US Industrial Production Index
Time Series Analysis and Forecasting

Bachelor Thesis

Georgij Alekseev

Ich versichere hiermit, dass ich die vorliegende Arbeit selbständig und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel verfasst habe. Wörtlich übernommene Sätze oder Satzteile sind als Zitat belegt, andere Anlehnungen, hinsichtlich Aussage und Umfang, unter Quellenangabe kenntlich gemacht. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen und ist nicht veröffentlicht. Sie wurde nicht, auch nicht auszugsweise, für eine andere Prüfungs- oder Studienleistung verwendet. Zudem versichere ich, dass die von mir abgegebenen schriftlichen (gebundenen) Versionen meiner Bachelorarbeit mit der abgegebenen elektronischen Version auf einem Datenträger inhaltlich übereinstimmen.

Ort, Datum

Unterschrift

Matriculation number: 5743249
Address: Röderbergweg 121, 60385, Frankfurt (Main)
Department: Economics and Business Administration
January 31, 2018

Supervisor:
Prof. Dr. M. Weba

Contents

Chapter 1

Prior Research, Purpose and Methodology

This thesis examines the time series of the US Industrial Production Index quantitatively. In order to identify the underlying processes, which drive the behaviour of the index, linear filters will be applied on the data until it exhibits stationarity. Then, proper autoregressive moving average (ARMA) models will be estimated and prospective values forecasted. The period on which the models were trained will be denoted as *training period*, whereas the period were the models will be evaluated will be denoted as *testing period*.

Various models will be compared and evaluated on their fit to the training period as well as forecast errors within the testing period. Main goal of this thesis is to infer potential suggestions regarding the application of robust models. Prior conducting this analysis, the required theoretical background for understanding the applied procedures will be briefly presented.

The comprehension of stationary time series and ARMA processes in particular, primarily relies on the work of Box and Jenkins [5], Brockwell and Davis [7], and Neusser [24]. Many definitions and notations are obtained from the latter two, however, in some cases the notations were adjusted for the sake of consistency within this thesis. Further, the original time series will be denoted as $\{Y_t\}$ and the transformed time series, to which ARMA models will be applied, as $\{X_t\}$.

The understanding of business cycles is grounded on the work of Lucas [21]. Since these regularly processes have to be extracted prior conducting further analysis, it will be necessary to have sound and easy smoothing filters at hand. Particularly, a two-sided [7] and the Hodrick-Prescott Filter will be applied according to Hodrick and Prescott [15], King and Rebelo [17], Ravn and Uhlig [25].

Numerical calculations are executed with the programming language R in RStudio version 1.1.383. Many instructions and commands were obtained from Cowpertwait and Metcalfe [8].

Applied data is based on the Federal Reserve releases with information available up to November 16, 2017 [13].

Chapter 2

ARIMA Models and Time Series Forecasting

2.1 General Time Series

In general, time series are premised on an underlying stochastic process. The following definition will be used throughout the thesis:

Definition (Stochastic Processes) [24]. *A stochastic process $\{X_t\}$ is a family of random variables indexed by $t \in \tau$ and defined on some given probability space.*

According to the Kolmogorov existence theorem, every stochastic process can be represented by specifying all finite-dimensional distributions [4]. Further, it is common to identify only the first two moments, mainly because the best linear predictors depend only on those two and specifying higher order moments is very complicated in practice.

The first two moments are given by the expected value $\mathbb{E}X_t$ and the variance $\mathbb{V}X_t$. For stochastic processes $\{X_t\}$ with $\mathbb{V}X_t < \infty$ can be specified as autocovariance function:

$$\gamma_x(t,s) = \operatorname{cov}(X_t, X_s) = \mathbb{E}[(X_t - \mathbb{E}X_t)(X_s - \mathbb{E}X_s)] = \mathbb{E}X_tX_s - \mathbb{E}X_t\mathbb{E}X_s.$$

One fundamental stochastic process is given by *white noise.* It will be used in later sections to describe more sophisticated processes and adopted as a goodness of fit measure for remaining residuals.

Definition (White noise) [24]. *$\{Z_t\}$ is called a white noise process if $\{Z_t\}$ satisfies:*

$$\gamma_x(t+h,t) = \begin{cases} \sigma^2, & \text{if } h = 0, \\ 0, & \text{if } h \neq 0. \end{cases} \tag{2.1}$$

We denote this by $Z_t \sim WN(0, \sigma^2)$.
If in addition Z_t is independently and identically distributed we write $Z_t \sim IID(0, \sigma^2)$.

2.2 Stationary Time Series

In order to conduct sound statistical procedures, algorithms usually rely on many independent and identically distributed (iid) samples. However, in time series analysis only one given sample from the distribution can be obtained. Loosely speaking, if it would be possible to go back in time and begin the stochastic process again, the time series would result in different outcomes. Therefore it is necessary to make assumptions regarding the persistence of the underlying process. This assumption is known as *stationarity* and will be essential throughout the remaining analysis.

Definition (Weak Stationarity) [24]. *A stochastic process* $\{X_t\}$ *is called* weakly stationary *if and only if for all integers* r, s *and* t *the following properties hold:*

(i) $\mathbb{E}X_t = \mu$ *constant*,

(ii) $\mathbb{V}X_t < \infty$,

(iii) $\gamma_x(t, s) = \gamma_x(t + r, s + r)$.

From these assumptions can be concluded that a weakly stationary time series is characterized by same mean vectors and covariance matrices of $(X_1, \ldots, X_n)'$ and the time shifted vector $(X_{1+r}, \ldots, X_{n+r})'$ for every integer r and positive integer n. Strictly stationary time series are characterized by identical whole joint distributions (not just means and covariances), i.e. $(X_1, \ldots, X_n)' \stackrel{d}{=} (X_{1+h}, \ldots, X_{n+h})'$. However, since this thesis will only have to rely on the weak stationary assumption, every weakly stationary time series will be denoted as stationary.

If $\{X_t\}$ is stationary, for the autocovariance function can be concluded:

$$\gamma_x(t, s) = \gamma_x(t - s, 0).$$

This enables the view of the autocovariance function for stationary processes as a function of just one argument. The autocovariance function will be denoted in this case by $\gamma_x(h), h \in \mathbb{Z}$. Additionally it is possible to conclude:

$$\gamma_x(h) = \gamma_x(-h) \text{ for all integers } h.$$

2.3 Classical Decomposition

It is common to describe many economic time series by the *classical decomposition model*. Thereby the observations are described as a realization of a *trend component* m_t, a *seasonal component* s_t and a *random noise component* X_t:

$$Y_t = m_t + s_t + X_t.$$

Seasonality arises because of fluctuating circumstances within a year, mainly due climatic alterations and holiday seasons. A time series trend arises because of technological progress as well as the accumulation of human and asset capital. Both effects typically have a distinct influence on the series and may overlay additional stochastic processes within it.

In the following sections it will be the goal to extract the stationary component X_t by applying linear filters to the data in order to identify and eliminate the deterministic components m_t and s_t. Then, a probabilistic model for X_t will be identified, which will enable the examination of its properties and forecasting prospective values.

2.4 Seasonal Adjustment

Data with seasonal variation has to be adjusted before it can be represented as a stationary process. In this section will only stochastic processes considered, which exhibit linear trends.

Introducing the backward shift operator B, defined as: $BX_t = X_{t-1}$, where powers are defined as following: $B^j X_t = X_{t-j}$, enables the notation of many linear filters more concisely. One linear filter to adjust for variation in quarterly observations is achieved by $X_t = \frac{1}{4}(1-B^4)Y_t$. Since $1-B^4 = (1-B)(1+B+B^2+B^3)$, this transformation does account for seasonality within quarterly data and also involves a first difference. First differences can be used to eliminate the remaining trend of the series. This section examines two more advanced linear filters.

Method 1: Two-Sided Filter

In general, sophisticated linear filters like the seasonal adjustment method of the U.S. Census Bureau (X-12 ARIMA) are two-sided, i.e. smoothing by using past and future values of the time series. In further applications a simple two-sided filter [7] will be considered for a time series with even periods of length d, where $d = 2q$. First a preliminary estimation of the trend will be obtained:

$$\hat{m}_t = (0.5Y_{t-q} + Y_{t-q+1} + \ldots + Y_{t+q-1} + 0.5Y_{t+q})/d, \quad q < t \leq T - q.$$

In the next step the average of the deviation of each observation, given a specific position in the period, from the seasonality adjusted trend component will be computed:

$$w_k = \textit{average of } \{(Y_{k+jd} - \hat{m}_{k+jd}), q < k + jd \leq T - q\} \textit{ for each } k = 1, \ldots, d.$$

Seasonal effect are usually defined cancelling each other out within a full period. The following step accounts for the fact that the average deviations do not necessarily sum to zero. Hence the seasonal component s_k will be defined as:

$$\hat{s}_k = w_k - \frac{\sum_{i=1}^{d} w_i}{d}, \quad k = 1, \ldots, d.$$

Finally the *deseasonalized* time series will be defined as the original time series diminished by the seasonal component:

$$d_t = Y_t - \hat{s}_t, \quad t = 1, \ldots, T.$$

The remaining trend will be removed by applying a first order or multiple order differences to obtain a stationary time series:

$$X_t = (1 - B)^j d_t, \quad j \geq 1.$$

Method 2: Hodrick-Prescott Filter
The *Hodrick-Prescott Filter* [15] [17] is an eminent filter in macroeconomic analysis. It can diminish or remove the trend fluctuations and business cycles. The Hodrick-Prescott filter is applied by solving the following minimization problem in respect to $\{G_t\}$ $(t = 1, \ldots, T)$:

$$\sum_{t=1}^{T} (Y_t - G_t)^2 + \lambda \sum_{t=2}^{T-1} [(G_{t+1} - G_t) - (G_t - G_{t-1})]^2 \longrightarrow min. \{G_t\}. \tag{2.2}$$

The first term penalizes deviations from the original data and therefore inhibits excessive smoothing. It is minimized when $G_t = Y_t$ for all t and then equal to zero. However, the second term penalizes fast fluctuations within the data and hence, smooths the trend as well as potential cycles. This term is minimized when the growth term does not change over time, i.e. when the new time series is a linear function. The meta-parameter λ controls between both opposite effects and has to be defined in advance. Ravn and Uhlig [25] developed a formula to adjust λ for given frequencies. It was constructed in such a way, that the trend component stays practically the same, whatever data frequency one may choose. Precisely, they proposed:

$$\lambda = \begin{cases} 6.25, & \textit{yearly observations}, \\ 1600, & \textit{quarterly observations}, \\ 129600, & \textit{monthly observations}. \end{cases}$$

Finally, the resulting series $\{G_t\}$ will be defined as preliminary estimation of the trend $\hat{m}_t$:

$$\hat{m}_t = G_t.$$

Afterwards, the same steps as described already for the two-sided filter will be applied to obtain the stationary series $\{X_t\}$.

2.5 ARIMA Model

The family of ARMA models are an important tool in time series analysis and forecasting. In fact, every stationary process can be replicated arbitrarily well by an autoregressive moving-average process [7]. Particularly, one can replicate any given autocovariance function γ by an ARMA process $\{X_t\}$ such that $\gamma_x(k) = \gamma(k)$ for any given $k \geq 0$, if $\lim_{h \longrightarrow \infty} \gamma(h) = 0$. Linear ARMA models also substantially simplify prediction methods.

Definition (ARMA Models) [24]. *A stochastic process $\{X_t\}$ with $t \in \mathbb{Z}$ is called an autoregressive moving-average process (ARMA process) of order (p,q), denoted by ARMA(p,q) process if the process is stationary and satisfies a linear stochastic difference equation of the form:*

$$X_t - \phi_1 X_{t-1} - \ldots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}, \tag{2.3}$$

with $Z_t \sim WN(0, \sigma^2)$ and $\phi_p \theta_p \neq 0$.

Using the backward shift operator B enables to rewrite the process more concisely in the form of:

$$\phi(B)X_t = \theta(B)Z_t,$$

where $\phi(B)X_t = \sum_{j=0}^{p} \phi_j X_{t-j}$ and $\theta(B)Z_t = \sum_{j=0}^{p} \theta_j Z_{t-j}$ with $\phi_0 = 1$ and $\theta_0 = 1$.

Even though economic time series generally do not fulfil the stationarity requirement, it is possible to apply some linear filters to achieve stationarity. Time series growing with a constant growth rate ε, can be transformed by taking the natural logarithm of each realization respectively, since: $\ln(1 + \varepsilon) \approx \varepsilon$ for small ε.

In general, a linear trend and seasonality will remain after this first transformation. It is common to describe the seasonally adjusted time series $\{Y_t\}$ by a linear trend $\mu_t = \alpha + \delta t$ and a function of past error terms as following:

$$Y_t = \alpha + \delta t + \Psi(B)Z_t.$$

The common notion to describe time series as difference stationary is primarily the achievement of Nelson and Plosser [23]. Thus, one approach is taking differences in order to eliminate the linear trend. In most cases taking the first difference will suffice, but sometimes it may not. Summarizing, the time series Y_t will be transformed to $X_t = (1 - B)^d Y_t$ ($d = 1, 2, \ldots$). The time series Y_t, which has to be differentiated d-times to achieve stationarity, is called integrated of order d. If $\{X_t\}$ now follows an ARMA(p,q) process, $\{Y_t\}$ is defined as autoregressive integrated moving average (ARIMA)(p, d, q) process:

Definition (ARIMA Models) [7]. *If d is a nonnegative integer, then $\{Y_t\}$ is an ARIMA(p,d,q) process if $X_t := (1 - B)^d Y_t$ is a causal ARMA(p,q) process.*

Since this thesis will restrict to order one integrated processes, this more precise definition will suffice for the following chapter:

Definition (First Order ARIMA Models) [24]. *The stochastic process $\{Y_t\}$ is called* integrated of order one *or* difference stationary, *denoted as $Y_t \sim I(1)$, if and only if $\Delta Y_t = Y_t - Y_{t-1}$ can be represented as:*

$$\Delta Y_t = (1 - B)Y_t = \delta + \Psi(B)Z_t, \quad \Psi(1) \neq 0, \tag{2.4}$$

with $\{Z_t\} \sim WN(0, \sigma^2)$ and $\sum_{j=0}^{\infty} |\Psi_j| < \infty$.

2.6 Properties of ARMA processes

For reasons previously explained, stationarity is an important assumption in time series analysis and thus, a part of the definition 2.3. In the following definitions complex z will be used, since the zeros of a polynomial of degree $p > 1$ can either be real or complex. Furthermore, the area defined by $|z| = 1$ will be denoted as unit circle.

Definition (Existence and Uniqueness of Stationary Solutions) [7]. *A stationary solution $\{X_t\}$ of equation (2.3) exists (and is also the unique stationary solution) if and only if:*

$$\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p \neq 0 \textit{ for all } |z| = 1,$$

and the polynomials $(1-\phi_1 z - \ldots - \phi_p z^p)$ and $(1+\theta_1 z + \ldots + \theta_q z^q)$ have no common factors.

In time series analysis it is important to determine whether today's state X_t can be represented as the result of current and past shocks $Z_t, Z_{t-1}, \ldots$. In this case, it is said that X_t is having a *causal representation*. Thus, current Z_t may be able to affect current X_t and also propagate to future X_t, but prospective Z_t cannot influence current or past X_t.

Definition (Causality) [7]. *An ARMA(p,q) process $\{X_t\}$ is* causal *if there exist constants $\{\psi_j\}$ such that $\sum_{j=0}^{\infty} |\psi_j| < \infty$ and:*

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \textit{ for all } t.$$

Causality is equivalent to the condition:

$$\phi(z) = 1 - \phi_1 z - \ldots - \phi_p z^p \neq 0 \textit{ for all } |z| \leq 1.$$

Stationarity and causality of an ARMA process ensure that the effect of Z_t vanishes eventually, i.e.:

$$\frac{\partial X_{t+j}}{\partial Z_t} \longrightarrow 0 \textit{ as } j \longrightarrow \infty.$$

Furthermore, it may be important to analyse current and past values of $\{Z_t\}$. Since solely realizations of $\{X_t\}$ are observed, without direct knowledge of the realizations of $\{Z_t\}$, a method of retrieving the unobserved shocks from $\{X_t\}$ is required. If it is feasible to do so, the ARMA process is called invertible.

Definition (Invertibility) [7]. *An ARMA(p,q) process $\{X_t\}$ is* invertible *if there exist constants $\{\pi_j\}$ such that $\sum_{j=0}^{\infty} |\pi_j| < \infty$ and:*

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} \textit{ for all } t.$$

Invertibility is equivalent to the condition:

$$\theta(z) = 1 + \theta_1 z + \ldots + \theta_q z^q \neq 0 \text{ for all } |z| < 1.$$

2.7 Auto- and Partial Autocorrelation Function

Autocorrelation Function

The autocorrelation function (ACF) can help determining characteristics of the stochastic process, the number of moving-average parameters that have to be specified and testing for white noise residuals. Whenever a moving-average process with order q is observed, the autocorrelation function would result in $\gamma(j) = 0$ for $j > q$. This characteristic makes it very important in time series analysis.

Since in practice the autocorrelation function γ and mean μ cannot be directly observed from the underlying data, it is necessary to estimate them before conducting further analysis. The following sections will rely on these specifications:

$$\begin{aligned}
\hat{\mu} &= \overline{X}_T = \frac{1}{T}(X_1 + X_2 + \ldots + X_T), \\
\hat{\gamma}(h) &= \frac{1}{T}\sum_{t=1}^{T-h}(X_t - \overline{X}_T)(X_{t+h} - \overline{X}_T), \\
\hat{\rho}(h) &= \frac{\hat{\gamma}(h)}{\hat{\gamma}(0)}.
\end{aligned}$$

The previous estimators for the autocovariance and the autocorrelation are obviously biased, because first the mean had to be estimated and the function still divides by T instead of $T - h$. Second, both multiplicands use the mean computed from the complete sample of observations and not the unbiased version with the mean computed from $X_1, \ldots, X_{t-h}$ and $X_{h+1}, \ldots, X_T$ in the first and second multiplicand, respectively.

However, these estimators guarantee that the covariance matrix $\hat{\Gamma}_T$ and the autocorrelation matrix $\hat{R}_T$ of $(X_1, \ldots, X_T)'$,

$$\begin{aligned}
\hat{\Gamma}_T &= \begin{pmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \cdots & \hat{\gamma}(T-1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \cdots & \hat{\gamma}(T-2) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\gamma}(T-1) & \hat{\gamma}(T-2) & \cdots & \hat{\gamma}(0) \end{pmatrix} \\
\hat{R}_T &= \frac{\hat{\Gamma}_t}{\hat{\gamma}(0)}
\end{aligned}$$

always result in non-negative definite and for $\hat{\gamma}(0) > 0$ in non-singular matrices. This property ensures mathematical convenience in further computations, especially the parameter estimation.

It can be shown that these estimators are consistent and asymptotically normally distributed. Box and Jenkins [5] recommended a sample size larger than 50 and an order of the autocorrelation coefficient smaller than $T/4$ to obtain proper estimates.

Partial Autocorrelation Function

Whenever a time series is modelled by past observations, the question whether a particular X_{T-j} $(j = 0, 1, \ldots)$ is of interest for forecasting X_{T+1} plays a crucial role in time series forecasting. It can help determining the number of autoregressive parameters that have to be specified and give further information about the underlying mechanisms. The *partial autocorrelation* answers this question by stating how much a particular X_{T-j} contributes to the forecast of X_{T+1} *controlling* for all X_i $(i \leq T$ and $j \neq i)$. Thus the partial autocorrelation function is defined as:

Definition (Partial autocorrelation function) [24]. *The* partial autocorrelation function (PACF) $\alpha(h)$, $h = 0, 1, 2, \ldots$, *of a stationary process is defined as following:*

$$\begin{aligned}
\alpha(0) &= 1 \\
\alpha(1) &= \operatorname{corr}(X_2, X_1) = \rho(1) \\
\alpha(h) &= \operatorname{corr}[X_{h+1} - \mathbb{P}(X_{h+1}|1, X_2, \ldots, X_h), X_1 - \mathbb{P}(X_1|1, X_2, \ldots, X_h)],
\end{aligned} \tag{2.5}$$

where $\mathbb{P}(X_{h+1}|1, X_2, \ldots, X_h)$ *and* $\mathbb{P}(X_1|1, X_2, \ldots, X_h)$ *denote the best, in the sense mean squared forecast errors, linear forecasts of* X_{h+1}, *respectively* X_1 *given* $\{1, X_2, \ldots, X_h\}$.

For example, an autoregressive process with p determining parameters may be considered. Thus, only the last p observations are required for forecasting X_{T+1} and the PACF α would correctly result in $\alpha(j) = 0$ for every $j > p$. It is worth noting that each of the last p observations is also a function of its own past p lags. An autocorrelation function γ would also account for this indirect correlations and would result in exponentially declining coefficients. Hence, the autocorrelation function γ would not be able to recommend a specific order of autogressive terms.

However, a moving-average process will always result in $\alpha(j) > 0$ for every $j \geq 0$, since Z_t is an infinite weighted sum of past X_t's and each new observation would further help predicting X_{T+1}. Thus, it can be concluded that *both* functions γ and α are necessary for conducting a thorough analysis of the underlying processes.

Moreover, the definition of partial autocorrelation suggests that it is strongly interconnected with the respective autocorrelation functions. Indeed, it is possible to express the PACF in terms of ACF's. The resulting computing technique is known as Durbin-Levinson algorithm, which will be applied in the following chapter [12]:

$$
\begin{aligned}
\alpha(0) &= 1, \\
\alpha(1) &= a_{11} = \rho(1), \\
\alpha(2) &= a_{22} = \frac{\rho(2) - \rho(1)^2}{1 - \rho(1)^2}, \\
&\dots \\
\alpha(h) &= a_{hh} = \frac{\rho(h) - \sum_{j=1}^{h-1} a_{h-1,j}\rho_{h-j}}{1 - \sum_{j=1}^{h-1} a_{h-1,j}\rho_j},
\end{aligned}
\tag{2.6}
$$

where $a_{h,j} = a_{h-1,j} - a_{hh}a_{h-1,h-j}$ for $j = 1, 2, \dots, h-1$.

Summarizing, the ACF and PACF can be used to identify the appropriate number of prediction parameters. Precisely, the ACF can be used to identify the number of moving average terms, whereas the PACF can help identifying the number of autoregressive terms. This approach is known as Box-Jenkins methodology [5][1].

2.8 Parameter Estimation

In this section, it will be assumed that a given order p and q has already been chosen and engaged only with the estimation of their coefficients. This is reasoned by the circumstance that the optimal order has to be identified by evaluating and comparing various combinations of previously specified orders.

Parameters of purely autoregressive models can be easily computed with algorithms like the Yule-Walker algorithm or ordinary least squares (OLS). However, if moving average terms want to be included, these procedures are not reliable any more, since it is not possible to directly observe the values of $\{Z_t\}$, in contrast to past $\{X_t\}$. The most common method for estimating p and q is known as *maximum likelihood method*. The maximum likelihood estimators *maximize* the probability that the actually observed sample will be received from the distribution, i.e. the ARMA model.

The maximum likelihood parameters for a Gaussian time series $\{X_t\}$ are obtained by maximizing the maximum likelihood function, which is defined as:

$$
L(\Gamma_n) = (2\pi)^{-n/2}(\det \Gamma_n)^{-1/2} \exp(-\frac{1}{2}\mathbf{X}_n'\Gamma_n^{-1}\mathbf{X}_n), \tag{2.7}
$$

where $\Gamma_n = \mathbb{E}(\mathbf{X}_n\mathbf{X}_n')$ and Γ_n is non-singular.

However, the computation of Γ_n is numerically arduous. The numerical calculations can be simplified by replacing Γ_n and Γ_n^{-1} in terms of the one-step least-squares prediction errors

[1] The Box-Jenkins methodology is further elaborated in the Appendix section Box-Jenkins methodology.

$X_j - \mathbb{P}_{j-1}X_j$ and their respective variances ν_{j-1} $(j = 1, \ldots, n)$. Since these can be easily calculated using the *innovation algorithm* [7], an efficient estimation is now facilitated[2]. The equivalent likelihood function is now represented as:

$$L(\Gamma_n) = \frac{1}{\sqrt{(2\pi)^n \nu_0 \cdots \nu_{n-1}}} \exp\left\{ -\frac{1}{2} \sum_{j=1}^{n} (X_j - \mathbb{P}_{j-1}X_j)^2 / \nu_{j-1} \right\}. \tag{2.8}$$

A justification for maximum likelihood is given by the property that the estimators are approximately normally distributed with variances less or equal as those of other asymptotically normally distributed estimators, if $X_1, X_2, \ldots, X_n$ are iid and n large [19].

Minimizing itself must be done numerically. In particular, many algorithms define preliminary values for ϕ and θ, for example by applying the Hannan-Rissanen and Innovation algorithm respectively. Preliminary estimates which are close to the maximum of the likelihood function can accelerate the computation tremendously. Afterwards the algorithms search systematically for values of ϕ and θ, which maximize the maximum likelihood function. It is worth noting that the maximum likelihood estimators result in the following estimate for the white noise variance, which will be used in the selection of a proper order p and q:

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{j=1}^{n} \left(X_j - \mathbb{P}_{j-1}X_j\right)^2 / r_{j-1}. \tag{2.9}$$

The variable r_{j-1} is defined as $r_{j-1} = \mathbb{E}(W_j - \hat{W}_j)^2$. $\{W_t\}$ is a transformed version of the process $\{X_t\}$, which is used for simplification issues. W_t is defined as $W_t = \sigma^{-1}X_t$ and $W_t = \sigma^{-1}\phi(B)X_t$ for $t = 1, \ldots, m$ and $t > m$ respectively. The variable r is obtained recursively by the innovation algorithm and is independent of σ by construction.

The maximization problem will be used for parameter estimation in the following chapter, whether or not the time series is Gaussian, since assuming specific distributions increases the difficulty of parameter estimation significantly. Resulting biases are negligible, one reason is that large-sample distribution of the estimators when $\{Z_t\} \sim \mathrm{IID}(0, \sigma^2)$ are identical with and without assuming Gaussian $\{Z_t\}$.

2.9 Order Selection

In practice, economic models mostly do not inherently suggest exact orders of p and q. To overcome this problem, identifying them by analysing the data is necessary. If the model is either over- or underfitted, i.e. it includes too many or not enough parameters respectively, the model will face several problems.

[2] A complete parameter identification can be obtained from the Appendix section Innovation Algorithm.

Overfitting will result in biased estimators θ_j and ϕ_j ($j = 0, 1, 2, \ldots$). However, it will not result in inconsistent coefficients for the causal representation ψ_j, where $\psi(z) = \frac{\theta(z)}{\phi(z)}$. For example, if one may try to fit an ARMA(1,1) process, i.e. $X_t - \phi X_{t-1} = Z_t + \theta Z_{t-1}$, to a white noise process, i.e. $X_t = Z_t \sim \mathrm{WN}(0, \sigma^2)$, the maximum likelihood estimators would not converge to their true values $\phi = \theta = 0$. Instead, they converge to $\phi = -\theta$, i.e. they converge to either ($\phi = 1; \theta = -1$) or ($\phi = -1; \theta = 1$) depending on the starting values. However, the causal representation would correctly converge to their true values $\psi_0 = 1$ and $\psi_j = 0$ for $j > 0$. Furthermore, very high order models generally result in small estimated white noise variance, but will result in high mean squared forecast errors. Thus, overfitted models are not suitable for forecasting purposes.

Therefore it can be concluded that overfitting inflates the model, disturbs the interpretability and increases forecast errors. Given the nature of the resulting biases for the underlying coefficients, overfitting can be detected by identifying autoregressive (AR) and moving average (MA) polynomials with similar roots, which are close to the unit circle in absolute value.

Underfitting will also result in biased estimators, due to the *omitted variable bias*. The estimators then converge as close as possible to the true parameters given the restricted parameter space. It is difficult to spot an underfitted model without trying various combinations of p and q parameters.

In order to identify the right order of p and q, multiple criteria and methods will be used. As previously mentioned, Box and Jenkins [5] suggested the analysis of the ACF and PACF. To determine the order solely with this approach requires according know-how, but it can be used very well as a first indication of the optimal parameters. Automatized order selection criteria are given by the Akaike information criterion (AIC), small sample size corrected version of the Akaike information criterion (AICc) and Bayesian information criterion (BIC). Each one has in common that a given criterion has to be minimized to obtain the appropriate order p and q. The first term of the criterion consists of the maximum likelihood estimate of the white noise variance $\hat{\sigma}^2$ (or the negative value of the logarithmized likelihood function $\ln L_{p,q}$ in other representations), which continuously decreases with increasing orders of p and q. However, the functions also include a penalty term, which compensates for the tendency of overfitting. Minimizing the occurring trade-off between variance of the residuals and overfitting shall result in appropriate order identification.

The *AIC* [2] is widely used in practice and was designed to result in approximately unbiased estimators. This statistic is defined as:

$$\mathrm{AIC}(p, q) = -2 \ln L_{p,q} + 2(p + q + 1). \tag{2.10}$$

The *AICc* [16] is a bias-corrected version of the AIC. It provides a higher penalty for models with large orders p and q and thus decreases the tendency of AICc to overfit the data. The

AICc statistic can be represented as:

$$\mathrm{AICc}(p,q) = -2\ln L_{p,q} + 2(p+q+1)\frac{T}{(T-p-q-2)}. \tag{2.11}$$

The *BIC* [3] is a consistent order-selection criterion in contrast to AIC and AICc. Thus, if the time series process $\{X_t\}$ is indeed a realization of an ARMA(p,q) process, the BIC estimated parameters $\hat{p}$ and $\hat{q}$ will converge to their true values, i.e. $\hat{p} \to p$ and $\hat{q} \to q$ as $n \to \infty$ [14]. The BIC statistic is given by:

$$\begin{aligned}\mathrm{BIC} = {} & (T-p-q)\ln[T\hat{\sigma}^2/(T-p-q)] + T(1+\ln\sqrt{2\pi}) \\ & + (p+q)\ln\left[\left(\sum_{t=1}^{T} X_t^2 - T\hat{\sigma}^2\right)/(p+q)\right].\end{aligned} \tag{2.12}$$

It is necessary to mention that in practice a *true* ARMA(p,q) model is generally not feasible, since there may be many processes that provide a good fit to the observed data. Further research on the probability of choosing the right order by utilization of AIC [11] suggested that all models within c of AIC_{min} could be considered in further evaluation. Whilst c should be adjusted to the sample size amongst others, a typical value of $c = 2$ is often used in practice. Final model selection is dependent on iid residuals as well as interpretability issues, such as simplicity.

2.10 Forecasting

After specifying a proper order and identifying the maximum likelihood estimators, the time series will be forecasted. The first step will be forecasting the underlying ARMA process:

$$\mathbb{P}_t X_{t+1} \begin{cases} \displaystyle\sum_{j=1}^{t} \theta_{tj}(X_{t+1-j} - \mathbb{P}_{t-j}X_{t+1-j}), & 1 \le t < m, \\ \\ \displaystyle\phi_1 X_t + \ldots + \phi_p X_{t+1-p} + \sum_{j=1}^{q} \theta_{tj}(X_{t+1-j} - \mathbb{P}_{t-1}X_{t+1-j}), & t \ge m. \end{cases} \tag{2.13}$$

Afterwards the undertaken transformations will be reversed. For the case of the previously discussed filters the series $\{X_t\}$ will be defined as resulting first order differences and hence:

$$\begin{aligned}\mathbb{P}d_T &= d_T, \\ \mathbb{P}_t d_{t+1} &= \mathbb{P}_{t-1}d_t + \mathbb{P}_t X_{t+1}, \quad t \ge T.\end{aligned}$$

Finally, the forecast of the original process $\{Y_t\}$ is obtained by adding the omitted seasonal effects:

$$\mathbb{P}_t Y_{t+1} = \mathbb{P}_t d_{t+1} + \hat{s}_{t+1}.$$

2.11 Goodness of Fit

In the following chapter various goodness of fit measures will be applied, which are introduced in this section. This thesis primarily distinguishes between *testing the residuals*, *unit root tests* and *forecasting suitability*. If multiple of these tests fail, the model will be rejected as suitable for the underlying stochastic process. However, it still has to be considered that the probability of failing one test of these increases with every further check conducted.

Testing the residuals
If an appropriate ARMA(p,q) model was chosen and the underlying stochastic process follows indeed an ARMA(p,q) process, the residuals of the model should behave characteristically. The residuals used in the following tests are defined as:

$$\hat{W}_t = \frac{X_t - \mathbb{P}_{t-1}X_t}{\sqrt{r_{t-1}}}.$$

These residuals approximate the white noise well, since $\mathbb{E}(W_t - Z_t)^2 \to 0$ as $t \to \infty$ [7]. Further, it can be concluded that the series $\{\hat{W}_t\}$ should behave approximately

(i) uncorrelated for $\{Z_t\} \sim \mathrm{WN}(0, \sigma^2)$,

(ii) independent for $\{Z_t\} \sim \mathrm{IID}(0, \sigma^2)$,

(iii) normally distributed for $\{Z_t\} \sim \mathrm{N}(0, \sigma^2)$.

The white noise standard deviation is defined as $\hat{\sigma}_W = \sqrt{(\sum_{t=1}^{n} W_t^2)/n}$ and the *rescaled residuals* $\hat{R}_t$ as:

$$\hat{R}_t = \frac{\hat{W}_t}{\hat{\sigma_W}}$$

For the following tests will be assumed that the fitted model is close to the true model and $\{Z_t\} \sim \mathrm{IID}(0, \sigma^2)$.

1. *The graph of the rescaled residuals $\hat{R}_t$.* Applying this procedure, the graph of $\hat{R}_t$ is analysed for obvious deviations from white noise with variance of 1. An obvious trend, cyclical component or nonconstant variance would lead to a rejection of the current model.

2. *The sample ACF of the residuals.* Plotting the sample autocorrelation function of the residuals should result in not more than two or three out of 40 values to fall outside the confidence bounds $\pm 1.96/\sqrt{n}$. However, it is necessary to mention that this test can be unreliable in specific circumstances, because the assumed standard error $\sqrt{n}$ overestimates the standard deviation of the residuals. Thus, this test may underestimate the statistical significance of deviations from zero and the confidence bounds should be seen as upper limit of the actual bounds [5][6].

3. *Portmanteau test.* The Portmanteau test also evaluates the sample autocorrelation functions, but relies on a single statistic, which is approximately chi-squared distributed with h degrees of freedom, assuming the residuals are iid. The original test formulation were introduced by Box and Jenkins [6]. Subsequently the formulation recommended by Ljung and Box [20] will be applied $Q = T(T+2)\sum_{j=1}^{h} \hat{\rho}^2(j)/(T-j)$ and the iid hypothesis rejected at level α if $Q > \chi^2_{1-\alpha}(h)$.

4. *Turning point test.* A turning point is defined at time i ($1 < i < T$), if $y_{i-1} < y_i$ and $y_i > y_{i+1}$ or if $y_{i-1} > y_i$ and $y_i < y_{i+1}$. The number of turning points within a period is defined as n. It is obvious that the probability of a turning point at time i for an iid sequence is $\frac{2}{3}$ and $\mu_n = \mathbb{E}(n) = 2(T-2)/3$. Further, n has a variance of $\sigma_n^2 = (16T - 29)/90$. Thus, n should behave approximately $\mathrm{N}(\mu_T, \sigma_T^2)$ and the hypothesis can be rejected at level α if $|n - \mu_n|/\sigma_n > \Phi_{1-\alpha/2}$ [7].

Unit root tests

Since the constructed ARMA models rely on the stationarity assumption it is important to determine if the underlying time series does indeed exhibit stationarity. These stationarity tests are called unit root tests. An eminemt example is the *Dickey-Fuller (DF) test* [9][10]. In the following the procedure will be briefly explained.

The DF model is defined as:

$$X_t = \alpha + \delta t + \beta X_{t-1} + \epsilon_t.$$

First the coefficients are regressed by OLS and the following t-statistic for the hypothesis $\beta = 1$ established:

$$t_{\hat{\beta}} = \frac{(\hat{\beta}_T - 1)}{\hat{\sigma}_{\hat{\beta}}}.$$

It is necessary to note that this test statistic is not asymptotically normally distributed and does require adjusted critical values. These have been tabulated for example by Mackinnon [22], which can be extrapolated to fit any given hypothesis. Thus, the hypothesis of this *one-sided test* is stated as:

$$\mathrm{H}_0: \ \beta = 1 \qquad \mathrm{H}_1: \ -1 < \beta < 1.$$

It can be concluded that evidence for stationarity is given, if H_0 can be rejected at a given confidence level α.

However, this DF test does not account for higher order AR terms and thus, is not reliable in this case. Therefore, the *augmented Dickey-Fuller (ADF) test* will be applied [26]. The standard testing procedures can be applied as previously described for the DF test and the ADF model is defined as:

$$X_t = \alpha + \beta X_{t-1} + \gamma_1 \Delta X_{t-1} + \ldots + \gamma_{p-1} \Delta X_{t-p+1} + \epsilon_t.$$

The hypothesis is still defined as:

$$\mathrm{H}_0: \ \beta = 1 \qquad \mathrm{H}_1: \ -1 < \beta < 1.$$

And again, it can be concluded that evidence for stationarity is given, if H_0 can be rejected at a given confidence level α.

Forecasting Suitability

The suitability of the process to result in predictions close to prospectively observed values is obviously very important in forecasting applications. However, the decision of an appropriate model should not be solely based on the accuracy within a given testing period, because this criteria could lead to an overfitted model, which will score poorly in another given testing period. The applied measure of forecast accuracy in the following chapter will be the *mean squared error* for a given time horizon N defined as:

$$MSE(N) = \frac{1}{N} \sum_{n=T+1}^{T+N} (X_n - \mathbb{P}_{n-1} X_n)^2.$$

It is clear by the structure of the MSE that a lower value is equivalent to higher forecast accuracy.

2.12 The Wold Decomposition

This chapter is closed by giving a further justification of the ARIMA models that will be used to predict a sophisticated stochastic process, precisely the US Industrial Production index.

In reality, most processes obviously do not follow simple linear models. However, the *Wold Decomposition Theorem* states that *any* arbitrary stationary process can be expressed by a linear combination of current and past errors $\{Z_t\}$ plus a deterministic component $\{V_t\}$.

Theorem (Wold Decomposition) [24]. *Every stationary stochastic process $\{X_t\}$ with mean zero and finite positive variance can be represented as*

$$X_t = \sum_{j=0}^{\infty} \phi_j Z_{t-j} + V_t = \Psi(B) Z_t + V_t, \tag{2.14}$$

where

(i) $Z_t = X_t - \tilde{\mathbb{P}}_{t-1} X_t = \tilde{\mathbb{P}}_t Z_t,$

(ii) $Z_t \sim WN(0, \sigma^2)$ *where* $\sigma^2 = \mathbb{E}(X_{t+1} - \tilde{\mathbb{P}}_t X_{t+1})^2 > 0,$

(iii) $\phi_0 = 1$ *and* $\sum_{j=0}^{\infty} \phi_j^2 < \infty,$

(iv) $\{V_t\}$ *is deterministic,*

(v) $\mathbb{E}(Z_t V_s) = 0$ *for all* $t, s \in \mathbb{Z}.$

The existence of the Wold Decomposition Theorem enables the justification of causal ARMA models for various time series. However, only a finite amount of past observations can be identified, whereas the theorem requires in general infinitely many past observations $(\phi_1, \phi_2, \ldots)$, which is clearly not feasible in practice. In order to achieve theoretically unbiased estimators, it will be further assumed that the transformed time series $\{X_t\}$ can be reconstructed from finite past parameters, i.e. the p and q parameters from a causal ARMA process:

$$\Psi(B) = \frac{\theta(B)}{\phi(B)} = \frac{1 + \theta_1 B + \theta_2 B^2 + \ldots + \theta_q B^q}{1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p},$$

and the purely deterministic part will be disregarded, i.e. $V_t = 0$.

Chapter 3

Applying ARIMA Models to the US Industrial Production Index

3.1 US Industrial Production Index

The US industrial production index is a monthly published statistical release, which measures the real output of manufacturing, mining, electric and gas utilities industries [13]. The reference period is defined as 2012. The industry definitions are primarily based on the North American Industry Classification System (NAICS) *plus* logging and newspaper, periodical, book and directory publishing.

The total index is an aggregation of 299 individual series, which measure different industries. Data is obtained from output measured in physical units as well as inputs to the production process. The aggregated index is determined by a version of the Fisher-ideal index formula, which weights each series according to its proportion of the total value-added output of all industries [1].

Data published around the 15th of the following month for the previous month is preliminary and may be revised in each of the following five subsequent months. In order to circumvent data, which is subject to revision, the following models rely only on data up to December 2016.

The Federal Reserve publishes original as well as seasonally adjusted data, utilizing the X-12 ARIMA method. Whenever this thesis relies on X-12 ARIMA adjusted data, the series made available by the Federal Reserve will be employed.

3.2 General Model

This analysis will be specified with monthly data of the US industrial production index ranging from 1964 to 2013 inclusive. Later, these models will be tested on the three year horizon from 2014 to 2016 inclusive. All Models will be based on an ARIMA(p,d,q) process according to definition 2.4, but will differ in respect to *training period*, *applied linear filters*, *outlier treatment* and *differencing*. The final models will consist of a combination of variant 1 or 2 for each characteristic respectively.

Characteristic	**Variant 1**	**Variant 2**
Training period	January 1964 - December 2013	January 1984 - December 2013
Applied linear filter	Two-Sided or Hodrick-Prescott (HP) Filter	X-12 Arima
Outlier treatment	winsorization	no treatment
Differencing	$X_t = (1 - B)Y_t$	$X_t = (1 - B^{12})Y_t$

The number of processed observations within the training period obviously differs in respect to the training period. The total sample numbers per alternative are illustrated below.

Training period	
1964 - 2013	1984 - 2013
600	360

Thus, at least 360 observations per model are obtained. These fairly exceed the proposed number of 100, which are required to obtain reasonable estimators [5].

A graphical illustration (Figure 3.1) of the original time series indicates non-stationarity as well as an obvious trend and seasonality. This first assumption is confirmed by examining the ACF (Figure 5.1), whose very slowly decaying values are a clear sign of non-stationarity. Economic reasoning and the plotted series suggest that the series exhibits a non-linear trend. Assuming non-linear growing residuals as well, the natural logarithm is applied in order to deal with these issues (Figure 3.2). In the following sections will be assumed that every series has been previously logarithmized, whether it is especially denoted or not.

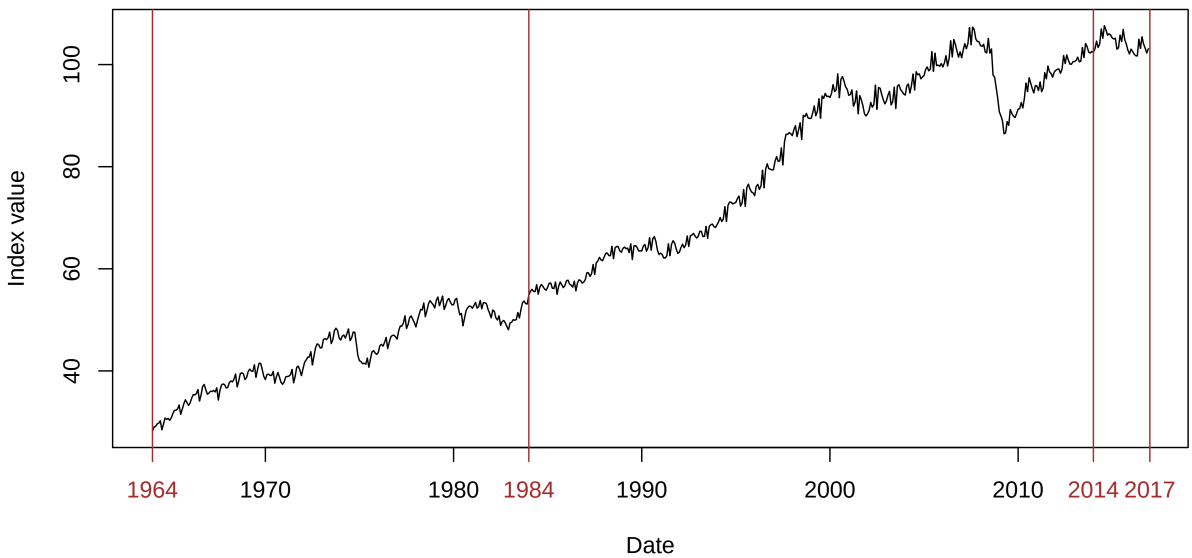

Figure 3.1: Original time series from January 1964 to December 2016. Time stamps for the training and testing periods are included.

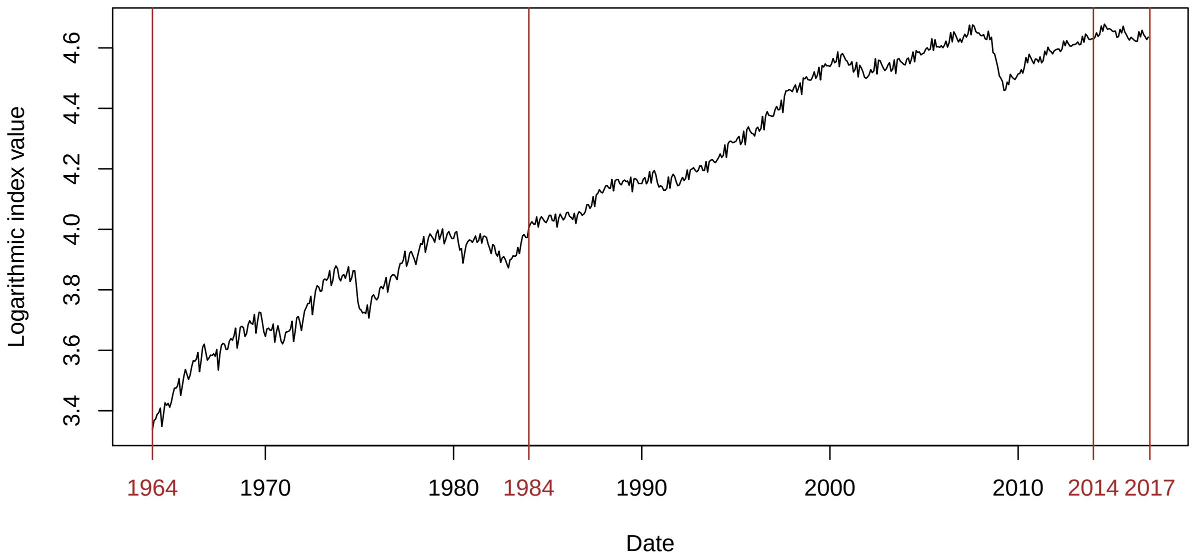

Figure 3.2: Logarithmized US industrial production index.

3.3 Data Manipulation

In order to achieve a stationary time series the two methods described in section 2.4 will be applied. The approach will be exemplary demonstrated on the 30 year training period. The preliminary trend estimation $\{\hat{m}\}$ of the two-sided filter and the Hodrick-Prescott filter is plotted in figure 3.3. The Hodrick-Prescott filter was applied with various values for λ, however, $\lambda = 14400$ resulted in best fit, which is common for monthly data and will be used throughout the remaining thesis.

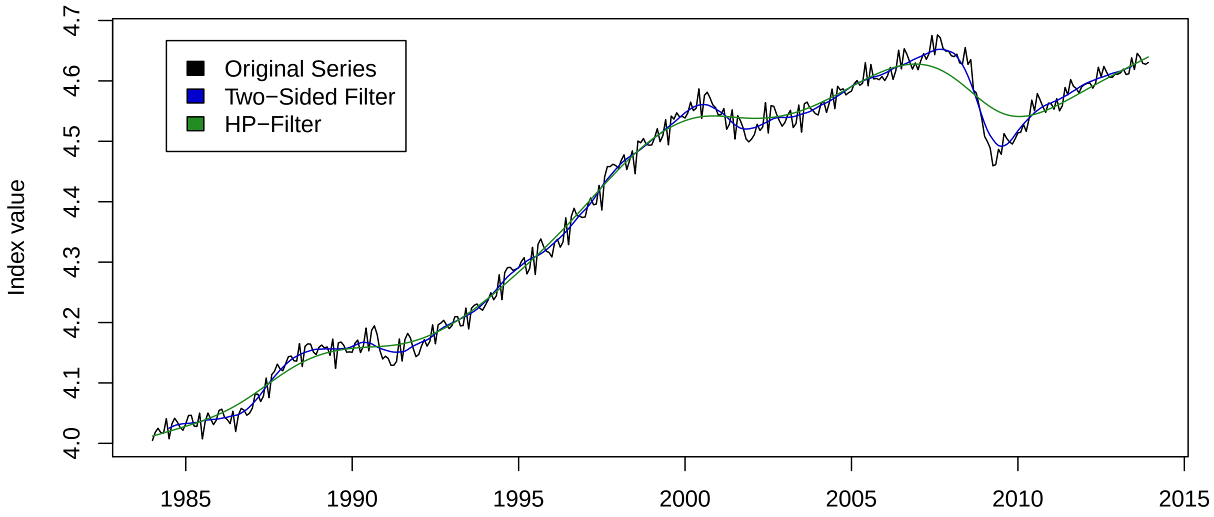

Figure 3.3: Comparison of the preliminary trend estimate of the two-sided filter and the Hodrick-Prescott filter.

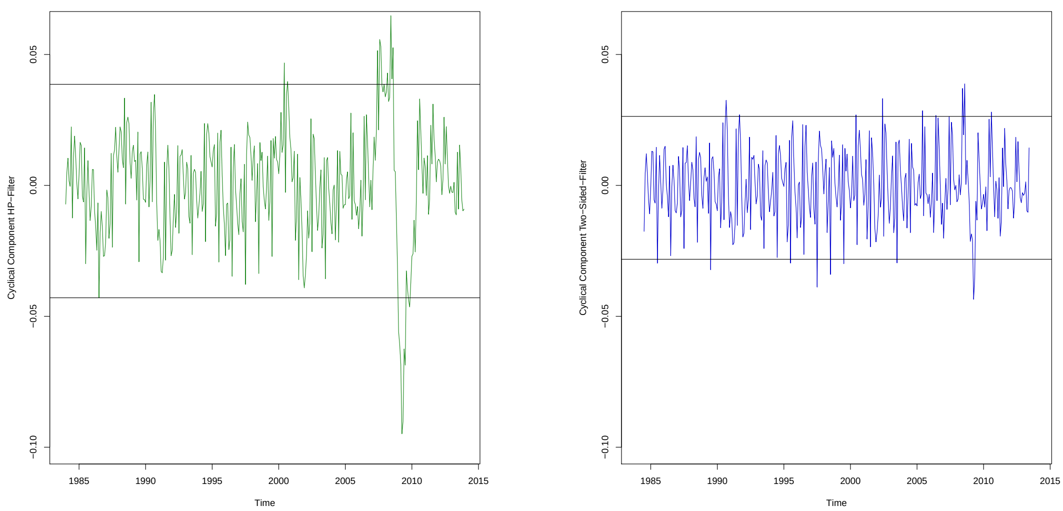

Figure 3.4: Remaining cyclical component of the Hodrick-Prescott and two-sided filter respectively.

An examination of the trend (Figure 3.3) and cyclical component (Figure 3.4) reveals that the applied two-sided filter results in closer proximity to the original data. This assumption is further confirmed by the density function of the cyclical components (5.2), where the cyclical components of the HP filter occupy more mass in the tails of the density function compared to the two-sided filter.

It is well known that unique outlier or an period with exceptional variance can result in biased estimators, if this period is not representative for the stochastic process and the training period not very large. Economic reasoning confirms this conjecture, because the outliers within past financial crisis may not reliably represent the current prudish economic state. These biases are especially perilous for the HP filter, because of the less aggressive smoothing approach. Thus, to the preliminary estimate of the cyclical component of some of the models a process known as *winsorization* will be applied. The idea behind winsorization is, to trim all outliers exceeding a given value to the maximum given value, which should represent usual fluctuations within the stochastic process. The lines plotted in figure 3.4 encompass 95% of the observations. Hence, every value exceeding the 97.5% highest value and each value that is below the 2.5% lowest value would be trimmed. The winsorized residuals of this example can be obtained from figure 5.3 and 5.4.

The next step of this analysis is extracting the seasonal component of the remaining cyclical component. Applying the algorithm elaborated in section 2.4 leads to figure 3.5. The final *deseasonalized* series can be obtained by deducing the seasonal effects from the original series, which is illustrated in figure 3.6.

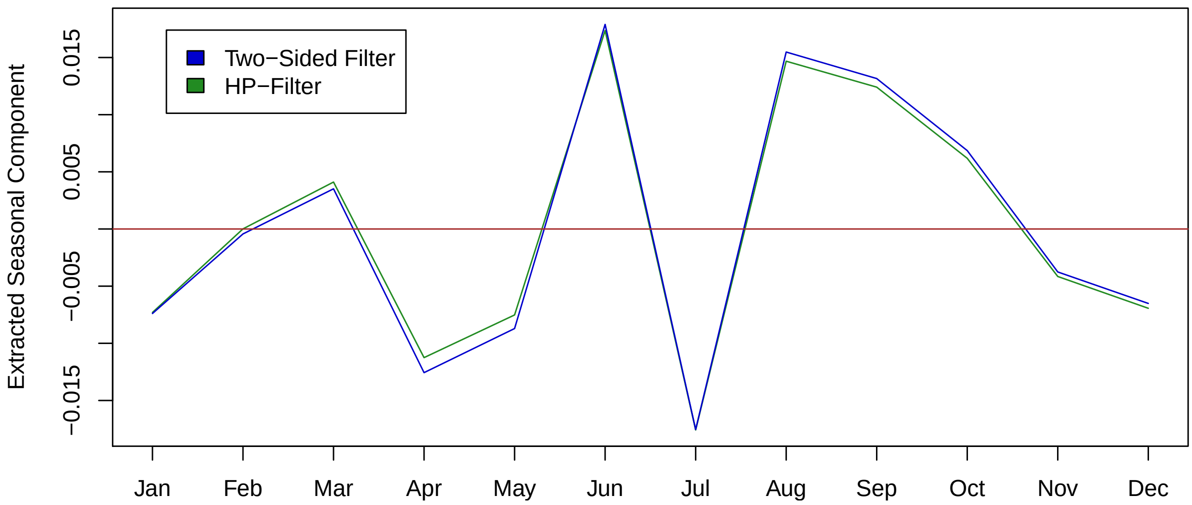

Figure 3.5: Comparison of the extracted seasonal component from the two-sided and HP filter for the 30 year horizon.

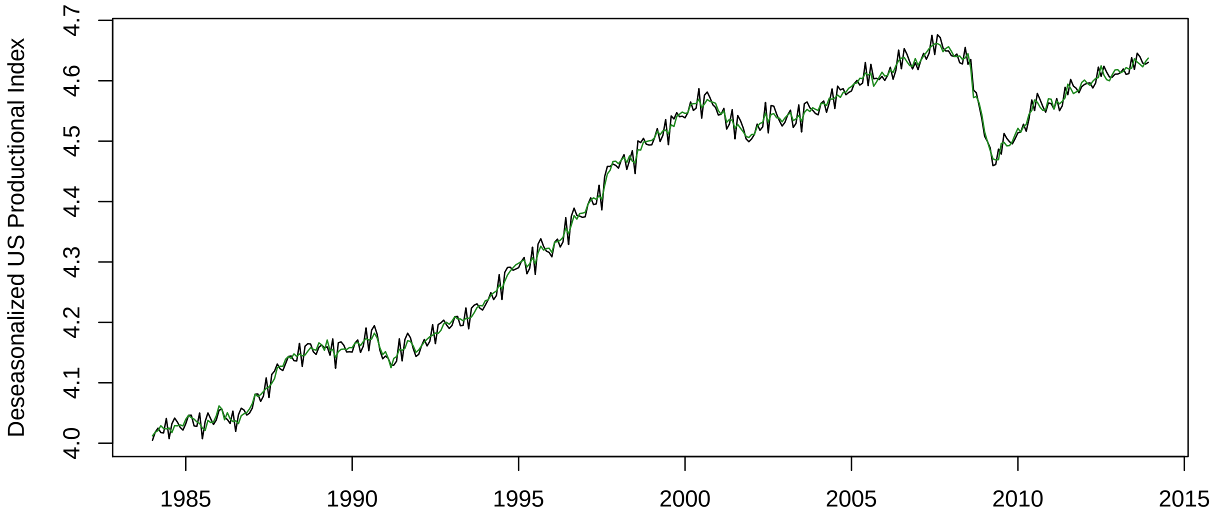

Figure 3.6: Comparison of original and deseasonalized time series.

Finally, one or multiple order differences are applied to obtain the ARIMA process. Examination of the first order difference ACF (Figure 5.5 and 5.6) shows dominant twelve months lags, which is an indication of seasonality. Thus, it can be concluded that the estimation of the seasonal effects did not vanish them entirely. As constants were assumed, one possible explanation is fluctuating seasonal variation over time, for example because of Februaries of differing lengths and leap years. However, application of the twelfth order difference results in ACF and PACF (Figure 3.7 and 3.8), which are typical for ARMA processes, because of its significant and rapidly declining ACF lags. Even though the seasonality did not disappear completely, its impact is considerably reduced and the series might prove as suitable for parameter estimation and forecasting. An application of another first order difference on the twelfth order differenced series eliminated almost all lags and reintroduced the dominant twelve month lag (Figure 5.7 and 5.8). Thus, it does not fit for forecasting purposes and this thesis will proceed with the twelfth order differenced seasonally adjusted series.

This transformed series will be compared with the twelfth order differenced original series (i.e. without previous seasonal adjustment) and the first order differenced X-12 Arima adjusted series in section 3.7. Their ACF and PACF can be obtained from figures 5.9 & 5.10 and 5.11 & 5.12 respectively.

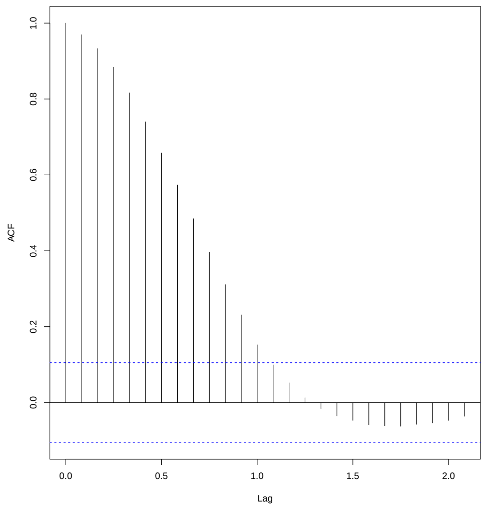

Figure 3.7: ACF of twelfth order difference two-sided filter seasonally adjusted series.

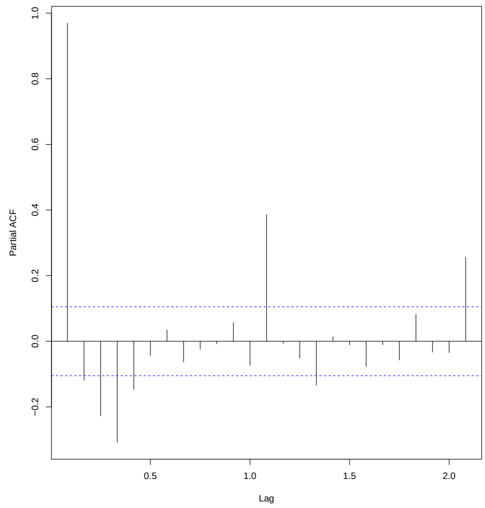

Figure 3.8: PACF of twelfth order difference two-sided filter seasonally adjusted series.

A further justification for the twelfth order differenced series is the significance level at which the ADF test rejects the unit root assumption. The computed p-values for the H_0 hypothesis suggest an appropriate purely autoregressive model with AR terms between 3rd and 11th order. This result is consistent with the plotted ACF, which exhibits non-significant lags past the 12th lag. However, this test does not account for additional MA coefficients, which will be involved in the following models. Thus, for now any model where $p + q \geq 4$ will be considered as competitive. Later, the remaining residuals will be tested for iid and inappropriate models discarded if necessary. The p-values and their respective assumed AR orders are tabulated below.

AR order	1	2	3	4	5	6	7	8	9	10	11	12	13
p-value	.3186	.0760	.01*	.01*	.01*	.01*	.01*	.01*	.01*	.01*	.01*	.2136	.2126

Table 3.1: ADF test p-values at different orders for the AR terms. Values denoted with an asterisk * are smaller than their printed number.

It is clear from examining the plotted twelfth order differences (Figure 3.9) that the 2009 financial crisis again distinctly exceeds the plotted 95% bounds and thus, further winsorization may be applied. Multiple models will be analysed in section 3.7, where the influences of winsorization *before identifying the seasonal influences*, *after seasonal adjustment*, *both* and *not at all* will be identified.

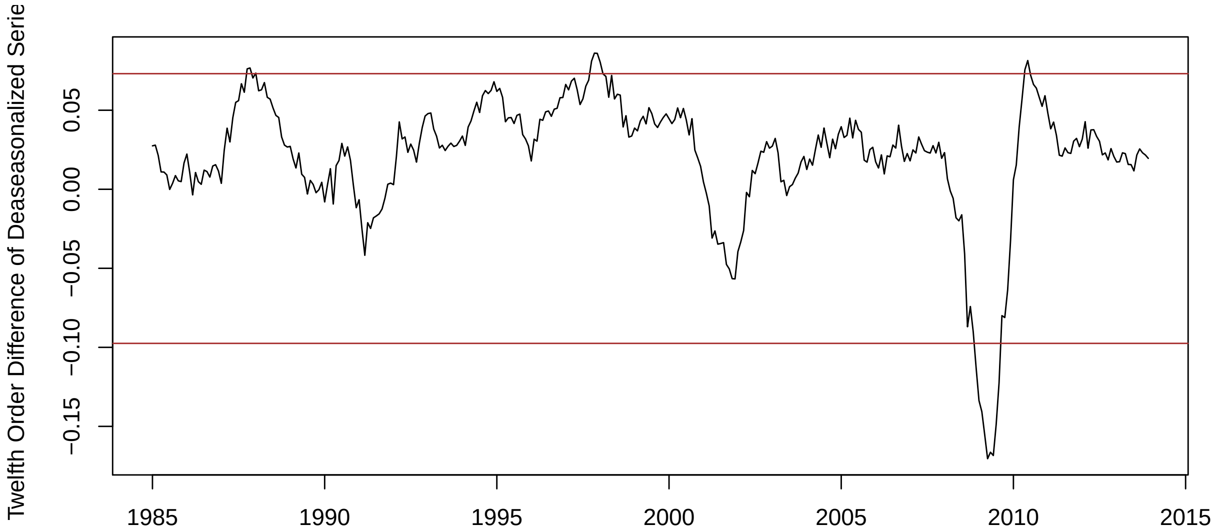

Figure 3.9: Deseasonalized series after taking the twelfth order difference.

3.4 Determining the Number of AR and MA Terms

First of all, it is important to acknowledge that there is in practice no *true* ARMA model, which does fit the transformed time series perfectly. Thus, it is necessary to make sophisticated guesses by relying on the ACF, PACF and information criteria. Afterwards, all sound models will be tested out-of-sample to conclude, which of these did perform best for the given test period.

In the following the deseasonalized and twelfth order differenced series will be exemplary considered, which was elaborated in section 3.3. Subsequently the potential numbers of AR and MA terms are tried to be identified. The ACF strongly suggests at least one AR term, whereas the PACF suggests multiple low-order MA terms. These conjectures can be further quantified by examining the AIC, BIC and AICc for multiple orders p and q. The resulting values for the BIC, calculated by maximum likelihood (ML) estimates are tabulated below. The suggested low order minimum $(p = 3, q = 3)$ is consistent with the ACF and the ADF, which suggests an AR order of at least 3. However, the BIC does also show a minimum at $(p = 4, q = 11)$, with a value of -2305.249.

		Order q				
		1	2	3	4	5
Order p	1	−2210.653	−2215.953	−2243.224	−2244.362	−2241.655
	2	−2212.873	−2208.410	−2247.603	−2250.371	−2247.205
	3	−2209.258	−2206.392	**−2259.737**	NA	NA
	4	NA	-2255.315	−2244.393	−2247.267	−2241.001
	5	NA	NA	−2238.781	NA	−2251.185

Table 3.2: BIC for differing orders p and q. Values denoted as NA did not result in stationary processes.

The resulting minimum AICc values, which are not tabulated for reasons of space, are -2367.787 and -2367.519 for $(p = 3, q = 13)$ and $(p = 4, q = 11)$ respectively. As previously explained, the AIC tends to specify higher models than the AICc. Because it is a common rule in case of doubt to prefer lower order models, the AIC will not be further elaborated. Thus, three potential models remain, whereas it is likely that $3 \leq p \leq 4$ would provide a reasonable fit.

3.5 Parameter Estimation and Forecasting

From now on, the estimated parameters will be denoted in the form of the following equation:

$$X_t = \phi_1 X_{t-1} + \ldots + \phi_p X_{t-p} + \theta_1 Z_{t-1} + \ldots + \theta_q Z_{t-q}.$$

Applying ML estimation, the parameters depicted in listing 5.1, 5.2 and 5.3 are estimated for each of the three previously identified models.

A preliminary examination of the coefficients and their respective standard errors suggests that Model 3 has a problem with insignificant coefficients, i.e. ar1 and ar2. This suggests, that the AR order of 3, which was recommended by BIC, may result in a better fit. However, insignificant estimators are not a problem per se in ARMA regression, since even low values determine cycles which are captured within the regression. Thus, all three models remain and the randomness of their residuals and the forecast suitability has to be determined.

A plot of the three-year horizon forecast reveals different approaches regarding the underlying stochastic process (Figure 3.10). The final forecast can be obtained by reversing the previous transformations, i.e. implementing the estimates as twelfth month differences, adding the seasonal components and applying the exponential function. These forecasts and the original time series are plotted in figure 3.11. The first impression indicates superiority of model 1. However, this assumption will be quantified and verified in the following section.

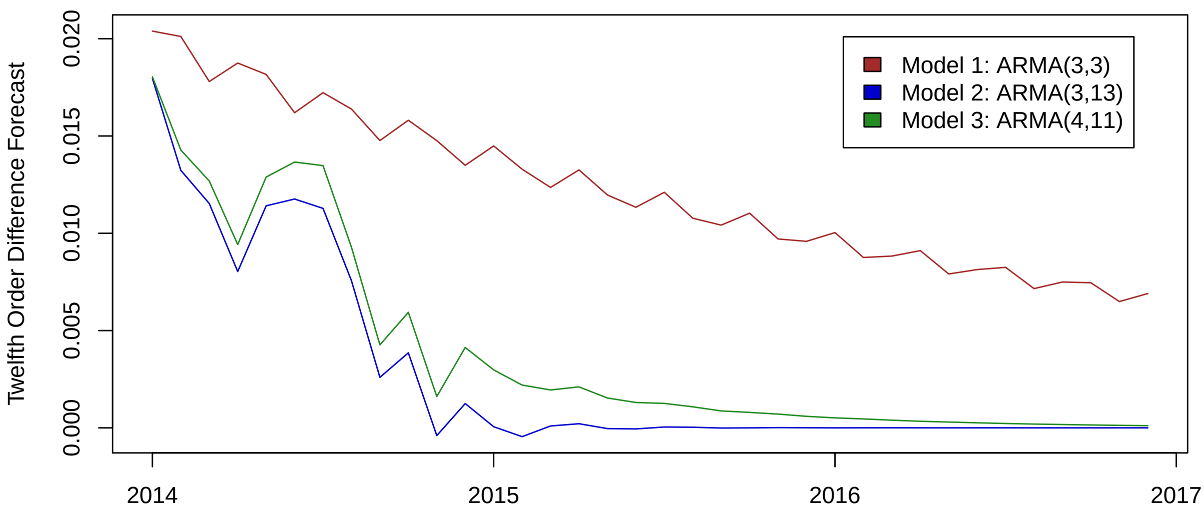

Figure 3.10: Twelfth order difference three-year horizon forecasts for Models 1, 2 and 3.

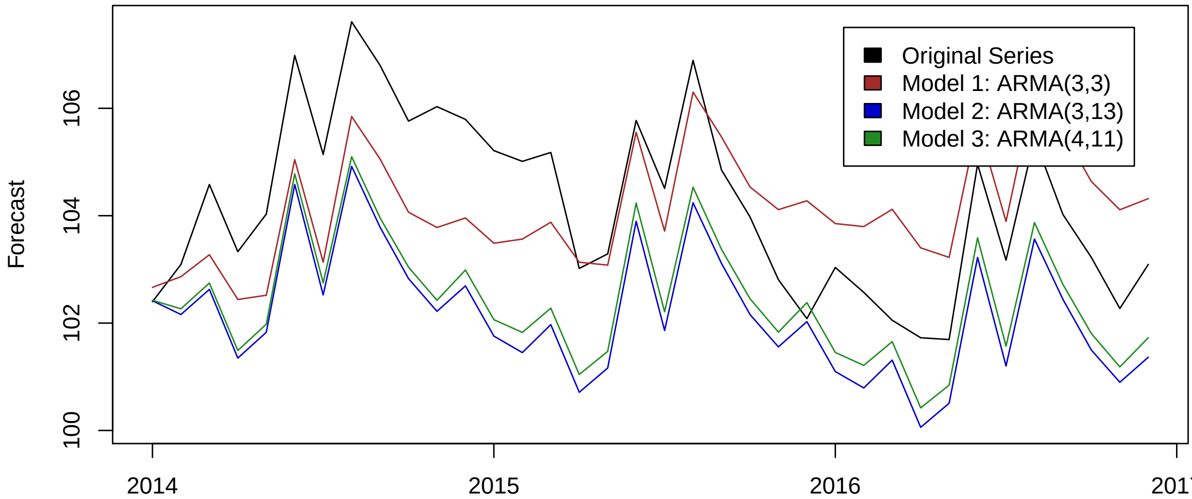

Figure 3.11: Three-year horizon forecasts for Models 1, 2 and 3.

3.6 Goodness of Fit

Three different approaches for measuring the goodness of fit were already discussed in section 2.11. Further, the ADF test was shown to reject the unit root hypothesis for AR orders 3 and 4 for any given significance level. Now residuals testing as well as MSE comparisons will be conducted. The diagnostic plots for model 2 and model 3 are very similar in nature, and thus every conjecture of model 2 does apply for model 3 as well. Model 3 diagnostics are pictured in figure 5.13 in the appendix chapter.

The graph of the rescaled residuals does not exhibit unusual trends or autocorrelations according to figure 3.12 and 3.13. However, the financial crisis of 2009 appears again as high variance aberration. Whereas the autocorrelation function of model 2 does not display any significant autocorrelations, the plot of model 1 is characterized by *multiple* exceedances of the plotted 95% confidence bounds. This sceptic is further justified by examining the p-vales of the Portmanteau test, which are below 0.05 for any lag $h \geq 2$. Thus, the hypothesis that the residuals of model 1 emerge from an iid distribution is rejected and genuine doubt put on the suitability of model 1. The p-values for Model 2 all boldly exceed the confidence bounds and do not raise any doubt on the iid hypothesis.

The outcome of the Portmanteau test is further strengthened by analysis of the turning point test. The p-values for model 1, model 2 and model 3 are 0.0677, 0.1486 and 0.1486 respectively, where H_0 is defined as iid residuals. Thus, the iid hypothesis can be rejected for model 1 at the 10% significance level, but cannot be rejected at any common significance level for model 2 and 3.

A final test is given by the MSE of the forecast horizon. Additional value will be attached on this test, since forecasting was defined as primary target. Anyway, it is obviously not the single criteria, because the model should not be overfitted to the chosen testing period. The MSE for model 1, model 2 and model 3 are 1.8903, 4.9395 and 3.8871 respectively. This result confirms the preliminary impression of the plotted forecast horizon in figure 3.11.

Summarizing, even though there are founded reasons to doubt the iid assumption of the residuals of model 1, it will not be discarded, because of its exceptional forecasting results. Furthermore, it is worth noting that the applied residual tests are highly restrictive by construction, since the confidence bounds rely on the iid assumption. Model 2 will be discarded because it is very similar to model 3, but does lead to inferior forecasting results. However, model 3 will be included in the further analysis, because of its in-sample adequacy and passing of all goodness of fit measures. These findings are pooled and tabulated below for reasons of clarity.

Model	p	q	Graph RR	ACF residuals	Portmanteau	turning point	MSE	Suitability
1	3	3	ordinary	significant	rejected	rejected	1.8903	eligible
2	3	13	ordinary	insignificant	not rejected	not rejected	4.9395	discarded
3	4	11	ordinary	insignificant	not rejected	not rejected	3.8871	eligible

Table 3.3: Summary of goodness of fit measures.

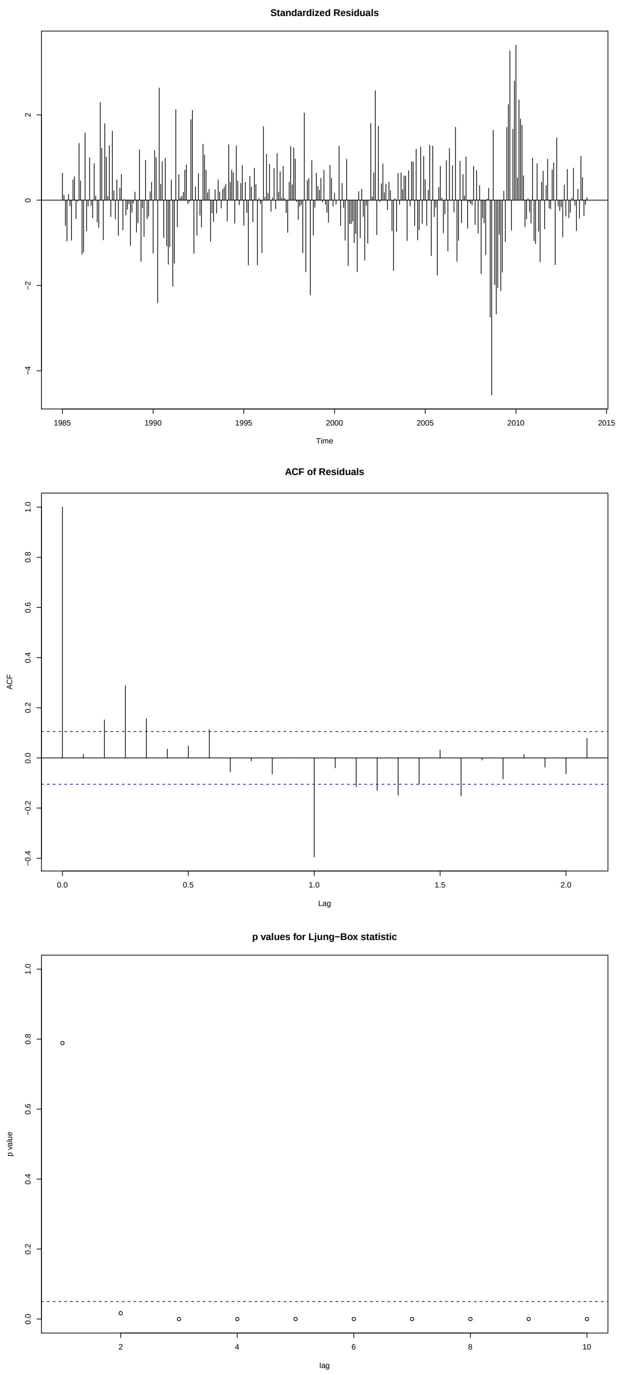

Figure 3.12: Diagnostics model 1.

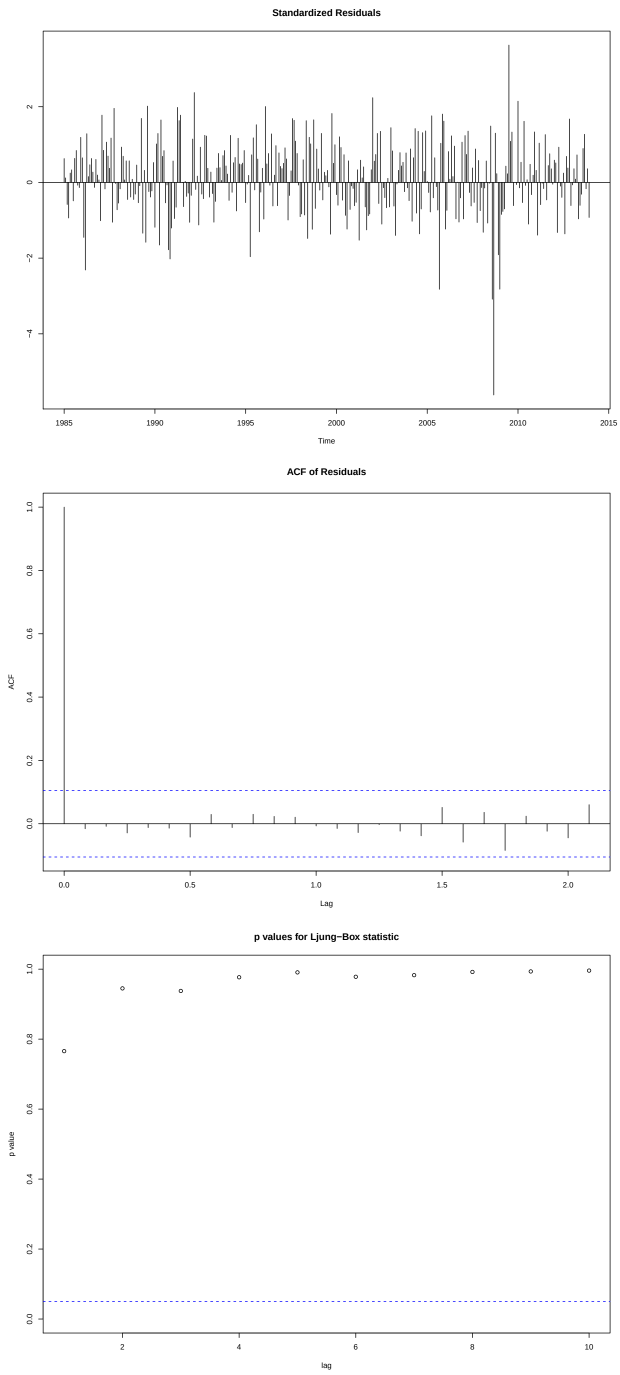

Figure 3.13: Diagnostics model 2.

3.7 Model Comparison

In this section 20 models will be scrutinized (Table 5.1), which have been constructed in a similar way as the two previously developed. First, specifications for the time horizon, winsorization and differencing have been fixed. Afterwards, two eligible models with different orders p and q for each combination of specifications were estimated. The models are chronologically sorted, i.e. model I and II, III and IV, V and VI etc. respectively differ only in respect to the order of AR and MA terms. The X-12-ARIMA deseasonalized models rely on data which was previously adjusted by the US census, all other models were adjusted with models previously demonstrated and explained. The testing period consists of unadjusted data from January 2014 to December 2016 for all but the X-12-ARIMA models, which were tested on X-12-ARIMA adjusted data in order to maintain comparability.

All orders p and q were chosen to constitute a BIC or AICc minimum, whereas one model constitutes a minimum for $(p \leq 16, q \leq 16)$ and the other one with the restriction $(p \leq 5, q \leq 5)$. The X-12-ARIMA adjusted models do not posses minima for orders higher than five and thus, are all within the stricter restriction.

Figure 3.14 demonstrates three deliberately chosen models, i.e. the model with the best fit overall VII, the best model result without previous seasonal adjustment X and the best model with double winsorization III. Inspection reveals that III is mostly driven by the seasonal components and thus does almost not all take the stochastic processes into account. Model X clearly underestimates the fluctuations within the first half of the time horizon. Model VII is the only model which provides a reasonable fit for the first 1.5 years, but also does slightly overestimate the fluctuations in the following months.

Figure 3.14: Three-year horizon forecasts for Models IV, VII and X.

Thorough examination of all models resulted in *six* hypotheses, which are applicable to ARIMA models of the US industrial production index for the inspected time horizon.

1. Low order models deliver better forecasting results.
Even though the information criteria resulted in lower values (i.e. preference) for the higher order models, there is no model in this sample with $p + q > 6$ *and* $MSE \leq 2$. Thus, it can be concluded that the underlying stochastic mechanism does follow a lower order process and information criteria tend to prefer models, which are overfitted to the training period. Further, it can be speculated if the AR order for the given test period is precisely 3, because the information criteria minima were in nine out of twelve cases at $p = 3$ for the 30 year horizon and these models scored the lowest MSE.

2. Simple seasonal adjustment methods improve forecasting results.
In the given sample, ARMA models fitted on time series, which were seasonally adjusted before taking the twelfth oder difference, were superior against unadjusted models, with the exception of the X-12-ARIMA adjustment method. Hence, taking differences does not suffice for eliminating disturbing seasonal components. As demonstrated in section 3.3, taking multiple differences is no solution either, because it suppresses the stochastic process entirely.

3. Applied winsorization methods delivered inferior forecasting results.
Thereby it did not matter, whether the winsorization was performed to extract the seasonal components or to eliminate additional noise within the series. Precisely, each winsorization procedure impaired the forecasting results ($MSE_V < MSE_I < MSE_{III}$). Therefore, all 'outliers' include information, which are representative for the whole stochastic process.

4. The 30 year horizon corresponds better than the 50 year horizon to the stochastic process.
No model fitted on the 50 year horizon delivered $MSE \leq 2.5$. Thus, some structural shifts within the period may be assumed, which disturb the estimation from the current underlying mechanism. The existence of past technological and political shifts strengthens this assumption.

5. The X-12-ARIMA adjustment is not suitable for ARMA forecasting.
It is obvious by examining the models XI, XII, XIX and XX that those delivered poor results. Further, the information criteria suggested very low orders of p and q. Consequently it is likely that this aggressive filter extracted most of the processes information and the remainder is almost following a white noise process.

6. There is no difference between the applied two-sided and HP filter seasonal adjustment for model building purposes.
All suitable models have identical MSE for the first decimal place, no matter which of these two linear filters were applied (V and VII, XIV and XVI). This discovery is not astonishing, because both filters led to very similar estimates of the seasonal component (Figure 3.5).

Chapter 4

Conclusion

The US industrial production index is a volatile time series dominated by trend and seasonality. However, trend and seasonality do not explain the entire historical development, because of the frequent presence of temporary shocks. Thus, this process is also driven by autocorrelated error terms, which can be modelled with autoregressive and moving-average parameters. In order to obtain an estimate of appropriate ARMA parameters for the US industrial production index multiple aspects have to be considered. First, models should not include too many parameters, to avoid the problem of overfitting. Further, seasonality should be removed by application of simple seasonal adjustment methods before the data will be differenced. This advise is especially important, because it is common practice to compare individually constructed forecasts with forecasts driven by ARMA models without prior seasonal adjustment. However, too sophisticated adjustment methods as well as winsorization can eliminate additional information included in the process. Finally, the time horizon on which the parameters are chosen has to be carefully selected to maximize obtained information of the trade-off between numerous past observations and potential structural shifts within the observation period.

It is important to acknowledge that the findings of this thesis rely solely on ARIMA processes for the monthly US industrial production index estimated within the period of January 1964 and December 2013, and assessed within January 2014 and December 2016. Other time horizons and data frequencies might result in other outcomes. Further, conditional heteroskedasticity, cointegrated and artificial neural network models often result in better fit [27]. Other models also describe the process as simple random walk [18]. Thus, these results may be further evaluated with different models as well as model specifications.

Chapter 5

Appendix

5.1 Plots and Outputs

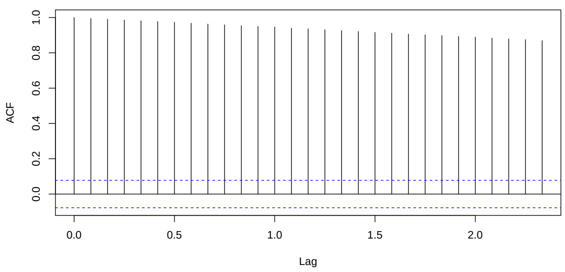

Figure 5.1: ACF of the untransformed US Industrial Production Index.

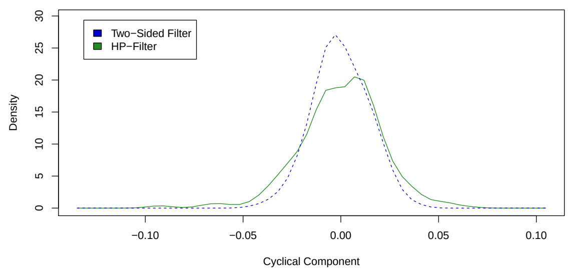

Figure 5.2: Density function of the preliminary cyclical component of the two-sided and HP filter.

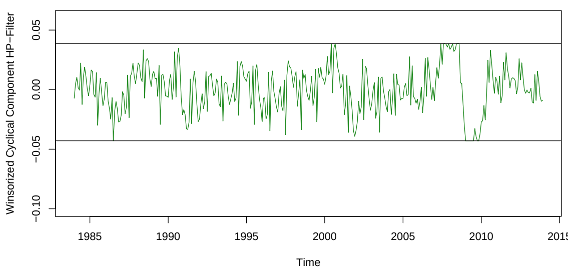

Figure 5.3: Winsorized preliminary cyclical component of the HP filter.

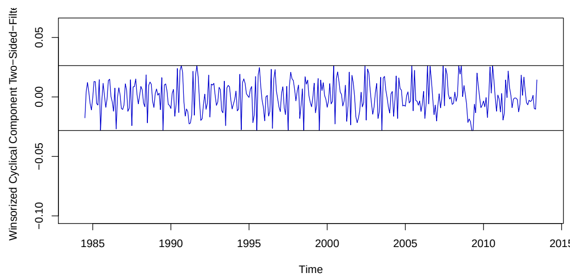

Figure 5.4: Winsorized preliminary cyclical component of the two-sided filter.

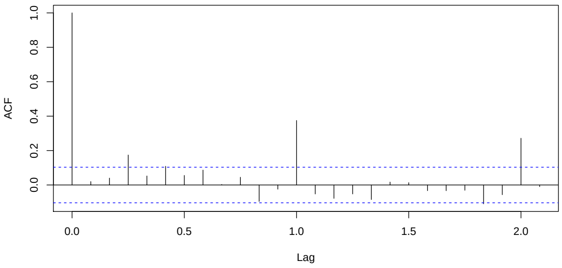

Figure 5.5: ACF of first order difference two-sided filter seasonally adjusted series.

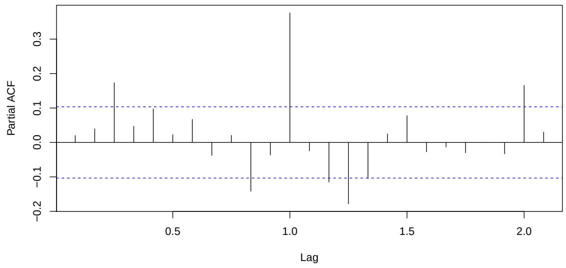

Figure 5.6: PACF of first order difference two-sided filter seasonally adjusted series.

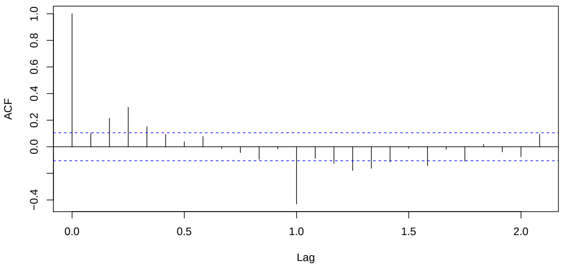

Figure 5.7: ACF of first and twelfth order difference two-sided filter seasonally adj. series.

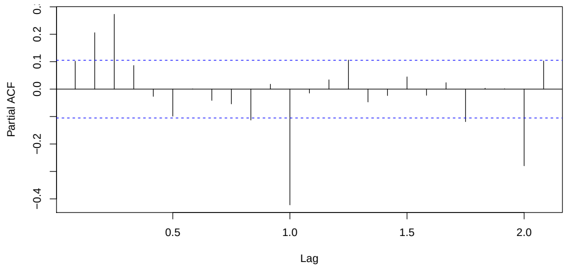

Figure 5.8: PACF of first and twelfth order difference two-sided filter seasonally adj. series.

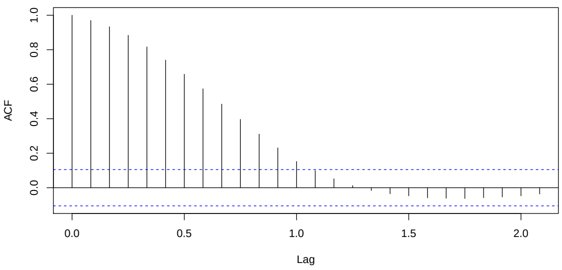

Figure 5.9: ACF of twelfth order difference original series.

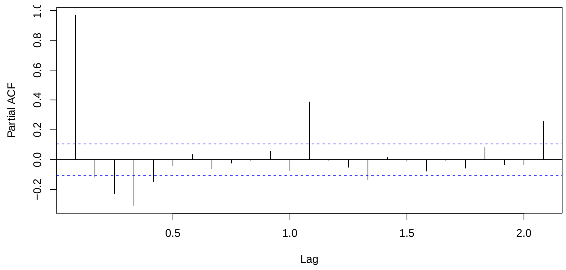

Figure 5.10: PACF of twelfth order difference original series.

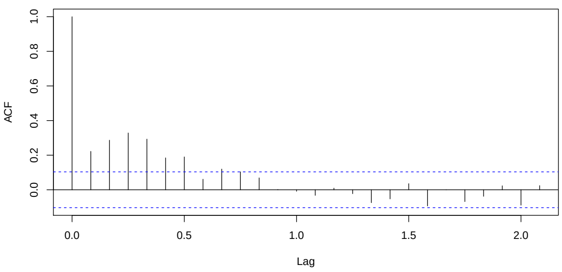

Figure 5.11: ACF of first order difference X-12 ARIMA adjusted series.

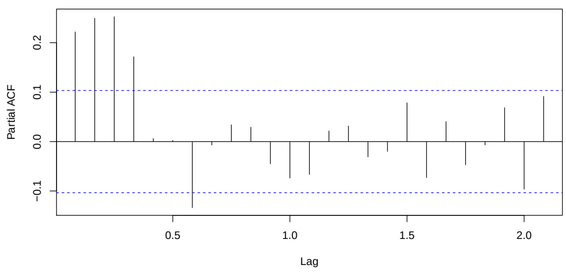

Figure 5.12: PACF of first order difference X-12 ARIMA adjusted series.

Listing 5.1: Model 1

```
Call:
arima(x = diff(deseason_mt_month_30 , 12), order = c(3, 0, 3), include.mean = FALSE)

Coefficients:
          ar1     ar2     ar3     ma1     ma2     ma3
      -0.1099  0.0979  0.9188  1.1851  1.1243  0.1164
s.e.   0.0241  0.0219  0.0216  0.0536  0.0595  0.0523

sigma^2 estimated as 9.061e-05:  log likelihood = 1122.23
```

Listing 5.2: Model 2

```
Call:
arima(x = diff(deseason_mt_month_30 , 12), order = c(3, 0, 13), include.mean = FALSE)

Coefficients:
         ar1      ar2     ar3     ma1     ma2     ma3     ma4     ma5     ma6     ma7
      0.6412  -0.2903  0.3229  0.3591  0.8175  0.6822  0.6136  0.6931  0.6501  0.6235
s.e.  0.1708   0.1869  0.1018  0.1703  0.1682  0.1474  0.1674  0.1857  0.1850  0.1914

         ma8     ma9    ma10    ma11     ma12    ma13
      0.6412  0.6102  0.5421  0.5414  -0.1373  0.2997
s.e.  0.1915  0.1778  0.1668  0.1537   0.1255  0.1323

sigma^2 estimated as 5.74e-05:  log likelihood = 1199.99
```

Listing 5.3: Model 3

```
Call:
arima(x = diff(deseason_mt_month_30 , 12), order = c(4, 0, 11), include.mean = FALSE)

Coefficients:
         ar1     ar2     ar3     ar4     ma1     ma2     ma3     ma4     ma5     ma6     ma7
     -0.0486  0.0904  0.3389  0.2398  1.0430  1.1067  1.0338  0.8575  0.8738  0.8587  0.8230
s.e.  0.0752  0.0814  0.0675  0.0851  0.0603  0.0911  0.0974  0.0812  0.0820  0.0913  0.1025

         ma8     ma9  ma10    ma11
      0.8279  0.80460.7242  0.6770
s.e.  0.1094  0.10210.0775  0.0468

sigma^2 estimated as 5.761e-05:  log likelihood = 1199.44
```

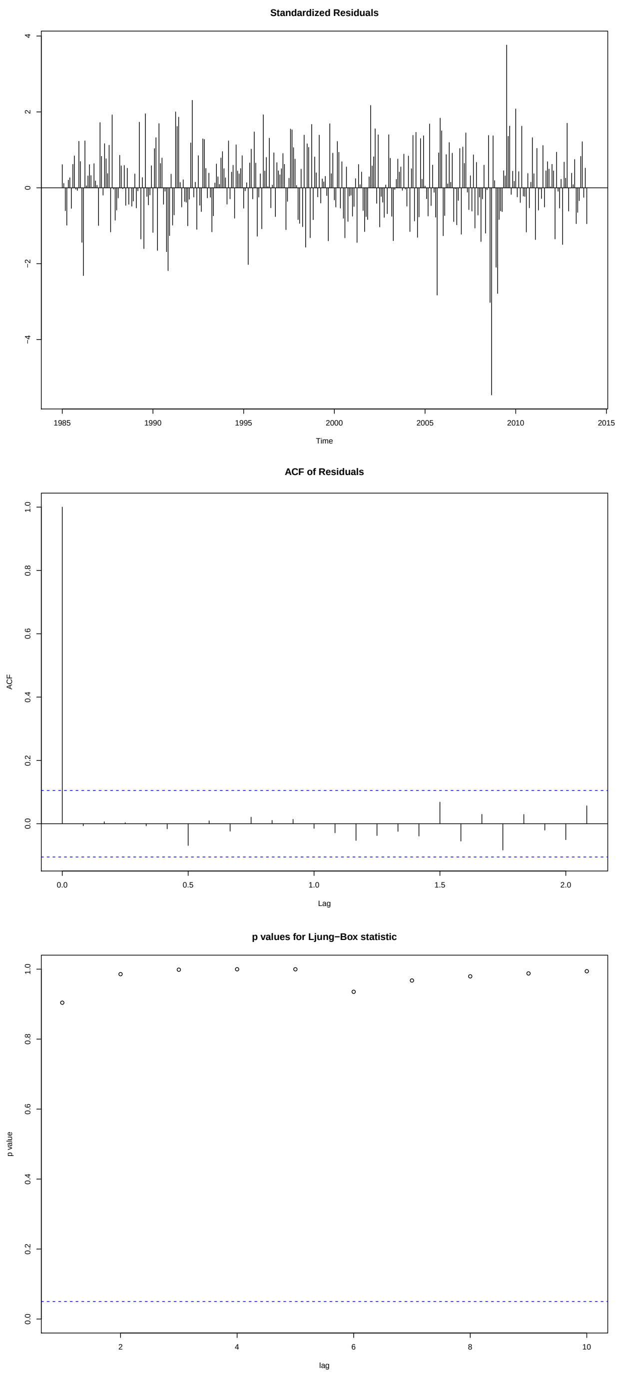

Figure 5.13: Plotted rescaled residuals, ACF of the residuals and p-values for the portmanteau test of model 3.

5.2 Models

Model	p	q	TH	Seasonal Filter	Wins. SF	Wins. Series	Differencing	ADF	Portmanteau	Turning point	BIC	MSE
I	3	3	30	TS	yes	no	$1-B^{12}$	.01*	$*$	.0677	-2203	1.8903
II	4	11	30	TS	yes	no	$1-B^{12}$	.01*	$**$	.1486	-2305	3.8871
III	3	3	30	TS	yes	yes	$1-B^{12}$	.01*	$**$	.3499	-2298	3.6917
IV	6	11	30	TS	yes	yes	$1-B^{12}$	.01*	$**$	.2342	-2307	4.5981
V	3	3	30	TS	no	no	$1-B^{12}$	.01*	$*$	.0677	-2203	1.8876
VI	3	11	30	TS	no	no	$1-B^{12}$	.01*	$**$	.1486	-2304	3.5896
VII	3	3	30	HP	no	no	$1-B^{12}$	.01*	$*$	.0677	-2203	1.8622
VIII	3	11	30	HP	no	no	$1-B^{12}$	.01*	$**$	.1486	-2304	3.9210
IX	3	3	30	none	no	no	$1-B^{12}$	.01*	$*$	.0677	-2203	3.8105
X	3	11	30	none	no	no	$1-B^{12}$	.01*	$**$	.1486	-2304	2.4108
XI	1	2	30	X-12-ARIMA	no	no	$1-B$	.01*	$**$	.1028	-2674	2.9989
XII	3	3	30	X-12-ARIMA	no	no	$1-B$	.01*	$**$	.1321	-2622	3.1962
XIII	4	2	50	TS	no	no	$1-B^{12}$	.01*	$**$	.6712	-3574	6.3568
XIV	1	15	50	TS	no	no	$1-B^{12}$	.01805	$**$	.7939	-3706	2.7269
XV	4	2	50	HP	no	no	$1-B^{12}$	.01*	$**$	.6712	-3574	6.3533
XVI	1	15	50	HP	no	no	$1-B^{12}$	.01805	$**$	.7939	-3706	2.7284
XVII	4	2	50	none	no	no	$1-B^{12}$	.01*	$**$	.6712	-3574	2.9845
XVIII	2	11	50	none	no	no	$1-B^{12}$	.01*	$**$	.7440	-3716	4.5187
XIX	1	2	50	X-12-ARIMA	no	no	$1-B$	.01*	$**$	.1028	-4236	3.1159
XX	2	2	50	X-12-ARIMA	no	no	$1-B$	.01*	$**$	.1321	-4230	3.1001

$p \mathrel{\widehat{=}}$ order AR terms; $q \mathrel{\widehat{=}}$ order MA terms; TH $\widehat{=}$ utilized time horizon for parameter estimation in years;
TS $\widehat{=}$ two-sided seasonal filter; HP $\widehat{=}$ Hodrick-Prescott derived seasonal filter;
Wins. SF $\widehat{=}$ seasonal filter has been determined with winsorized series; Wins. Series $\widehat{=}$ seasonally adjusted series has been winsorized;
ADF $\widehat{=}$ ADF test applied with p AR terms; Portmanteau $\widehat{=}$ Portmanteau test has been significant for at least one lag ($*$)
or at least last ten ($**$); turning point $\widehat{=}$ p-value of turning point test; values denoted with an asterisk * are smaller than their printed value

Table 5.1: Summary of 20 ARIMA models with different approaches regarding time horizon, winsorization and differencing. Orders p and q have been chosen in order to obtain a local BIC minimum as well as providing forecast suitability.

5.3 Box-Jenkins methodology

In this section will be the Box-Jenkins methodology briefly examined on four examples in order to identify proper ARMA models to the underlying stochastic process [5]. The analysis will be conducted on a generated sample size of 100 observations of each given process. The Box-Jenkins methodology can be extended to more sophisticated process, but nevertheless prerequisites sound knowledge of fundamental AR and MA processes. The following conclusions will rely on the properties of the theoretical ACF and PACF summarized below.

Process	ACF	PACF
AR(p)	(exponentially) declining	$\alpha(h) = 0$ for $h > p$
MA(q)	$\rho(h) = 0$ for $h > q$	(exponentially) declining
ARMA(p,q)	(exponentially) declining	(exponentially) declining

White noise process

The first process (5.14) is defined as $Z_t \sim \text{IID}(0,1)$ and can be easily identified by autocorrelation and partial autocorrelation function values inside the 95% confidence bounds $\pm 1.96/\sqrt{n}$ for lags $h > 0$. The same realization of $\{Z\}$ will be used in the following examples.

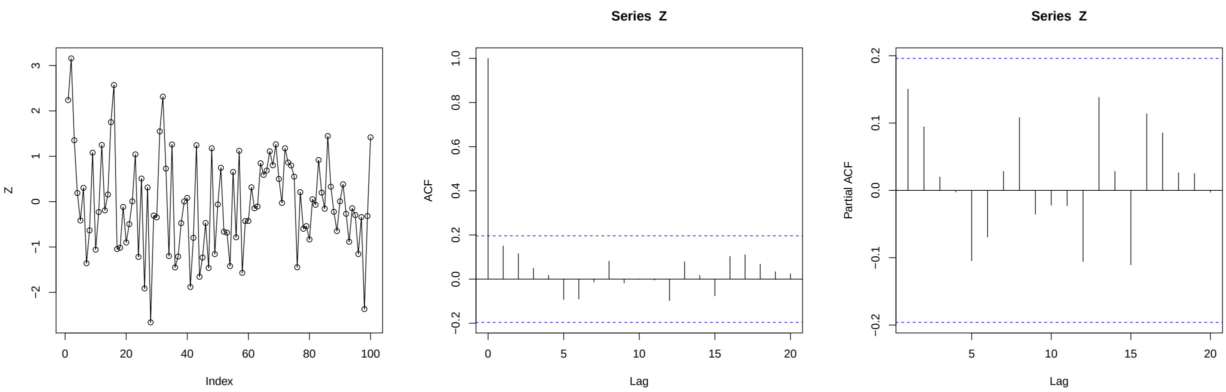

Figure 5.14: $Z_t \sim \text{IID}(0,1)$: Plot, ACF and PACF

AR(1) process

The second process (5.15) is defined as $X_t = 0.8X_{t-1} + Z_t$ where $Z_t \sim \text{IID}(0,1)$. It it obvious to assume an AR(1) process, because of the rapidly declining ACF and non-significant values for the PACF past the first lag.

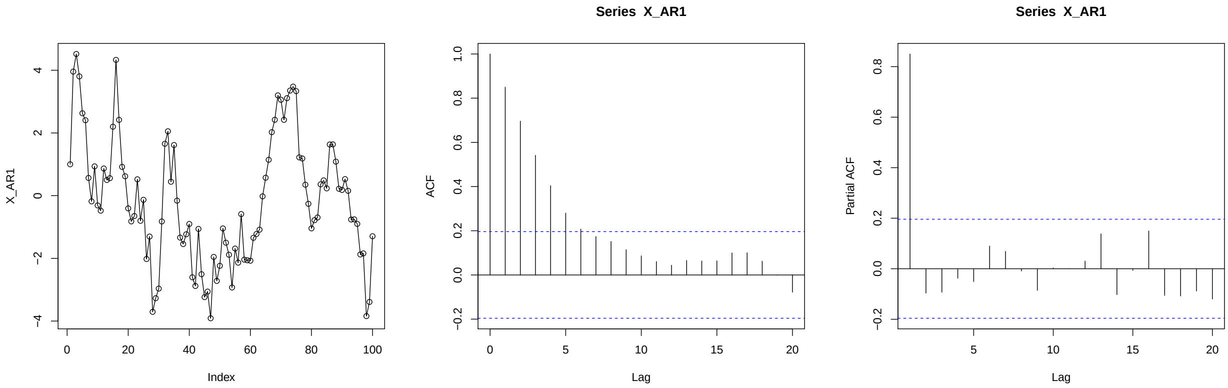

Figure 5.15: $X_t = 0.8X_{t-1} + Z_t$: Plot, ACF and PACF

MA(1) process

The third process (5.16) is defined as $X_t = Z_t + 0.9Z_{t-1}$ where $Z_t \sim \text{IID}(0,1)$. It makes sense to assume an MA(1) process, because of the rapidly declining PACF and non-significant values for the ACF past the first lag. However, the significant lags of the PACF may be further examined.

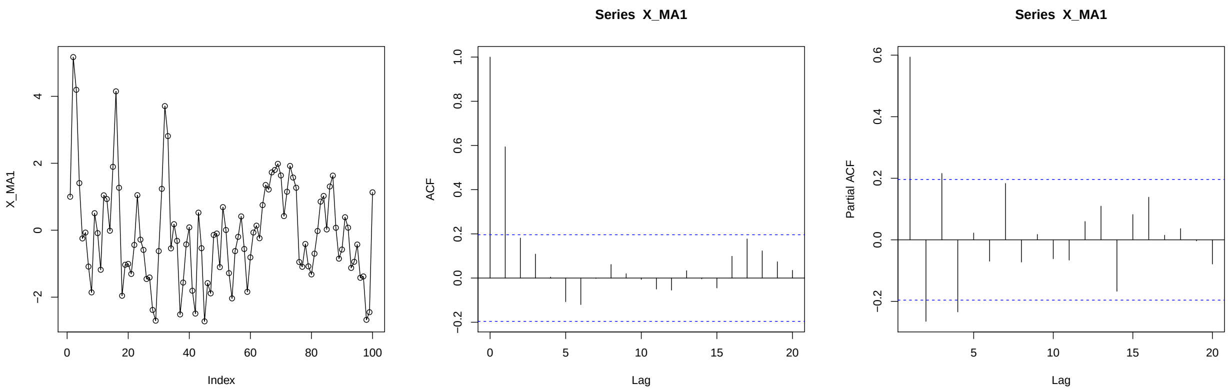

Figure 5.16: $X_t = Z_t + 0.9Z_{t-1}$: Plot, ACF and PACF

ARMA(1,1) process
The fourth process (5.17) is defined as $X_t = Z_t + 0.9Z_{t-1} + 0.8X_{t-1}$ where $Z_t \sim \text{IID}(0, 1)$. It makes sense to assume an ARMA(1,1) process, because this time series exhibits an ACF characteristic for an AR(1) process as well as a PACF characteristic for an MA(1) process.

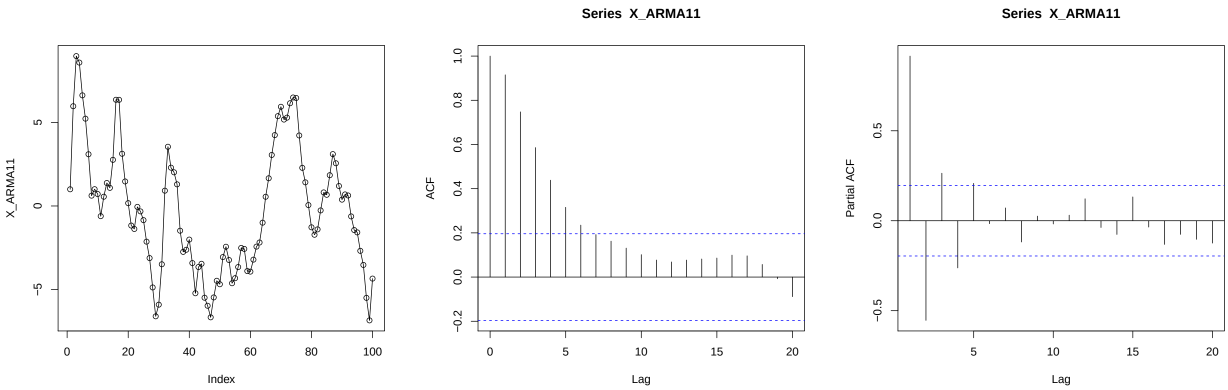

Figure 5.17: $X_t = Z_t + 0.9Z_{t-1} + 0.8X_{t-1}$ where $Z_t \sim \text{IID}(0, 1)$: Plot, ACF and PACF

Further processes
Finally, it is worth considering some findings [5] on more elaborated processes:

1. Mixed processes are indicated by exponentially declining autocorrelation and partial autocorrelation functions. Precisely, after lag $q - p$ the theoretical autocorrelations of the mixed process are similar to those of a pure autoregressive process.

2. Nonstationary processes, i.e. processes with a root close to unity, can be identified by an autocorrelation function that does not decrease quickly. Thus, it may be required to apply first or higher order differences on the data until the autocorrelation function dies out rapidly.

5.4 Innovation Algorithm

In this section a generic parameter estimation applying the innovation algorithm will be conducted. The innovation algorithm can be applied to all stochastic processes with finite second moments and thus, does not require the stationarity assumption. However, some notations will be simplified for the sake of ARMA parameter estimation according to definition 2.3. For notational convenience the notation for linear predictors in this section is simplified as following: $\hat{X}_n = \mathbb{P}_{n-1}X_n$.

The causal ARMA(p,q) process will be assumed:

$$\phi(B)X_t = \theta(B)Z_t, \quad \{Z_t\} \sim \mathrm{WN}(0, \sigma^2).$$

The causality assumption implies that it can be represented as following:

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j}.$$

where $\sum_{j=0}^{\infty} \psi_j z^j = \theta(z)/\phi(z), \; |z| \leq 1$.
Further, it will be defined that $m = \max(p, q)$. The algorithm can be simplified by applying it to $\{W_t\}$ instead of $\{X_t\}$ directly:

$$W_t = \begin{cases} \sigma^{-1}X_t, & t = 1, \ldots, m, \\ \sigma^{-1}\phi(B)X_t, & t > m. \end{cases}$$

The innovation algorithm relies on the autocovariance function of the underlying process. Thus, it will be specified that $E(W_iW_j) = \kappa(i,j)$:

$$\kappa(i,j) = \begin{cases} \sigma^{-2}\gamma_x(i-j), & 1 \leq i, \; j \leq m, \\ \sigma^{-2}\left[\gamma_x(i-j) - \sum_{r=1}^{p} \phi_r \gamma_x(r - |i-j|)\right], & \min(i,j) \leq m < \max(i,j) \leq 2m, \\ \sum_{r=0}^{q} \theta_r \theta_{r+|i-j|}, & \min(i,j) > m, \\ 0, & \textit{otherwise}. \end{cases}$$

Where the coefficients ϕ have been previously estimated, for example by the Hannan-Rissannen algorithm. The autocovariance function of $\{X_t\}$ can be easily obtained by:

$$\gamma_x(h) = \mathbb{E}(X_{t+h}X_t) = \sigma^2 \sum_{j=0}^{\infty} \psi_j \psi_{j+|h|}.$$

It is obvious by the structure of the autocovariance function of $\{X_t\}$ that σ^2 cancels out within the autocovariance function of $\{W_t\}$. Applying the recursive innovation algorithm results in the coefficients $\theta_{n1}, \ldots, \theta_{nn}$ and the mean squared errors $r_n = \mathbb{E}(W_{n+1} - \hat{W}_{n+1})^2$:

$$
\begin{aligned}
r_0 &= \kappa(1,1), \\
\theta_{n,n-k} &= r_k^{-1}\left(\kappa(n+1,k+1) - \sum_{j=0}^{k-1} \theta_{k,k-j}\theta_{n,n-j}r_j\right), \quad 0 \le k < n, \\
r_n &= \kappa(n+1,n+1) - \sum_{j=0}^{n-1} \theta_{n,n-j}^2 r_j.
\end{aligned}
$$

Thus, first r_0 would be computed and then proceeded with θ_{11}, r_1; θ_{22}, θ_{21}, r_2 and so forth.

Finally, the stochastic process can be estimated:

$$
\hat{X}_{n+1} = \begin{cases} \displaystyle\sum_{j=1}^{n} \theta_{nj}(X_{n+1-j} - \hat{X}_{n+1-j}), & 1 \le n < m, \\ \displaystyle\phi_1 X_n + \ldots + \phi_p X_{n+1-p} + \sum_{j=1}^{n} \theta_{nj}(X_{n+1-j} - \hat{X}_{n+1-j}), & n \ge m. \end{cases}
$$

The variance of the one step prediction errors ν_n is therefore:

$$
\nu_n = \mathbb{E}(X_{n+1} - \hat{X}_{n+1})^2 = \sigma^2 \mathbb{E}(W_{n+1} - \hat{W}_{n+1})^2 = \sigma^2 r_n.
$$

5.5 Table of Abbreviations

ACF	autocorrelation function
AIC	Akaike information criterion
AICc	small sample size corrected version of the Akaike information criterion
AR	autoregressive
ARMA	autoregressive moving average
ARIMA	autoregressive integrated moving average
ADF	augmented Dickey-Fuller
BIC	Bayesian information criterion
DF	Dickey-Fuller
HP	Hodrick-Prescott
iid	independent and identically distributed
MA	moving average
ML	maximum likelihood
OLS	ordinary least squares
PACF	partial autocorrelation function
X-12 ARIMA	seasonal adjustment method of the U.S. Census Bureau

Bibliography

[1] Federal reserve bulletin, March 2001.

[2] H. Akaike. Information theory and an extension of the maximum likelihood principle. *2nd International Symposium on Information Theory*, pages B.N. Petrov and F. Csaki (eds.), Akademiai Kiado, Budapest, 267–281, 1973.

[3] H. Akaike. Time series analysis and control through parametric models. In *Applied Time Series Analysis I*, pages 1–23. Elsevier, 1978.

[4] P. Billingsley. *Probability and Measure, 3rd ed. (Wiley Series in Probability and Statistics).* 1995.

[5] G. E. P. Box, G. M. Jenkins, and G. C. Reinsel. *Time series analysis: Forecasting and control / George E.P. Box, Gwilym M. Jenkins, Gregory C. Reinsel.* Wiley series in probability and statistics. Wiley, Oxford, 4th ed. edition, 2008.

[6] G. E. P. Box and D. A. Pierce. Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. *Journal of the American Statistical Association*, 65(332):1509, 1970.

[7] P. J. Brockwell and R. A. Davis. *Introduction to Time Series and Forecasting.* Springer Texts in Statistics. Springer International Publishing and Imprint: Springer, Cham, 3rd ed. 2016 edition, 2016.

[8] P. S. P. Cowpertwait and A. V. Metcalfe. *Introductory time series with R.* Springer, Dordrecht, 2009.

[9] D. A. Dickey and W. A. Fuller. Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74(366):427, 1979.

[10] D. A. Dickey and W. A. Fuller. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, 49(4):1057, 1981.

[11] Q. P. Duong. On the choice of the order of autoregressive models: a ranking and selection approach. *Journal of Time Series Analysis*, 5(3):145–157, 1984.

[12] J. Durbin. The fitting of time-series models. *Revue de l'Institut International de Statistique / Review of the International Statistical Institute*, 28(3):233, 1960.

[13] Federal Reserve statistical release. Industrial production and capacity utilization, November 16, 2017.

[14] E. J. Hannan. The estimation of the order of an arma process. *Ann. Stat., 8*, pages 1071–1081, 1980.

[15] R. J. Hodrick and E. C. Prescott. Postwar u.s. business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, 29(1):1, 1997.

[16] C. M. Hurvich and C.-L. Tsai. Regression and time series model selection in small samples. *Biometrika*, 76(2):297, 1989.

[17] R. G. King and S. T. Rebelo. Low frequency filtering and real business cycles. *Journal of Economic Dynamics and Control*, 17(1-2):207–231, 1993.

[18] R. Krol. Trends, random walks and persistence: An empirical study of disaggregated u.s. industrial production. *Review of Economics and Statistics*, 74(1):154, 1992.

[19] E. L. Lehmann and G. Casella. *Theory of point estimation*. Springer Texts in Statistics. Springer, New York and London, 2nd ed. / e.l. lehmann, george casella edition, 1998.

[20] G. M. Ljung and G. E. P. Box. On a measure of lack of fit in time series models. *Biometrika*, 65(2):297, 1978.

[21] R. E. Lucas. Methods and problems in business cycle theory. *Journal of Money, Credit and Banking*, 12(4):696, 1980.

[22] J. G. MacKinnon. Numerical distribution functions for unit root and cointegration tests. *Journal of Applied Econometrics*, 11(6):601–618, 1996.

[23] C. R. Nelson and C. R. Plosser. Trends and random walks in macroeconmic time series. *Journal of Monetary Economics*, 10(2):139–162, 1982.

[24] K. Neusser. *Time Series Analysis in Economics*. Bern, 2015.

[25] M. O. Ravn and H. Uhlig. On adjusting the hodrick-prescott filter for the frequency of observations. *Review of Economics and Statistics*, 84(2):371–376, 2002.

[26] S. E. Said and D. A. Dickey. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71(3):599–607, 1984.

[27] N. R. Swanson and H. White. A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks. *Review of Economics and Statistics*, 79(4):540–550, 1997.