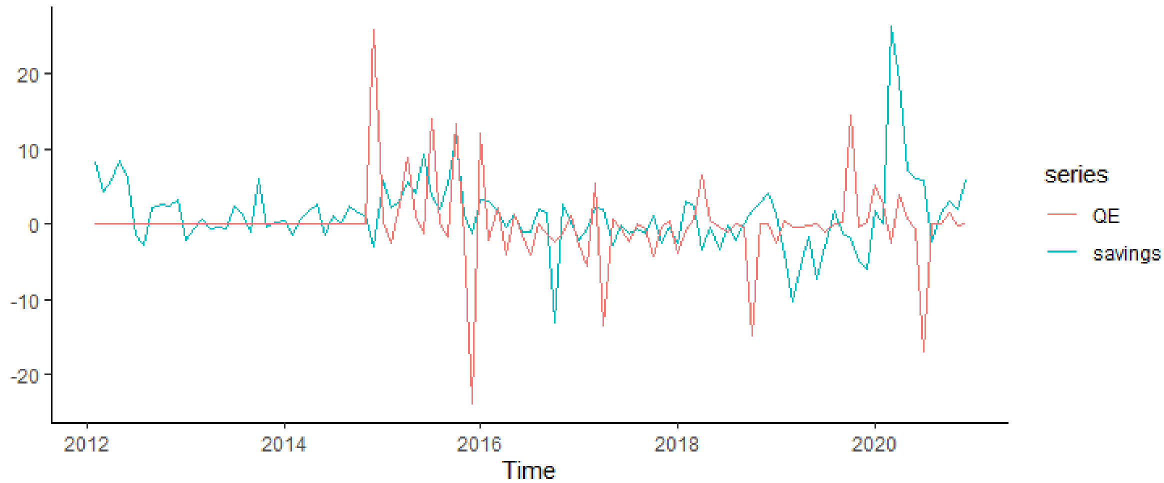

Figure 23: First order and seasonally differenced savings deposits compared to first order and seasonally differenced values of quantitative easing

The estimated linear regression model with non-differenced variables was

$$X_t = 1.4514V_t + \eta_t,$$

where X_t is savings deposit volume at time t, V_t is quantitative easing at time t and η_t is model residual at time t. The residuals from linear regression model were stationary after first order and seasonal difference. Multiple models were fitted, and the comparison of the statistics is summarized in Table 24.

Table 24: Comparison of fitted models for residuals from linear regression model with QE

Model	p-value of L-B test	AIC	AICc	BIC
$ARIMA(0,1,1)(0,1,0)_{12}$	0.0490	242.9	243.13	250.92
$ARIMA(0,1,3)(0,1,1)_{12}$	0.6119	227.38	228.22	243.42
$ARIMA(1,1,0)(0,1,0)_{12}$	0.1274	237.33	237.56	245.35
$ARIMA(1,1,0)(1,1,0)_{12}$	0.1958	225.02	225.41	235.71
$ARIMA(1,1,0)(0,1,1)_{12}$	0.4283	223.40	223.79	234.09
$ARIMA(2,1,1)(0,1,1)_{12}$	0.5915	222.48	223.32	238.52
$ARIMA(0,1,1)(1,1,0)_{12}$	0.1180	230.94	231.34	241.64

Linear regression with ARIMA$(2,1,1)(0,1,1)_{12}$ errors was showing the best results. The formula is

$$X_t = 0.0046V_t + \eta_t,$$

where

$$\eta_t = 2.3645(\eta_{t-1} - \eta_{t-13}) - 1.7359(\eta_{t-2} - \eta_{t-14}) + 0.3714(\eta_{t-3} - \eta_{t-15}) + \eta_{t-12} + Z_t - 0.9200Z_{t-1} - 0.6267Z_{t-12} + 0.5766Z_{t-13}$$

and $Z_t \sim \text{WN}(0,\ 0.6453^2)$.

Stibor 1-month, CCI and GDP

In addition to ARIMAX models with single regressor, models with multiple independent variables were also tried out. To find the appropriate list of variables to the linear regression, a stepwise procedure using p-values was performed. The variables in the final model with p-value less than 0.05 were Stibor 1-month, consumer confidence indicator and gross domestic product. The estimated simple linear regression was

$$\nabla\nabla_{12}X_t \ = \ -1.6555\nabla V_{Stibor,t} - 0.0881V_{CCI,t} - 0.0244\nabla\nabla_{12}V_{GDP,t} + \eta_t,$$

where $\nabla\nabla_{12}X_t$ is first order and seasonally differenced savings deposit volume at time t, $\nabla V_{Stibor,t}$ is first order differenced Stibor 1-month value at time t, $V_{CCI,t}$ is consumer confidence indicator at time t, $\nabla\nabla_{12}V_{GDP,t}$ is first order and seasonally differenced volume of GDP and η_t is model residual at time t. The residuals from linear regression model were stationary. Based on ACF and PACF plots, different models were fitted, and the comparison of the models based on Ljung-Box test and goodness of fit statistics is summarized in Table 25.

Table 25: Comparison of fitted models for residuals from linear regression model with Stibor 1-month, CCI and GDP

Model	p-value of L-B test	AIC	AICc	BIC
ARIMA$(0,0,1)(0,0,0)_{12}$	0.0437	238.08	238.67	251.44
ARIMA$(0,0,1)(0,0,1)_{12}$	0.2709	229.11	229.95	245.14

$\text{ARIMA}(1,0,0)(0,0,0)_{12}$	0.0512	233.88	234.47	247.24
$\text{ARIMA}(1,0,0)(1,0,0)_{12}$	0.3216	221.99	222.83	238.03
$\text{ARIMA}(1,0,0)(0,0,1)_{12}$	0.4026	223.08	223.92	239.11
$\text{ARIMA}(1,0,1)(0,0,1)_{12}$	0.3631	222.52	223.65	241.23
$\text{ARIMA}(0,0,1)(1,0,0)_{12}$	0.2297	226.83	227.67	242.87
$\text{ARIMA}(1,0,1)(1,0,0)_{12}$	0.3132	221.34	222.48	240.05

Linear regression with $\text{ARIMA}(1,0,0)(1,0,0)_{12}$ errors was having the lowest AIC and AICc, and highest p-value of Ljung-Box test. The formula is

$$\nabla\nabla_{12}X_t \ = \ -0.3905\nabla V_{Stibor,t} - 0.0611V_{CCI,t} - 0.0161\nabla\nabla_{12}V_{GDP,t} + \eta_t,$$

where $\eta_t = 0.4526\eta_{t-1} - 0.4295\eta_{t-12} + 0.1944\eta_{t-13} + Z_t$ and $Z_t \sim \text{WN}(0, 0.6529^2)$.

PMI and GDP

As it was seen, the ARIMAX model with Stibor 1-month, CCI and GDP was not having better goodness of fit statistics than some of the ARIMAX models with single regressor variable. Therefore, additional model with multiple regressors was tried out. As ARIMAX model with PMI services was showing good results, several models with combination of this variable were fitted. The best one was with purchasing managers' index (services) and gross domestic product. The estimated simple linear regression was

$$\nabla\nabla_{12}X_t \ = \ 0.0034V_{PMI,t} - 0.0260\nabla\nabla_{12}V_{GDP,t} + \eta_t,$$

where $\nabla\nabla_{12}X_t$ is first order and seasonally differenced savings deposit volume at time t, $V_{PMI,t}$ is purchasing managers index at time t, $\nabla\nabla_{12}V_{GDP,t}$ is first order and seasonally differenced volume of GDP and η_t is model residual at time t. The residuals from linear regression model were stationary. Based on the sample ACF and PACF, several models were fitted to the residuals. The results are summarized in Table 26.

Table 26: Comparison of fitted models for residuals from linear regression model with PMI and GDP

Model	p-value of L-B test	AIC	AICc	BIC
$\text{ARIMA}(0,0,2)(0,0,0)_{12}$	0.0550	238.05	238.64	251.41
$\text{ARIMA}(0,0,2)(0,0,1)_{12}$	0.4689	221.24	222.08	237.28
$\text{ARIMA}(1,0,0)(0,0,0)_{12}$	0.0770	234.18	234.57	244.87
$\text{ARIMA}(1,0,0)(1,0,0)_{12}$	0.4651	219.97	220.56	233.33
$\text{ARIMA}(1,0,0)(0,0,1)_{12}$	0.6777	217.29	217.88	230.65
$\text{ARIMA}(1,0,1)(0,0,1)_{12}$	0.6451	218.73	219.57	234.77
$\text{ARIMA}(0,0,1)(1,0,0)_{12}$	0.1980	223.9	224.49	237.26
$\text{ARIMA}(1,0,1)(1,0,0)_{12}$	0.4558	220.72	221.56	236.76

The best model was linear regression with $\text{ARIMA}(1,0,0)(0,0,1)_{12}$ errors

$$\nabla\nabla_{12}X_t = 0.0028V_{PMI,t} - 0.0135\nabla\nabla_{12}V_{GDP,t} + \eta_t,$$

where $\eta_t = 0.4457\eta_{t-1} + Z_t - 0.5597Z_{t-12}$, $Z_t \sim \text{WN}(0, 0.6355^2)$.

3.3 Validation

In this section the accuracy of the models from chapter 3.2 was examined. For this purpose, the savings deposits' volumes from the test period were compared to the model forecasts.

Holt and Holt-Winters

The first time series method investigated was the Holt's linear trend method. As it was shown in paragraph 3.2.1, the h-step-ahead forecast $\hat{x}_{t+h|t}$ is calculated as the last estimated level l_t plus h times the last estimated trend b_t. Therefore, the forecasts are a linear function of h (see Figure 27).

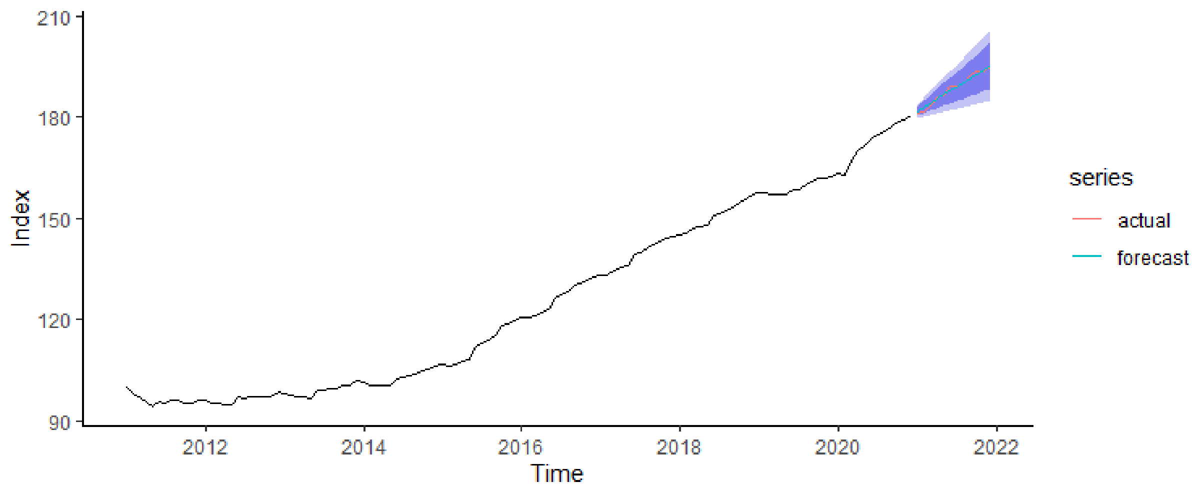

Figure 27: The actual savings deposits' volumes compared to the forecasts from Holt's linear trend method, 80% and 95% CI

ARIMA model

The forecasting formula for ARIMA$(1,1,0)(0,1,1)_{12}$ is

$$\hat{X}_{t+1} = 1.5227(X_t - X_{t-12}) - 0.5227(X_{t-1} - X_{t-13}) + X_{t-11} + \hat{Z}_{t+1} - 0.5023Z_{t-11}.$$

For example, the prediction for January 2021 was

$$\begin{aligned}\hat{X}_{Jan\,2021} &= 1.5227(X_{Dec\,2020} - X_{Dec\,2019}) - 0.5227(X_{Nov\,2020} - X_{Nov\,2019}) + X_{Jan\,2020} \\ &\quad + \hat{Z}_{Jan\,2021} - 0.5023Z_{Jan\,2020} \\ &= 1.5227(180.6125 - 162.5998) - 0.5227(179.1499 - 162.1103) \\ &\quad + 163.3414 - 0.5023 \cdot 0.6882 = 181.5171,\end{aligned}$$

while the actual value was 181.3418. The 80% and 95% prediction intervals for the same period were accordingly

$$\hat{X}_{Jan\,2021} \pm 1.282\sqrt{{\sigma_1}^2} = 181.5171 \pm 1.282 \cdot 0.6576 = 181.5171 \pm 1.2890,$$

$$\hat{X}_{Jan\,2021} \pm 1.960\sqrt{{\sigma_1}^2} = 181.5171 \pm 1.960 \cdot 0.6576 = 181.5171 \pm 0.8430.$$

Hence, the actual value stayed within the confidence intervals. All the values from the test period compared to the model forecasts can be seen in Figure 28. The confidence intervals from ARIMA models with $d \geq 1$ and/or $D \geq 1$ increase in time.

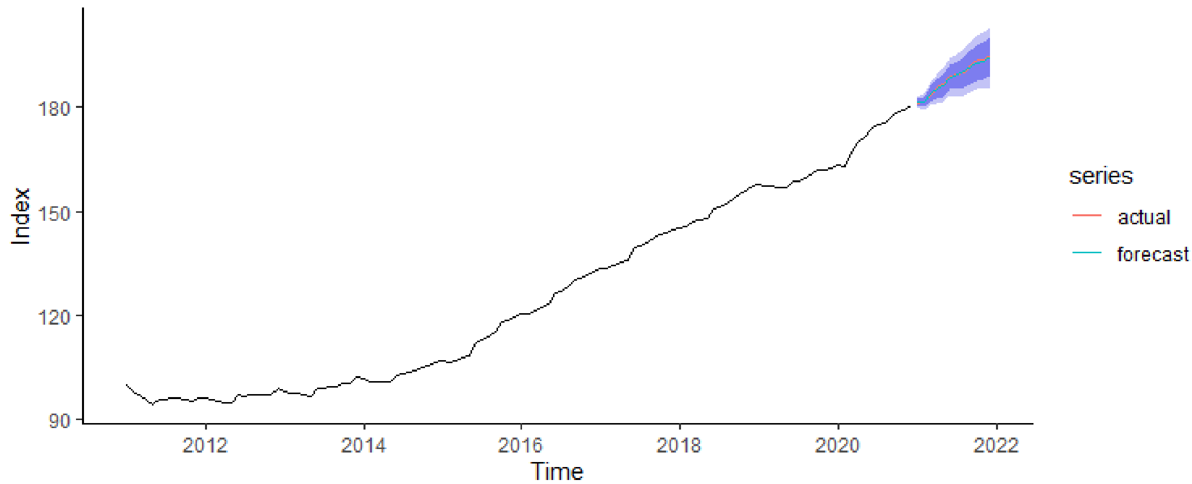

Figure 28: The actual savings deposits' volumes compared to the forecasts from ARIMA$(1,1,0)(0,1,1)_{12}$, 80% and 95% CI

ARIMAX models

Stibor 1-month

The prediction formula for linear regression with Stibor 1-month and ARIMA$(1,0,2)(0,0,1)_{12}$ errors is

$$\begin{aligned}\nabla\nabla_{12}\hat{X}_{t+1} &= 0.4574\nabla\hat{V}_{t+1} + 0.9797\eta_t + \hat{Z}_{t+1} - 0.5213Z_t - 0.2686Z_{t-1} - 0.6456Z_{t-11} \\ &\quad + 0.3366Z_{t-12} + 0.1734Z_{t-13}.\end{aligned}$$

In case the actual values of Stibor 1-month were used, the point forecast for January 2021 was

$$\begin{aligned}\nabla\nabla_{12}\hat{X}_{Jan\ 2021} &= 0.4574\nabla\hat{V}_{Jan\ 2021} + 0.9797\eta_{Dec\ 2020} + 0 - 0.5213Z_{Dec\ 2020} \\ &\quad - 0.2686Z_{Nov\ 2020} - 0.6456Z_{Jan\ 2020} + 0.3366Z_{Dec\ 2019} + 0.1734Z_{Nov\ 2019} \\ &= 0.4574 \cdot 0.0538 + 0.9797 \cdot 0.9916 + 0 - 0.5213 \cdot 0.4250 \\ &\quad - 0.2686(-0.3653) - 0.6456 \cdot 0.4787 + 0.3366(-0.1729) \\ &\quad + 0.1734(-0.2921) = 0.4547.\end{aligned}$$

The observed value was -0.0123. The 80% and 95% prediction intervals for January were accordingly

$$\hat{X}_{Jan\ 2021} \pm 1.282\sqrt{{\sigma_1}^2} = 0.4547 \pm 1.282 \cdot 0.6425 = 0.4547 \pm 0.8237,$$

$$\hat{X}_{Jan\ 2021} \pm 1.960\sqrt{{\sigma_1}^2} = 0.4547 \pm 1.960 \cdot 0.6425 = 0.4547 \pm 1.2593.$$

Thus, the actual value stayed within the prediction intervals. All the actual values in comparison of model forecasts are shown in Figure 29.

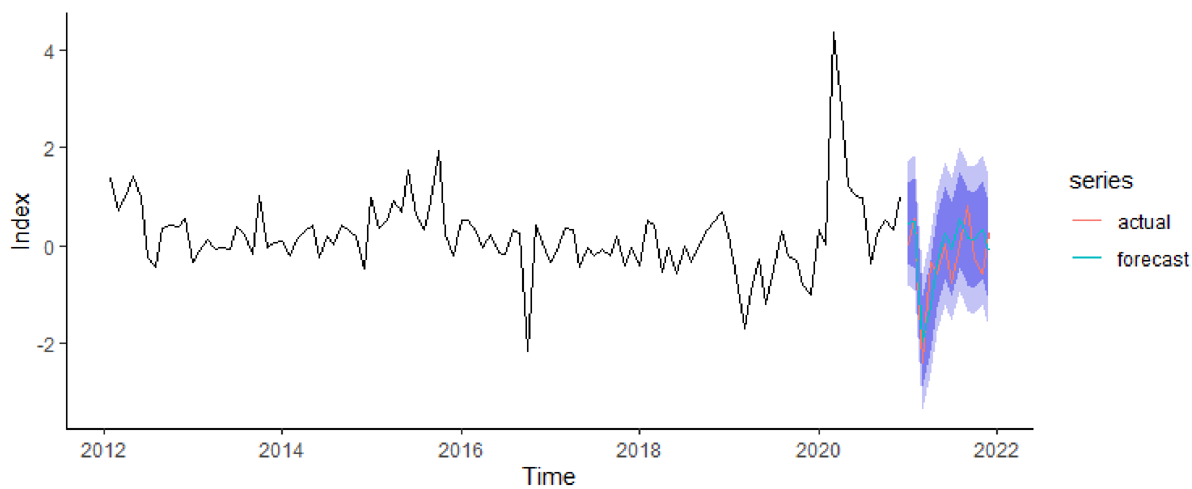

Figure 29: The actual differenced savings deposits volumes compared to the forecasts from linear regression with Stibor 1-month and ARIMA$(1,0,2)(0,0,1)_{12}$ errors, 80% and 95% CI

After converting the differenced savings volumes back to original values, the forecast for January 2021 was

$$\begin{aligned}\hat{X}_{Jan\ 2021} &= \nabla\nabla_{12}\hat{X}_{Jan\ 2021} + X_{Dec\ 2020} + X_{Jan\ 2020} - X_{Dec\ 2019} \\ &= 0.4547 + 180.6125 + 163.3414 - 162.5998 = 181.8088.\end{aligned}$$

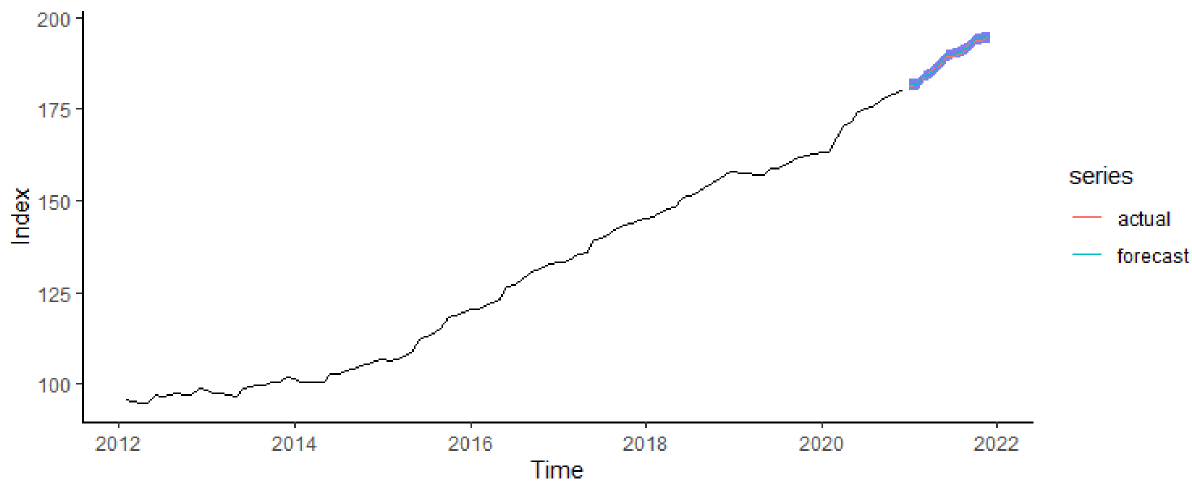

Figure 30: The actual savings deposits volumes compared to the converted forecasts from linear regression with Stibor 1-month and ARIMA$(1,0,2)(0,0,1)_{12}$ errors

Consumer confidence indicator

The formula for calculating forecasts from linear regression with CCI and ARIMA$(1,0,0)(1,0,0)_{12}$ errors is

$$\nabla\nabla_{12}\hat{X}_{t+1} = -0.0688\nabla\hat{V}_{t+1} + 0.4712\eta_t - 0.4265\eta_{t-11} + 0.2010\eta_{t-12} + \hat{Z}_{t+1}.$$

In case the observed values of consumer confidence indicator were used, then the forecast for January 2021 was

$$\begin{aligned}\nabla\nabla_{12}\hat{X}_{Jan\ 2021} &= -0.0688\hat{V}_{Jan\ 2021} + 0.4712\eta_{Dec\ 2020} - 0.4265\eta_{Jan\ 2020} + 0.2010\eta_{Dec\ 2019} \\ &= -0.0688 \cdot 2.7 + 0.4712 \cdot 1.0350 - 0.4265 \cdot 0.0543 + 0.2010 \cdot (-1.2418) \\ &= 0.0292\end{aligned}$$

The actual value was -0.0123. All forecasts compared to observed values can be seen in Figure 31.

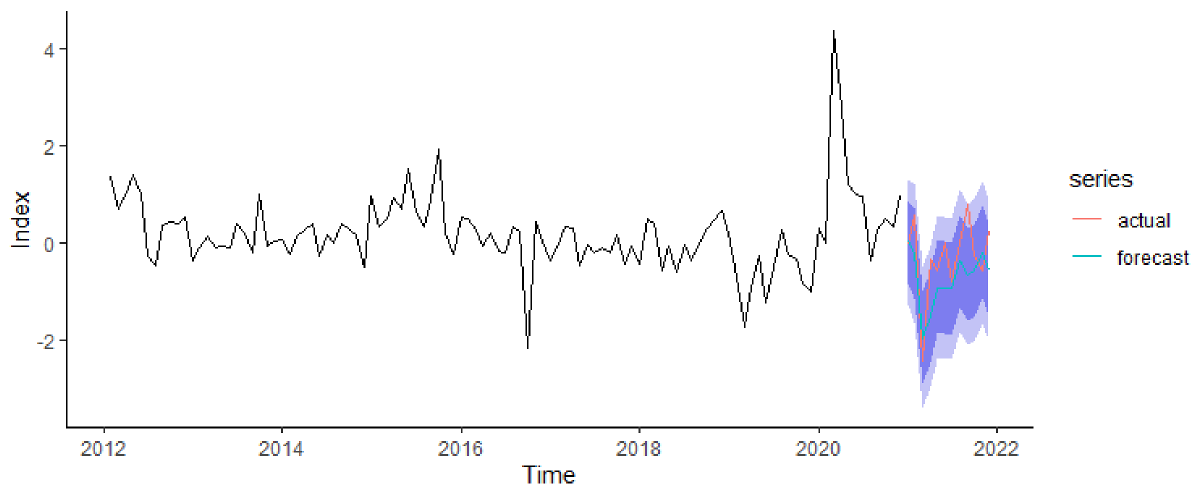

Figure 31: The actual differenced savings deposits volumes compared to the forecasts from linear regression with CCI and ARIMA$(1,0,0)(1,0,0)_{12}$ errors, 80% and 95% CI

After transforming the differenced savings volumes back to original, the forecast for January 2021 was

$$\begin{aligned}\hat{X}_{Jan\ 2021} &= \nabla\nabla_{12}\hat{X}_{Jan\ 2021} + X_{Dec\ 2020} + X_{Jan\ 2020} - X_{Dec\ 2019} \\ &= 0.0292 + 180.6125 + 163.3414 - 162.5998 = 181.3833.\end{aligned}$$

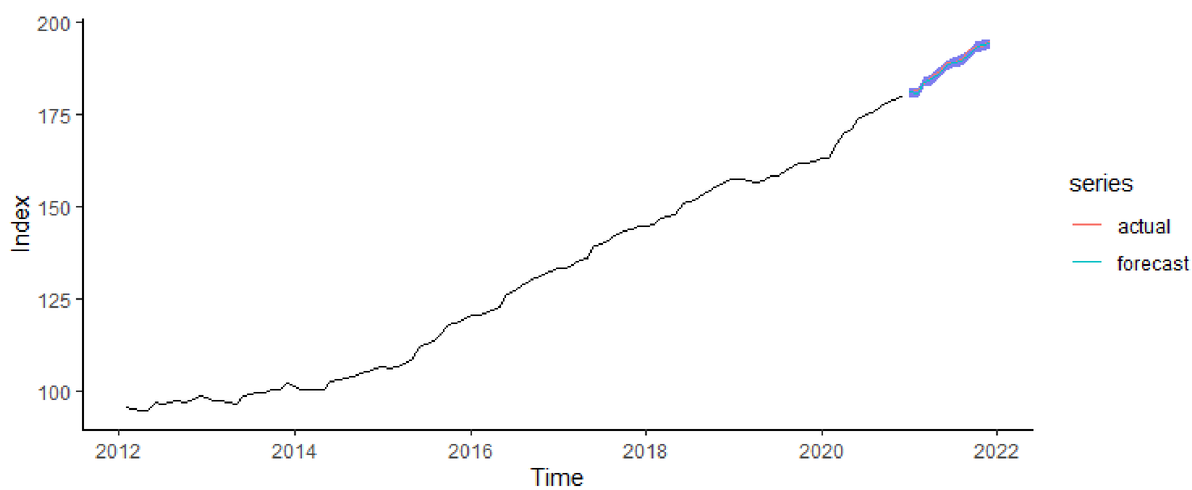

Figure 32: The actual savings deposits volumes compared to the converted forecasts from linear regression with CCI and ARIMA$(1,0,0)(1,0,0)_{12}$ errors

Purchasing managers' index

The prediction formula for linear regression with PMI and $\text{ARIMA}(1,0,0)(0,0,1)_{12}$ errors is

$$\nabla\nabla_{12}\hat{X}_{t+1} = 0.0027\hat{V}_{t+1} + 0.4517\eta_t + \hat{Z}_{t+1} - 0.5935Z_{t-11}.$$

In case the actual values of purchasing managers' index were used, the forecast for January 2021 was

$$\begin{aligned}\nabla\nabla_{12}\hat{X}_{Jan\ 2021} &= 0.0027\hat{V}_{Jan\ 2021} + 0.4517\eta_{Dec\ 2020} - 0.5935Z_{Jan\ 2020}\\ &= 0.0027 \cdot 63.0 + 0.4517 \cdot 0.8092 - 0.5935 \cdot 0.4385 = 0.2753.\end{aligned}$$

In Figure 33 all the values from the test period compared to the model forecasts are shown.

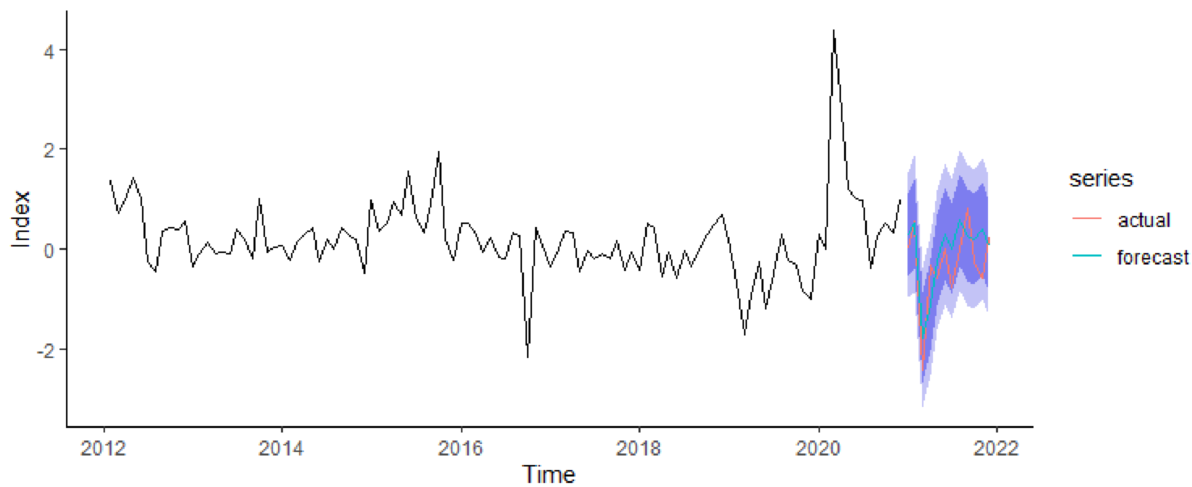

Figure 33: The actual differenced savings deposits volumes compared to the forecasts from linear regression with PMI and $\text{ARIMA}(1,0,0)(0,0,1)_{12}$ errors, 80% and 95% CI

By converting the differenced savings volumes back to original, the prediction for January 2021 was

$$\begin{aligned}\hat{X}_{Jan\ 2021} &= \nabla\nabla_{12}\hat{X}_{Jan\ 2021} + X_{Dec\ 2020} + X_{Jan\ 2020} - X_{Dec\ 2019}\\ &= 0.2753 + 180.6125 + 163.3414 - 162.5998 = 181.6294.\end{aligned}$$

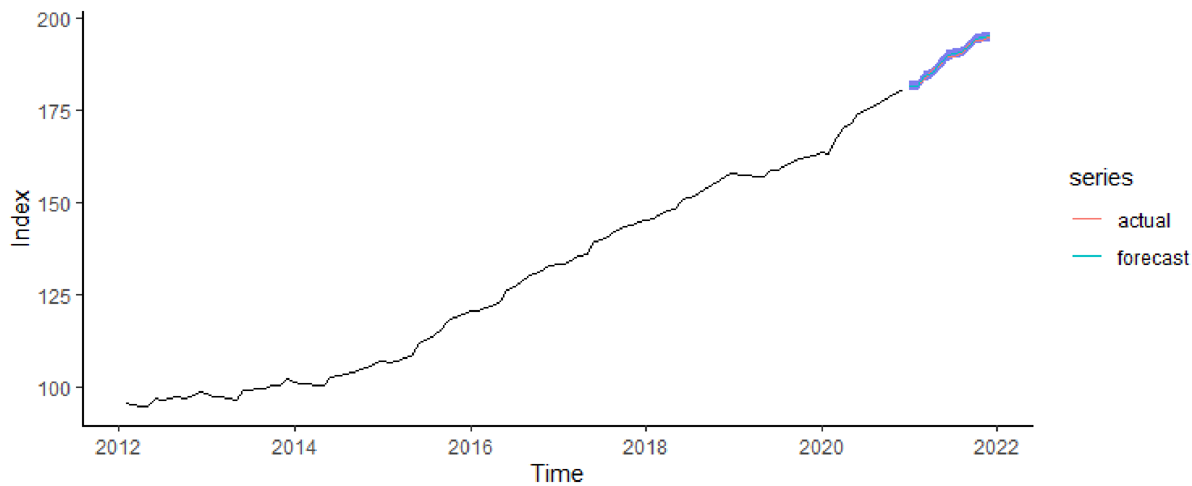

Figure 34: The actual savings deposits volumes compared to the converted forecasts from linear regression with PMI and $\text{ARIMA}(1,0,0)(0,0,1)_{12}$ errors

Gross domestic product

The forecasting formula for linear regression with GDP and $\text{ARIMA}(2,1,1)(0,1,1)_{12}$ errors is

$$\begin{aligned}\hat{X}_{t+1} &= -0.0147\hat{V}_{t+1} + 2.3210(\eta_t - \eta_{t-12}) - 1.6542(\eta_{t-1} - \eta_{t-13}) \\ &\quad + 0.3332(\eta_{t-2} - \eta_{t-14}) + \eta_{t-11} + \hat{Z}_{t+1} - 0.8926Z_t - 0.5812Z_{t-11} \\ &\quad + 0.5188Z_{t-12}.\end{aligned}$$

When using the actual values of gross domestic product, the forecast for January 2021 was

$$\begin{aligned}\hat{X}_{Jan\ 2021} &= -0.0147\hat{V}_{Jan\ 2021} + 2.3210(\eta_{Dec\ 2020} - \eta_{Dec\ 2019}) \\ &\quad - 1.6542(\eta_{Nov\ 2020} - \eta_{Nov\ 2019}) + 0.3332(\eta_{Oct\ 2020} - \eta_{Oct\ 2019}) + \eta_{Jan\ 2020} \\ &\quad - 0.8926Z_{Dec\ 2020} - 0.5812Z_{Jan\ 2020} + 0.5188Z_{Dec\ 2019} \\ &= -0.0147 \cdot 435.1818 + 2.3210(187.1654 - 169.2590) \\ &\quad - 1.6542(185.4914 - 168.5547) + 0.3332(185.1971 - 168.5934) \\ &\quad + 169.6803 + 0 - 0.8926 \cdot 0.4155 - 0.5812 \cdot 0.5995 + 0.5188(-0.2980) \\ &= 181.4856.\end{aligned}$$

All the observed values in comparison of model forecasts are shown in Figure 35.

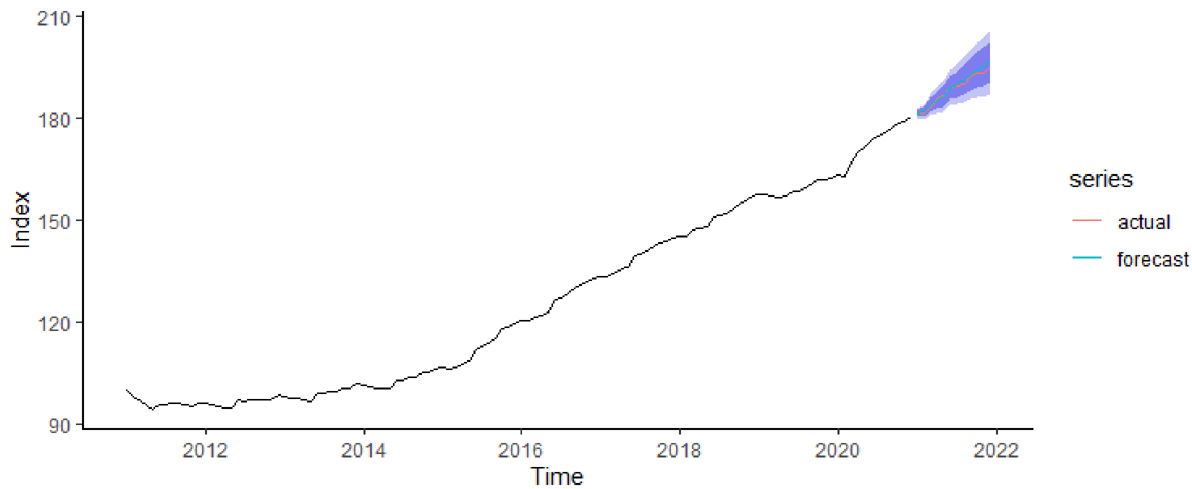

Figure 35: The actual savings deposits volumes compared to the forecasts from linear regression with GDP and ARIMA$(2,1,1)(0,1,1)_{12}$ errors

House price index

The formula for calculating forecasts from linear regression with HPI and ARIMA$(1,1,0)(0,1,1)_{12}$ errors is

$$\hat{X}_{t+1} = 0.2141\hat{V}_{t+1} + 1.4995(\eta_t - \eta_{t-12}) - 0.4995(\eta_{t-1} - \eta_{t-13}) + \eta_{t-11} + \hat{Z}_{t+1} - 0.5048Z_{t-11}.$$

In case the observed values of house price index were used, the forecast for January 2021 was

$$\begin{aligned}\hat{X}_{Jan\ 2021} = {} & 0.2141\hat{V}_{Jan\ 2021} + 1.4995(\eta_{Dec\ 2020} - \eta_{Dec\ 2019}) - 0.4995(\eta_{Nov\ 2020} \\ & - \eta_{Nov\ 2019}) + \eta_{Jan\ 2020} + \hat{Z}_{Jan\ 2021} - 0.5048Z_{Jan\ 2020} \\ & = 0.2141 \cdot 170.4347 + 1.4995(144.4313 - 128.2321) \\ & - 0.4995(143.1832 - 127.7756) + 128.8399 + 0 - 0.5048 \cdot 0.5800 \\ & = 181.6323.\end{aligned}$$

The actual values in parallel with model forecasts are shown in Figure 36.

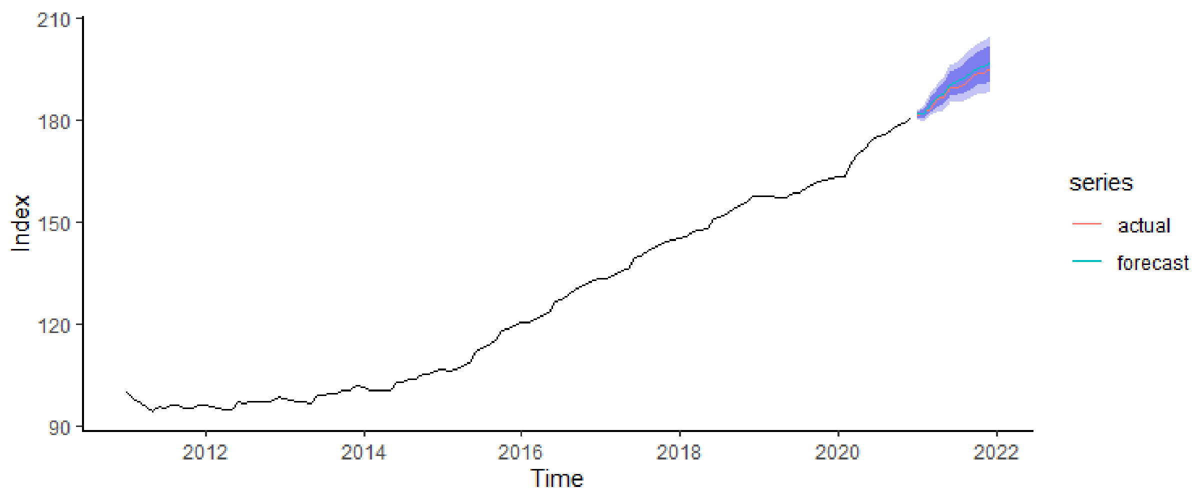

Figure 36: The actual savings deposits volumes compared to the forecasts from linear regression with HPI and ARIMA$(1,1,0)(0,1,1)_{12}$ errors

Quantitative easing

The forecasting formula for linear regression with QE and ARIMA$(2,1,1)(0,1,1)_{12}$ errors is

$$\hat{X}_{t+1} = 0.0046\hat{V}_{t+1} + 2.3645(\eta_t - \eta_{t-12}) - 1.3645(\eta_{t-1} - \eta_{t-13}) + 0.3714(\eta_{t-2} - \eta_{t-14}) + \eta_{t-11} + \hat{Z}_{t+1} - 0.9200Z_t - 0.6267Z_{t-11} + 0.5766Z_{t-12}.$$

In case the actual volumes of quantitative easing were used, the forecast for January 2021 was

$$\begin{aligned}\hat{X}_{Jan\ 2021} = {} & 0.0046\hat{V}_{Jan\ 2021} + 2.3645(\eta_{Dec\ 2020} - \eta_{Dec\ 2019}) \\ & - 1.7359(\eta_{Nov\ 2020} - \eta_{Nov\ 2019}) + 0.3714(\eta_{Oct\ 2020} - \eta_{Oct\ 2019}) + \eta_{Jan\ 2020} \\ & - 0.9200Z_{Dec\ 2020} - 0.6267Z_{Jan\ 2020} + 0.5766Z_{Dec\ 2019} \\ & = 0.0046 \cdot 125.1814 + 2.3645(180.0681 - 162.0678) \\ & - 1.7359(178.6231 - 161.5955) + 0.3714(178.0998 - 161.4022) \\ & + 162.7707 - 0.9200 \cdot 0.4050 - 0.6267 \cdot 0.7019 + 0.5766(-0.3562) \\ & = 181.5337\end{aligned}$$

All the model forecasts in comparison of observed values can be seen in Figure 37.

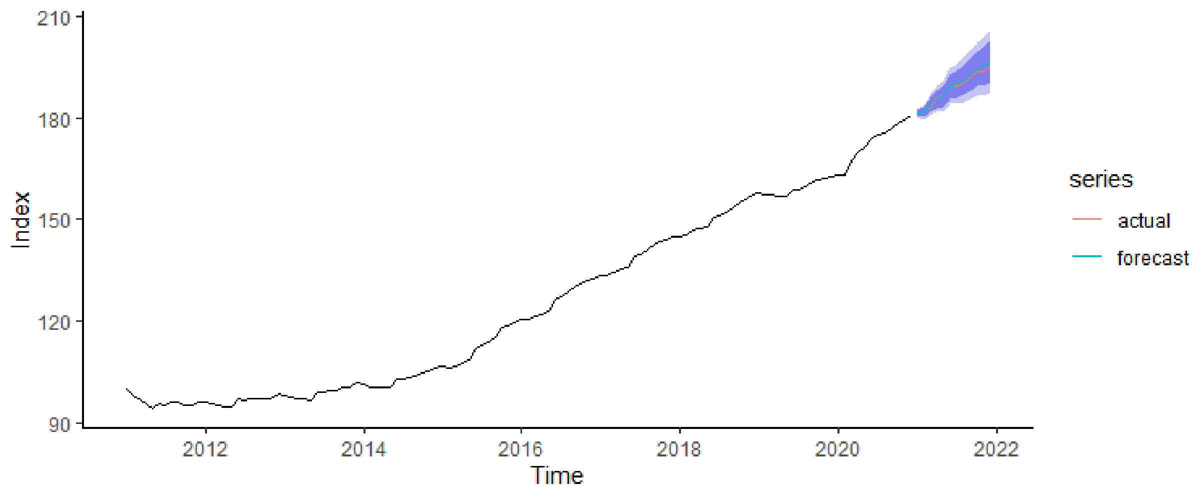

Figure 37: The actual savings deposits volumes compared to the forecasts from linear regression with QE and ARIMA$(2,1,1)(0,1,1)_{12}$ errors

Stibor 1-month, CCI and GDP

The formula for calculating forecasts from linear regression with Stibor 1-month, CCI, GDP and ARIMA$(1,0,0)(1,0,0)_{12}$ errors is

$$\begin{aligned}\nabla\nabla_{12}\hat{X}_{t+1} &= -0.3905\nabla\hat{V}_{Stibor,t+1} - 0.0611\hat{V}_{CCI,t+1} - 0.0161\nabla\nabla_{12}\hat{V}_{GDP,t+1} + 0.4526\eta_t \\ &\quad - 0.4295\eta_{t-11} + 0.1944\eta_{t-12} + \hat{Z}_{t+1}.\end{aligned}$$

When using the actual values of Stibor 1-month, CCI and GDP, the forecast for January 2021 was

$$\begin{aligned}\nabla\nabla_{12}\hat{X}_{Jan\,2021} &= -0.3905\nabla\hat{V}_{Stibor,Jan\,2021} - 0.0611\hat{V}_{CCI,Jan\,2021} - 0.0161\nabla\nabla_{12}\hat{V}_{GDP,Jan\,2021} \\ &\quad + 0.4526\eta_{Dec\,2020} - 0.4295\eta_{Jan\,2020} + 0.1944\eta_{Dec\,2019} \\ &= -0.3905\cdot 0.0539 - 0.0611\cdot 2.7 - 0.0161\cdot 10.4454 + 0.4526\cdot 1.0086 \\ &\quad - 0.4295\cdot 0.0067 + 0.1944(-1.2460) = -0.1424.\end{aligned}$$

All the observed values in comparison of model forecasts are shown in Figure 38.

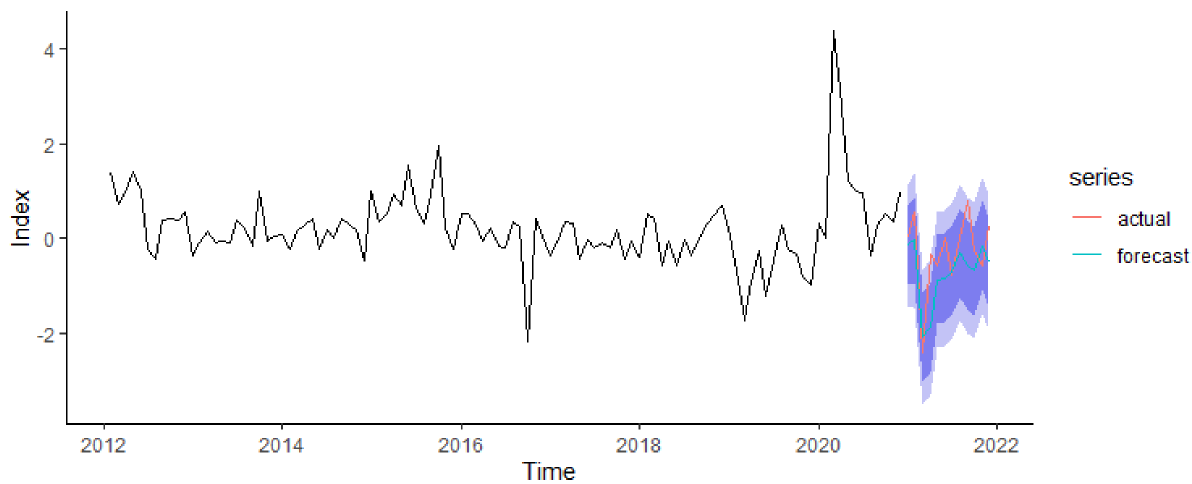

Figure 38: The actual differenced savings deposits volumes compared to the forecasts from linear regression with Stibor 1-month, CCI, GDP and ARIMA$(1,0,0)(1,0,0)_{12}$ errors, 80% and 95% CI

After transforming the differenced savings volumes back to original values, the forecast for January 2021 was

$$\begin{aligned}\hat{X}_{Jan\,2021} &= \nabla\nabla_{12}\hat{X}_{Jan\,2021} + X_{Dec\,2020} + X_{Jan\,2020} - X_{Dec\,2019} \\ &= -0.1424 + 180.6125 + 163.3414 - 162.5998 = 181.2117.\end{aligned}$$

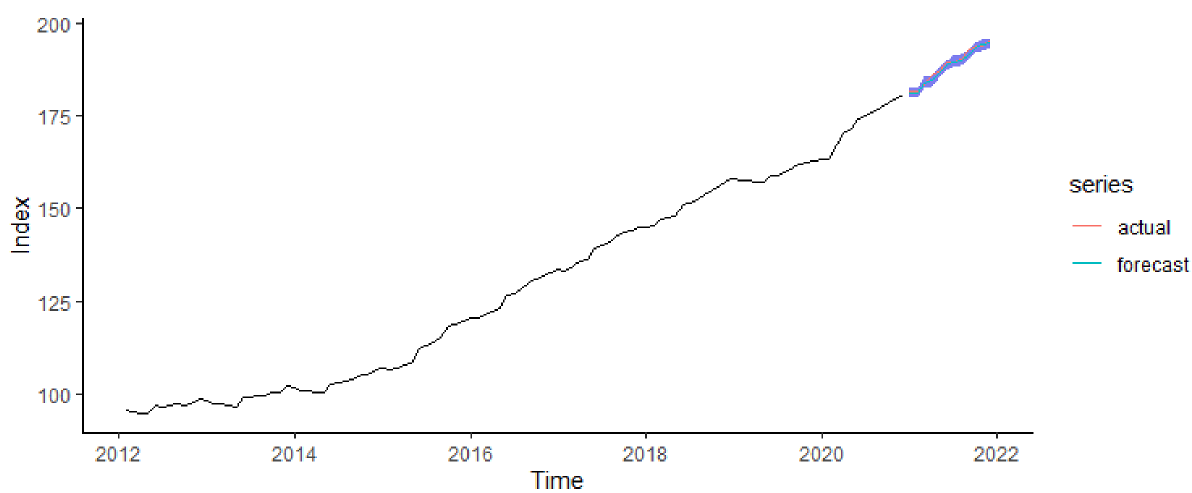

Figure 39: The actual savings deposits volumes compared to the converted forecasts from linear regression with Stibor 1-month, CCI, GDP and ARIMA$(1,0,0)(1,0,0)_{12}$ errors

PMI and GDP

The forecasting formula for linear regression with PMI, GDP and ARIMA$(1,0,0)(0,0,1)_{12}$ errors is

$$\nabla\nabla_{12}\hat{X}_{t+1} = 0.0028\hat{V}_{PMI,t+1} - 0.0135\nabla\nabla_{12}\hat{V}_{GDP,t+1} + 0.4457\eta_t + \hat{Z}_{t+1} - 0.5597Z_{t-11}.$$

In case the actual values of PMI services and GDP were used, the point forecast for January 2021 was

$$\begin{aligned}\nabla\nabla_{12}\hat{X}_{Jan\ 2021} &= 0.0028V_{PMI,Jan\ 2021} - 0.0135\nabla\nabla_{12}V_{GDP,Jan\ 2021} + 0.4457\eta_{Dec\ 2020} \\ &\quad - 0.5597Z_{Jan\ 2020} \\ &= 0.0028 \cdot 63.0 - 0.0135 \cdot 10.4454 + 0.4457 \cdot 0.8032 - 0.5597 \cdot 0.2998 \\ &= 0.2233.\end{aligned}$$

All the observed values in comparison of model forecasts are shown in Figure 40.

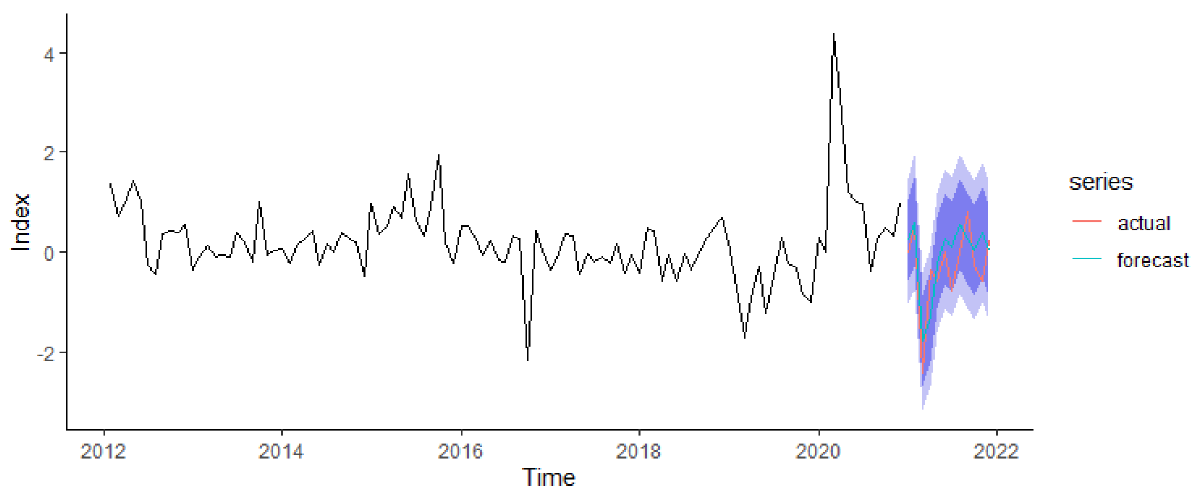

Figure 40: The actual differenced savings deposits volumes compared to the forecasts from linear regression with PMI services, GDP and ARIMA$(1,0,0)(0,0,1)_{12}$ errors, 80% and 95% CI

By converting the differenced savings volumes back to original, the prediction for January 2021 was

$$\hat{X}_{Jan\ 2021} = \nabla\nabla_{12}\hat{X}_{Jan\ 2021} + X_{Dec\ 2020} + X_{Jan\ 2020} - X_{Dec\ 2019}$$
$$= 0.2233 + 180.6125 + 163.3414 - 162.5998 = 181.5774.$$

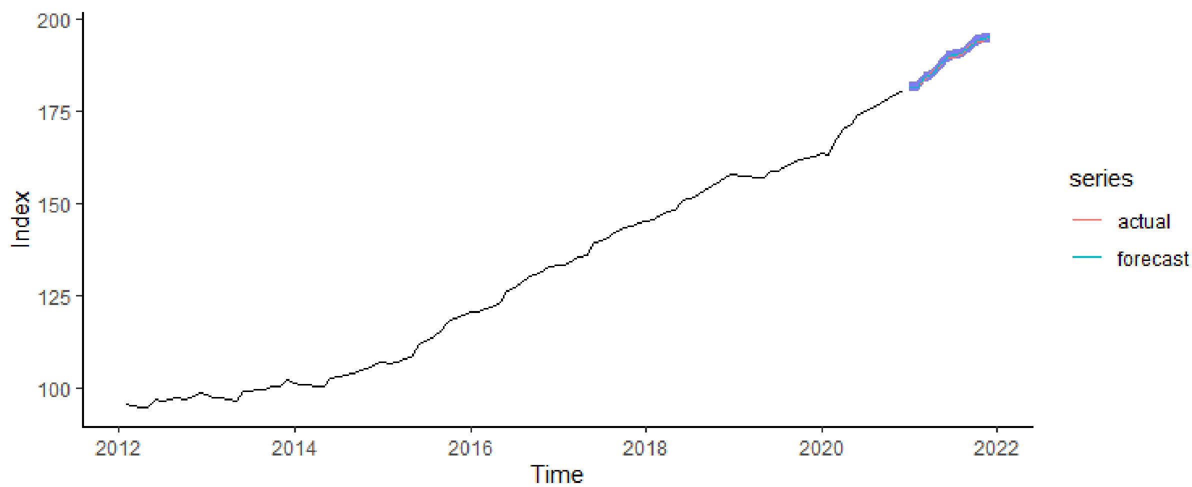

Figure 41: The actual savings deposits volumes compared to the converted forecasts from linear regression with PMI, GDP and ARIMA$(1{,}0{,}0)(0{,}0{,}1)_{12}$ errors

3.4 Results

In this paragraph the results presented in subsections 3.2-3.3 are summarized.

To evaluate forecasting accuracy, it is important to compare model performance on completely unseen data, i.e. the test set. Overall, the validations in paragraph 3.3 imply that all the models were performing well during the test period and that time series models are appropriate for forecasting deposit volumes. To find the best models, error measures such as

$$MAE = \frac{1}{h}\sum_{t}^{h}|\hat{x}_t - x_t|$$

and

$$RMSE = \sqrt{\frac{1}{h}\sum_{t}^{h}(\hat{x}_t - x_t)^2}$$

are used for measuring the accuracy of model forecasts over the test period. MAE measures the absolute magnitude of errors, while $RMSE$ captures the mean square magnitude of errors and gives more weight to large deviations such as outliers. The results are presented in Table 42.

Table 42: Error measures for test dataset

Model		$\boldsymbol{MAE(\widehat{X}_{test})}$	$\boldsymbol{RMSE(\widehat{X}_{test})}$
Holt		0.590	0.738
ARIMA		0.381	0.458
ARIMAX	Stibor	0.488	0.545
	CCI	0.599	0.725
	PMI	0.488	0.553
	GDP	0.641	0.798
	HPI	1.148	1.278
	QE	0.596	0.757
	Stibor, CCI, GDP	0.595	0.742
	PMI, GDP	0.490	0.575

By looking at the error metrics, it turns out that the ARIMA model performs the best with MAE of 0.381 and RMSE of 0.458. One could expect that adding explanatory variables to a model would increase the accuracy, i.e. lower error measures. However, this is not the case in this study. ARIMAX models with the best accuracy include Stibor 1-month, purchasing managers' index, and combination of purchasing managers' index and gross domestic product. All three models have similar error measures with MAE between 0.488-0.490 and RMSE between 0.545-0.575. The measures are around 20-30% higher than for the ARIMA model.

The least performing model is ARIMAX with house price index with considerably higher MAE (1.1487) and RMSE (1.278). Moreover, the ARIMAX models fitted in paragraph 3.2 are more complex compared to the ARIMA model from a perspective of having the necessity of forecasting explanatory variables first. On the other hand, using transfer function model instead of ARIMAX would eliminate the need. In some cases, the impact of a predictor in a regression model is not straightforward and immediate, and using a model which allows lagged effects could potentially give better results.

Altogether, ARIMA model was performing very good during the time scope of this thesis when economy was rising. During economic depression, the model might act differently and using exogenous variables could be useful. Therefore, additional investigation on possible predictor variables or concentration on ARIMAX models with combination of several predictors is recommended. As well, investigation on transfer function models and machine learning models could be considered.

Conclusion

With increasing regulatory supervision, financial institutions are required to carefully monitor their liquidity risk. Forecasting deposit volumes allows financial institutions to improve their risk assessment and allocate funds more efficiently with regards to maturity.

The objective of the thesis was to conduct an empirical study to find the best model to forecast the volume of savings deposits of one undisclosed financial institution in Sweden. Time series models such as Holt, ARIMA and ARIMAX were applied to the deposit volumes. The explanatory variables included in ARIMAX models were Stibor 1-month, consumer confidence indicator, purchasing managers' index, gross domestic product, house price index and quantitative easing. Results imply that ARIMA$(1,1,0)(0,1,1)_{12}$ model performs the best, resulting in approximately 20% lower errors measures compared to the best three ARIMAX models – Stibor 1-month, purchasing managers' index, and combination of purchasing managers' index and gross domestic product.

The thesis has provided insight to how accurate some macroeconomic measures are in forecasting savings volumes. As ARIMAX models were not showing better accuracy than regular ARIMA model, then further research could include additional variables or focus more on ARIMAX models with combination of several explanatory variables. Secondly, investigation on transfer function models or more advanced modelling techniques, like machine learning models, could be considered. It would be interesting to see if the accuracy is higher compared to the time series models proposed and investigated in this study. In addition, future study could analyse corporate customers' savings deposits. This part of the portfolio is presumed to act differently to the changes in the market and economic variables.

References

Bank of Japan (2014). *Survey on Core Deposit Modeling in Japan: Toward Enhancing Asset Liability Management*. Financial System and Bank Examination Department. URL: https://www.boj.or.jp/en/research/brp/ron_2014/data/ron140331a.pdf

Basel Committee on Banking Supervision (2016). *Standards: Interest Rate Risk in the Banking Book*. URL: https://www.bis.org/bcbs/publ/d368.pdf

Brockwell, P. J. and David, R. A. (2002). *Introduction to Time Series and Forecasting. (2nd ed)*. New York: Springer.

Carroll, C. D., and Samwick, A. A. (1998). *How Important Is Precautionary Saving?*. Review of Economics and Statistics, 80(3), 410-419.

Carroll, C., Slacalek, J. and Sommer, M. (2012). *Dissecting Saving Dynamics: Measuring Wealth, Precautionary, and Credit Effects*. Working Paper Series No. 1474, European Central Bank. URL: https://www.ecb.europa.eu/pub/pdf/scpwps/ecbwp1474.pdf

Chatfield, C. (2004). *The Analysis of Time Series: An Introduction (6th ed)*. Chapman & Hall. London.

Chaturvedi, V., Brajesh, K. and Dholakia, R. H. (2009). *Inter-Relationship Between Economic Growth, Savings and Inflation in Asia*. Journal of International Economic Studies, No.23, 1–22.

Directorate General for Economic and Financial Affairs (2016). *The Joint Harmonised EU Programme of Business and Consumer Surveys - User Guide*. URL: https://ec.europa.eu/economy_finance/db_indicators/surveys/documents/bcs_user_guide_en.pdf

Eurostat (2022a). *House price index (2015 = 100) – quarterly data*. Retrieved on March 6, 2022, from https://ec.europa.eu/eurostat/databrowser/view/PRC_HPI_Q__custom_2220583/default/table?lang=en

Eurostat (2022b). *Sentiment indicators - monthly data*. Retrieved on January 11, 2022, from https://ec.europa.eu/eurostat/databrowser/view/EI_BSSI_M_R2/default/table?lang=en

Grimaldi, M. B., Crosta, A. and Zhang, D. (2021). *The Liquidity of the Government Bond Market – What Impact Does Quantitative Easing Have? Evidence from Sweden.* Sveriges Riksbank Working Paper Series No. 402, 2021. URL: https://www.riksbank.se/globalassets/media/rapporter/working-papers/2019/no.-381--how-much-information-do-monetary-policy-committees-disclose-evidence-from-the-fomcs-minutes-and-transcripts.pdf

Hyndman, R.J., and Athanasopoulos, G. (2018) *Forecasting: principles and practice, 2nd edition*. OTexts: Melbourne, Australia. URL: www.OTexts.com/fpp2

Kangro, R. (2016). *Aegridade Analüüs* [Course material]. MTMS.01.023 Time Series Analysis. University of Tartu.

Kennedy, N. and Andersen, P. S. (1994). *Household Saving and Real House Prices: An International Perspective*. BIS Working Paper No. 20. URL: https://www.bis.org/publ/work20.pdf

Klerck, H. and Salame, D (2017). *The Impact of House Price Changes on Household Savings*.

Kłopocka, A. M. (2016). *Does Consumer Confidence Forecast Household Saving and Borrowing Behavior? Evidence for Poland.* Springer. URL: https://d-nb.info/1103344269/34

Masson, P. R., T. Bayoumi, and H. Samiei (1998). *International Evidence on the Determinants of Private Saving*. The World Bank Economic Review, Vol. 12, No. 3: 483-501.

Mody, A., Ohnsorge, F. and Sandri, D. (2012). *Precautionary Savings in the Great Recession*. IMF Working Paper. URL: https://www.imf.org/external/pubs/ft/wp/2012/wp1242.pdf

Phillips, P. C. B. and Perron, P. (1988). *Testing for a Unit Root in Time Series Regression*. Biometrika, Vol. 75, No. 2: 335–346.

Statistics Sweden (2022). *GDP Quarterly 1993–2022:4 (publ: 2022-02-28)*. Retrieved on March 21, 2022, from https://www.scb.se/en/finding-statistics/statistics-by-subjectarea/national-accounts/national-accounts/national-accounts-quarterly-and-annual-estimates/pong/tables-and-graphs/tables/gdp-quarterly/

Sveriges Riksbank (2021a). *Purchases of Government Bonds*. URL: https://www.riksbank.se/en-gb/monetary-policy/monetary-policy-instruments/purchases-of-government-bonds/

Sveriges Riksbank (2022a). *Monetary Policy Report, September 2021*. Retrieved on February 21, 2022, from https://www.riksbank.se/en-gb/monetary-policy/monetary-policy-report/2022/monetary-policy-report-february-2022/

Sveriges Riskbank (2022b). *Search interest & exchange rates*. Retrieved on January 25, 2022, from https://www.riksbank.se/en-gb/statistics/search-interest--exchange-rates/

Swedbank (2022). *Purchasing Manager's Index – Manufacturing.* URL: https://www.swedbank.com/corporate/swedbank-research/purchasing-managers-index-manufacturing.html

Swedish Financial Benchmark Facility (2022). *STIBOR Rates*. Retrieved on January 25, 2022, from https://swfbf.se/sfbf-benchmarks/rates/

Appendixes

Appendix 1. Selection of ARIMA model

The following figures and R outputs illustrate how the best ARIMA model was chosen.

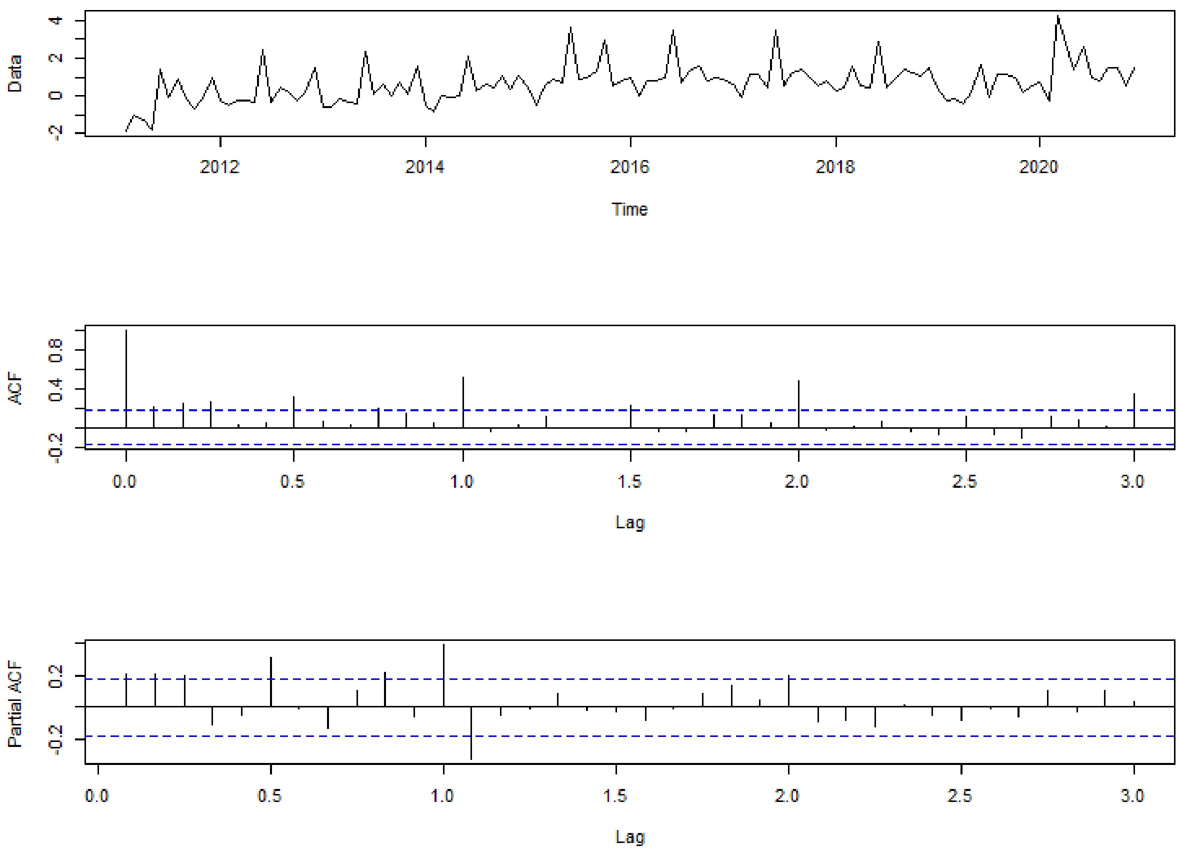

Figure X: First order differenced savings deposits, ACF and PACF plots

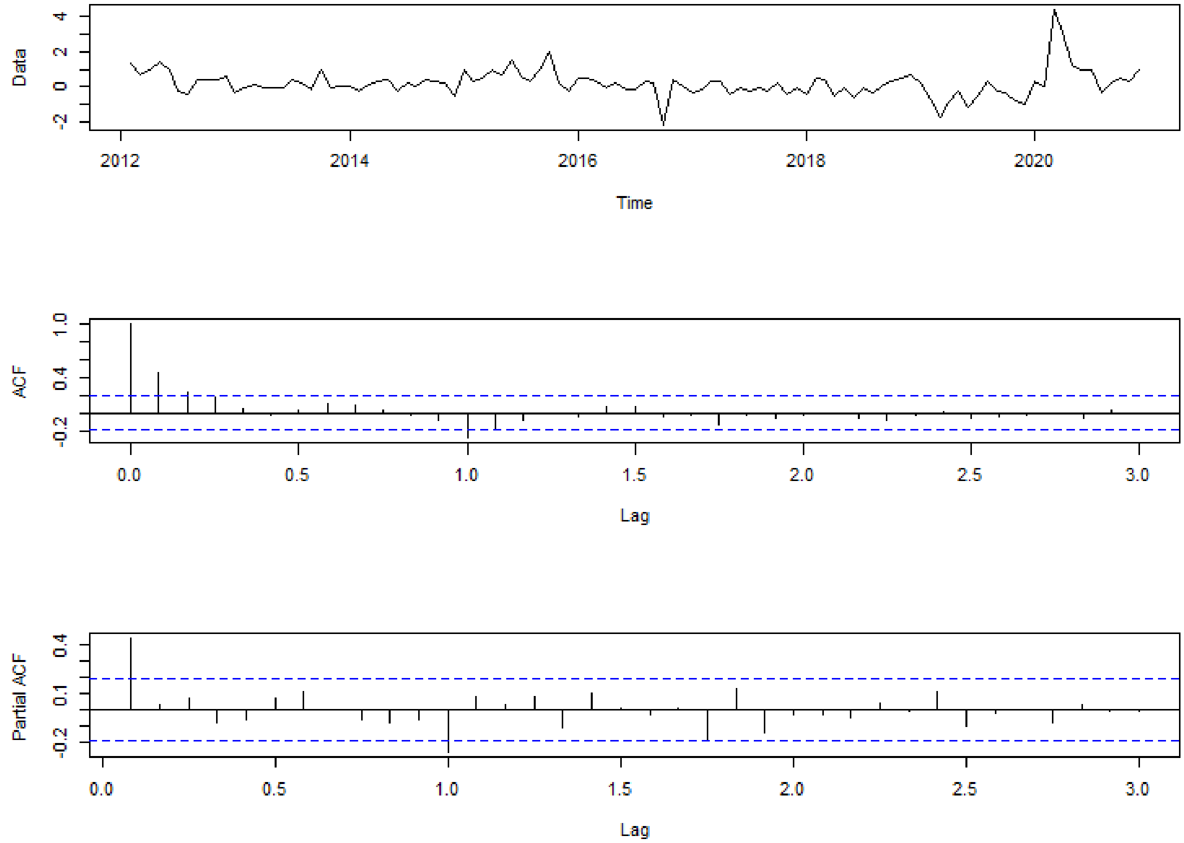

Figure X: First order and seasonally differenced savings deposits, ACF and PACF plots

R output of ARIMA$(1,1,0)(0,1,1)_{12}$ model:

```
Series: savings_ts
ARIMA(1,1,0)(0,1,1)[12]

Coefficients:
         ar1     sma1
      0.5227  -0.5023
s.e.  0.0850   0.1126

sigma^2 estimated as 0.4325:  log likelihood=-107.73
AIC=221.46   AICc=221.69   BIC=229.48
```

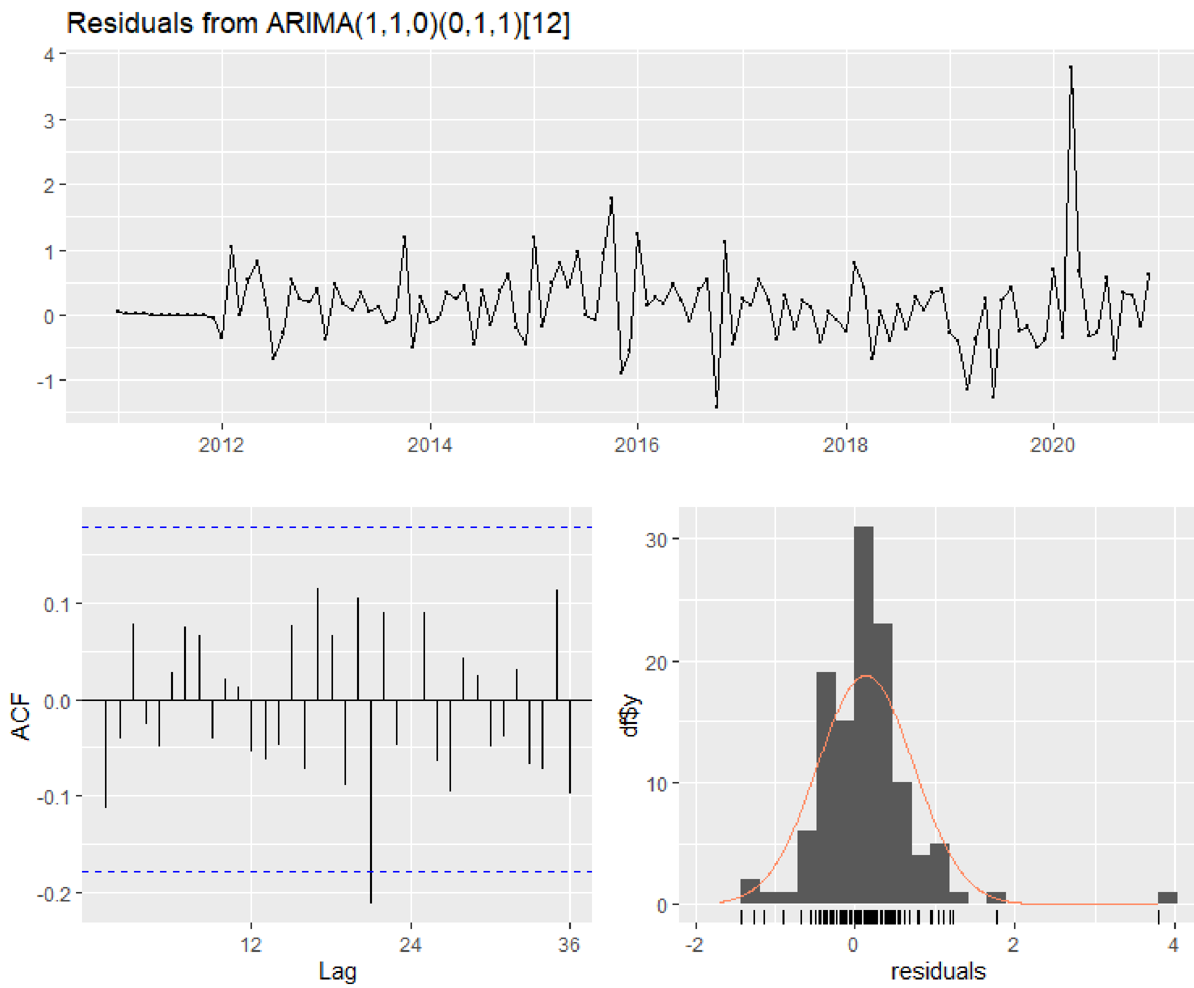

Figure X: Residuals from ARIMA$(1,1,0)(0,1,1)_{12}$

Appendix 2. Selection of ARIMAX model with Stibor 1-month

The following figures and R outputs illustrate how the best ARIMAX model was chosen.

R output of linear regression with Stibor 1-month:

```
Call:
tslm(formula = diff(diff(savings_ts, 12)) ~ window(diff(rates_ts),
    start = c(2012, 2)) - 1)

Residuals:
    Min      1Q  Median      3Q     Max
-2.1296 -0.2165  0.1403  0.4762  4.4610

Coefficients:
                                         Estimate Std. Error t value Pr(>|t|)
window(diff(rates_ts), start = c(2012, 2))  -1.3339     0.8653  -1.542    0.126

Residual standard error: 0.8102 on 106 degrees of freedom
Multiple R-squared:  0.02193,   Adjusted R-squared:  0.0127
F-statistic: 2.376 on 1 and 106 DF,  p-value: 0.1262
```

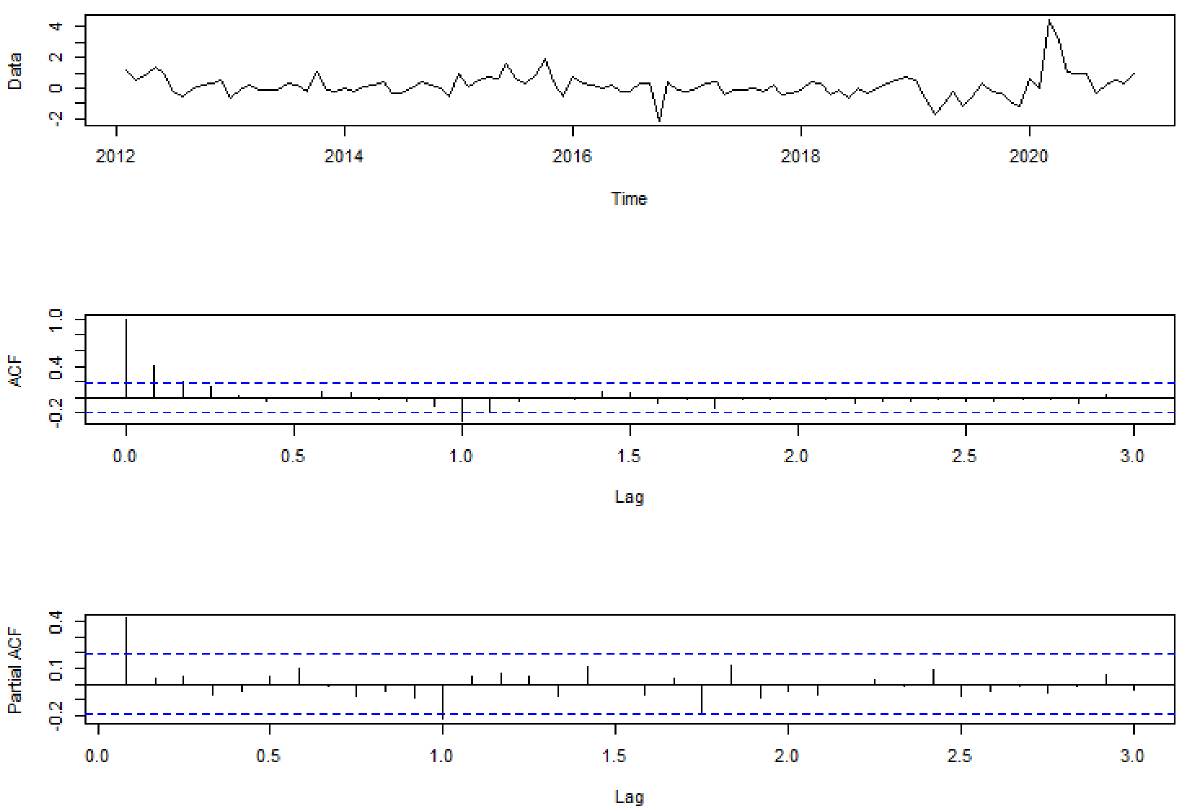

Figure X: Residuals from linear regression model with Stibor 1-month, ACF and PACF plots

R output of linear regression with Stibor 1-month and ARIMA errors:

```
Series: diff(diff(savings_ts, 12))
Regression with ARIMA(1,0,2)(0,0,1)[12] errors

Coefficients:
         ar1      ma1      ma2     sma1    xreg
      0.9797  -0.5213  -0.2686  -0.6456  0.4574
s.e.  0.0380   0.1065   0.1083   0.1240  0.4254

sigma^2 estimated as 0.4128:  log likelihood=-105.12
AIC=222.24   AICc=223.08   BIC=238.27
```

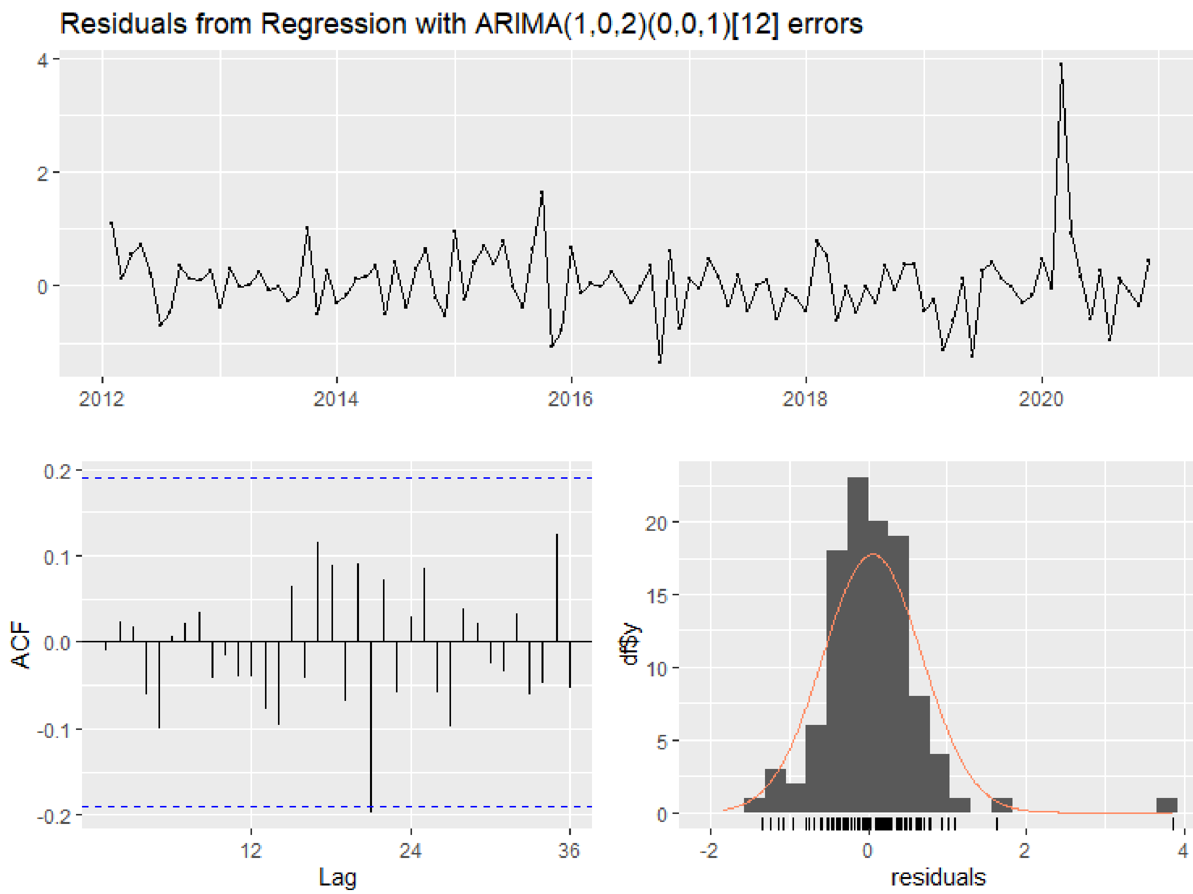

Figure X: Residuals from linear regression model with Stibor 1-month and ARIMA$(1,0,2)(0,0,1)_{12}$ errors

Appendix 3. Selection of ARIMAX model with CCI

The following figures and R outputs illustrate how the best ARIMAX model was chosen.

R output of linear regression with consumer confidence indicator:

```
Call:
tslm(formula = diff(diff(savings_ts, 12)) ~ window(cci_ts, start = c(2012,
    2)) - 1)

Residuals:
    Min      1Q  Median      3Q     Max
-2.1054 -0.1866  0.1478  0.4771  3.9148

Coefficients:
                                 Estimate Std. Error t value Pr(>|t|)
window(cci_ts, start = c(2012, 2)) -0.09933    0.03358  -2.958  0.00382 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7874 on 106 degrees of freedom
Multiple R-squared:  0.07627,   Adjusted R-squared:  0.06755
F-statistic: 8.752 on 1 and 106 DF,  p-value: 0.003815
```

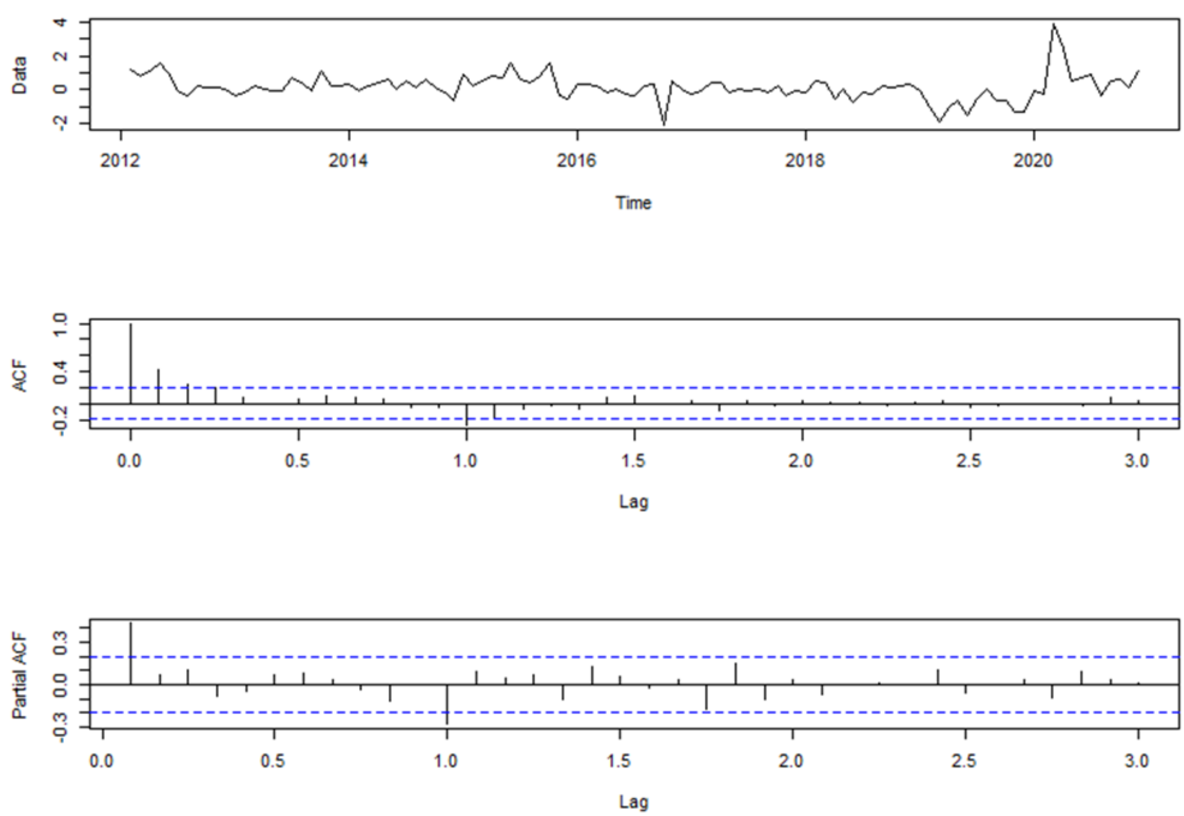

Figure X: Residuals from linear regression model with consumer confidence index, ACF and PACF plots

R output of linear regression with consumer confidence indicator and ARIMA errors:

```
Series: diff(diff(savings_ts, 12))
Regression with ARIMA(1,0,0)(1,0,0)[12] errors

Coefficients:
         ar1     sar1     xreg
      0.4712  -0.4265  -0.0688
s.e.  0.0860   0.1096   0.0357

sigma^2 estimated as 0.4316:  log likelihood=-106.68
AIC=221.35   AICc=221.74   BIC=232.04
```

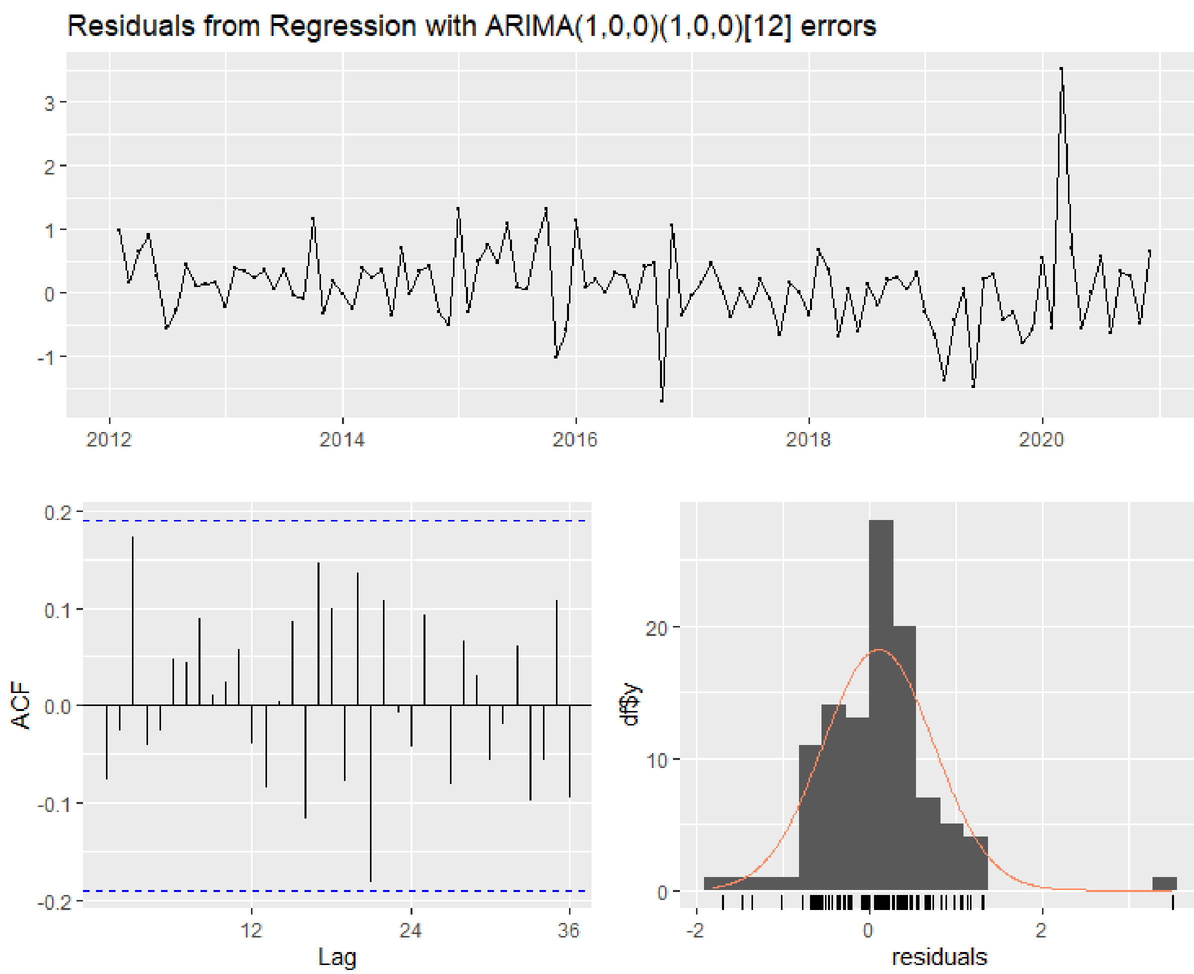

Figure X: Residuals from linear regression model with CCI and ARIMA$(1,0,0)(1,0,0)_{12}$ errors

Appendix 4. Selection of ARIMAX model with PMI

The following figures and R outputs illustrate how the best ARIMAX model was chosen.

R output of linear regression with PMI manufacturing:

```
Call:
tslm(formula = diff(diff(savings_ts, 12)) ~ window(pmi_ts, start = c(2012,
    2)) - 1)

Residuals:
    Min      1Q  Median      3Q     Max
-2.3667 -0.3862  0.0030  0.2900  4.2604

Coefficients:
                                   Estimate Std. Error t value Pr(>|t|)
window(pmi_ts, start = c(2012, 2)) 0.003376   0.001433   2.356   0.0203 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7986 on 106 degrees of freedom
Multiple R-squared:  0.04977,   Adjusted R-squared:  0.0408
F-statistic: 5.552 on 1 and 106 DF,  p-value: 0.0203
```

R output of linear regression with PMI services:

```
Call:
tslm(formula = diff(diff(savings_ts, 12)) ~ window(pmi2_ts, start = c(2012,
    2)) - 1)

Residuals:
    Min      1Q  Median      3Q     Max
-2.3727 -0.3916  0.0082  0.2828  4.2527

Coefficients:
                                    Estimate Std. Error t value Pr(>|t|)
window(pmi2_ts, start = c(2012, 2)) 0.003341   0.001396   2.393   0.0185 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.798 on 106 degrees of freedom
Multiple R-squared:  0.05124,   Adjusted R-squared:  0.04229
F-statistic: 5.724 on 1 and 106 DF,  p-value: 0.01849
```

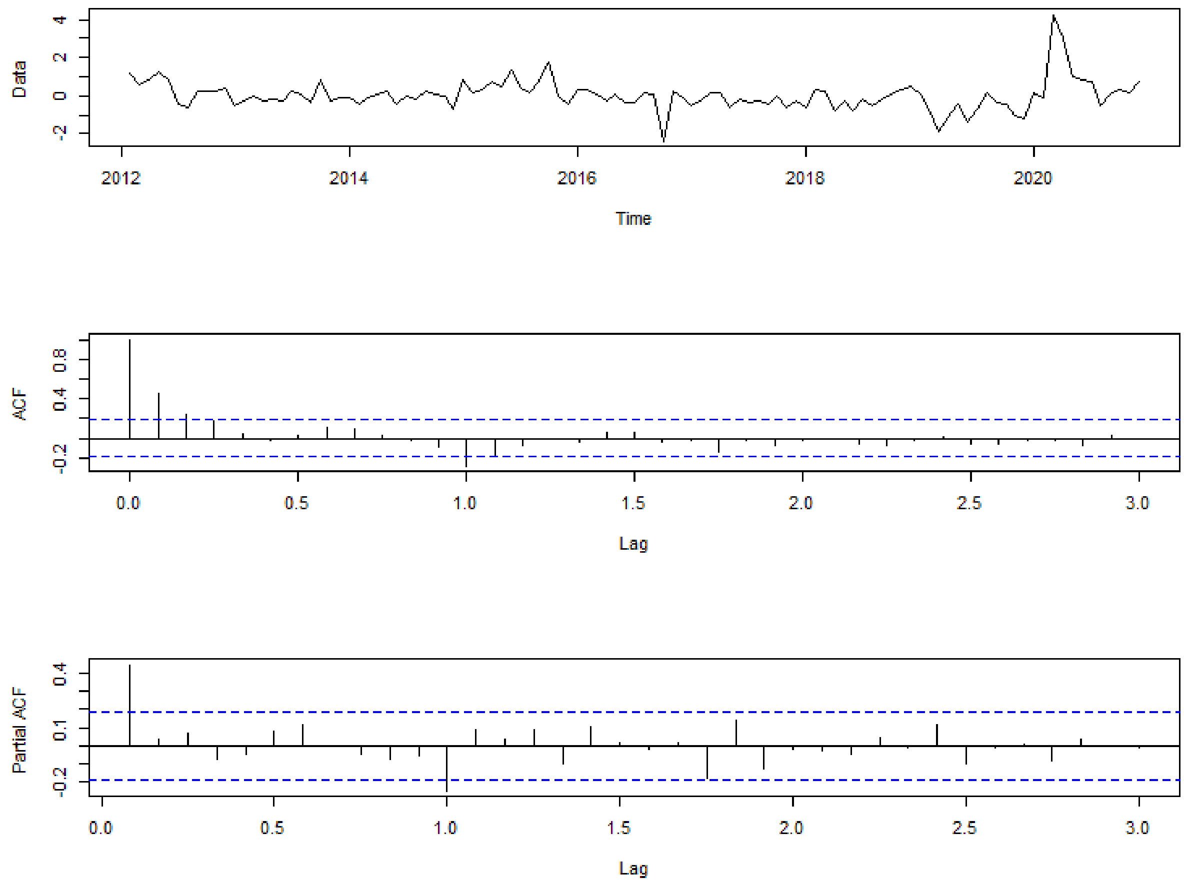

Figure X: Residuals from linear regression model with PMI services, ACF and PACF plots

R output of linear regression with PMI and ARIMA errors:

```
Series: diff(diff(savings_ts, 12))
Regression with ARIMA(1,0,0)(0,0,1)[12] errors

Coefficients:
         ar1     sma1    xreg
      0.4517  -0.5935  0.0027
s.e.  0.0865   0.1187  0.0010

sigma^2 estimated as 0.4064:  log likelihood=-104.85
AIC=217.7   AICc=218.09   BIC=228.39
```

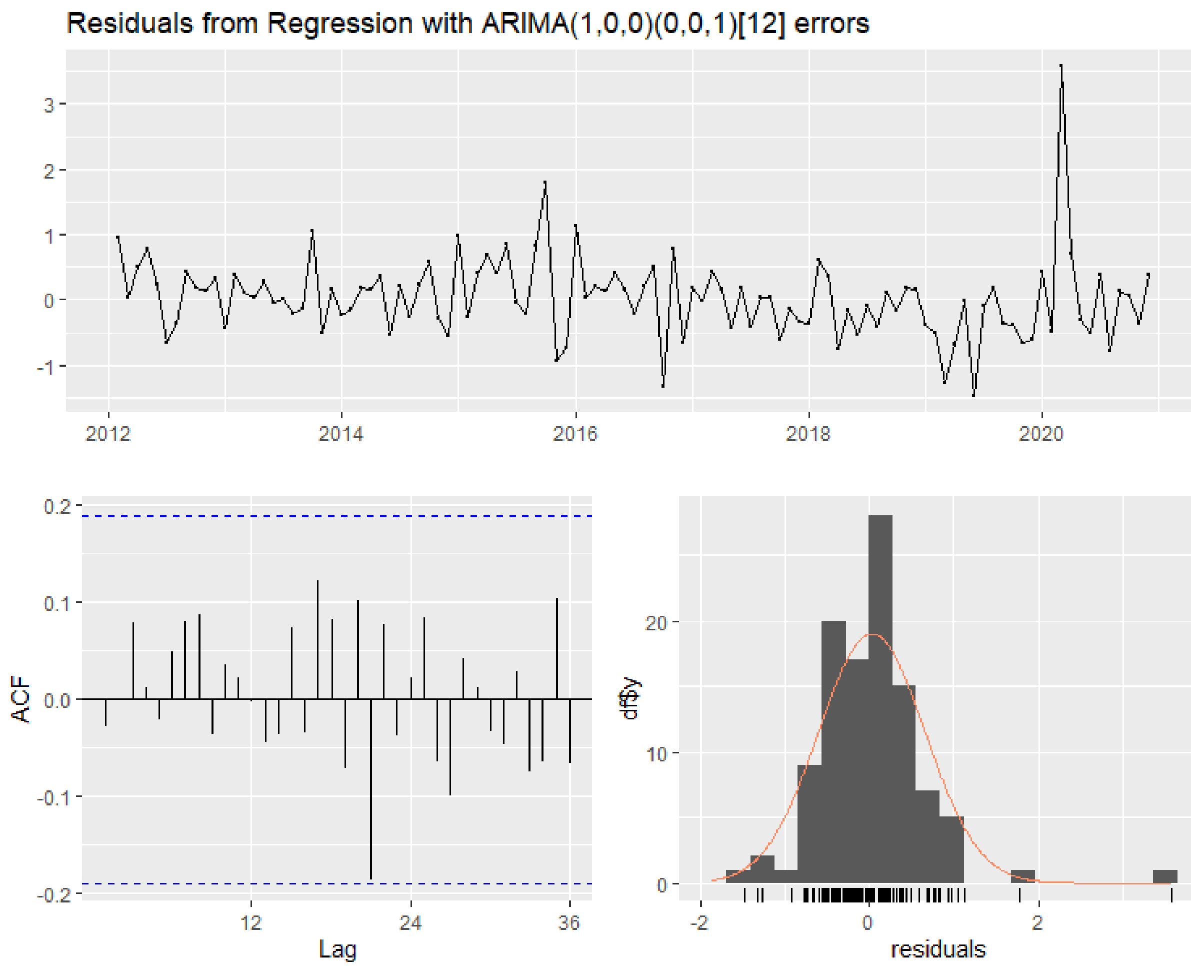

Figure X: Residuals from linear regression model with PMI services and ARIMA$(1,0,0)(0,0,1)_{12}$ errors

Appendix 5. Selection of ARIMAX model with GDP

The following figures and R outputs illustrate how the best ARIMAX model was chosen.

R output of linear regression with GDP:

```
Call:
tslm(formula = savings_ts ~ gdp_ts - 1)

Residuals:
    Min      1Q  Median      3Q     Max 
-28.825 -19.828  -6.934  15.665  54.819 

Coefficients:
       Estimate Std. Error t value Pr(>|t|)    
gdp_ts  0.32038    0.00494   64.86   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 21.39 on 119 degrees of freedom
Multiple R-squared:  0.9725,    Adjusted R-squared:  0.9723 
F-statistic:  4207 on 1 and 119 DF,  p-value: < 2.2e-16
```

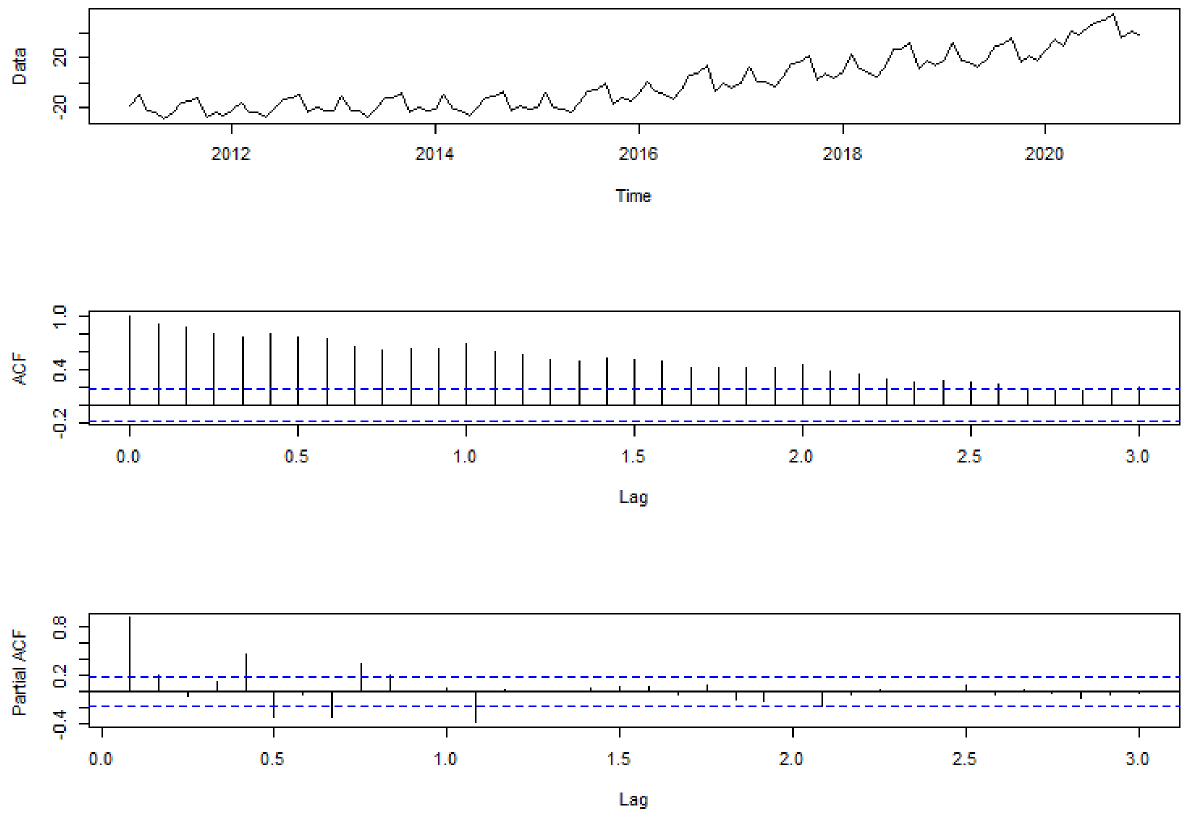

Figure X: Residuals from linear regression model with GDP, ACF and PACF plots

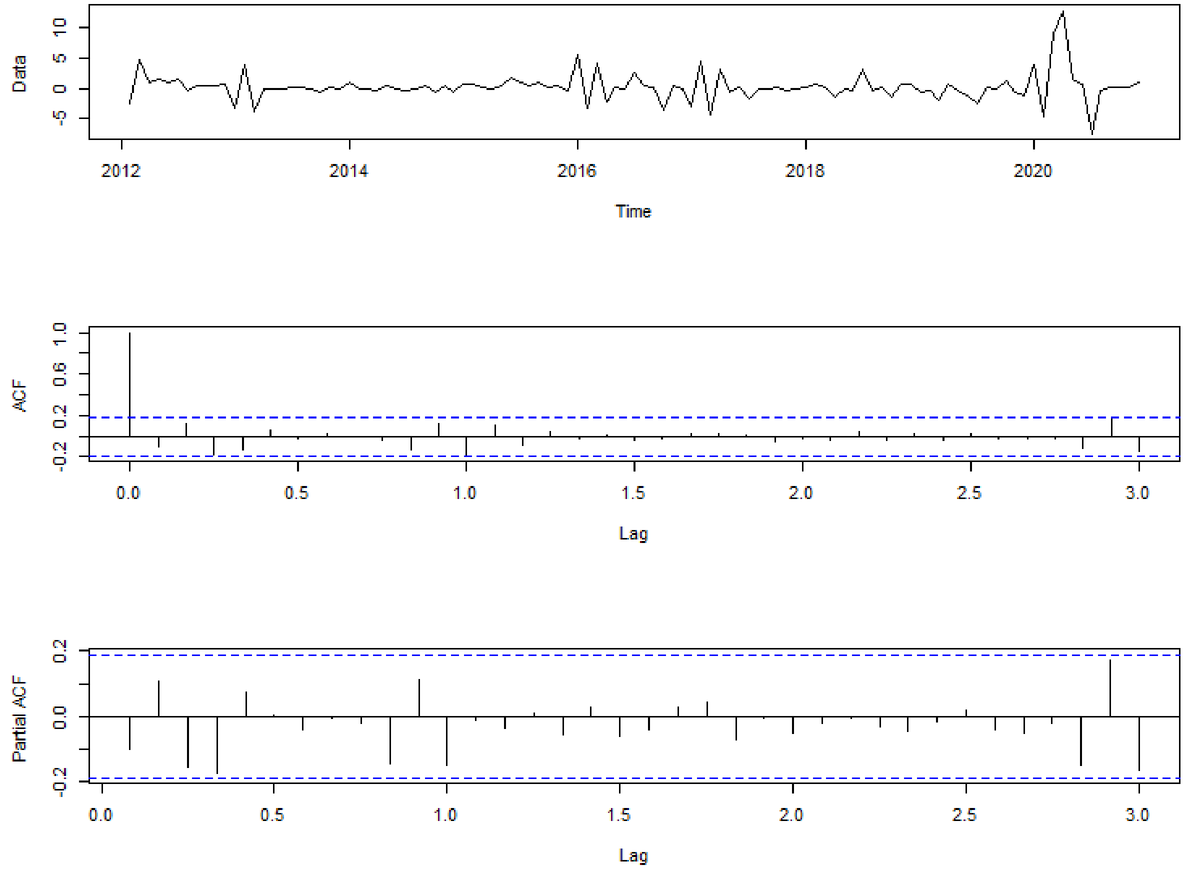

Figure X: Residuals from linear regression model with GDP after first order difference and seasonal difference, ACF and PACF plots

R output of the linear regression with GDP and ARIMA errors:

```
Series: savings_ts
Regression with ARIMA(2,1,1)(0,1,1)[12] errors

Coefficients:
         ar1      ar2      ma1     sma1     xreg
      1.3210  -0.3332  -0.8926  -0.5812  -0.0147
s.e.  0.1585   0.1420   0.1113   0.1172   0.0089

sigma^2 estimated as 0.4107:  log likelihood=-103.95
AIC=219.89   AICc=220.73   BIC=235.93
```

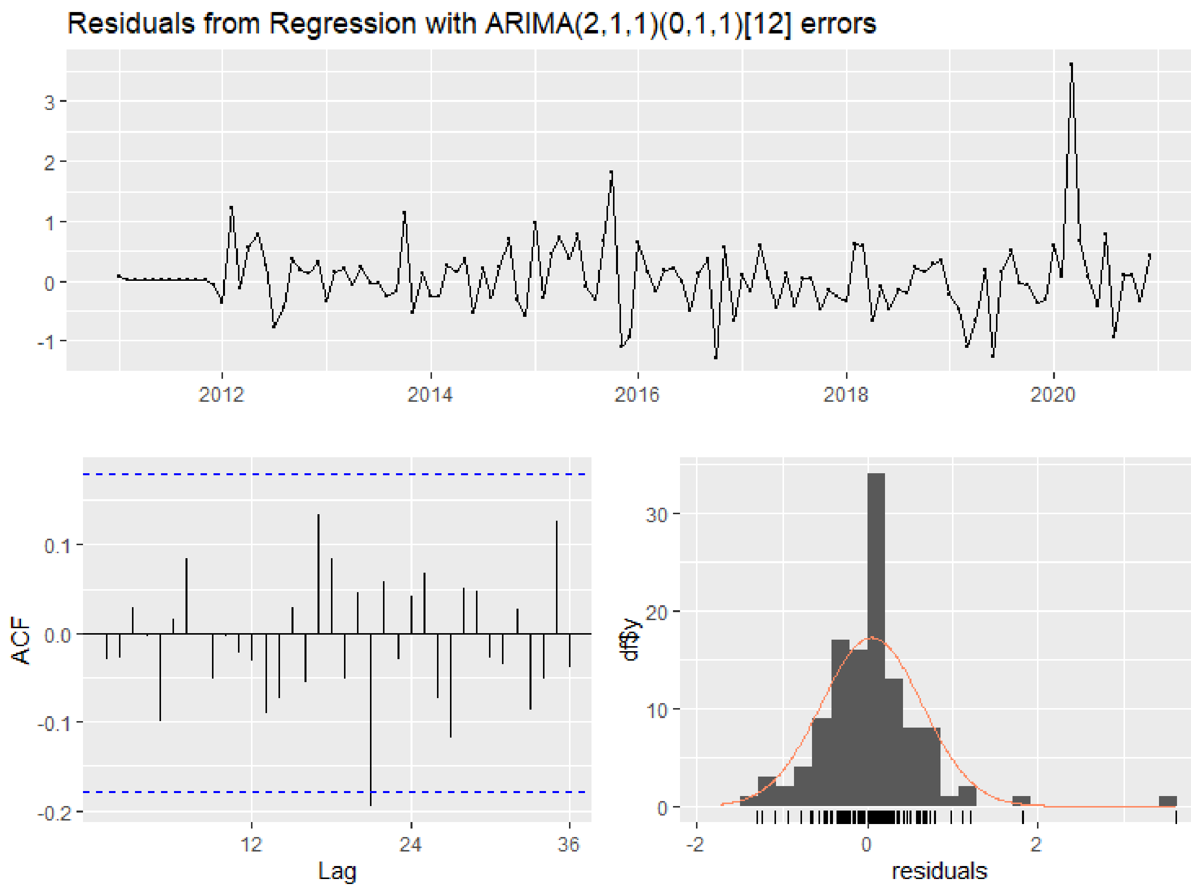

Figure X: Residuals from linear regression model with GDP and ARIMA$(2,1,1)(0,1,1)_{12}$ errors

Appendix 6. Selection of ARIMAX model with HPI

The following figures and R outputs illustrate how the best ARIMAX model was chosen.

R output of linear regression with house price index:

```
Call:
tslm(formula = savings_ts ~ hpi_ts - 1)

Residuals:
    Min      1Q  Median      3Q     Max 
-14.990  -8.774  -1.521   6.145  22.102 

Coefficients:
       Estimate Std. Error t value Pr(>|t|)    
hpi_ts 0.937990   0.006609   141.9   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.884 on 119 degrees of freedom
Multiple R-squared:  0.9941,    Adjusted R-squared:  0.9941 
F-statistic: 2.014e+04 on 1 and 119 DF,  p-value: < 2.2e-16
```

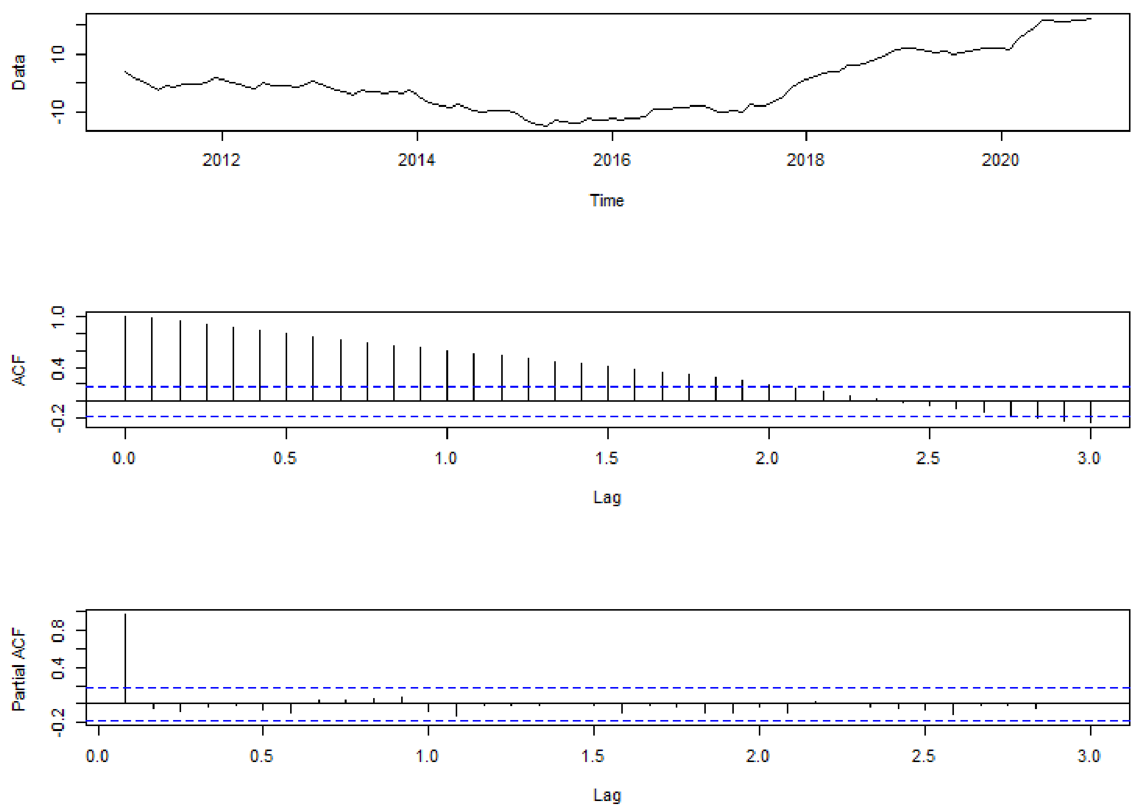

Figure X: Residuals from linear regression model with house price index, ACF and PACF plots

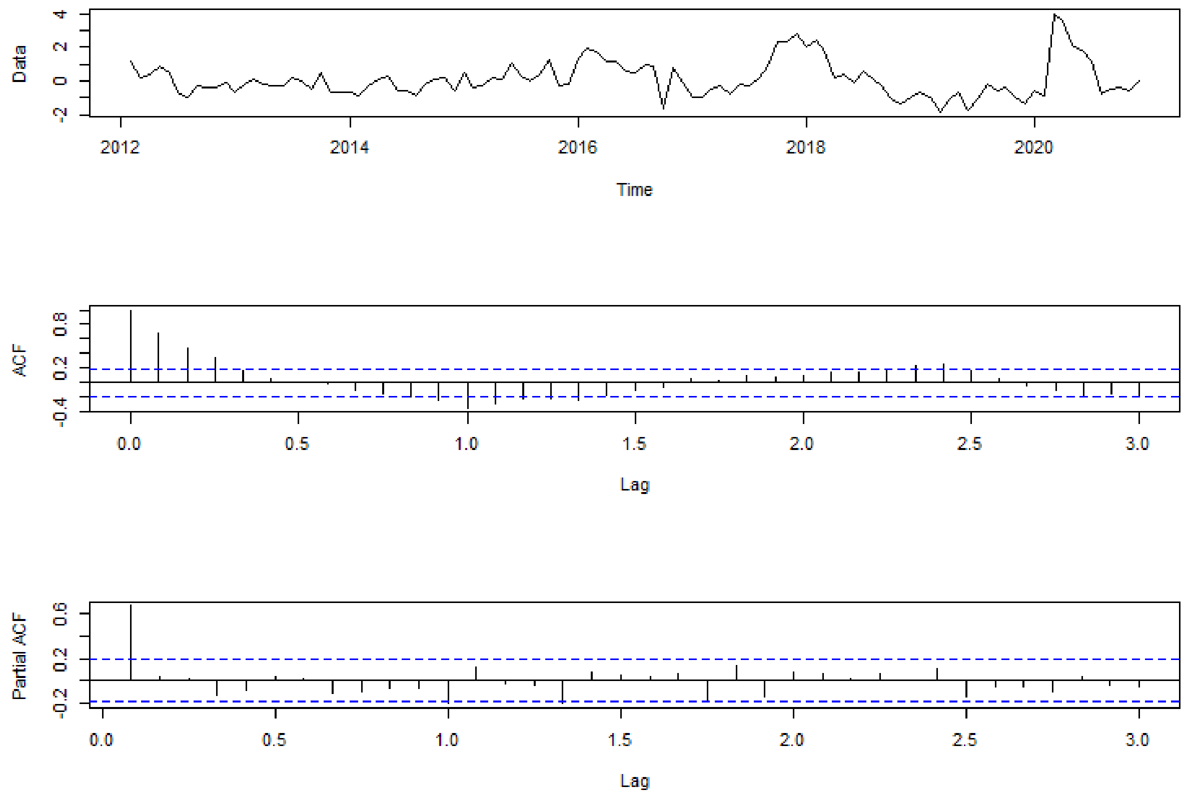

Figure X: Residuals from linear regression model with house price index after first order difference and seasonal difference, ACF and PACF plots

R output of linear regression with housing price index and ARIMA errors:

```
Series: savings_ts
Regression with ARIMA(1,1,0)(0,1,1)[12] errors

Coefficients:
         ar1     sma1    xreg
      0.4995  -0.5048  0.2141
s.e.  0.0857   0.1096  0.1564

sigma^2 estimated as 0.4289:  log likelihood=-106.83
AIC=221.67   AICc=222.06   BIC=232.36
```

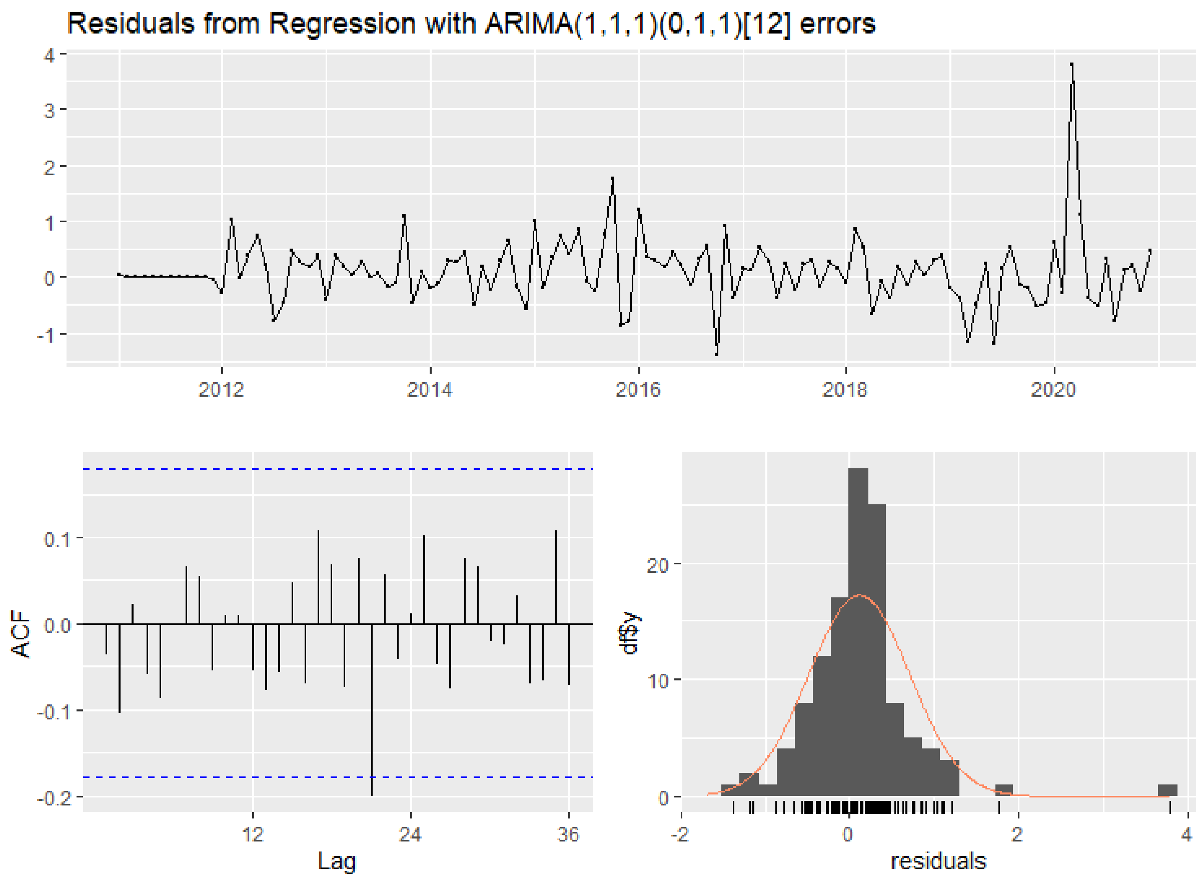

Figure X: Residuals from linear regression model with house price index and $\text{ARIMA}(1{,}1{,}0)(0{,}1{,}1)_{12}$ errors

Appendix 7. Selection of ARIMAX model with QE

The following figures and R outputs illustrate how the best ARIMAX model was chosen.

R output of linear regression with QE:

```
Call:
tslm(formula = savings_ts ~ qe_ts - 1)

Residuals:
    Min      1Q  Median      3Q     Max 
-20.287  -3.951  25.733  97.017 105.330 

Coefficients:
      Estimate Std. Error t value Pr(>|t|)    
qe_ts  1.45138    0.07756   18.71   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 64.95 on 119 degrees of freedom
Multiple R-squared:  0.7464,	Adjusted R-squared:  0.7442 
F-statistic: 350.2 on 1 and 119 DF,  p-value: < 2.2e-16
```

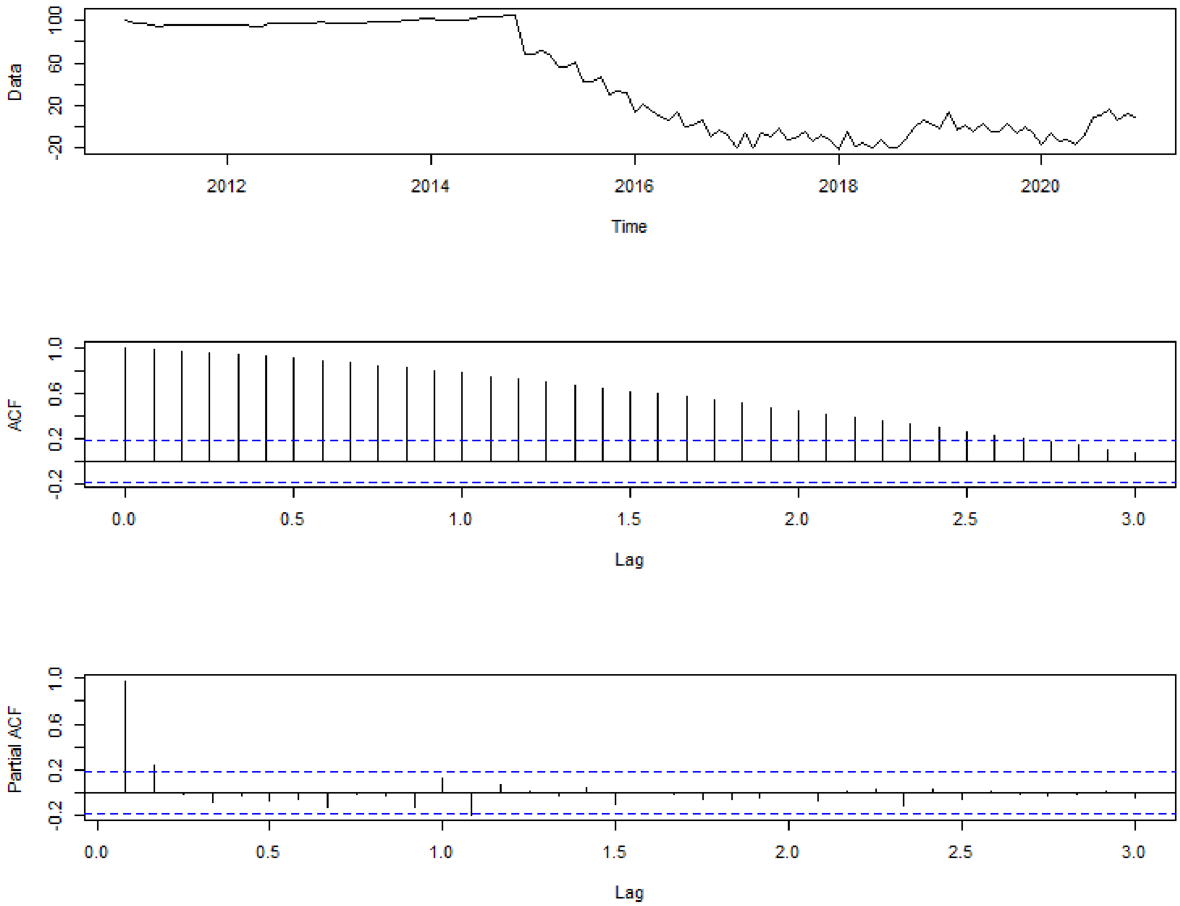

Figure X: Residuals from linear regression model with quantitative easing, ACF and PACF plots

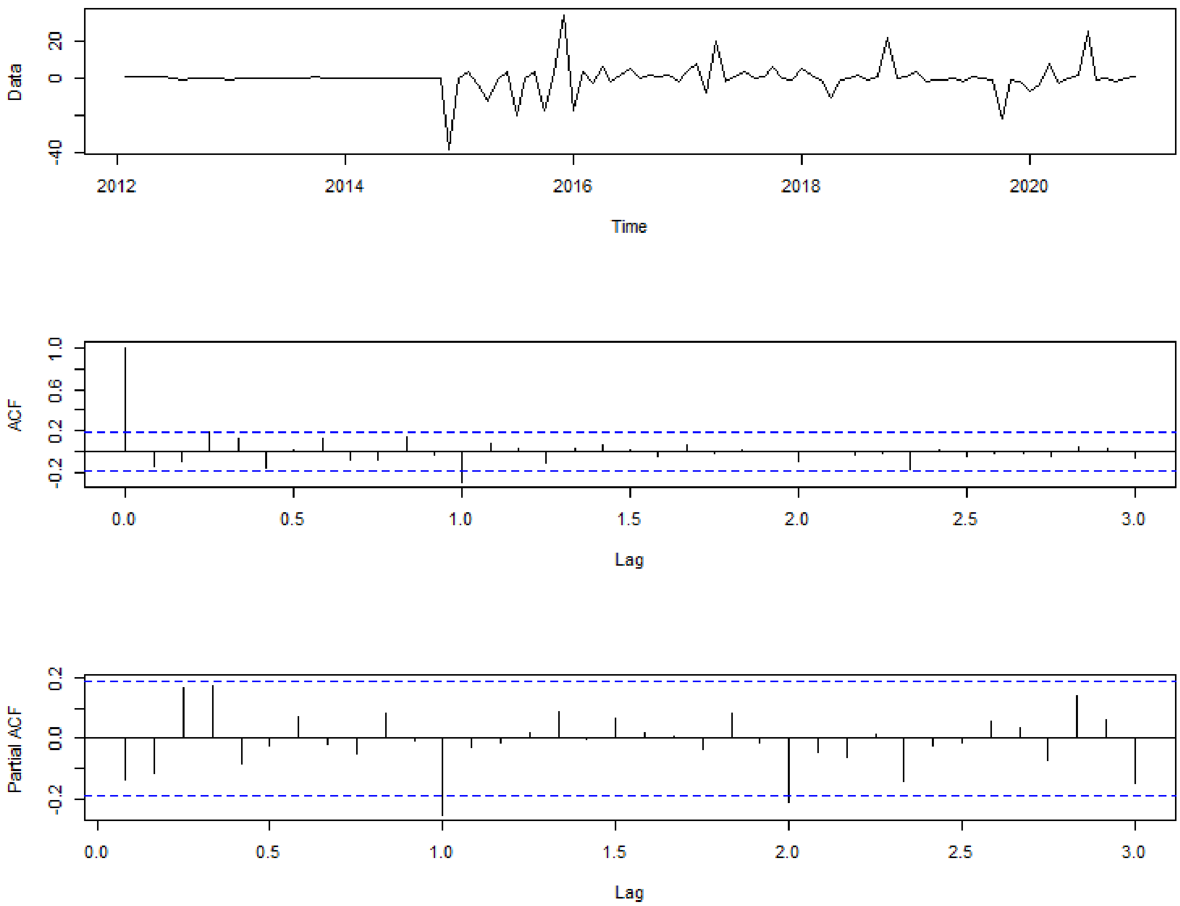

Figure X: Residuals from linear regression model with quantitative easing after first order difference and seasonal difference, ACF and PACF plots

R output of linear regression with QE and ARIMA errors:

```
Series: savings_ts
Regression with ARIMA(2,1,1)(0,1,1)[12] errors

Coefficients:
         ar1      ar2      ma1     sma1    xreg
      1.3645  -0.3714  -0.9200  -0.6267  0.0046
s.e.  0.1298   0.1222   0.0826   0.1203  0.0113

sigma^2 estimated as 0.4164:  log likelihood=-105.24
AIC=222.48   AICc=223.32   BIC=238.52
```

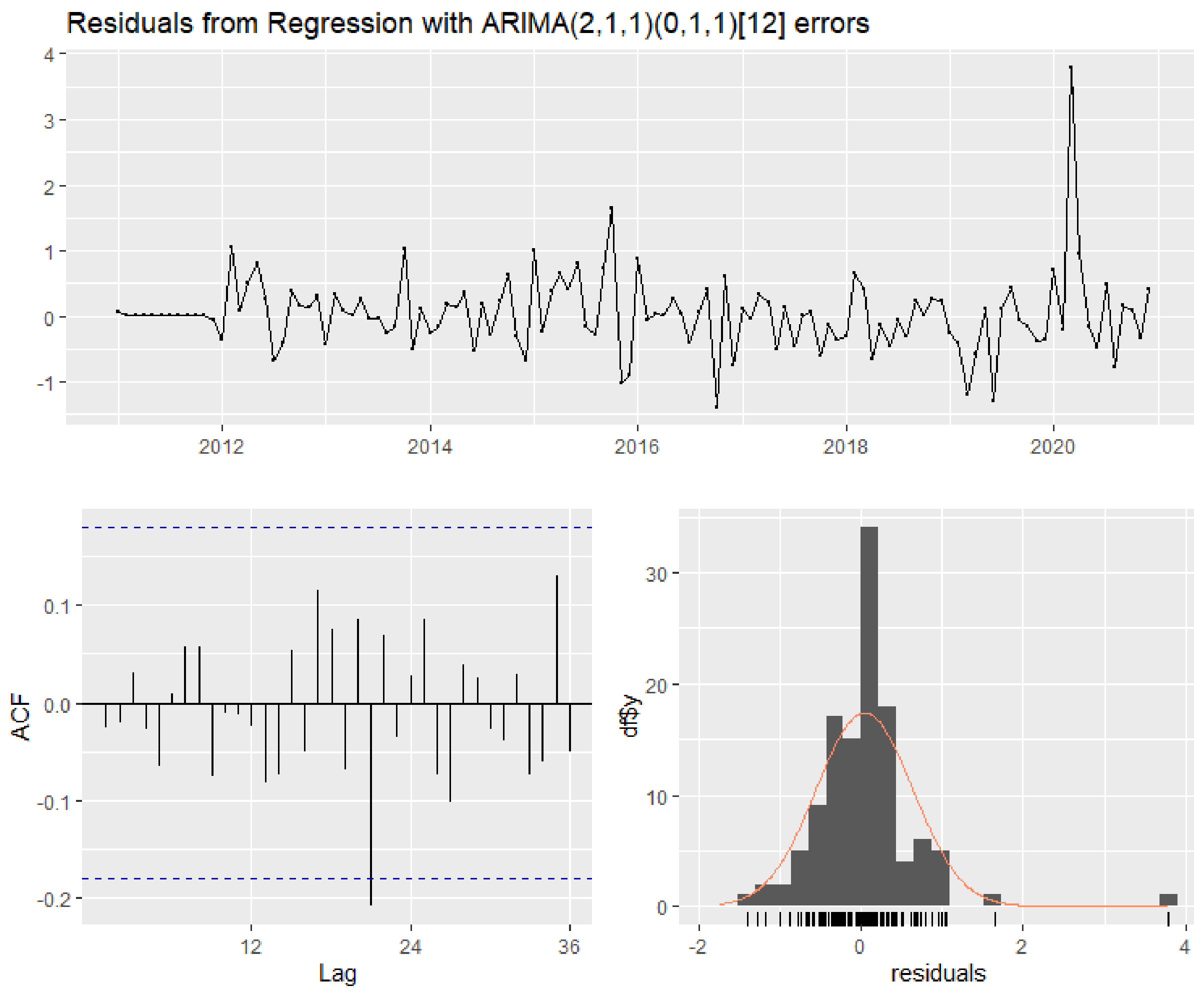

Figure X: Residuals from linear regression model with quantitative easing and ARIMA$(2,1,1)(0,1,1)_{12}$ errors

Appendix 8. Selection of ARIMAX model with best regressors

The following figures and R outputs illustrate how the best regressors and the best ARIMAX model were chosen.

R outputs of stepwise procedure to find the best regressors:

```
Call:
lm(formula = diff(diff(savings_ts, 12)) ~ window(diff(rates_ts),
    start = c(2012, 2)) + window(cci_ts, start = c(2012, 2)) +
    window(pmi2_ts, start = c(2012, 2)) + diff(diff(gdp_ts, 12)) +
    diff(diff(hpi_ts, 12)) + diff(diff(qe_ts, 12)) - 1)

Residuals:
    Min      1Q  Median      3Q     Max
-2.0211 -0.3265 -0.0721  0.4046  3.6143

Coefficients:
                                         Estimate Std. Error t value Pr(>|t|)
window(diff(rates_ts), start = c(2012, 2)) -1.276609   0.858316  -1.487   0.1400
window(cci_ts, start = c(2012, 2))        -0.074253   0.033965  -2.186   0.0311 *
window(pmi2_ts, start = c(2012, 2))        0.002013   0.001420   1.417   0.1595
diff(diff(gdp_ts, 12))                    -0.025605   0.011857  -2.159   0.0332 *
diff(diff(hpi_ts, 12))                     0.138554   0.098166   1.411   0.1612
diff(diff(qe_ts, 12))                      0.001008   0.014106   0.071   0.9432
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7644 on 101 degrees of freedom
Multiple R-squared:  0.1705,    Adjusted R-squared:  0.1212
F-statistic: 3.461 on 6 and 101 DF,  p-value: 0.003764


Call:
lm(formula = diff(diff(savings_ts, 12)) ~ window(diff(rates_ts),
    start = c(2012, 2)) + window(cci_ts, start = c(2012, 2)) +
    window(pmi2_ts, start = c(2012, 2)) + diff(diff(gdp_ts, 12)) +
    diff(diff(hpi_ts, 12)) - 1)

Residuals:
    Min      1Q  Median      3Q     Max
-2.0228 -0.3264 -0.0723  0.4061  3.6093

Coefficients:
                                         Estimate Std. Error t value Pr(>|t|)
window(diff(rates_ts), start = c(2012, 2)) -1.273465   0.852997  -1.493    0.139
window(cci_ts, start = c(2012, 2))        -0.074347   0.033774  -2.201    0.030 *
window(pmi2_ts, start = c(2012, 2))        0.002012   0.001413   1.424    0.158
diff(diff(gdp_ts, 12))                    -0.025722   0.011687  -2.201    0.030 *
diff(diff(hpi_ts, 12))                     0.138854   0.097597   1.423    0.158
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7607 on 102 degrees of freedom
Multiple R-squared:  0.1705,    Adjusted R-squared:  0.1298
F-statistic: 4.192 on 5 and 102 DF,  p-value: 0.001662
```

```
Call:
lm(formula = diff(diff(savings_ts, 12)) ~ window(diff(rates_ts),
    start = c(2012, 2)) + window(cci_ts, start = c(2012, 2)) +
    window(pmi2_ts, start = c(2012, 2)) + diff(diff(gdp_ts, 12)) -
    1)

Residuals:
    Min      1Q  Median      3Q     Max
-2.0996 -0.3869 -0.0407  0.3510  3.6635

Coefficients:
                                            Estimate Std. Error t value Pr(>|t|)
window(diff(rates_ts), start = c(2012, 2)) -1.324210   0.856478  -1.546   0.1251
window(cci_ts, start = c(2012, 2))         -0.076752   0.033899  -2.264   0.0257 *
window(pmi2_ts, start = c(2012, 2))         0.002172   0.001416   1.534   0.1281
diff(diff(gdp_ts, 12))                     -0.024514   0.011714  -2.093   0.0388 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7644 on 103 degrees of freedom
Multiple R-squared:  0.154,     Adjusted R-squared:  0.1212
F-statistic: 4.688 on 4 and 103 DF,  p-value: 0.001622


Call:
lm(formula = diff(diff(savings_ts, 12)) ~ window(diff(rates_ts),
    start = c(2012, 2)) + window(cci_ts, start = c(2012, 2)) +
    diff(diff(gdp_ts, 12)) - 1)

Residuals:
    Min      1Q  Median      3Q     Max
-1.9523 -0.2627  0.0854  0.4223  3.7152

Coefficients:
                                           Estimate Std. Error t value Pr(>|t|)
window(diff(rates_ts), start = c(2012, 2)) -1.65545    0.83418  -1.985  0.04983 *
window(cci_ts, start = c(2012, 2))         -0.08807    0.03330  -2.645  0.00945 **
diff(diff(gdp_ts, 12))                     -0.02442    0.01179  -2.071  0.04081 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7694 on 104 degrees of freedom
Multiple R-squared:  0.1347,    Adjusted R-squared:  0.1097
F-statistic: 5.396 on 3 and 104 DF,  p-value: 0.00172
```

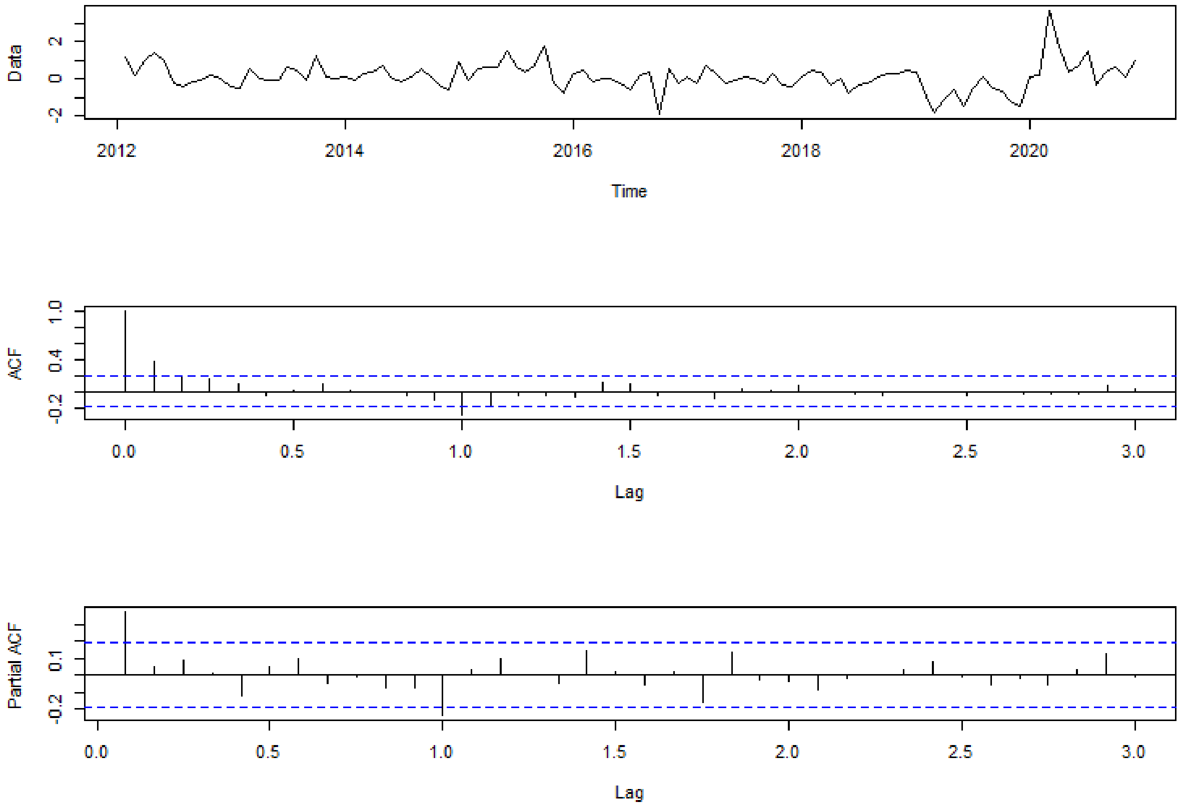

Figure X: Residuals from linear regression model with Stibor 1-month, CCI and GDP, ACF and PACF plots

R output of linear regression with Stibor 1-month, CCI, GDP and ARIMA errors:

```
Series: diff(diff(savings_ts, 12))
Regression with ARIMA(1,0,0)(1,0,0)[12] errors

Coefficients:
         ar1      sar1  window(diff(rates_ts), start = c(2012, 2))  window(cci_ts, start = c(2012, 2))
      0.4526  -0.4295                                      -0.3905                              -0.0611
s.e.  0.0927   0.1076                                       0.5856                               0.0350
      diff(diff(gdp_ts, 12))
                     -0.0161
s.e.                  0.0088

sigma^2 estimated as 0.4263:  log likelihood=-105
AIC=221.99   AICc=222.83   BIC=238.03
```

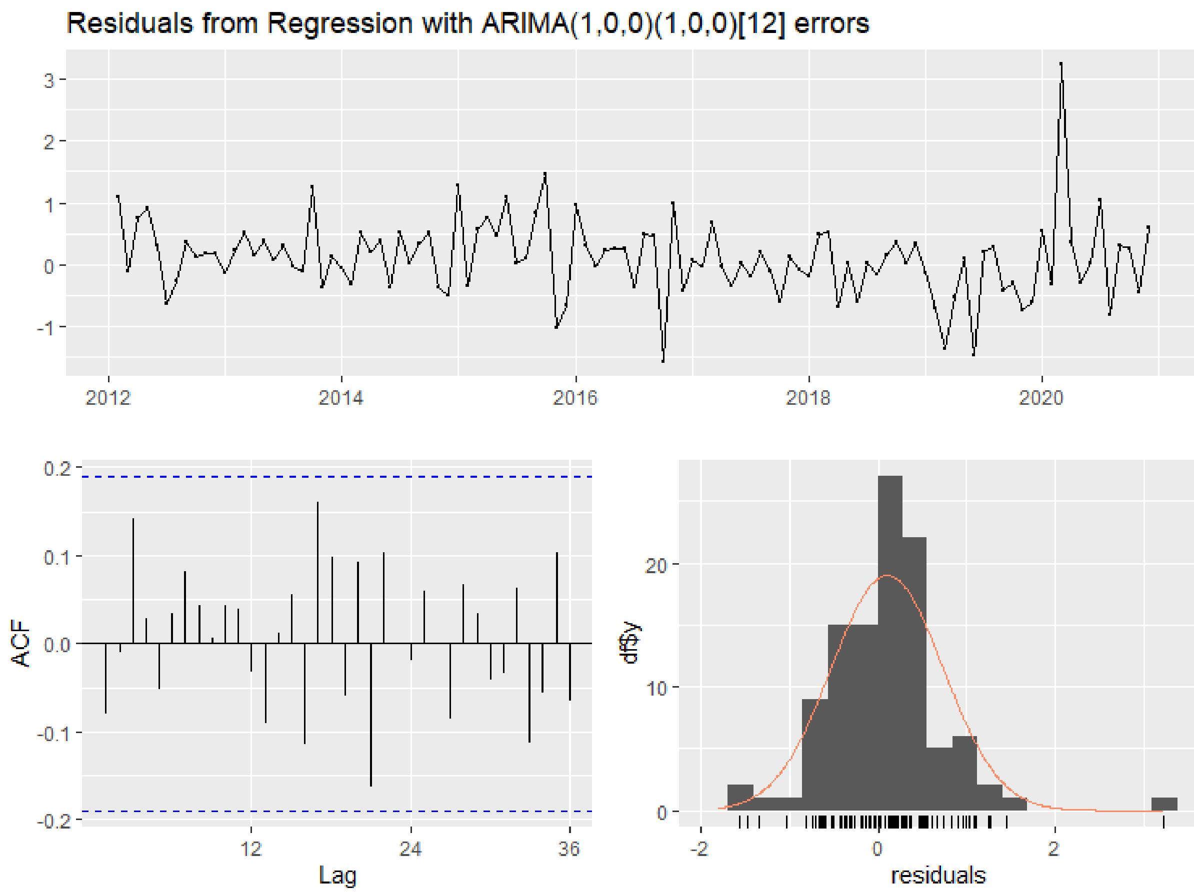

Figure X: Residuals from linear regression model with Stibor 1-month, CCI, GDP and ARIMA$(1,0,0)(1,0,0)_{12}$ errors

Appendix 9. Selection of ARIMAX models with PMI and GDP

The following figures and R outputs illustrate how the best ARIMA model was chosen.

R output of linear regression with PMI and GDP:

```
Call:
lm(formula = diff(diff(savings_ts, 12)) ~ diff(diff(gdp_ts, 12)) +
    window(pmi2_ts, start = c(2012, 2)) + -1)

Residuals:
    Min      1Q  Median      3Q     Max
-2.2632 -0.4137 -0.0310  0.2801  3.8772

Coefficients:
                                     Estimate Std. Error t value Pr(>|t|)
diff(diff(gdp_ts, 12))              -0.026028   0.011654  -2.234   0.0276 *
window(pmi2_ts, start = c(2012, 2))  0.003377   0.001371   2.463   0.0154 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.7834 on 105 degrees of freedom
Multiple R-squared:  0.09427,   Adjusted R-squared:  0.07702
F-statistic: 5.464 on 2 and 105 DF,  p-value: 0.005527
```

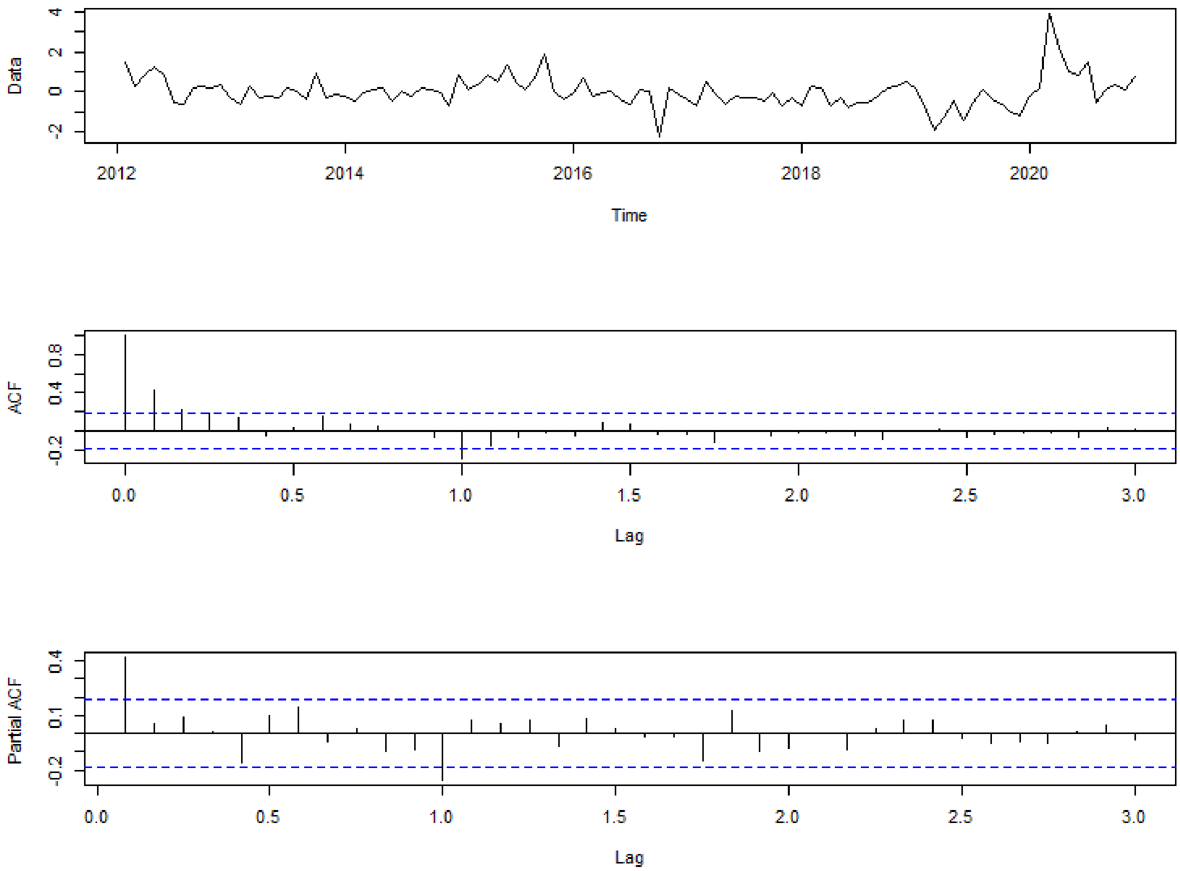

Figure X: Residuals from linear regression model with PMI and GDP, ACF and PACF plots

R output of linear regression with PMI, GDP and ARIMA errors:

```
Series: diff(diff(savings_ts, 12))
Regression with ARIMA(1,0,0)(0,0,1)[12] errors

Coefficients:
         ar1     sma1  diff(diff(gdp_ts, 12))  window(pmi2_ts, start = c(2012, 2))
      0.4457  -0.5597                 -0.0135                              0.0028
s.e.  0.0876   0.1198                  0.0087                              0.0010

sigma^2 estimated as 0.4038:  log likelihood=-103.64
AIC=217.29   AICc=217.88   BIC=230.65
```

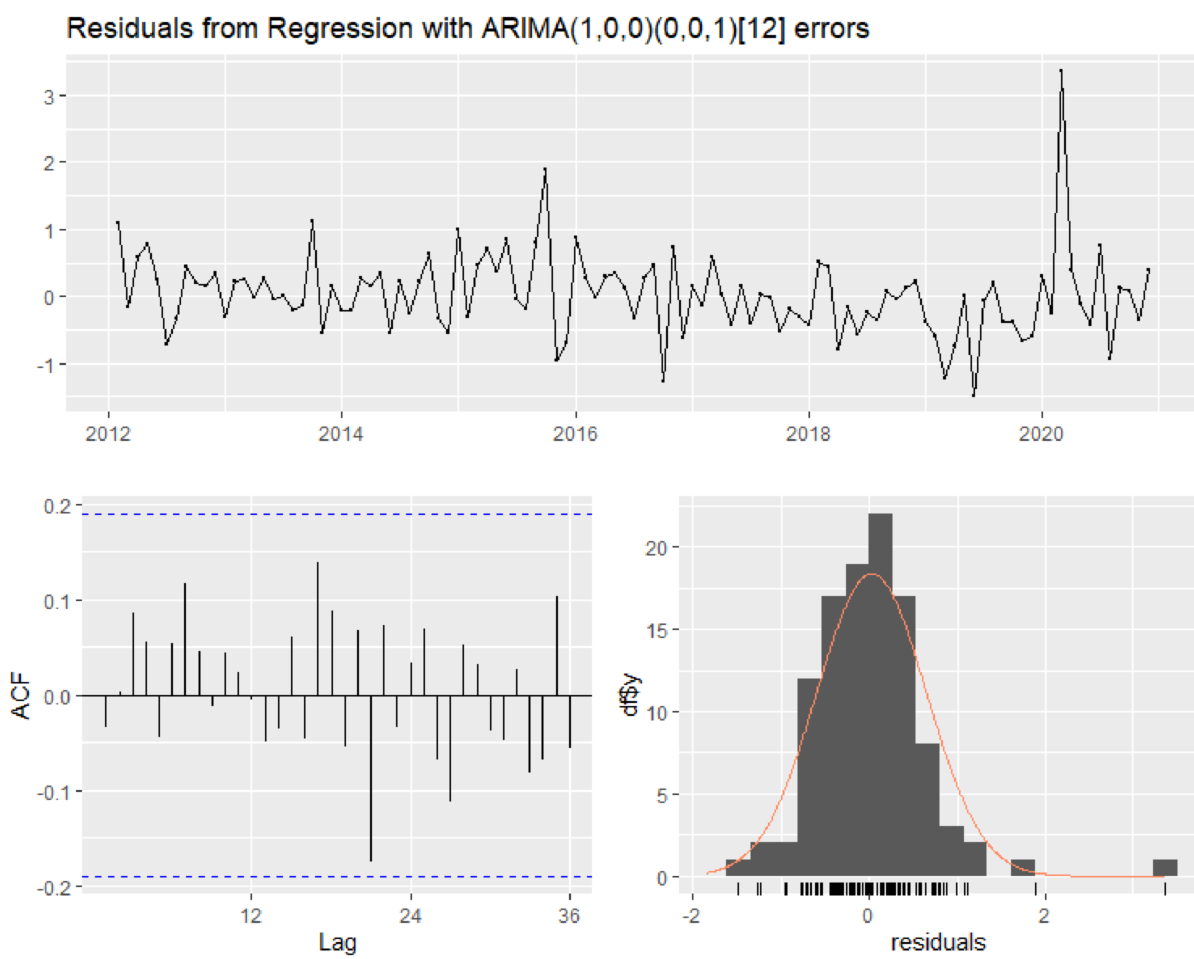

Figure X: Residuals from linear regression model with PMI, GDP and ARIMA$(1,0,0)(0,0,1)_{12}$ errors

Non-exclusive licence to reproduce thesis and make thesis public

I, Enelin Haviko,

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright, **Modelling volume of savings deposits**,
   supervised by Kalev Pärna and Jonathan Diamant.

2. I grant the University of Tartu a permit to make the thesis specified in point 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in points 1 and 2.

4. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Enelin Haviko

17/05/2022