Tomáš Polák

Predikce transformovaných časových řad

Katedra pravděpodobnosti a matematické statistiky
Vedoucí diplomové práce: Prof. RNDr. Jiří Anděl, DrSc.
Studijní program: Matematika
Studijní obor: Pravděpodobnost, matematická statistika a ekonometrie
Studijní plán: Matematická statistika
Rád bych poděkoval vedoucímu mé diplomové práce Prof. RNDr. Jiřímu Andělovi, DrSc. za pomoc při vedení této práce a za cenné rady a připomínky.

Prohlašuji, že jsem svou diplomovou práci napsal samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce.

V Praze dne 18. dubna 2008

Tomáš Polák
Prediction of transformed time series

Department of Probability and Mathematical Statistics
Supervisor: Prof. RNDr. Jiří Anděl, DrSc.
Study program: Matematika
Study specialization: Probability, Mathematical Statistics and Econometrics
Study plan: Mathematical Statistics
CONTENTS

4.3.3 Cubic transformation ................................................. 40
4.4 ARMA(1, 1) process ..................................................... 41
  4.4.1 Exponential transformation ........................................ 42
  4.4.2 Quadratic transformation .......................................... 43
  4.4.3 Cubic transformation ................................................ 43

5 Empirical part .............................................................. 45

6 Conclusion ......................................................................... 49

7 Appendix A - Properties of Hermite polynomials ................... 50

8 Appendix B - Best linear prediction using spectral density ...... 53
Název práce: Predikce transformovaných časových řad  
Autor: Tomáš Polák  
Katedra: Katedra pravděpodobnosti a matematické statistiky  
Vedoucí diplomové práce: Prof. RNDr. Jiří Anděl, DrSc.  
e-mail vedoucího: jiri.andel@mff.cuni.cz  
Abstrakt: Cílem této práce je najít predikci pro nelineární transformaci časové řady. Za jistých předpokladů týkajících se původní časové řady je nejprve zkoumána autokovarianční funkce a spektrální hustota transformované řady. Obecná tvrzení jsou aplikována na konkrétní příklady ARMA procesů. Dále jsou uvedeny obecné vzorce pro predikce transformované časové řady, pro které není nutná znalost autokovarianční funkce transformované řady ani její spektrální hustoty. Tyto vzorce jsou aplikovány na tři konkrétní transformace a jsou odvozeny explicitní vzorce pro ARMA procesy. Tři typy predikcí (optimální, naivní a lineární) jsou porovnány ve smyslu proporcionalního nárůstu střední kvadratické chyby. Explicitní vzorce pro ARMA procesy jsou ověřeny pomocí simulace.  
Klíčová slova: Predikce, časové řady, nelineární transformace, Hermitovské polynomy, ARMA procesy

Title: Prediction of transformed time series  
Author: Tomáš Polák  
Department: Department of Probability and Mathematical Statistics  
Supervisor: Prof. RNDr. Jiří Anděl, DrSc.  
Supervisor’s e-mail address: jiri.andel@mff.cuni.cz  
Abstract: The aim of this thesis is to find prediction for non-linear transformation of time series. First, under certain assumptions regarding the original time series, the autocovariance function and spectral density of the transformed time series are studied. General theorems are applied to concrete ARMA processes. Then general formulas for predictions of the transformed time series, which do not require knowledge of the autocovariance function of the transformed series nor its spectral density are presented. These formulas are applied to three concrete transformations and explicit formulas for ARMA processes are derived. Three types of predictions (optimal, naive and linear) are compared in the terms of proportional increase of mean square prediction error. Explicit formulas for ARMA processes are verified by a simulation.  
Keywords: Prediction, time series, non-linear transformation, Hermite polynomials, ARMA processes
Suppose we observe time series \( X_t \), but we wish to find prediction for non-linear transformation of this time series, say \( Y_t = T(X_t) \). We will assume that the series \( X_t \) follows some simple model, such as ARMA, which enables us to find prediction \( \hat{X}_{t+h} \) for the original series. The obvious method of simply transforming this prediction to obtain prediction \( \hat{Y}_{t+h} = T(\hat{X}_{t+h}) \) for the new series \( Y_t \) may sometimes lead to very poor results. It is the so-called “naive prediction”.

The aim of this diploma thesis is to describe the methods, which provide better results than the naive prediction and to evaluate the mean square errors for these predictions.

In general it can be said that the problem of finding the best prediction, or at least the best linear prediction, of the transformed series, is the problem of understanding the “generating mechanism” of the transformed series. Once we have enough information about the autocovariance structure of the transformed series or about its spectral density there are well-developed methods for using this knowledge to find the best linear prediction. However, under certain circumstances (Gaussian processes, simple transformations) it is possible to use direct formulas that greatly facilitate finding the prediction and estimating its mean square error.

Granger and Newbold [6] developed a direct method of finding the optimal and other predictions for transformations, which can be expressed by means of Hermite polynomials. The authors restricted themselves to Gaussian series only, on the other hand the results are very simple to implement and offer very quick yet accurate way of finding good predictions.

Somewhat similar approach via Hermite polynomials can be found in Hannan [8]. The author proves formula for the autocovariance function and spectral density of processes transformed from the original series, with known spectral density, by a linear combination of Hermite polynomials. Once we know the autocovariance function or the spectral density of transformed process we can use this information for making the prediction, see e.g. Grenander and Rosenblatt [7]. However no “ready-to-go” results were provided.

Choi and Taniguchi [5] studied the naive and bias-adjusted predictions for the square-transformed processes using the spectral density. The results they obtained are, as expected, in accordance with the results obtained by Granger and Newbold [6] via the autocovariance function. It is worth to mention that Choi and Taniguchi in their article
derived the expected square prediction error of the naive prediction also for non-Gaussian processes and made somewhat surprising finding that in certain cases the expected square error is smaller for non-Gaussian than for Gaussian series.

Palma and Zevallos [10] considered only the square transformation, but with more general innovations. They provided specific formula for the autocorrelation of the square-transformed processes.

When dealing with transformations for which it is impossible or impractical to use expansion with Hermite polynomials, such as fractional powers, one of the possibilities is to study similar transformation that can be expressed using the Hermite polynomials. In case that more accurate results are required Abadir and Talmain [1] offer a very general theorem, which creates a link between the autocovariance functions of two transforms of a process. As a special case the theorem establishes a link between autocovariance functions of a process and its transform. Once the autocovariance structure of the transformed process is known standard prediction methods can be used for finding the optimal prediction.
Chapter 2

Autocovariance and spectral density

In this chapter we will look at the methods of finding the autocovariance and spectral density of the transformed series.

2.1 Autocovariance of transformed Gaussian processes

Granger and Newbold [6] derived exact formulas for general class of instantaneous transformations of a stationary Gaussian time series. The formulas use expansion of the transformation with Hermite polynomials, which can theoretically be found for almost any real function in the form

\[ T(x) = \sum_{j=0}^{\infty} \alpha_j H_j(x), \]

using Formula (7.6) (see Appendix A - Properties of Hermite polynomials), however the derivation is in most cases very complicated. Expansion for the exponential function can be found from the generating function (7.4), and the expansion for polynomial functions can be found using e.g. the table in Abramowitz[2], p. 801 or Formulas (7.7) and (7.8).

**Theorem 2.1.** Let \( X_t \) be a stationary Gaussian time series with mean \( \mu \), variance \( \sigma^2 \) and autocorrelation function \( \text{corr}(X_t, X_{t+\tau}) = \rho_\tau \). Define

\[ Z_t = \frac{X_t - \mu}{\sigma} \text{ for all } t. \]

Then \( Z_t \) and \( Z_{t+\tau} \) have bivariate normal distribution with zero means, unit variance and correlation \( \rho_\tau \). Define series \( Y_t = T(Z_t) \), where \( T(Z_t) = \sum_{j=0}^{M} \alpha_j H_j(Z_t) \) and \( H_j(Z_t) \) are Hermite polynomials. Then the mean of the transformed series is

\[ \mathbb{E}(Y_t) = \alpha_0, \quad (2.1) \]

the covariance between the original and transformed series is

\[ \text{cov}(X_t, Y_{t+\tau}) = \alpha_1 \rho_\tau \sigma, \quad (2.2) \]

and finally the autocovariance function of the transformed series is

$$\text{cov}(Y_t, Y_{t+\tau}) = \sum_{j=1}^{M} j! \alpha_j^2 \rho_{\tau}^j. \quad (2.3)$$

**Proof.** The first statement follows from the basic properties of normal distribution. Now consider the orthogonality of the system of Hermite polynomials (7.3), which implies

$$E[H_j(Z_t)H_k(Z_{t+\tau})] = \begin{cases} 
0, & j \neq k, \\
j! \rho_{\tau}^j, & j = k.
\end{cases}$$

For the formula (2.1) we write

$$E(Y_t) = E[H_0(Z_t)Y_t] = E \left[ H_0(Z_t) \sum_{j=0}^{M} \alpha_j H_j(Z_t) \right]$$

$$= \sum_{j=0}^{M} \alpha_j E[H_0(Z_t)H_j(Z_t)] = \alpha_0 E[H_0(Z_t)H_0(Z_t)] = \alpha_0.$$

Now we have

$$\text{cov}(X_t, Y_{t+\tau}) = E(X_tY_{t+\tau}) - E(X_t)E(Y_{t+\tau})$$

$$= E[(\sigma Z_t + \mu)Y_{t+\tau}] - \mu \alpha_0$$

$$= \sigma E(Z_tY_{t+\tau}) + \mu E(Y_{t+\tau}) - \mu \alpha_0$$

$$= \sigma \alpha_1 \rho_{\tau} + \mu \alpha_0 - \mu \alpha_0 = \alpha_1 \rho_{\tau}$$

and finally

$$\text{cov}(Y_t, Y_{t+\tau}) = E \left[ \sum_{j=0}^{M} \alpha_j H_j(Z_t) \sum_{i=0}^{M} \alpha_i H_i(Z_{t+\tau}) \right]$$

$$= \sum_{j=1}^{M} \sum_{i=1}^{M} \alpha_j \alpha_i E[H_j(Z_t)H_i(Z_{t+\tau})]$$

$$= \sum_{j=1}^{M} j! \alpha_j^2 \rho_{\tau}^j.$$
2.1.1 Application to three transformations

**Corollary 2.2. (Exponential transformation)** Using the expression for the generating function of Hermite polynomials (7.4) we can write for the exponential transformation

\[ Y_t = \exp(X_t) = \exp \left( \sigma Z_t + \mu - \frac{1}{2} \sigma^2 + \frac{1}{2} \sigma^2 \right) = \exp \left( \mu + \frac{1}{2} \sigma^2 \right) \exp \left( \sigma Z_t - \frac{1}{2} \sigma^2 \right) = \exp \left( \mu + \frac{1}{2} \sigma^2 \right) \sum_{j=0}^{\infty} \frac{\sigma^j}{j!} H_j(Z_t). \]

Define \( \alpha_j = \exp \left( \mu + \frac{1}{2} \sigma^2 \right) \frac{\sigma^j}{j!} \). Then

\[ Y_t = \exp(X_t) = \sum_{j=0}^{\infty} \alpha_j H_j(Z_t). \]

From formulas (2.1), (2.2) and (2.3) it follows that the mean of the transformed series is

\[ \mathbb{E}(Y_t) = \alpha_0 = \exp \left( \mu + \frac{1}{2} \sigma^2 \right), \]

the covariance between the original and the transformed series is

\[ \text{cov}(X_t, Y_{t+\tau}) = \alpha_1 \rho \sigma = \exp \left( \mu + \frac{1}{2} \sigma^2 \right) \sigma^2 \rho, \]

and finally the autocovariance function of the transformed series is

\[ \text{cov}(Y_t, Y_{t+\tau}) = \sum_{j=1}^{M} j! \alpha_j^2 \rho_j^2 = \exp \left( 2\mu + \sigma^2 \right) \sum_{j=1}^{\infty} \frac{(\sigma^2 \rho \sigma)^j}{j!} = \exp \left( 2\mu + \sigma^2 \right) \left( \exp(\sigma^2 \rho \sigma) - 1 \right). \]

(2.4)

**Corollary 2.3. (Quadratic transformation)** Recall that

\[
\begin{align*}
H_0(x) &= 1, \\
H_1(x) &= x, \\
H_2(x) &= x^2 - 1, \\
H_3(x) &= x^3 - 3x.
\end{align*}
\]

Consider the transformation

\[ Y_t = X_t^2 = (\sigma Z_t + \mu)^2 = \mu^2 + 2\mu \sigma Z_t + \sigma^2 Z_t^2 = \mu^2 + \sigma^2 + 2\sigma \mu Z_t + \sigma^2 Z_t^2 - \sigma^2. \]
CHAPTER 2. AUTOCOVARIANCE AND SPECTRAL DENSITY

Denoting
\[
\begin{align*}
\alpha_0 &= \mu^2 + \sigma^2, \\
\alpha_1 &= 2\sigma\mu, \\
\alpha_2 &= \sigma^2
\end{align*}
\]
we obtain the Hermite expansion for the quadratic transformation
\[
Y_t = X_t^2 = \alpha_0 + \alpha_1 H_1(Z_t) + \alpha_2 H_2(Z_t).
\]
Using again formulas (2.1), (2.2) and (2.3) we write the mean of the transformed series
\[
E(Y_t) = \alpha_0 = \mu^2 + \sigma^2,
\]
the covariance between the original and the transformed series
\[
\text{cov}(X_t, Y_{t+\tau}) = \alpha_1 \rho \sigma = 2\sigma^2 \mu \rho \tau
\]
and finally the autocovariance function of the transformed series
\[
\text{cov}(Y_t, Y_{t+\tau}) = \sum_{j=1}^{M} j! \alpha_j^2 \rho^j \tau = \alpha_1^2 \rho \tau + 2\alpha_2^2 \rho^2 \tau = 4\sigma^2 \mu^2 \rho \tau + 2\sigma^4 \rho^2 \tau.
\]
(2.5)

Corollary 2.4. (Cubic transformation) Consider the transformation
\[
Y_t = X_t^3 = (\sigma Z_t + \mu)^3
\]
\[
= \mu^3 + 3\sigma \mu^2 Z_t + 3\sigma^2 \mu Z_t^2 + 3\sigma^3 Z_t^3
\]
\[
= \mu^3 + 3\sigma^2 \mu + (3\sigma^2 + 3\sigma^3) Z_t + 3\sigma^2 \mu Z_t^2 - 3\sigma^2 \mu + 3\sigma^3 Z_t^3 - 3\sigma^3 Z_t.
\]
Denoting
\[
\begin{align*}
\alpha_0 &= \mu^3 + 3\sigma^2 \mu, \\
\alpha_1 &= 3\sigma \mu^2 + 3\sigma^3, \\
\alpha_2 &= 3\sigma^2 \mu, \\
\alpha_3 &= \sigma^3
\end{align*}
\]
we obtain the Hermite expansion for the cubic transformation
\[
\alpha_0 + \alpha_1 H_1(Z_t) + \alpha_2 H_2(Z_t) + \alpha_3 H_3(Z_t).
\]
The mean of the transformed series is then
\[
E(Y_t) = \alpha_0 = \mu^3 + 3\sigma^2 \mu,
\]
the covariance between the original and the transformed series is
\[
\text{cov}(X_t, Y_{t+\tau}) = \alpha_1 \rho \sigma = 3\sigma^2 \mu^2 \rho \tau + 3\sigma^4 \rho \tau
\]
and finally the autocovariance function of the transformed series is
\[
\text{cov}(Y_t, Y_{t+\tau}) = \sum_{j=1}^{M} j! \alpha_j^2 \rho^j \tau = \alpha_1^2 \rho \tau + 2\alpha_2^2 \rho^2 \tau + 6\alpha_3^2 \rho^3 \tau
\]
(2.6)
\[
= (3\sigma^3 \mu^2 + 3\sigma^3)^2 \rho \tau + 2(3\sigma^2 \mu)^2 \rho^2 \tau + 6(\sigma^3)^2 \rho^3 \tau.
\]
(2.7)
2.1.2 Application to ARMA processes

In order to apply general formulas to a given process we need to know its mean $\mu$ and variance $\sigma^2$, as well as its autocorrelation function $\rho_\tau$. This is very easy to calculate for MA($\infty$) process.

Example 2.5. Consider a MA($\infty$) process $X_t$ defined by

$$X_t = \sum_{i=0}^{\infty} \theta_i \epsilon_{t-i},$$

where $\theta_0 = 1$, $\sum_{i=0}^{\infty} \theta_i^2 < \infty$ and $\epsilon_t \sim N(0, \sigma_\epsilon^2)$. The mean of this process is obviously zero, its variance is

$$\sigma^2 = \text{var}(X_t) = \sum_{i=0}^{\infty} \theta_i^2 \text{var}(\epsilon_{t-i}) = \sigma_\epsilon^2 \sum_{i=0}^{\infty} \theta_i^2,$$

the autocovariance function is

$$\text{cov}(X_t, X_{t+\tau}) = \sigma_\epsilon^2 \sum_{i=0}^{\infty} \theta_i \theta_{i+\tau}$$

and the autocorrelation function is

$$\rho_\tau = \frac{\text{cov}(X_t, X_{t+\tau})}{\text{var}(X_t)} = \frac{\sum_{i=0}^{\infty} \theta_i \theta_{i+\tau}}{\sum_{i=0}^{\infty} \theta_i^2}.$$

Using the derived formulas we can write for the exponential transformation $Y_t = \exp(X_t)$:

$$\mathbb{E}(Y_t) = \exp \left( \frac{1}{2} \sigma_\epsilon^2 \sum_{i=0}^{\infty} \theta_i^2 \right),$$

$$\text{cov}(X_t, Y_{t+\tau}) = \exp \left( \frac{1}{2} \sigma_\epsilon^2 \sum_{i=0}^{\infty} \theta_i^2 \right) \left( \sigma_\epsilon^2 \sum_{i=0}^{\infty} \theta_i \theta_{i+\tau} \right),$$

$$\text{cov}(Y_t, Y_{t+\tau}) = \exp \left( \sigma_\epsilon^2 \sum_{i=0}^{\infty} \theta_i^2 \right) \left[ \exp \left( \sigma_\epsilon^2 \sum_{i=0}^{\infty} \theta_i \theta_{i+\tau} \right) - 1 \right],$$

for the square transformation $Y_t = X_t^2$:

$$\mathbb{E}(Y_t) = \sigma_\epsilon^2 \sum_{i=0}^{\infty} \theta_i^2,$$

$$\text{cov}(X_t, Y_{t+\tau}) = 0,$$

$$\text{cov}(Y_t, Y_{t+\tau}) = 2\sigma_\epsilon^4 \left( \sum_{i=0}^{\infty} \theta_i \theta_{i+\tau} \right)^2.$$
and finally for $Y_t = X_t^3$:

$$
E(Y_t) = 0,
$$

$$\operatorname{cov}(X_t, Y_{t+\tau}) = 3\sigma^4 \left( \sum_{i=0}^{\infty} \theta_i^2 \right) \left( \sum_{t=0}^{\infty} \theta_t \theta_{t+\tau} \right),
$$

$$\operatorname{cov}(Y_t, Y_{t+\tau}) = 9\sigma^6 \left( \sum_{i=0}^{\infty} \theta_i^2 \right)^2 \left( \sum_{t=0}^{\infty} \theta_t \theta_{t+\tau} \right) + 6\sigma^6 \left( \sum_{t=0}^{\infty} \theta_t \theta_{t+\tau} \right)^3.
$$

\textbf{Example 2.6.} As a special case of the previous example consider the following AR(1) process

$$X_t = \varphi X_{t-1} + \epsilon_t,$$

where $\epsilon_t \sim \mathcal{N}(0, 1)$ are i.i.d. This process can also be written in the form

$$X_t = \sum_{j=0}^{\infty} \varphi^j \epsilon_{t-j}.
$$

The mean of this process is zero, $\mu = 0$, and its variance is

$$\sigma^2 = \operatorname{var}(X_t) = \sum_{j=0}^{\infty} \varphi^{2j} = \frac{1}{1 - \varphi^2},
$$

taking into consideration that $\epsilon_t$ are i.i.d. The autocovariance function of this process is

$$R_X(\tau) = \operatorname{cov}(X_t, X_{t+\tau}) = \operatorname{cov} \left( \sum_{j=0}^{\infty} \varphi^j \epsilon_{t-j}, \sum_{j=0}^{\infty} \varphi^j \epsilon_{t+\tau-j} \right)
$$

$$= \sum_{j=0}^{\infty} \varphi^{j+\tau} \varphi^j \operatorname{var}(\epsilon_{t-j+\tau} \epsilon_{t+\tau-j}) = \sum_{j=0}^{\infty} \varphi^{2j+\tau} = \frac{\varphi^\tau}{1 - \varphi^2}.
$$

The autocorrelation function is then

$$\rho_\tau = \frac{R_X(\tau)}{\operatorname{var}(X_t)} = \varphi^\tau.
$$

Using formula (2.4) we can find the autocovariance function of the transformed series $Y_t = \exp(X_t)$

$$R_Y^{(1)}(\tau) = \operatorname{cov}(Y_t, Y_{t+\tau}) = \exp \left( 2\mu + \sigma^2 \right) \left( \exp(\sigma^2 \rho_\tau) - 1 \right)
$$

$$= \exp \left( \frac{1}{1 - \varphi^2} \right) \left[ \exp \left( \frac{\varphi^\tau}{1 - \varphi^2} \right) - 1 \right].$$
Using formula (2.5) we can find the autocovariance function of the transformed series \( Y_t = X_t^2 \)

\[
R_Y^{(2)}(\tau) = \text{cov}(Y_t, Y_{t+\tau}) = 4\sigma^2 \mu^2 \rho_\tau + 2\sigma^4 \rho_\tau^2
\]

\[
= 2\varphi^\tau \left( \frac{1}{1-\varphi^2} \right)^2.
\]

Finally, using formula (2.6) we can find the autocovariance function of the transformed series \( Y_t = X_t^3 \)

\[
R_Y^{(3)}(\tau) = \text{cov}(Y_t, Y_{t+\tau}) = (3\sigma \mu^3 + 3\sigma^3)^2 \rho_\tau + 2(3\sigma^2 \mu)^2 \rho_\tau^2 + 6\sigma^6 \rho_\tau^3
\]

\[
= 9\varphi^\tau \left( \frac{1}{1-\varphi^2} \right)^3 + 6\left( \frac{\varphi^\tau}{1-\varphi^2} \right)^3 = (9\varphi^\tau + 6\varphi^{3\tau}) \left( \frac{1}{1-\varphi^2} \right)^3.
\]

In Figure 2.1 there is autocovariance function of AR(1) process, as well as the autocovariance functions of the transformed series \( \exp(X_t), X_t^2 \) and \( X_t^3 \).

**Example 2.7.** Finally, let us consider the following MA(1) process

\[ X_t = \epsilon_t + \theta \epsilon_{t-1}, \]

where \( \epsilon_t \sim \mathcal{N}(0, 1) \) are i.i.d. The mean of this process is zero, \( \mu = 0 \), and its variance is

\[ \sigma^2 = \text{var}(X_t) = (1 + \theta^2). \]

The autocovariance function of this process is

\[
R_X(\tau) = \text{cov}(X_t, X_{t+\tau}) = \text{cov}(\epsilon_t + \theta \epsilon_{t-1}, \epsilon_{t+\tau} + \theta \epsilon_{t+\tau-1})
\]

\[
= \begin{cases} 
1 + \theta^2, & \text{if } \tau = 0, \\
\theta, & \text{if } \tau = 1 \text{ or } \tau = -1, \\
0, & \text{otherwise}
\end{cases}
\]

and its the autocorrelation function is

\[
\rho_X(\tau) = \begin{cases} 
1, & \text{if } \tau = 0, \\
\frac{\theta}{1+\theta^2}, & \text{if } \tau = 1 \text{ or } \tau = -1, \\
0, & \text{otherwise}
\end{cases}
\]

The autocovariance function of the transformed process \( Y_t = \exp(X_t) \) is from (2.4) calculated using the formula

\[
R_Y^{(1)}(\tau) = \exp(1 + \theta^2) \left[ \exp((1 + \theta^2)\rho_X(\tau)) - 1 \right]
\]

and so

\[
R_Y^{(1)}(\tau) = \begin{cases} 
\exp(1 + \theta^2) \left[ \exp(1 + \theta^2) - 1 \right], & \text{if } \tau = 0, \\
\exp(1 + \theta^2) (\exp(\theta) - 1), & \text{if } \tau = 1 \text{ or } \tau = -1, \\
0, & \text{otherwise}
\end{cases}
\]
Figure 2.1: Autocovariance functions of the original process AR(1) and its transforms.
The autocovariance function of the transformed process $Y_t = X_t^2$ is from (2.5) calculated using the formula

$$R_Y^{(2)}(\tau) = 2(1 + \theta^2)^2 \rho_X^2(\tau)$$

and so

$$R_Y^{(2)}(\tau) = \begin{cases} 
2(1 + \theta^2)^2, & \text{if } \tau = 0, \\
2\theta^2, & \text{if } \tau = 1 \text{ or } \tau = -1, \\
0, & \text{otherwise}.
\end{cases}$$

Finally, the autocovariance function of the transformed process $Y_t = X_t^3$ is from (2.6)

$$R_Y^{(3)}(\tau) = 9(1 + \theta^2)^3 \rho_X(\tau) + 6(1 + \theta^2)^3 \rho_X^3(\tau)$$

and so

$$R_Y^{(3)}(\tau) = \begin{cases} 
9(1 + \theta^2)^3 + 6(1 + \theta^2)^3 \rho_X^3(\tau), & \text{if } \tau = 0, \\
9\theta(1 + \theta^2)^2 + 6\theta^3, & \text{if } \tau = 1 \text{ or } \tau = -1, \\
0, & \text{otherwise}.
\end{cases}$$

For illustration table 2.1 contains the values of autocovariance of the original and transformed series for the MA(1) process.

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>$\tau = 0$</th>
<th>$\tau = 1$</th>
<th>$\theta$</th>
<th>$\tau = 0$</th>
<th>$\tau = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_t$</td>
<td>1.063</td>
<td>0.250</td>
<td>$X_t$</td>
<td>1.562</td>
<td>0.750</td>
</tr>
<tr>
<td>$\exp(X_t)$</td>
<td>5.479</td>
<td>0.723</td>
<td>$\exp(X_t)$</td>
<td>17.989</td>
<td>3.578</td>
</tr>
<tr>
<td>$X_t^2$</td>
<td>2.258</td>
<td>0.265</td>
<td>$X_t^2$</td>
<td>4.882</td>
<td>1.172</td>
</tr>
<tr>
<td>$X_t^3$</td>
<td>10.795</td>
<td>2.540</td>
<td>$X_t^3$</td>
<td>34.332</td>
<td>16.479</td>
</tr>
</tbody>
</table>

Table 2.1: Autocovariance of MA(1) process and its transforms.

### 2.2 General innovations and the square transformation

Palma and Zevallos in [10] studied the behavior of the autocorrelation function of the square of a time series with the following expansion

$$X_t = \Psi(B)\epsilon_t,$$

where

$$\Psi(B) = \sum_{i=0}^{\infty} \psi_i B^i, \quad \psi_0 = 1, \quad \sum_{i=0}^{\infty} \psi_i^2 < \infty,$$

$\epsilon_t$ has finite kurtosis $\eta$ and $B$ is the lag operator. In the following we will assume that $\epsilon_t$ are uncorrelated, but not necessarily independent.
Theorem 2.8. Let us assume that \( \epsilon_t \) are i.i.d. random variables with zero mean, finite kurtosis, \( \eta = \mathbb{E}(\epsilon_t^4)/[\mathbb{E}(\epsilon_t^2)^2] \). Let us assume that

\[
\mathbb{E}(\epsilon_u\epsilon_v) = \begin{cases} 
\sigma_e^2, & \text{if } u = v, \\
0, & \text{otherwise},
\end{cases}
\]

where \( \sigma_e \) is the standard deviation of \( \epsilon_t \). The autocorrelation function function of \( \epsilon_t^2 \) is \( \rho_e(\tau) \) and

\[
\mathbb{E}(\epsilon_u\epsilon_v) = \begin{cases} 
[1 + (\eta - 1)\rho_e(s-v)]\sigma_e^4, & \text{if } s = t, \ u = v \text{ or } s = u, \ t = v, \\
[1 + (\eta - 1)\rho_e(s-t)]\sigma_e^4, & \text{if } s = v, \ t = u, \\
0, & \text{otherwise}.
\end{cases}
\]

Consider now process \( X_t \) defined by formula

\[
X_t = \sum_{i=0}^{\infty} \theta_i \epsilon_{t-i},
\]

where \( \theta_0 = 1 \) and \( \sum_{i=0}^{\infty} \theta_i^2 < \infty \). The autocorrelation function of the process \( X_t \) is

\[
\text{corr}(X_t, X_{t+\tau}) = \rho_X(\tau) = \frac{\sum_{i=0}^{\infty} \theta_i \theta_{i+\tau}}{\sum_{i=0}^{\infty} \theta_i^2}.
\]

Let us define

\[
\alpha(\tau) = \frac{\sum_{i=0}^{\infty} \theta_i^2 \theta_{i+\tau}}{\sum_{i=0}^{\infty} \theta_i^4},
\]

\[
\Delta(\tau) = \sum_{s=0}^{\infty} \sum_{t=0}^{\infty} \theta_i \theta_s \theta_{i+s} \rho_e(t-s),
\]

\[
\phi(\tau) = \frac{\sum_{s=0}^{\infty} \sum_{t=0}^{\infty} \theta_i^2 \theta_s \rho_e(t+s)}{(\sum_{i=0}^{\infty} \theta_i^2)^2},
\]

\[
\kappa = 3 - 2\eta \frac{\sum_{i=0}^{\infty} \theta_i^4}{(\sum_{i=0}^{\infty} \theta_i^2)^2} + 3(\eta - 1)\Delta(0),
\]

where \( \kappa \) is kurtosis of \( Y_t \). Then the autocorrelation function of the squared process \( Y_t = (X_t)^2 \) is given by

\[
\text{corr}(Y_t, Y_{t+\tau}) = \rho_Y(\tau) = \frac{2}{\kappa - 1} \rho_X^2(\tau) + \frac{\kappa - 3}{\kappa - 1} \alpha(\tau) + \frac{\eta - 1}{\kappa - 1} [\phi(\tau) + 2\Delta(\tau) - 3\Delta(0)\alpha(\tau)].
\]

Proof. See Palma and Zevallos [10].

Corollary 2.9. \textbf{(Linear process)}

Let \( \epsilon_t \) be i.i.d. random variables with zero mean and finite kurtosis \( \eta \), i.e. \( \epsilon_t \) is strict white noise. Then

\[
\rho_Y(\tau) = \frac{2}{\kappa - 1} \rho_X^2(\tau) + \frac{\kappa - 3}{\kappa - 1} \alpha(\tau),
\]

where \( \kappa \) is the kurtosis of \( Y_t \) given by

\[
\kappa = (\eta - 3) \frac{\sum_{i=0}^{\infty} \theta_i^4}{(\sum_{i=0}^{\infty} \theta_i^2)^2} + 3.
\]
Proof. In this case we can write

\[ \phi(\tau) = \Delta(\tau) = \Delta(0)\alpha(\tau) = \sum_{t=0}^{\infty} \theta_t^2 \theta_{t+\tau}^2 \overline{\sum_{i=0}^{\infty} \theta_i^2}^2 \]

and therefore \( \phi(\tau) + 2\Delta(\tau) - 3\Delta(0)\alpha(\tau) = 0 \), also

\[ \Delta(0) = \sum_{i=0}^{\infty} \theta_i^4 \overline{\sum_{i=0}^{\infty} \theta_i^2}^2. \]

The expected results follow immediately.

The above general theorem and corollary are useful generalizations, however the principal focus of this work is on ARMA processes and in the following we show that when considering the MA(\( \infty \)) as a special case we obtain the same results as with the method developed earlier.

Example 2.10. Consider the MA(\( \infty \)) process

\[ X_t = \sum_{i=0}^{\infty} \theta_i \epsilon_{t-i}, \]

with \( \theta_0 = 1 \), \( \sum_{i=0}^{\infty} \theta_i^2 < \infty \) and \( \epsilon_t \sim N(0, \sigma^2) \). Autocorrelation sequence of \( \epsilon_t^2 \) is \( \rho_{\epsilon^2}(t) = 1 \) for \( t = 0 \) and 0 otherwise. From the basic properties of the normal distribution we know that kurtosis of \( \epsilon_t \) is \( \eta = 3 \). Hence

\[ \rho_{Y}(\tau) = \rho_X^2(\tau), \]

where the autocorrelation of the original series (see Example (2.5)) is

\[ \rho(\tau) = \text{corr}(X_t, X_{t+\tau}) = \frac{\text{cov}(X_t, X_{t+\tau})}{\sqrt{\text{var}(X_t)^2 \text{var}(X_{t+\tau})^2}} = \frac{\sum_{t=0}^{\infty} \theta_t \theta_{t+\tau}}{\sum_{i=0}^{\infty} \theta_i^4}. \]

This result is in accordance with the results obtained in example 2.5, because

\[ \rho_{Y}(\tau) = \frac{\text{cov}(Y_t, Y_{t+\tau})}{\text{var}(Y_t)} = \frac{2\sigma_t^4 (\sum_{i=0}^{\infty} \theta_i \theta_{i+\tau})^2}{2\sigma_t^4 (\sum_{i=0}^{\infty} \theta_i^2)^2} = \rho_X^2(\tau). \]

2.3 Generalized transformations

A different approach was offered by Abadir and Talmain [1], who established a link between the autocovariance functions of two transforms of a process. The problem investigated in this work can be considered a special case, one of the transforms being the identical transformation. However, more detailed description of this rather general theorem exceeds the scope of this work and will not be included.
2.4 Spectral representation of transformed processes

Hannan [8] derived the same results regarding the autocovariance of transformed processes as Granger and Newbold [6], moreover he derived also a formula for the spectral density of the transformed process.

**Theorem 2.11.** Consider a Gaussian process $X_t$ with zero mean, unit variance, autocovariance function $R(\tau)$ and spectral density $f(\lambda)$. Then the autocovariance of the transformed series

$$Y_t = \sum_{j=0}^{M} \alpha_j H_j(X_t)$$

is

$$R_Y(\tau) = \text{cov}(Y_t, Y_{t+\tau}) = \sum_{j=0}^{M} j! \alpha_j^2 R^j(\tau)$$ (2.8)

and

$$R^j(\tau) = \int_{-\infty}^{\infty} f^{*j}(\lambda) d\lambda,$$

where $f^{*j}(\lambda)$ is the $j$-fold convolution of $f(\lambda)$ with itself, $f^{*0}(\lambda)$ is defined as $\delta(\lambda)$, the Dirac delta function. The spectral density of the transformed process $Y_t$ is

$$f_Y(\lambda) = \sum_{j=0}^{M} j! \alpha_j^2 f^{*j}(\lambda),$$ (2.9)

**Proof.** The idea behind the proof of (2.8) is the same as that of (2.3) in Theorem 2.1, note only that we assume unit variance of the process, hence

$$\rho_\tau = \text{corr}(Y_t, Y_{t+\tau}) = \text{cov}(Y_t, Y_{t+\tau}) = R(\tau).$$

For the rest of the proof see Hannan [8], p. 82. \qed

As an illustration of application of this Theorem we consider MA(1) process.

**Example 2.12.** Consider the following MA(1) process

$$X_t = \epsilon_t + \theta \epsilon_{t-1},$$

where $\epsilon_t \sim \mathcal{N}(0, 1)$ are i.i.d. Recall that we have already studied this case in Example 2.7. In the following we will show that the previous theorem leads to the same results. The density of this MA(1) process is

$$f(\lambda) = \frac{1}{2\pi} |1 + \theta e^{-i\lambda}|^2, \text{ where } \lambda \in [-\pi, \pi].$$
Consider the square transformation of this process, i.e. process
\[ Y_t = X_t^2 = \alpha_0 H_0(X_t) + \alpha_2 H_2(X_t), \]
where \( \alpha_0 = 1 \) and \( \alpha_2 = 1 \). Using (2.9) the density of this transformed process is
\[ f_Y(\lambda) = \alpha_0^2 f^x + 2! \alpha_2^2 f^{x^2} = \delta(\lambda) + 2 f^{x^2} \]
and
\[ f^{x^2} = \int_{-\pi}^{\pi} f(\lambda - a) f(a) \, da = \int_{-\pi}^{\pi} \frac{1}{2\pi} \left| 1 + \theta e^{-i(\lambda - a)} \right|^2 \frac{1}{2\pi} \left| 1 + \theta e^{-ia} \right|^2 \, da \]
\[ = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} (1 + \theta e^{-i(\lambda - a)}) (1 + \theta e^{i(\lambda - a)}) (1 + \theta e^{-ia}) (1 + \theta e^{ia}) \, da \]
\[ = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} (1 + \theta^2 + 2\theta \cos(\lambda - a)) (1 + \theta^2 + 2\theta \cos(a)) \, da \]
\[ = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} [(1 + \theta^2)^2 + 4\theta^2 \cos(\lambda - a)] \, da \]
\[ = \frac{1}{4\pi^2} \left[ 2\pi(1 + \theta^2)^2 + 4\pi \theta^2 \cos(\lambda) \right] \]
\[ = \frac{1}{2\pi} (1 + \theta^2)^2 + \frac{1}{\pi} \theta^2 \cos(\lambda). \]

Hence the spectral density of the transformed process is
\[ f_Y(\lambda) = \delta(\lambda) + \frac{1}{\pi} [(1 + \theta^2)^2 + 2\theta^2 \cos(\lambda)]. \]

To show that this is indeed in accordance with the results in example 2.7 we will use the following relationship between the autocovariance function and spectral density (see e.g. Prášková [11], p. 28).
\[ R(t) = \int_{-\pi}^{\pi} e^{it\lambda} f(\lambda) \, d\lambda. \]
In our case
\[ R^{(2)}_Y(\tau) = \frac{1}{\pi} \int_{-\pi}^{\pi} e^{it\lambda} \left[ \pi \delta(\lambda) + (1 + \theta^2)^2 + 2\theta^2 \cos(\lambda) \right] \, d\lambda. \]
Calculating this integral for concrete values of \( \tau \) we have
\[ R^{(2)}_Y(0) = 2(1 + \theta^2)^2, \]
\[ R^{(2)}_Y(1) = R^{(2)}_Y(-1) = 2\theta^2, \]
\[ R^{(2)}_Y(\tau) = 0, \text{ for } \tau \geq 2, \]
which is in accordance with the results derived in example 2.7.
Chapter 3

Predictions

We have already addressed the issue of autocovariance and spectral properties of transformed time series and thus obtained tools for predicting future realizations of the transformed series. All that is required is to apply the standard methods for predictions. However, Granger and Newbold [6] derived explicit formulas that can be used without the need of investigating the autocovariance or spectral properties of the transformed time series.

3.1 Optimal, naive and linear predictions

3.1.1 Optimal prediction

Theorem 3.1. (Optimal prediction) Let $X_t$ be a Gaussian stationary process with mean $\mu$ and variance $\sigma^2$, and $\hat{X}_{t+h}$ be the optimal $h$-step ahead prediction of $X_{t+h}$, given the information set $I_t$ (previous observations $X_t, X_{t-1}, \ldots$), i.e. $\hat{X}_{t+h} = E(X_{t+h}|I_t)$. Assume that the error of prediction $e_{t+h} = X_{t+h} - \hat{X}_{t+h}$ given $I_t$ has conditionally normal distribution with zero mean and variance $S^2(h)$. Define $Z_t = \frac{X_t - \mu}{\sigma} \sim N(0, 1)$ and $A = \sqrt{1 - \frac{S^2(h)}{\sigma^2}}$. Consider the following transformation

$$Y_t = T(Z_t) = \sum_{i=0}^{M} \alpha_i H_i(Z_t).$$

Define random variable

$$W_{t+h} = \frac{X_{t+h} - \hat{X}_{t+h}}{S(h)}.$$

Then constants $\gamma_i$ can be found such that $Y_t = \sum_{i=0}^{M} \gamma_i H_i(W_t)$. The $h$-step ahead optimal prediction of $Y_{t+h}$ is

$$\hat{Y}_{t+h}^{(1)} = \gamma_0 \quad (3.1)$$
and the unconditional mean square error of this prediction is

$$V^{(1)}(h) = \mathbb{E} \left( X_{t+h} - \hat{X}_{t+h} \right)^2 = \sum_{j=1}^{M} j! \alpha_j^2 (1 - A^2).$$

Proof. The h-step ahead prediction of $Y_{t+h}$ is in fact the conditional expectation of this random variable. Hence we can use the results from theorem 2.1 and the formula (3.1) follows immediately.

Random variable $X_{t+h}$ has conditionally normal distribution with mean $\hat{X}_{t+h}$ and variance $S^2(h)$ and therefore random variable $W_{t+h}$, is conditionally, given $I_t$, distributed as a standard normal variable. Define $B = \sqrt{1 - A^2} = \sqrt{\frac{S^2(h)}{\sigma^2}}$ and $C = \frac{\hat{X}_{t+h} - \mu}{\sqrt{\sigma^2 - S^2(h)}}$. We have

$$Z_{t+h} = \frac{X_{t+h} - \mu}{\sigma} = \frac{W_{t+h}S(h) + \hat{X}_{t+h} - \mu}{\sigma}$$
$$= \frac{S(h)}{\sigma}W_{t+h} + \sqrt{1 - \frac{S^2(h)}{\sigma^2}} \frac{\hat{X}_{t+h} - \mu}{\sqrt{\sigma^2 - S^2(h)}}$$
$$= BW_{t+h} + AC.$$ 

Note that $A^2 + B^2 = 1$, hence we can use the summation formula (7.5). For each $j \in \mathbb{N}$ we can write

$$H_j(Z_{t+h}) = H_j(BW_{t+h} + AC)$$
$$= \sum_{k=0}^{j} \binom{j}{k} B^k A^{j-k} \mathbb{E}[H_k(W_{t+h})].$$

Recall that the Hermite polynomials have the following property (7.1)

$$\mathbb{E}[H_k(W_{t+h})] = \begin{cases} 
0, & k > 0, \\
1, & k = 0.
\end{cases}$$

Now we have the optimal prediction in the quadratic loss sense:

$$\hat{Y}^{(1)}_{t+h} = \mathbb{E}(Y_{t+h}) = \mathbb{E} \left[ \sum_{j=0}^{M} \alpha_j H_j(Z_{t+h}) \right] = \mathbb{E} \left[ \sum_{j=0}^{M} \alpha_j H_j(BW_{t+h} + AC) \right]$$
$$= \sum_{j=0}^{M} \alpha_j \sum_{k=0}^{j} \binom{j}{k} B^k A^{j-k} \mathbb{E}[H_k(W_{t+h})] = \sum_{j=0}^{M} \alpha_j A^j H_j(C).$$

To calculate the unconditional mean square error we first write

$$V^{(1)}(h) = \mathbb{E} \left[ Y_{t+h} - \hat{Y}^{(1)}_{t+h} \right]^2 = \mathbb{E} \left[ Y_{t+h} \right]^2 - 2\mathbb{E} \left[ Y_{t+h} \hat{Y}^{(1)}_{t+h} \right] + \mathbb{E} \left[ \hat{Y}^{(1)}_{t+h} \right]^2. \quad (3.2)$$
Recall from the properties of Hermite polynomials, formula (7.3), that

\[ E\{H_i(Z_{t+h})H_j(Z_{t+h})\} = \begin{cases} 0, & i \neq j, \\ j!1^j, & i = j. \end{cases} \]

Therefore

\[ E[Y_{t+h}]^2 = E\left[ \sum_{j=0}^{M} \alpha_j H_j(Z_{t+h}) \right]^2 = \sum_{i=0}^{M} \sum_{j=0}^{M} \alpha_i \alpha_j E\{H_i(Z_{t+h})H_j(Z_{t+h})\} = \sum_{i=0}^{M} j! \alpha_i^2. \]

Similarly,

\[ E\left[ \hat{S}_{t+h}^{(1)} \right]^2 = E\left[ \sum_{j=0}^{M} \alpha_j A^j H_j(C) \right]^2 = \sum_{i=0}^{M} \sum_{j=0}^{M} \alpha_i \alpha_j A^{i+j} E\{H_i(C)H_j(C)\} = \sum_{i=0}^{M} j! \alpha_j^2 A^{2j}, \]

because \( C \sim N(0,1) \). Since \( \text{cov}(\hat{X}_{t+h}, e_{t+h}) = 0 \) and \( \text{cov}(e_{t+h}, e_{t+h}) = S^2(h) \) we can write

\[
\begin{align*}
\text{cov}(X_{n+h}, \hat{X}_{t+h}) &= \text{cov}(X_{t+h}, X_{t+h} - e_{t+h}) = \text{cov}(X_{t+h}, X_{t+h}) - \text{cov}(X_{t+h}, e_{t+h}) \\
&= \sigma^2 - \text{cov}(\hat{X}_{t+h} + e_{t+h}, e_{t+h}) = \sigma^2 - \text{cov}(\hat{X}_{t+h}, e_{t+h}) - \text{cov}(e_{t+h}, e_{t+h}) \\
&= \sigma^2 - S^2(h)
\end{align*}
\]

and so the autocorrelation between \( Z_{t+h} \) and \( C \) is

\[
\rho = \text{corr}(Z_{t+h}, C) = \frac{\text{cov}(Z_{t+h}, C)}{\sqrt{\text{var}(Z_{t+h})\text{var}(C)}} = \text{cov}(Z_{t+h}, C)
\]

\[
\begin{align*}
&= \text{cov} \left[ \frac{X_{t+h} - \mu}{\sigma}, \frac{\hat{X}_{t+h} - \mu}{\sqrt{\sigma^2 - S^2(h)}} \right] \\
&= \frac{1}{\sigma \sqrt{\sigma^2 - S^2(h)}} \text{cov}(X_{n+h}, \hat{X}_{t+h}) \\
&= \frac{1}{\sigma \sqrt{\sigma^2 - S^2(h)}} (\sigma^2 - S^2(h)) = \sqrt{\frac{\sigma^2 - S^2(h)}{\sigma^2}} = A.
\end{align*}
\]

Now we can write

\[
2E\left[ \hat{Y}_{t+h}^{(1)} Y_{t+h} \right] = 2E \left[ \sum_{i=0}^{M} \alpha_i H_i(Z_{t+h}) \right] \left[ \sum_{j=0}^{M} \alpha_j A^j H_j(P) \right] = 2E \left[ \sum_{i=0}^{M} \sum_{j=0}^{M} \alpha_i \alpha_j A^j H_i(Z_{t+h}) H_j(C) \right] = 2 \sum_{j=0}^{M} j! \alpha_j^2 A^j \rho^j = 2 \sum_{j=0}^{M} j! \alpha_j^2 A^{2j}. 
\]
Returning back to formula (3.2) we derived the expected expression for the forecast error

\[ V^{(1)}(h) = \sum_{j=0}^{M} \alpha_j^2 j! - 2 \sum_{j=0}^{M} \alpha_j^2 A^2 j! + \sum_{j=0}^{M} \alpha_j^2 A^2 j! \]

\[ = \sum_{j=0}^{M} \alpha_j^2 j! (1 - A^2j). \]

\[ = \sum_{j=0}^{M} \alpha_j^2 j! (1 - A^2j). \]

3.1.2 Naive prediction

In the following we will be using the notation \( \lfloor X \rfloor \) for the integer part of a real number.

**Theorem 3.2. (Naive prediction)** Under the same assumptions as in Theorem 3.1 consider the naive \( h \)-step ahead prediction of \( Y_{t+h} \)

\[ \hat{Y}^{(2)}_{t+h} = T \left( \frac{\hat{X}_{t+h} - \mu}{\sigma} \right). \]

Then the mean square error of this prediction is

\[ V^{(2)}(h) = \sum_{j=1}^{M} \alpha_j^2 j!(1 - A^2j) + \sum_{j=0}^{M} A^2 j! \sum_{k=1}^{\lfloor \frac{1}{2}(M-j) \rfloor} \alpha_{j+2k} \frac{(j+2k)!}{k!} \left( -\frac{1}{2} B^2 \right)^k. \] (3.3)

**Proof.** We have

\[ V^{(2)}(h) = E \left[ Y_{t+h} - \hat{Y}^{(2)}_{t+h} \right]^2 = E \left[ \left( Y_{t+h} - \hat{Y}^{(1)}(h) \right) - \left( Y_{t+h} - Y^{(1)}_{t+h} \right) \right]^2 \]

\[ = E \left( Y_{t+h} - \hat{Y}^{(1)}(h) \right)^2 - 2E \left( Y_{t+h} - \hat{Y}^{(1)}(h) \right) \left( Y^{(2)}_{t+h} - Y^{(1)}_{t+h} \right) + E \left( Y_{t+h} - Y^{(1)}_{t+h} \right)^2. \]

The first term in (3.3) is \( V^{(1)}(h) \) from Theorem 3.1. The second term is derived in similar manner using the properties of Hermite polynomials. See also Granger and Newbold [6].

3.1.3 Linear prediction

**Theorem 3.3. (Linear prediction)** Under the same assumptions as in Theorem 3.1 consider \( h \)-step ahead prediction of \( Y_{t+h} \), which is optimal in the quadratic loss sense in the class of forecasts that are linear in \( X_{t-j}, j \geq 0 \). Then this prediction can be written as follows:

\[ \hat{Y}^{(3)}_{t+h} = \alpha_0 + \alpha_1 \frac{\hat{X}_{t+h} - \mu}{\sigma}. \]
The expected square error of this prediction is

\[ V^{(3)}(h) = \sum_{j=2}^{M} j!\alpha_j^2 + \frac{\alpha_1^2 S^2(h)}{\sigma^2}. \]

**Proof.** The covariance between \(X_t\) and \(Y_{t+h}\) is given by (2.2). The cross spectrum between \(X_t\) and \(Y_t\) will be therefore a constant times the spectrum of \(X_t\). Hence

\[ \hat{Y}^{(3)}_{t+h} = \alpha_0 + \alpha_1 \frac{\hat{X}_{t+h} - \mu}{\sigma}. \]

The mean square prediction error is

\[ V^{(3)}(h) = \mathbb{E} \left( Y_{t+h} - \hat{Y}^{(3)}_{t+h} \right)^2 = \mathbb{E} \left( Y_{t+h} \right)^2 - 2\mathbb{E} \left( Y_{t+h} \hat{Y}^{(3)}_{t+h} \right) + \mathbb{E} \left( \hat{Y}^{(3)}_{t+h} \right)^2 \]

where

\[ \mathbb{E} \left[ Y_{t+h} \right]^2 = \sum_{j=0}^{M} j!\alpha_j^2. \]

Notice that the remaining two terms are special cases of the corresponding terms in (3.2), because

\[ \hat{Y}^{(3)}_{t+h} = \sum_{j=0}^{1} \alpha_j H_j \left( \frac{\hat{X}_{t+h} - \mu}{\sigma} \right) = \sum_{j=0}^{1} \alpha_j A^j H_j (C) = a_0 + a_1 A H_1 (C), \]

see Theorem 3.1. Hence we have

\[ \mathbb{E} \left( Y_{t+h} \hat{Y}^{(3)}_{t+h} \right)^2 = \mathbb{E} \left( \hat{Y}^{(3)}_{t+h} \right)^2 = \sum_{j=0}^{1} j!\alpha_j^2 A^{2j}. \]

And so the mean square prediction error is

\[
V^{(3)}(h) = \sum_{j=0}^{M} j!\alpha_j^2 - \frac{1}{\sigma^2} \sum_{j=0}^{M} j!\alpha_j^2 A^{2j} = \sum_{j=0}^{M} j!\alpha_j^2 - \alpha_0^2 A^2 = \sum_{j=1}^{M} j!\alpha_j^2 - \alpha_1^2 A^2 = \\
= \sum_{j=2}^{M} j!\alpha_j^2 + \alpha_1^2 (1 - A^2) = \sum_{j=2}^{M} j!\alpha_j^2 + \alpha_1^2 B^2 = \sum_{j=2}^{M} j!\alpha_j^2 + \frac{\alpha_1^2 S^2(h)}{\sigma^2}.
\]

\[ \square \]
3.2 Three transformations

In this section three transformations are considered and all three predictions are calculated for each of the transformation, together with the mean square prediction errors of these predictions. The predictions are compared on the basis of the proportional increase in mean square error, defined as follows

\[ G^{(2)}(h) = \frac{V^{(2)}(h) - V^{(1)}(h)}{V^{(1)}(h)}, \quad (3.4) \]

\[ G^{(3)}(h) = \frac{V^{(3)}(h) - V^{(1)}(h)}{V^{(1)}(h)}. \quad (3.5) \]

3.2.1 Exponential transformation

Let us first consider the exponential transformation and apply the formulas derived in the previous section to find the optimal, naive and linear predictions and their mean square errors.

Corollary 3.4. Consider the transformation

\[ Y_t = \exp(X_t) = \sum_{j=0}^{\infty} \alpha_j H_j(Z_t) = \sum_{j=0}^{\infty} \gamma_j H_j(W_t), \]

where

\[ \alpha_j = \exp\left[ \mu + \frac{1}{2} \sigma^2 \right] \frac{\sigma^j}{j!}, \]

\[ \gamma_j = \exp\left[ \hat{X}_{t+h} + \frac{1}{2} S^2(h) \right] \frac{S_j(h)}{j!}. \]

Then the predictions for the transformed series are

\[ \hat{Y}_{t+h}^{(1)} = \exp\left[ \hat{X}_{t+h} + \frac{1}{2} S^2(h) \right], \]

\[ \hat{Y}_{t+h}^{(2)} = \exp\left[ \hat{X}_{t+h} \right], \]

\[ \hat{Y}_{t+h}^{(3)} = \exp\left[ \mu + \frac{1}{2} \sigma^2 \right] \left( 1 + \hat{X}_{t+h} - \mu \right), \]

their mean square prediction errors

\[ V^{(1)}(h) = \exp\left[ 2(\mu + \sigma^2) \right] \left[ 1 - \exp(-S^2(h)) \right], \]

\[ V^{(2)}(h) = \exp\left[ 2(\mu + \sigma^2) \right] \left[ 1 - 2 \exp\left( -\frac{3}{2} S^2(h) \right) + \exp(-2S^2(h)) \right], \]

\[ V^{(3)}(h) = \exp(2\mu + \sigma^2) \left[ \exp(\sigma^2) - 1 - \sigma^2 + S^2(h) \right]. \]
and the proportional increase in the mean square error for the naive and linear predictions

\[
G^{(2)}(h) = \frac{\left[ \exp\left(-\frac{1}{2}S^2(h)\right) - \exp(-S^2(h)) \right]^2}{1 - \exp(-S^2(h))},
\]
\[
G^{(3)}(h) = \frac{\exp(-\sigma^2) [S^2(h) - 1 - \sigma^2] + \exp(-S^2(h))}{1 - \exp(-S^2(h))}.
\]

**Proof.** The coefficients of Hermite expansion \(\alpha_i\) were derived using the generating function of Hermite polynomials (formula (7.4)), see Corollary 2.2. To show the expressions for \(\gamma_i\) recall from Theorem 3.1 that by definition \(W_{t+h} = (X_{t+h} - \hat{X}_{t+h})/S(h)\) and \(Z_t = (X_t - \mu)/\sigma\), where \(Z_t \sim N(0, 1)\) and \((X_{t+h} - \hat{X}_{t+h})\) has conditionally normal distribution with zero mean and variance \(S^2(h)\). Hence if we know the coefficients \(\alpha_j\) from Corollary 2.2 we also know the coefficients \(\gamma_j\); they will have \(\hat{X}_{t+h}\) instead of \(\mu\) and \(S^2(h)\) instead of \(\sigma^2\).

Using Theorem 3.1 we have

\[
\hat{Y}^{(1)}_{t+h} = \gamma_0 = \exp\left[\hat{X}_{t+h} + \frac{1}{2}S^2(h)\right]
\]

and

\[
V^{(1)}(h) = \sum_{j=1}^{\infty} j! \alpha_j^2 (1 - A^{2j})
\]

\[
= \exp\left(2\mu + \sigma^2\right) \sum_{j=1}^{\infty} j! \frac{\sigma^{2j}}{(j!)^2} (1 - A^{2j})
\]

\[
= \exp\left(2\mu + \sigma^2\right) \left[\sum_{j=1}^{\infty} \frac{\sigma^{2j}}{j!} - \sum_{j=1}^{\infty} \frac{\sigma^{2j}}{j!} A^{2j}\right]
\]

\[
= \exp\left(2\mu + \sigma^2\right) \left[\exp(\sigma^2) - 1 - \exp(\sigma^2 A^2) + 1\right]
\]

\[
= \exp\left[2(\mu + \sigma^2)\right] \left[1 - \exp(\sigma^2 A^2 - \sigma^2)\right]
\]

\[
= \exp\left[2(\mu + \sigma^2)\right] \left[1 - \exp(-S^2(h))\right].
\]

Using Theorem 3.2 we have

\[
\hat{Y}^{(2)}_{t+h} = \exp(X_{t+h})
\]
and to evaluate $V^{(1)}(h)$ we need first to calculate the second term in (3.3)

$$
\lim_{M \to \infty} \sum_{j=0}^{M} A^{2j} (j!)^{-1} \left[ \left( \frac{1}{2} (M-j) \right) \sum_{k=1}^{j} \frac{(j+2k)!}{k!} \left( -\frac{1}{2} B^2 \right)^k \right]
$$

$$
= \lim_{M \to \infty} \exp(2\mu + \sigma^2) \sum_{j=0}^{M} \frac{(A^2 \sigma^2)^j}{j!} \left[ \sum_{k=1}^{j} \frac{1}{k!} \left( -\frac{1}{2} B^2 \sigma^2 \right)^k \right]^2
$$

$$
= \exp(2\mu + \sigma^2) \exp(\sigma^2 A^2) \left[ \exp \left( -\frac{1}{2} B^2 \sigma^2 \right) - 1 \right]^2
$$

$$
= \exp(2\mu + \sigma^2) \exp(\sigma^2 - S^2(h)) \left[ 1 - \exp \left( -\frac{1}{2} S^2(h) \right) \right]^2
$$

$$
= \exp \left[ 2(\mu + \sigma^2) \right] \left[ \exp \left( -\frac{1}{2} S^2(h) \right) - \exp(-S^2(h)) \right]^2.
$$

Now we can write

$$
V^{(2)}(h) = \exp \left[ 2(\mu + \sigma^2) \right] \left[ 1 - \exp(-S^2(h)) + \left[ \exp \left( -\frac{1}{2} S^2(h) \right) - \exp(-S^2(h)) \right] \right]^2
$$

$$
= \exp \left[ 2(\mu + \sigma^2) \right] \left( 1 - 2 \exp \left( -\frac{3}{2} S^2(h) \right) + \exp(-2S^2(h)) \right).
$$

Using Theorem 3.3 we have

$$
\hat{Y}_{t+h}^{(3)} = \alpha_0 + \alpha_1 \frac{\hat{X}_{t+h} - \mu}{\sigma} = \exp \left[ \mu + \frac{1}{2} \sigma^2 \right] \left( 1 + \frac{\hat{X}_{t+h} - \mu}{\sigma} \right)
$$

$$
= \exp \left[ \mu + \frac{1}{2} \sigma^2 \right] \left( 1 + \hat{X}_{t+h} - \mu \right)
$$

and

$$
V^{(3)}(h) = \lim_{M \to \infty} \sum_{j=2}^{M} \frac{j! \alpha_j^2 S^2(h)}{\sigma^2}
$$

$$
= \exp \left( \mu + \frac{1}{2} \sigma^2 \right)^2 \left[ \sum_{j=0}^{\infty} \frac{\sigma^2 j!}{(j!)^2} - 1 - \sigma^2 + S^2(h) \right]
$$

$$
= \exp(2\mu + \sigma^2) \left[ \exp(\sigma^2) - 1 - \sigma^2 + S^2(h) \right].
$$

The proportional increase in mean square prediction error for the naive prediction and for the linear prediction follow immediately from the definition ((3.4) and (3.5)) and the derived expressions for $V^{(1)}(h)$, $V^{(2)}(h)$ and $V^{(3)}(h)$. \qed
3.2.2 Quadratic transformation

Let us now consider the quadratic transformation \( Y_t = X_t^2 \). We know from Corollary 2.3 that only the first three coefficients of the Hermite expansions (\( \alpha_0, \alpha_1 \) and \( \alpha_2 \)) are greater than zero, therefore finding the predictions and their errors will be much easier than in the case of the exponential transformation.

Corollary 3.5. (Quadratic transformation) Consider the transformation

\[
Y_t = X_t^2 = \sum_{i=0}^{2} \alpha_i H_i(Z_t) = \sum_{i=0}^{2} \gamma_i H_i(W_t),
\]

where

\[
\begin{align*}
\alpha_0 &= \mu^2 + \sigma^2, \\
\alpha_1 &= 2\sigma\mu, \\
\alpha_2 &= \sigma^2, \\
\gamma_0 &= \hat{X}_{t+h}^2 + S^2(h), \\
\gamma_1 &= 2\hat{X}_{t+h}S(h), \\
\gamma_2 &= S^2(h),
\end{align*}
\]

see corollary 2.3. Then the predictions for the transformed series are

\[
\begin{align*}
\hat{Y}_{t+h}^{(1)} &= \hat{X}_{t+h}^2 + S^2(h), \\
\hat{Y}_{t+h}^{(2)} &= \hat{X}_{t+h}^2, \\
\hat{Y}_{t+h}^{(3)} &= \sigma^2 - \mu^2 + 2\mu\hat{X}_{t+h},
\end{align*}
\]

their mean square expected errors

\[
\begin{align*}
V^{(1)}(h) &= 4(\mu^2 + \sigma^2)S^2(h) - 2S^4(h), \\
V^{(2)}(h) &= 4(\mu^2 + \sigma^2)S^2(h) - S^4(h), \\
V^{(3)}(h) &= 2\sigma^4 + 4\mu^2S^2(h),
\end{align*}
\]

and the proportional increase in the mean square error for the naive and linear predictions

\[
\begin{align*}
G^{(2)}(h) &= \frac{S^4(h)}{4(\mu^2 + \sigma^2) - 2S^4(h)}, \\
G^{(3)}(h) &= \frac{[\sigma^2 - S^2(h)]^2}{2(\mu^2 + \sigma^2)S^2(h) - S^4(h)}. \\
\end{align*}
\]

Proof. The coefficients of the Hermite expansion \( \alpha_i \) were derived in Corollary 2.3. Using the same reasoning as in the previous proof can say that the coefficients \( \gamma_i \) will be the same as \( \alpha_i \), but with \( \hat{X}_{t+h} \) instead of \( \mu \) and \( S^2(h) \) instead of \( \sigma^2 \). Using Theorem 3.1 we write

\[
\hat{Y}_{t+h}^{(1)} = \gamma_0 = \hat{X}_{t+h}^2 + S^2(h)
\]
and
\[ V^{(1)}(h) = \alpha_2^2(1 - A^2) + \alpha_2^2(1 - A^4) \]
\[ = 4\sigma^2 \mu^2 \left( \frac{S^2(h)}{\sigma^2} \right) + 2\sigma^4 \left[ 1 - \left( 1 - \frac{S^2(h)}{\sigma^2} \right)^2 \right] \]
\[ = 4\mu^2 S^2(h) + 2\left[ \sigma^4 - (\sigma^2 - S^2(h))^2 \right] \]
\[ = 4(\mu^2 + \sigma^2) S^2(h) - 2S^4(h). \]

Using Theorem 3.2 we write
\[ \hat{Y}^{(2)}_{t+h} = \hat{X}^2_{t+h} \]
and the second term in (3.3) (only the term with \( k = 1 \) is greater than zero)
\[ \sum_{j=0}^{M} \alpha_j^2 j! (1 - A^2^j) + \sum_{j=0}^{M} A^2^j (j!)^{-1} \left[ \sum_{k=1}^{1/2(M-j)} \alpha_{j+2k} (j + 2k)! \left( -\frac{1}{2} B^2 \right)^k \right] \]
\[ = \left[ 2\alpha_2 \left( -\frac{1}{2} B^2 \right) \right]^2 = \left[ \sigma^2 \frac{S^2(h)}{\sigma^2} \right]^2 = S^4(h), \]

hence
\[ V^{(2)}(h) = 4(\mu^2 + \sigma^2) S^2(h) - 2S^4(h) + S^4(h). \]

Finally using Theorem 3.3 we have
\[ \hat{Y}^{(3)}_{t+h} = \alpha_0 + \alpha_1 \frac{\hat{X}_{t+h} - \mu}{\sigma} \]
\[ = \mu^2 + \sigma^2 + 2\sigma \mu \frac{\hat{X}_{t+h} - \mu}{\sigma} = \sigma^2 - \mu^2 + 2\mu \hat{X}_{t+h} \]
and
\[ V^{(3)}(h) = 2\alpha_2^2 + \frac{\alpha_1^2 S^2(h)}{\sigma^2} = 2\sigma^2 + 4\mu^2 S^2(h). \]

Again, the proportional increase in mean square prediction error for the naive prediction and for the linear prediction follow immediately from the definition ((3.4) and (3.5)) and the derived expressions for \( V^{(1)}(h), V^{(2)}(h) \) and \( V^{(3)}(h) \).

### 3.2.3 Cubic transformation

Finally let us look at the cubic transformation. From Corollary 2.4 we know that only the first four coefficients (\( \alpha_0, \alpha_1, \alpha_2 \) and \( \alpha_3 \)) in the Hermite expansion are greater than zero. The application of the Theorems 3.1, 3.2 and 3.3 is very similar to the case of the quadratic transformation, although a little more tedious.
Corollary 3.6. (Cubic transformation) Consider the transformation

\[ Y_i = X_i^3 = \sum_{i=0}^{3} \alpha_i H_i(Z_i) = \sum_{i=0}^{3} \gamma_i H_i(W_i), \]

where

\[ \alpha_0 = \mu^3 + 3\sigma^2\mu, \]
\[ \alpha_1 = 3\sigma\mu^2 + 3\sigma^3, \]
\[ \alpha_2 = 3\sigma^2\mu, \]
\[ \alpha_3 = \sigma^3, \]
\[ \gamma_0 = \hat{X}_{t+h}^3 + 3S^2(h)\hat{X}_{t+h}, \]
\[ \gamma_1 = 3\hat{X}_{t+h}^2S(h) + 3S^3(h), \]
\[ \gamma_2 = 3S^2(h)\hat{X}_{t+h}, \quad \gamma_3 = S^3(h), \]

see corollary 2.4. Then the predictions for the transformed series are Then

\[ \hat{Y}_{t+h}^{(1)} = \hat{X}_{t+h}^3 + 3S^2(h)\hat{X}_{t+h}, \]
\[ \hat{Y}_{t+h}^{(2)} = \hat{X}_{t+h}, \]
\[ \hat{Y}_{t+h}^{(3)} = 3(\mu^3 + \sigma^2)\hat{X}_{t+h} - 2\mu^3 \]

their mean square errors

\[ V^{(1)}(h) = S^2(h) \left[ 9(\mu^2 + \sigma^2)^2 + 6S^4(h) - 18\mu^2S^2(h) + 18\sigma^2(2\mu^2 + \sigma^2 - S^2(h)) \right], \]
\[ V^{(2)}(h) = V^{(1)}(h) + 9S^4(h)(\mu^2 + \sigma^2 - S^2(h)), \]
\[ V^{(3)}(h) = 6\sigma^4(3\mu^2 + \sigma^2) + 9S^2(h)(\mu^2 + \sigma^2)^2, \]

and the proportional increase in the mean square error for the naive and linear predictions

\[ G^{(2)}(h) = \frac{3S^2(h)(\mu^2 + \sigma^2 - S^2(h))}{3(\mu^2 + \sigma^2)^2 + 2S^3(h) - 6\mu^2S^2(h) + 6\sigma^2(2\mu^2 + \sigma^2 - S^2(h))}; \]
\[ G^{(3)}(h) = \frac{6\sigma^4(3\mu^2 + \sigma^2) - S^2(h)(6S^4(h) - 18\mu^2S^2(h) + 18\sigma^2(2\mu^2 + \sigma^2 - S^2(h))}{S^2(h) \left[ 9(\mu^2 + \sigma^2)^2 + 6S^4(h) - 18\mu^2S^2(h) + 18\sigma^2(2\mu^2 + \sigma^2 - S^2(h)) \right]}. \]

Proof. To see that the coefficients \( \alpha_j \) are calculated correctly we write:

\[ Y_i = \sum_{j=0}^{4} \alpha_j H_j(Z_i) \]
\[ = \mu^3 + 3\sigma^2\mu + (3\sigma\mu^2 + 3\sigma^3) \frac{X_i - \mu}{\sigma} + \\
+ 3\sigma^2\mu \left[ \left( \frac{X_i - \mu}{\sigma} \right)^2 - 1 \right] + \sigma^3 \left[ \left( \frac{X_i - \mu}{\sigma} \right)^3 - 3 \frac{X_i - \mu}{\sigma} \right] \]
\[ = X_i^3. \]
The coefficients $\gamma_j$ are obtained in similar way as in the case previous transformations; with $\hat{X}_{t+h}$ instead of $\mu$ and $S^2(h)$ instead of $\sigma^2$. Using Theorem 3.1 we write

$$\hat{Y}_{t+h}^{(1)} = \gamma_0 = \hat{X}_{t+h}^3 + 3S^2(h)\hat{X}_{t+h}$$

and

$$V^{(1)}(h) = \sum_{j=1}^{4} j!\alpha_j^2 (1 - A^{2j})
= (3\sigma\mu^2 + 3\sigma^3)^2 \left[ 1 - \frac{\sigma^2 - S^2(h)}{\sigma^2} \right] + 2(3\sigma^2\mu)^2 \left[ 1 - \left( \frac{\sigma^2 - S^2(h)}{\sigma^2} \right)^2 \right] +
+6(\sigma^3)^2 \left[ 1 - \left( \frac{\sigma^2 - S^2(h)}{\sigma^2} \right)^3 \right]
= S^2(h) \left\{ 9(\mu^2 + \sigma^2)^2 + 6S^4(h) - 18\mu^2S^2(h) + 18\sigma^2(2\mu^2 + \sigma^2 - S^2(h)) \right\}.$$

Using Theorem 3.2 we write

$$\hat{Y}_{t+h}^{(2)} = \hat{X}_{t+h}^3.$$

The second term in (3.3) the only non-zero terms are

$$j = 0 \text{ and } k = 1 : \left[ 2\alpha_2 \left( -\frac{1}{2}B^2 \right) \right]^2 = \alpha_2^2 B^4
j = 0 \text{ and } k = 1 : A^2 \left[ 3!\alpha_3 \left( -\frac{1}{2}B^2 \right) \right]^2 = 9\alpha_3^2 A^2 B^4.$$

Hence

$$V^{(2)}(h) = V^{(1)}(h) + B^4(\alpha_2^2 + 9\alpha_3^2 A^2)
= V^{(1)}(h) + \frac{S^4(h)}{\sigma^4} \left[ 9\sigma^4 \mu^2 + 9\sigma^6 \left( \frac{\sigma^2 - S^2(h)}{\sigma^2} \right) \right]
= V^{(1)}(h) + 9S^4(h)(\mu^2 + \sigma^2 - S^2(h)).$$

Finally, using Theorem 3.3 we have

$$\hat{Y}_{t+h}^{(3)} = \alpha_0 + \alpha_1 \left( \frac{\hat{X}_{t+h} - \mu}{\sigma} \right) = \mu^3 + 3\sigma^2 \mu + (3\sigma^2 + 3\sigma^3) \left( \frac{\hat{X}_{t+h} - \mu}{\sigma} \right)
= 3(\mu^3 + \sigma^2)\hat{X}_{t+h} - 2\mu^3.$$
and

\[ V^{(3)}(h) = \sum_{j=2}^{3} j! \alpha_j^2 + \alpha_1^2 \frac{S^2(h)}{\sigma^2} \]
\[ = 18 \sigma^4 \mu^2 + 6 \sigma^6 + (3 \sigma \mu^2 + 3 \sigma^3)^2 \frac{S^2(h)}{\sigma^2} \]
\[ = 6 \sigma^4(3 \mu^2 + \sigma^2) + 9 S^2(h)(\mu^2 + \sigma^2)^2. \]

Again, the proportional increase in mean square prediction error for the naive prediction and for the linear prediction follow immediately from the definition ((3.4) and (3.5)) and the derived expressions for \( V^{(1)}(h), V^{(2)}(h) \) and \( V^{(3)}(h) \).

\[ \square \]

### 3.3 Predictions using spectral density

Theorem 2.11 provides the framework for finding the spectral density of a transformation of a Gaussian process. Once we know the spectral density of the transformed series, we can find the best linear predictor for the transformed series. Methods for finding the optimal linear prediction using the spectral density are described in Anděl [3] and in Grenander and Rosenblatt [7]. General Theorem is given in Appendix B. This method, however general, is not very practical. In special cases much simpler formulas can be derived.

Choi and Taniguchi [5] derived formula for the mean square prediction error of the naive prediction for square-transformed Gaussian process and compared the naive prediction with the bias-adjusted prediction.

**Theorem 3.7. (Naive prediction)** Let \( X_t \) be Gaussian stationary process with zero mean and variance \( \sigma^2 \) and spectral density

\[ g(\lambda) = \frac{1}{2\pi} |c(e^{-i\lambda})|^2, \quad |c(0)|^2 = \sigma_e^2, \]

where \( c(z) \neq 0 \) for \( |z| \leq 1 \). Then the mean square prediction error of the naive prediction is

\[ \mathbb{E}[X_{t+1}^2 - \hat{X}_{t+1}^2] = 4 \sigma_e^2 \sigma^2 - \sigma_e^4, \]

where \( \hat{X}_{t+1} \) is the optimal predictor of \( X_{t+1} \).

**Proof.** Process \( X_t \) can be expressed using its spectral density (see e.g. Anděl [3] or Prášková [11]) in the following way

\[ X_t = \int_{-\pi}^{\pi} e^{it\lambda} d\lambda. \]

The best linear prediction (see Theorem 8.1 in Appendix B and Grenander and Rosenblatt [7]) is

\[ \hat{X}_{t+1} = \int_{-\pi}^{\pi} e^{i(t+1)\lambda} \frac{c(e^{-i\lambda}) - c(0)}{c(e^{-i\lambda})} d\lambda. \]
Define \( e_{t+1} = X_{t+1} - \hat{X}_{t+1} \). Then \( E(e_{t+1}) = 0 \) and
\[
\epsilon_{t+1} = \int_{-\pi}^{\pi} e^{i(t+1)\lambda} \, dz(\lambda) - \int_{-\pi}^{\pi} e^{i(t+1)\lambda} \frac{c(e^{-i\lambda}) - c(0)}{c(e^{-i\lambda})} \, dz(\lambda) = c(0) \int_{-\pi}^{\pi} e^{i(t+1)\lambda} \frac{1}{c(e^{-i\lambda})} \, dz(\lambda)
\]
and so
\[
E(e_{t+1})^2 = \sigma^2_e.
\]
Because we assume that \( X_t \) is Gaussian also \( \epsilon_{t+1} \) is Gaussian and from the basic properties of the normal distribution (see e.g. Andel [4]) we have \( E\epsilon_{t+1}^3 = 0 \) and \( E\epsilon_{t+1}^4 = 3\sigma_e^4 \). Hence we can write
\[
EX_{t+1}^2 = E(\epsilon_{t+1} + \hat{X}_{t+1})^2 = E\epsilon_{t+1}^2 + 2E\epsilon_{t+1}\hat{X}_{t+1} + E\hat{X}_{t+1}^2 = E\hat{X}_{t+1}^2 + \sigma_e^2.
\]
The mean square prediction error is then
\[
E[X_{t+1}^2 - \hat{X}_{t+1}^2]^2 = E[(\epsilon_{t+1} + \hat{X}_{t+1})^2 - \hat{X}_{t+1}^2]^2 = E[\epsilon_{t+1}^2 + 2\epsilon_{t+1}\hat{X}_{t+1} + \hat{X}_{t+1}^2]^2 = E[\epsilon_{t+1}^2 + 4\epsilon_{t+1}^2\hat{X}_{t+1} + 4\epsilon_{t+1}\hat{X}_{t+1}^2 + 4\epsilon_{t+1}\hat{X}_{t+1}^2] = 3\sigma_e^4 + 4\sigma_e^3E\hat{X}_{t+1} + 4\sigma_e^2E\hat{X}_{t+1}^2 = 3\sigma_e^4 + 4\sigma_e^2(3\sigma_e^2 - \sigma_e^2) = 3\sigma_e^4 + 4\sigma_e^2EX_{t+1}^2 - 4\sigma_e^4 = 4\sigma_e^2EX_{t+1}^2 - \sigma_e^4
\]
Notation \( V(2)(1) \) was chosen in accordance with previously used notation.

**Theorem 3.8. (The bias-adjusted prediction)** Under the same assumptions as in the Theorem 3.7 consider now bias-adjusted transformed series \( \tilde{V}_t = Y_t - EY_t = X_t^2 - EX_t^2 \) and bias-adjusted prediction \( \tilde{Y}_{t+1} = \hat{Y}_{t+1} - E\hat{Y}_{t+j} = \hat{X}_{t+1}^2 - E\hat{X}_{t+1}^2 \). The expected error of the bias-adjusted prediction is
\[
E \left( \tilde{V}_{t+1} - \tilde{Y}_{t+1} \right)^2 = V(2)(1) - (\sigma^2 - \sigma_e^2)^2.
\]

**Proof.** We have
\[
E \left( V_{t+1} - \hat{V}_{t+1} \right)^2 = E \left[ (X_{t+1}^2 - EX_{t+1}^2) - (\hat{X}_{t+1}^2 - E\hat{X}_{t+1}^2) \right]^2 = E \left[ (X_{t+1}^2 - \hat{X}_{t+1}^2) - (EX_{t+1}^2 - E\hat{X}_{t+1}^2) \right]^2 = E \left( U_{t+1} - EU_{t+1} \right)^2 = EU_{t+1}^2 - (EU_{t+1})^2,
\]
where \( U_t = X_t^2 - \hat{X}_t^2 \). Now we have
\[
EU_{t+1}^2 = E \left( X_{t+1}^2 - \hat{X}_{t+1}^2 \right)^2 = V(2)(1)
\]
\[
EU_{t+1} = EX_{t+1} - E\hat{X}_{t+1}^2 = \sigma^2 - \sigma_e^2.
\]
Further comparison of the naive and bias-adjusted predictions in case of the square-transformed process can be found in Choi and Taniguchi [5].
Chapter 4

Application to ARMA processes

In this chapter we will apply the derived formulas to the three predictions (optimal, naive and linear) and their mean square errors on each of the three transformations (exponential, quadratic and cubic) of processes MA(1), AR(1) and ARMA(1, 1). From Corollaries 3.4, 3.5 and 3.6 we know that the predictions and their mean square errors are functions of the following four parameters: the mean $\mu$ and variance $\sigma^2$ of the original series $X_t$, the optimal prediction in the original series $\hat{X}_{t+h}$ and its error $S^2(h)$.

4.1 MA(1) process

Consider the following MA(1) process

$$X_t = \epsilon_t + \theta \epsilon_{t-1},$$

where $\epsilon_t \sim N(0, \sigma^2)$ are i.i.d. Obviously $\mu = 0$. Variance of the process is

$$\sigma^2 = \text{var}(X_t) = \sigma^2(1 + \theta^2)$$

The optimal prediction is

$$\hat{X}_{t+1} = \theta \epsilon_t = \sum_{j=1}^{\infty} (-\theta)^j X_{t+1-j}$$

and the error of the 1-step ahead prediction is

$$S^2(1) = \sigma^2.$$

Now we can use the general formulas to find the predictions and their mean square errors for this process.
4.1.1 Exponential transformation
Consider the transformation \( Y_t = \exp(X_t) \).

From corollary 3.4 the predictions for this transformation are

\[
\hat{Y}^{(1)}_{t+1} = \exp\left(-\sum_{j=1}^{\infty} (-\theta)^j X_{t+1-j} + \frac{1}{2} \sigma^2 \right),
\]

\[
\hat{Y}^{(2)}_{t+1} = \exp\left[-\sum_{j=1}^{\infty} (-\theta)^j X_{t+1-j} \right],
\]

\[
\hat{Y}^{(3)}_{t+1} = \exp\left[\frac{1}{2} \sigma^2 (1 + \theta^2) \right] \left[1 - \sum_{j=1}^{\infty} (-\theta)^j X_{t+1-j} \right]
\]

and the mean square errors of these predictions are

\[
V^{(1)}(1) = \exp[2\sigma^2 (1 + \theta^2)] \left[1 - \exp(\sigma^2) \right],
\]

\[
V^{(2)}(1) = \exp[2\sigma^2 (1 + \theta^2)] \left[1 - 2 \exp\left(-\frac{3}{2} \sigma^2 \right) + \exp(-2\sigma^2) \right],
\]

\[
V^{(3)}(1) = \exp \left[\sigma^2 (1 + \theta^2) \right] \left[\exp[\sigma^2 (1 + \theta^2)] - 1 - \sigma^2 \theta^2 \right].
\]

4.1.2 Quadratic transformation
Consider the transformation \( Y_t = X_t^2 \).

From corollary 3.5 the predictions for this transformation are

\[
\hat{Y}^{(1)}_{t+1} = \left[-\sum_{j=1}^{\infty} (-\theta)^j X_{t+1-j} \right]^2 + \sigma^2,
\]

\[
\hat{Y}^{(2)}_{t+1} = \left[-\sum_{j=1}^{\infty} (-\theta)^j X_{t+1-j} \right]^2,
\]

\[
\hat{Y}^{(3)}_{t+1} = \sigma^2 (1 + \theta^2)
\]

and the mean square errors of these predictions are

\[
V^{(1)}(1) = 4 \left[\sigma^2 (1 + \theta^2) \right] \sigma^2 - 2(\sigma^2)^2,
\]

\[
V^{(2)}(1) = 4 \left[\sigma^2 (1 + \theta^2) \right] \sigma^2 - (\sigma^2)^2,
\]

\[
V^{(3)}(1) = 2 \left[\sigma^2 (1 + \theta^2) \right]^2.
\]
4.1.3 Cubic transformation

Consider the transformation $Y_t = X_t^3$.

From corollary 3.6 the predictions for this transformation are

$$
\hat{Y}_{t+1}^{(1)} = \left[ -\sum_{j=1}^{\infty} (-\theta)^j X_{t+1-j} \right]^3 + 3\sigma^2 \left( -\sum_{j=1}^{\infty} (-\theta)^j X_{t+1-j} \right),
$$

$$
\hat{Y}_{t+1}^{(2)} = \left[ -\sum_{j=1}^{\infty} (-\theta)^j X_{t+1-j} \right]^3,
$$

$$
\hat{Y}_{t+1}^{(3)} = 3\sigma^2 (1 + \theta^2) \left( -\sum_{j=1}^{\infty} (-\theta)^j X_{t+1-j} \right)
$$

and the mean square prediction errors

$$
V^{(1)}(1) = 6\sigma^6 \left[ 4 + 6\theta^2 + 3\theta^4 \right]
$$

$$
V^{(2)}(1) = V^{(1)}(h) + 9\theta^2 \sigma^6,
$$

$$
V^{(3)}(1) = 3\sigma^6 (1 + \theta^2)^3 \left( 5 + 3\theta^2 \right).
$$

<table>
<thead>
<tr>
<th>$Y_t = \exp(X_t)$</th>
<th>$h = 1$</th>
<th>$h = 2$</th>
<th>$h = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta = 0.25$</td>
<td>$G^{(2)}(h)$</td>
<td>0.015</td>
<td>0.015</td>
</tr>
<tr>
<td>$\theta = 0.25$</td>
<td>$G^{(3)}(h)$</td>
<td>4.097</td>
<td>4.097</td>
</tr>
<tr>
<td>$\theta = 0.75$</td>
<td>$G^{(2)}(h)$</td>
<td>0.080</td>
<td>0.080</td>
</tr>
<tr>
<td>$\theta = 0.75$</td>
<td>$G^{(3)}(h)$</td>
<td>0.350</td>
<td>0.350</td>
</tr>
</tbody>
</table>

Table 4.1: MA(1), $Y_t = \exp(X_t)$, proportional increase in mean square prediction error

<table>
<thead>
<tr>
<th>$Y_t = X_t^2$</th>
<th>$h = 1$</th>
<th>$h = 2$</th>
<th>$h = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\theta = 0.25$</td>
<td>$G^{(2)}(h)$</td>
<td>0.015</td>
<td>0.015</td>
</tr>
<tr>
<td>$\theta = 0.25$</td>
<td>$G^{(3)}(h)$</td>
<td>7.758</td>
<td>7.758</td>
</tr>
<tr>
<td>$\theta = 0.75$</td>
<td>$G^{(2)}(h)$</td>
<td>0.110</td>
<td>0.110</td>
</tr>
<tr>
<td>$\theta = 0.75$</td>
<td>$G^{(3)}(h)$</td>
<td>0.694</td>
<td>0.694</td>
</tr>
</tbody>
</table>

Table 4.2: MA(1), $Y_t = X_t^2$, proportional increase in mean square prediction error
\[
Y_t = X_t^3
\]

<table>
<thead>
<tr>
<th>( Y_t = X_t^3 )</th>
<th>( h = 1 )</th>
<th>( h = 2 )</th>
<th>( h = 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta = 0.25 )</td>
<td>( G^{(2)}(h) )</td>
<td>0.019</td>
<td>0.019</td>
</tr>
<tr>
<td>( \theta = 0.25 )</td>
<td>( G^{(3)}(h) )</td>
<td>3.275</td>
<td>3.275</td>
</tr>
<tr>
<td>( \theta = 0.75 )</td>
<td>( G^{(2)}(h) )</td>
<td>0.097</td>
<td>0.097</td>
</tr>
<tr>
<td>( \theta = 0.75 )</td>
<td>( G^{(3)}(h) )</td>
<td>0.205</td>
<td>0.205</td>
</tr>
</tbody>
</table>

Table 4.3: MA(1), \( Y_t = X_t^3 \), proportional increase in mean square prediction error

### 4.2 MA(\( \infty \)) process

It is very simple to extend the previous case to MA(\( \infty \)) process

\[
X_t = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} \cdots,
\]

where again \( \epsilon_t \sim N(0, \sigma^2) \) are i.i.d. The variance of this process is

\[
\sigma^2 = \text{var}(X_t) = \sigma^2 \sum_{j=0}^{\infty} \theta_j^2,
\]

where \( \theta_0 = 1 \). The optimal prediction can be calculated by means of standard methods, we will use the notation \( \hat{X}_{t+1} \). The error of this prediction is, as in the previous case,

\[
S^2(1) = \sigma^2.
\]

Now we can use the general formulas to find the predictions and their mean square errors for this process.

#### 4.2.1 Exponential transformation

Consider the transformation

\[
Y_t = \exp(X_t).
\]

From corollary 3.4 the predictions for this transformation are

\[
\hat{Y}^{(1)}_{t+1} = \exp \left( \hat{X}_{t+1} + \frac{1}{2} \sigma^2 \right),
\]

\[
\hat{Y}^{(2)}_{t+1} = \exp \left[ \hat{X}_{t+1} \right],
\]

\[
\hat{Y}^{(3)}_{t+1} = \exp \left[ \frac{1}{2} \sigma^2 \sum_{j=0}^{\infty} \theta_j^2 \right] \left[ 1 + \hat{X}_{t+1} \right]
\]
and the mean square errors of these predictions are

\[
V^{(1)}(1) = \exp \left( 2\sigma^2 \sum_{j=0}^{\infty} \theta^2_j \right) \left[ 1 - \exp(\sigma^2_\epsilon) \right],
\]

\[
V^{(2)}(1) = \exp \left( 2\sigma^2 \sum_{j=0}^{\infty} \theta^2_j \right) \left[ 1 - 2 \exp \left( -\frac{3}{2} \sigma^2_\epsilon \right) + \exp(-2\sigma^2_\epsilon) \right],
\]

\[
V^{(3)}(1) = \exp \left[ \sigma^2_\epsilon \sum_{j=0}^{\infty} \theta^2_j \right] \left[ \exp \left( \sigma^2_\epsilon \sum_{j=0}^{\infty} \theta^2_j \right) - 1 - \sigma^2_\epsilon \sum_{j=0}^{\infty} \theta^2_j + \sigma^2_\epsilon \right].
\]

### 4.2.2 Quadratic transformation

Consider the transformation

\[ Y_t = \exp(X_t). \]

From corollary 3.5 the predictions for this transformation are

\[
\hat{Y}^{(1)}_{t+1} = \hat{X}^2_{t+1} + \sigma^2_\epsilon,
\]

\[
\hat{Y}^{(2)}_{t+1} = \hat{X}^2_{t+1},
\]

\[
\hat{Y}^{(3)}_{t+1} = \sigma^2_\epsilon \sum_{j=0}^{\infty} \theta^2_j
\]

and the mean square errors of these predictions are

\[
V^{(1)}(1) = 4 \left[ \sigma^2_\epsilon \sum_{j=0}^{\infty} \theta^2_j \right] \sigma^2_\epsilon - 2\sigma^4_\epsilon,
\]

\[
V^{(2)}(1) = 4 \left[ \sigma^2_\epsilon \sum_{j=0}^{\infty} \theta^2_j \right] \sigma^2_\epsilon - \sigma^4_\epsilon,
\]

\[
V^{(3)}(1) = 2 \left[ \sigma^2_\epsilon \sum_{j=0}^{\infty} \theta^2_j \right]^2.
\]

### 4.2.3 Cubic transformation

Consider the transformation

\[ Y_t = X^3_t. \]

From corollary 3.6 the predictions for this transformation are

\[
\hat{Y}^{(1)}_{t+1} = \hat{X}^3_{t+1} + 3\sigma^2_\epsilon \hat{X}_{t+1},
\]

\[
\hat{Y}^{(2)}_{t+1} = \hat{X}^3_{t+1},
\]

\[
\hat{Y}^{(3)}_{t+1} = 3\sigma^2_\epsilon \hat{X}_{t+1}
\]
and the mean square prediction errors are

\[
V^{(1)}(1) = 3\sigma_\epsilon^6 \left[ 9 \left( \sum_{j=0}^{\infty} \theta_j^2 \right)^2 + 6 \sum_{j=0}^{\infty} \theta_j^2 + 2 \right]
\]

\[
V^{(2)}(1) = V^{(1)}(1) + 9\sigma_\epsilon^6 \left[ \left( \sum_{j=0}^{\infty} \theta_j^2 \right) - 1 \right],
\]

\[
V^{(3)}(1) = 3\sigma_\epsilon^6 \left( \sum_{j=0}^{\infty} \theta_j^2 \right)^3 \left( 2 + 3 \sum_{j=0}^{\infty} \theta_j^2 \right).
\]

### 4.3 AR(1) process

Consider now the AR(1) process

\[ X_t = \varphi X_{t-1} + \epsilon_t, \]

where \( \epsilon_t \sim N(0, \sigma_\epsilon^2) \) are i.i.d. This process can be also written in the following form

\[ X_t = \sum_{j=0}^{\infty} \varphi^j \epsilon_{t-j} \]

from which it is obvious that the variance of the process is

\[ \sigma^2 = \text{var}(X_t) = \sigma_\epsilon^2 \sum_{j=0}^{\infty} \varphi^{2j} = \frac{\sigma_\epsilon^2}{1 - \varphi^2}. \]

The optimal prediction is

\[ \hat{X}_{t+1} = \varphi X_t \]

and the mean square prediction error is

\[ S^2(1) = \sigma_\epsilon^2. \]

Now we can use the general formulas to find the predictions and their mean square errors for this process.

#### 4.3.1 Exponential transformation

Consider the transformation

\[ Y_t = \exp(X_t). \]
The predictions for this transformation are

\[ \hat{Y}_{t+1}^{(1)} = \exp \left[ \varphi X_t + \frac{1}{2} \sigma^2 \right], \]
\[ \hat{Y}_{t+1}^{(2)} = \exp \left[ \varphi X_t \right], \]
\[ \hat{Y}_{t+1}^{(3)} = \exp \left[ \frac{\sigma^2}{2(1 - \varphi^2)} \right] (1 + \varphi X_t) \]

and the mean square errors of these predictions are

\[ V^{(1)}(1) = \exp \left( \frac{2 \sigma^4}{1 - \varphi^2} \right) [1 - \exp(-\sigma^2)], \]
\[ V^{(2)}(1) = \exp \left( \frac{2 \sigma^4}{1 - \varphi^2} \right) \left[ 1 - 2 \exp \left( -\frac{3}{2} \sigma^2 \right) + \exp \left( -2\sigma^2 \right) \right], \]
\[ V^{(3)}(1) = \exp \left( \frac{\sigma^2}{1 - \varphi^2} \right) \left[ \exp \left( \frac{\sigma^2}{1 - \varphi^2} \right) - 1 - \frac{\sigma^2}{1 - \varphi^2} + \sigma^2 \right]. \]

**4.3.2 Quadratic transformation**

Consider the transformation

\[ Y_t = X_t^2 \]

The predictions for this transformation are

\[ \hat{Y}_{t+1}^{(1)} = \varphi X_t^2 + \sigma^2, \]
\[ \hat{Y}_{t+1}^{(2)} = \varphi X_t^2, \]
\[ \hat{Y}_{t+1}^{(3)} = \frac{\sigma^2}{1 - \varphi^2} \]

and the mean square errors of these predictions are

\[ V^{(1)}(1) = 4 \left( \frac{\sigma^4}{1 - \varphi^2} \right) - 2\sigma^4, \]
\[ V^{(2)}(1) = 4 \left( \frac{\sigma^4}{1 - \varphi^2} \right) - \sigma^4, \]
\[ V^{(3)}(1) = 2 \left( \frac{\sigma^2}{1 - \varphi^2} \right)^2. \]

**4.3.3 Cubic transformation**

Consider the transformation

\[ Y_t = X_t^3 \]
The predictions for this transformation are

\[
\hat{Y}_{t+1}^{(1)} = \varphi X_t^3 + 3\sigma^2 \hat{X}_{t+h},
\]

\[
\hat{Y}_{t+1}^{(2)} = \varphi X_t^3,
\]

\[
\hat{Y}_{t+1}^{(3)} = 3 \left( \sum_{j=0}^{\infty} \varphi^2 \right) \hat{X}_{t+1}
\]

and the mean square errors of these predictions are

\[
V^{(1)}(1) = 3\sigma^6 \left[ 9 \frac{1}{1-\varphi^2} + 6 \frac{1}{1-\varphi^2} + 2 \right],
\]

\[
V^{(2)}(1) = V^{(1)} + 9\sigma^6 \frac{\varphi^2}{1-\varphi^2},
\]

\[
V^{(3)}(1) = 3\sigma^6 \left[ \frac{1}{1-\varphi^2} \right]^2 \left[ \frac{2}{1-\varphi^2} + 3 \right].
\]

<table>
<thead>
<tr>
<th>(Y_t = \exp(X_t))</th>
<th>(h = 1)</th>
<th>(h = 2)</th>
<th>(h = 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\varphi = 0.25)</td>
<td>(G^{(2)}(h))</td>
<td>0.015</td>
<td>0.016</td>
</tr>
<tr>
<td>(\varphi = 0.25)</td>
<td>(G^{(3)}(h))</td>
<td>4.120</td>
<td>3.850</td>
</tr>
<tr>
<td>(\varphi = 0.75)</td>
<td>(G^{(2)}(h))</td>
<td>0.080</td>
<td>0.090</td>
</tr>
<tr>
<td>(\varphi = 0.75)</td>
<td>(G^{(3)}(h))</td>
<td>0.681</td>
<td>0.292</td>
</tr>
</tbody>
</table>

Table 4.4: AR(1), \(Y_t = \exp(X_t)\), proportional increase in mean square prediction error

<table>
<thead>
<tr>
<th>(Y_t = X_t^2)</th>
<th>(h = 1)</th>
<th>(h = 2)</th>
<th>(h = 3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(\varphi = 0.25)</td>
<td>(G^{(2)}(h))</td>
<td>0.015</td>
<td>0.016</td>
</tr>
<tr>
<td>(\varphi = 0.25)</td>
<td>(G^{(3)}(h))</td>
<td>7.791</td>
<td>7.289</td>
</tr>
<tr>
<td>(\varphi = 0.75)</td>
<td>(G^{(2)}(h))</td>
<td>0.070</td>
<td>0.119</td>
</tr>
<tr>
<td>(\varphi = 0.75)</td>
<td>(G^{(3)}(h))</td>
<td>1.317</td>
<td>0.610</td>
</tr>
</tbody>
</table>

Table 4.5: AR(1), \(Y_t = X_t^2\), proportional increase in mean square prediction error

### 4.4 ARMA(1, 1) process

Consider now the ARMA(1, 1) process

\[X_t = \varphi X_{t-1} + \epsilon_t + \theta \epsilon_{t-1},\]
**Chapter 4. Application to ARMA Processes**

\[ Y_t = X_t^3 \]

<table>
<thead>
<tr>
<th>( \varphi )</th>
<th>( G^{(2)}(h) )</th>
<th>( h = 1 )</th>
<th>( h = 2 )</th>
<th>( h = 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.25</td>
<td>0.019</td>
<td>0.020</td>
<td>0.020</td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>3.290</td>
<td>3.068</td>
<td>3.055</td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td>0.073</td>
<td>0.102</td>
<td>0.112</td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td>0.456</td>
<td>0.174</td>
<td>0.101</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.6: AR(1), \( Y_t = X_t^3 \), proportional increase in mean square prediction error

where \( \epsilon_t \sim \mathcal{N}(0, \sigma^2_\epsilon) \) are i.i.d. If \( |\varphi| < 1 \) then this process can be also written in the following form

\[
X_t = \sum_{j=0}^{\infty} c_j \epsilon_{t-j},
\]

where

\[
c_0 = 1, \quad c_j = \varphi^{j-1}(\theta + \varphi), \quad j \geq 1,
\]

see for example Prášková [11], p. 69. The variance of this process is

\[
\sigma^2 = \sigma^2_\epsilon \frac{1 + 2\varphi\theta + \theta^2}{1 - \varphi^2}
\]

The optimal prediction can be calculated by means of standard methods, we will use the notation \( \hat{X}_{t+1} \). The error of this prediction is, as in the previous case, and the mean square prediction error is

\[
S^2(1) = \sigma^2_\epsilon
\]

Now we can use the general formulas to find the predictions and their mean square errors for this process.

### 4.4.1 Exponential transformation

Consider the transformation

\[ Y_t = \exp(X_t) \]

The predictions for this transformation are

\[
\begin{align*}
\hat{Y}_{t+1}^{(1)} &= \exp\left(\hat{X}_{t+1} + \frac{1}{2}\sigma^2_\epsilon\right), \\
\hat{Y}_{t+1}^{(2)} &= \exp\left(\hat{X}_{t+1}\right), \\
\hat{Y}_{t+1}^{(3)} &= \exp\left(\frac{1}{2}\sigma^2_\epsilon \frac{1 + \varphi\theta + \theta^2}{1 - \varphi^2}\right) (1 + \hat{X}_{t+1})
\end{align*}
\]
and the mean square errors of these predictions are

\[ V^{(1)}(1) = \exp \left( 2\sigma^2 \frac{1 + \varphi\theta + \theta^2}{1 - \varphi^2} \right) \left[ 1 - \exp(-\sigma^2) \right], \]
\[ V^{(2)}(1) = \exp \left( 2\sigma^2 \frac{1 + \varphi\theta + \theta^2}{1 - \varphi^2} \right) \left[ 1 - 2 \exp \left( -\frac{3}{2} \sigma^2 \right) + \exp \left( -2\sigma^2 \right) \right], \]
\[ V^{(3)}(1) = \exp \left[ \sigma^2 \frac{1 + 2\varphi\theta + \theta^2}{1 - \varphi^2} \right] \left[ \exp \left( \sigma^2 \frac{1 + 2\varphi\theta + \theta^2}{1 - \varphi^2} \right) - 1 - \sigma^2 \frac{1 + 2\varphi\theta + \theta^2}{1 - \varphi^2} + \sigma^2 \right]. \]

### 4.4.2 Quadratic transformation

Consider the transformation

\[ Y_t = X_t^2 \]

The predictions for this transformation are

\[ \hat{Y}_{t+1}^{(1)} = \varphi X_t^2 + \sigma^2, \]
\[ \hat{Y}_{t+1}^{(2)} = \varphi X_t^2, \]
\[ \hat{Y}_{t+1}^{(3)} = \sigma^2 \frac{1 + 2\varphi\theta + \theta^2}{1 - \varphi^2}, \]

and the mean square errors of these predictions are

\[ V^{(1)}(1) = 4\sigma^4 \frac{1 + 2\varphi\theta + \theta^2}{1 - \varphi^2} - 2\sigma^2, \]
\[ V^{(2)}(1) = 4\sigma^4 \frac{1 + 2\varphi\theta + \theta^2}{1 - \varphi^2} - \sigma^2, \]
\[ V^{(3)}(1) = 2\sigma^4 \left( \frac{1 + 2\varphi\theta + \theta^2}{1 - \varphi^2} \right)^2. \]

### 4.4.3 Cubic transformation

Consider the transformation

\[ Y_t = X_t^3 \]

The predictions for this transformation are

\[ \hat{Y}_{t+1}^{(1)} = \varphi X_t^3 + 3\sigma^2 \hat{X}_{t+1}, \]
\[ \hat{Y}_{t+1}^{(2)} = \varphi X_t^3, \]
\[ \hat{Y}_{t+1}^{(3)} = 3 \left( \sigma^2 \frac{1 + 2\varphi\theta + \theta^2}{1 - \varphi^2} \right) \hat{X}_{t+1}. \]
and the mean square errors of these predictions are

\[ V^{(1)}(1) = 3\sigma^6 \left[ 9 \left( \frac{1 + 2\varphi \theta + \theta^2}{1 - \varphi^2} \right)^2 + 6 \left( \frac{1 + 2\varphi \theta + \theta^2}{1 - \varphi^2} \right) + 2 \right], \]

\[ V^{(2)}(1) = V^{(1)} + 9\sigma^6 \left( \frac{1 + 2\varphi \theta + \theta^2}{1 - \varphi^2} - 1 \right), \]

\[ V^{(3)}(1) = 3\sigma^6 \left( \frac{1 + 2\varphi \theta + \theta^2}{1 - \varphi^2} \right)^2 \left[ 2\frac{1 + 2\varphi \theta + \theta^2}{1 - \varphi^2} + 3 \right]. \]
Chapter 5
Empirical part

The formulas derived in the previous chapter were verified in a simulation. Tested were three models; MA(1), AR(1) and ARMA(1,1) with various parameters, but constant variance of innovations $\text{var}(\epsilon) = \sigma^2 = 1$. The observed average squared errors for each prediction

\[
\begin{align*}
V^{(1)}(1) &= \frac{\sum_{i=1}^{n}(\hat{Y}^{(1)}_{t+1} - Y^{(1)}_{t+1})^2}{n} \\
V^{(2)}(1) &= \frac{\sum_{i=1}^{n}(\hat{Y}^{(2)}_{t+1} - Y^{(2)}_{t+1})^2}{n} \\
V^{(3)}(1) &= \frac{\sum_{i=1}^{n}(\hat{Y}^{(3)}_{t+1} - Y^{(3)}_{t+1})^2}{n}
\end{align*}
\]

were compared with the theoretical values $V^{(1)}(1)$, $V^{(2)}(1)$ and $V^{(3)}(1)$. The variance of innovations is in all cases $\sigma^2 = 1$ and therefore error of the 1-step ahead prediction $X_{t+1}$ is in all cases $S^2(1) = 1$. The simulation was carried out for $n = 20,000$ observations.

Note that in each row the observed average squared error of the optimal prediction is better than that of the naive and linear predictions, which is in accordance with the expectations. For each transformation it can be said that with increasing coefficient $\theta$ also the error of the prediction increases, which is also in accordance with the expectations. In general, exponential transformation leads to very large errors for large AR coefficients ($\varphi$ close to 1), but the prediction for cubic-transformed process has larger error for small values of the AR coefficient.

We can also see significant difference between the theoretical mean square error and calculated average square error and that especially in case of the exponential transformation (see e.g. table 5.7) for large values of $\varphi$. Similar discrepancies can be observed in case of the cubic transformation, while in case of the quadratic transformation the theoretical and observed errors are relatively close. After running the simulation several times we recorded very large positive as well as negative differences between the theoretical and calculated errors. Together with the fact that simulations with more observations lead in general to reduction and gradual elimination of these differences we came to the conclusion that there
is no systematic error. It seems that in case of large values of $\varphi$ any randomly occurring perturbations are projected within the process very far, thus distorting the calculated average square errors of predictions.

In general it can be said that the simulation verified the results derived in the previous chapters.

$$\begin{array}{cccccccc}
\theta & \sigma^2 & V^{(1)}(1) & V^{(1)}(1) & V^{(2)}(1) & V^{(2)}(1) & V^{(3)}(1) & V^{(3)}(1) \\
0.1 & 1.010 & 4.702 & 4.765 & 5.150 & 5.194 & 4.702 & 4.765 \\
0.5 & 1.250 & 7.586 & 7.701 & 8.243 & 8.395 & 7.671 & 7.920 \\
0.9 & 1.810 & 23.629 & 23.602 & 25.671 & 25.728 & 26.331 & 26.278 \\
\end{array}$$

Table 5.1: MA(1) process, $Y_t = \exp(X_t)$

$$\begin{array}{cccccccc}
\theta & \sigma^2 & V^{(1)}(1) & V^{(1)}(1) & V^{(2)}(1) & V^{(2)}(1) & V^{(3)}(1) & V^{(3)}(1) \\
0.1 & 1.010 & 2.053 & 2.040 & 3.079 & 3.040 & 2.054 & 2.040 \\
0.5 & 1.250 & 3.102 & 3.000 & 4.119 & 4.000 & 3.239 & 3.125 \\
0.9 & 1.810 & 5.212 & 5.240 & 6.198 & 6.240 & 6.519 & 6.552 \\
\end{array}$$

Table 5.2: MA(1) process, $Y_t = X_t^2$

$$\begin{array}{cccccccc}
\theta & \sigma^2 & V^{(1)}(1) & V^{(1)}(1) & V^{(2)}(1) & V^{(2)}(1) & V^{(3)}(1) & V^{(3)}(1) \\
0.1 & 1.010 & 15.829 & 15.363 & 15.946 & 15.453 & 15.829 & 15.363 \\
0.5 & 1.250 & 27.595 & 25.688 & 30.100 & 27.938 & 27.722 & 25.781 \\
0.9 & 1.810 & 67.526 & 61.875 & 75.532 & 69.165 & 71.205 & 65.063 \\
\end{array}$$

Table 5.3: MA(1) process, $Y_t = X_t^3$
<table>
<thead>
<tr>
<th>$\varphi$</th>
<th>$\sigma^2$</th>
<th>$\nabla^{(1)}(1)$</th>
<th>$V^{(1)}(1)$</th>
<th>$\nabla^{(2)}(1)$</th>
<th>$V^{(2)}(1)$</th>
<th>$\nabla^{(3)}(1)$</th>
<th>$V^{(3)}(1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>1.010</td>
<td>4.518</td>
<td>4.766</td>
<td>4.935</td>
<td>5.196</td>
<td>4.518</td>
<td>4.766</td>
</tr>
<tr>
<td>0.9</td>
<td>5.263</td>
<td>8.203</td>
<td>23.567</td>
<td>8.811</td>
<td>25.691</td>
<td>13.629</td>
<td>36.267</td>
</tr>
</tbody>
</table>

Table 5.4: AR(1) process, $Y_t = \exp(X_t)$

<table>
<thead>
<tr>
<th>$\varphi$</th>
<th>$\sigma^2$</th>
<th>$\nabla^{(1)}(1)$</th>
<th>$V^{(1)}(1)$</th>
<th>$\nabla^{(2)}(1)$</th>
<th>$V^{(2)}(1)$</th>
<th>$\nabla^{(3)}(1)$</th>
<th>$V^{(3)}(1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>1.010</td>
<td>2.019</td>
<td>2.040</td>
<td>3.011</td>
<td>3.040</td>
<td>2.020</td>
<td>2.041</td>
</tr>
<tr>
<td>0.5</td>
<td>1.333</td>
<td>3.389</td>
<td>3.333</td>
<td>4.408</td>
<td>4.333</td>
<td>3.611</td>
<td>3.55600</td>
</tr>
<tr>
<td>0.9</td>
<td>5.263</td>
<td>19.634</td>
<td>19.053</td>
<td>20.669</td>
<td>20.053</td>
<td>57.215</td>
<td>55.402</td>
</tr>
</tbody>
</table>

Table 5.5: AR(1) process, $Y_t = X_t^2$

<table>
<thead>
<tr>
<th>$\varphi$</th>
<th>$\sigma^2$</th>
<th>$\nabla^{(1)}(1)$</th>
<th>$V^{(1)}(1)$</th>
<th>$\nabla^{(2)}(1)$</th>
<th>$V^{(2)}(1)$</th>
<th>$\nabla^{(3)}(1)$</th>
<th>$V^{(3)}(1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>1.010</td>
<td>15.875</td>
<td>15.366</td>
<td>15.997</td>
<td>15.457</td>
<td>15.875</td>
<td>15.366</td>
</tr>
<tr>
<td>0.5</td>
<td>1.333</td>
<td>34.177</td>
<td>30.000</td>
<td>37.795</td>
<td>33.000</td>
<td>34.482</td>
<td>30.222</td>
</tr>
<tr>
<td>0.9</td>
<td>5.263</td>
<td>771.326</td>
<td>659.186</td>
<td>819.446</td>
<td>697.554</td>
<td>1471.947</td>
<td>1124.071</td>
</tr>
</tbody>
</table>

Table 5.6: AR(1) process, $Y_t = X_t^3$

<table>
<thead>
<tr>
<th>$(\varphi, \theta)$</th>
<th>$\sigma^2$</th>
<th>$\nabla^{(1)}(1)$</th>
<th>$V^{(1)}(1)$</th>
<th>$\nabla^{(2)}(1)$</th>
<th>$V^{(2)}(1)$</th>
<th>$\nabla^{(3)}(1)$</th>
<th>$V^{(3)}(1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(0.25,0.25)$</td>
<td>1.267</td>
<td>9.220</td>
<td>7.962</td>
<td>10.033</td>
<td>8.679</td>
<td>9.383</td>
<td>8.100</td>
</tr>
<tr>
<td>$(0.75,0.25)$</td>
<td>3.286</td>
<td>388.534</td>
<td>451.580</td>
<td>418.022</td>
<td>492.268</td>
<td>534.034</td>
<td>626.569</td>
</tr>
<tr>
<td>$(0.25,0.75)$</td>
<td>2.067</td>
<td>49.236</td>
<td>39.435</td>
<td>53.018</td>
<td>42.988</td>
<td>55.565</td>
<td>46.062</td>
</tr>
<tr>
<td>$(0.75,0.75)$</td>
<td>6.142</td>
<td>56.338</td>
<td>136.904</td>
<td>32.703</td>
<td>149.239</td>
<td>40.816</td>
<td>213.721</td>
</tr>
</tbody>
</table>

Table 5.7: ARMA(1, 1) process, $Y_t = \exp(X_t)$

<table>
<thead>
<tr>
<th>$(\varphi, \theta)$</th>
<th>$\sigma^2$</th>
<th>$\nabla^{(1)}(1)$</th>
<th>$V^{(1)}(1)$</th>
<th>$\nabla^{(2)}(1)$</th>
<th>$V^{(2)}(1)$</th>
<th>$\nabla^{(3)}(1)$</th>
<th>$V^{(3)}(1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(0.25,0.25)$</td>
<td>1.267</td>
<td>3.211</td>
<td>3.067</td>
<td>4.234</td>
<td>4.067</td>
<td>3.375</td>
<td>3.209</td>
</tr>
<tr>
<td>$(0.75,0.25)$</td>
<td>3.286</td>
<td>12.152</td>
<td>11.143</td>
<td>13.233</td>
<td>12.143</td>
<td>24.935</td>
<td>21.592</td>
</tr>
<tr>
<td>$(0.25,0.75)$</td>
<td>2.067</td>
<td>6.637</td>
<td>6.267</td>
<td>7.668</td>
<td>7.267</td>
<td>9.080</td>
<td>8.542</td>
</tr>
<tr>
<td>$(0.75,0.75)$</td>
<td>6.142</td>
<td>22.976</td>
<td>22.571</td>
<td>23.978</td>
<td>23.571</td>
<td>77.575</td>
<td>75.469</td>
</tr>
</tbody>
</table>

Table 5.8: ARMA(1, 1) process, $Y_t = X_t^2$
Table 5.9: ARMA(1, 1) process, $Y_t = X_t^3$

<table>
<thead>
<tr>
<th>$(\varphi, \theta)$</th>
<th>$\sigma^2$</th>
<th>$V^{(1)}(1)$</th>
<th>$V^{(1)}(1)$</th>
<th>$V^{(2)}(1)$</th>
<th>$V^{(2)}(1)$</th>
<th>$V^{(3)}(1)$</th>
<th>$V^{(3)}(1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$(0.25, 0.25)$</td>
<td>1.267</td>
<td>29.205</td>
<td>26.520</td>
<td>31.974</td>
<td>28.920</td>
<td>29.353</td>
<td>26.634</td>
</tr>
<tr>
<td>$(0.75, 0.25)$</td>
<td>3.286</td>
<td>290.404</td>
<td>238.347</td>
<td>317.687</td>
<td>258.918</td>
<td>405.426</td>
<td>309.997</td>
</tr>
<tr>
<td>$(0.25, 0.75)$</td>
<td>2.067</td>
<td>94.258</td>
<td>84.120</td>
<td>105.163</td>
<td>93.720</td>
<td>102.832</td>
<td>91.402</td>
</tr>
<tr>
<td>$(0.75, 0.75)$</td>
<td>6.142</td>
<td>882.115</td>
<td>914.265</td>
<td>924.160</td>
<td>960.551</td>
<td>1637.966</td>
<td>1730.405</td>
</tr>
</tbody>
</table>
Chapter 6

Conclusion

The aim of this diploma thesis was to find and evaluate predictions for transformed time series. We considered a class of transformations that can be written as linear combinations of Hermite polynomials. This enabled derivation of explicit formulas for the general class of Gaussian processes. The main focus was on situations when the original series can be described by a simple ARMA process. Explicit formulas were derived for three concrete transformations.

In the second chapter we investigated the autocovariance function and spectral density of the transformed series, specific results were derived and some further topics were outlined.

In the third chapter general theorems for three types of predictions were stated and applied to three concrete transformations. Also some results were derived using the spectral density. Theorems from the third chapter were further developed in the fourth chapter and derived were explicit results for simple ARMA models.

Finally in the fifth chapter the formulas for the simple ARMA processes were verified in a simulation.
Chapter 7

Appendix A - Properties of Hermite polynomials

Hermite polynomials $H_n(x)$ are orthogonal polynomials over the domain $(-\infty, \infty)$ with weighting function $\exp(-x^2)$.

**Definition 7.1. (Hermite polynomials)** The system of Hermite polynomials $H_n(x)$ is defined in terms of the standard normal distribution as

$$H_n(x) = (-1)^n \exp \left( \frac{x^2}{2} \right) \frac{d^n}{dx^n} \exp \left( \frac{x^2}{2} \right),$$

$$H_n(x) = (-1)^n \frac{\phi^{(n)}(x)}{\phi(x)}$$

where $\phi(x)$ is the standard normal probability density function.

Explicitly, we can write

$$H_n(x) = n! \sum_{m=0}^{\lfloor n/2 \rfloor} (-1)^m \frac{2^m m!(n-2m)!}{x^{n-2m}}$$

where $\lfloor N \rfloor$ is the integer part of $N$.

We have

$$H_0(x) = 1,$$

$$H_1(x) = x,$$

$$H_2(x) = x^2 - 1,$$

$$H_3(x) = x^3 - 3x,$$

$$H_4(x) = x^4 - 6x^2 + 3,$$

$$H_5(x) = x^5 - 10x^3 + 15x$$
and so on. Define operators $E_0$ and $E$ by

\[
E_0 \{ \psi(x) \} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \psi(x) \exp \left( \frac{-x^2}{2} \right) \, dx,
\]
\[
E \{ \psi(x) \} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} \psi(x) \exp \left( \frac{-(x-\mu)^2}{2\sigma^2} \right) \, dx.
\]

The Hermite polynomials constitute an orthogonal system with respect to the standard normal probability density function, so that

\[
E_0 \{ H_n(x) H_k(x) \} = \begin{cases} 0, & n \neq k, \\ n!, & n = k, \end{cases}
\]

and since $H_0(x) = 1$, it follows that

\[
E_0 \{ H_n(x) \} = 0 \text{ for } n > 0.
\]

If $X$ and $Y$ have bivariate normal distribution with zero means, unit variances and correlation coefficient $\rho$ then

\[
E \{ H_n(X) | Y = y \} = \rho^n H_n(y)
\]

and

\[
E \{ H_n(X) H_k(Y) \} = \begin{cases} 0, & n \neq k, \\ n! \rho^n, & n = k. \end{cases}
\]

The Hermite polynomials obey the recursion formula

\[
H_{n+1}(x) - xH_n(x) + nH_{n-1}(x) = 0
\]

and have the following generating function:

\[
\exp \left( tx - \frac{t^2}{2} \right) = \sum_{n=0}^{\infty} H_n(x) \frac{t^n}{n!}.
\]

For all $x$, $y$ and $A$, $B$ such that $A^2 + B^2 = 1$ we have

\[
H_n(Ax + By) = \sum_{k=0}^{n} \binom{n}{k} A^k B^{n-k} H_k(x) H_{n-k}(y).
\]

In general, the coefficients $\alpha_j$ for Hermite polynomial expansion of a function

\[
T(x) = \sum_{j=0}^{\infty} \alpha_j H_j(x)
\]
are given by
\[ \alpha_n = E_0 \left\{ \frac{d^n}{dx^n} T(x) \right\} n!. \] (7.6)

To find Hermite expansion polynomial functions it is possible to use the following formulas:

\[ x^{2p} = \frac{(2p)!}{2^{2p}} \sum_{j=0}^{p} \frac{H_{2j}(x)}{(2j)!(p-j)!}, \quad p = 0, 1, 2, \ldots \] (7.7)

and

\[ x^{2p+1} = \frac{(2p + 1)!}{2^{2p+1}} \sum_{j=0}^{p} \frac{H_{2j+1}(x)}{(2j + 1)!(p-j)!}, \quad p = 0, 1, 2, \ldots \] (7.8)

More examples can be found in [9], along with comprehensive account of the properties of Hermite polynomials. For more information see also [2].
Chapter 8

Appendix B - Best linear prediction using spectral density

The following Theorem provides a method for finding the optimal linear prediction when the spectral density is known. More on this topic can be found in Anděl [3] and in Grenander and Rosenblatt [7].

**Theorem 8.1. (Optimal linear predictor)** Let $X_t$ be a process with spectral density $f(\lambda)$ and absolutely continuous spectral distribution function

$$F(\lambda) = \int_{-\pi}^{\lambda} f(x) \, dx,$$

where

$$\int_{-\pi}^{\pi} \log f(\lambda) \, d\lambda > -\infty.$$

Let $z(\lambda)$ be the random measure corresponding to the process $X_t$ and

$$c(z) = \sum_{j=0}^{\infty} c_j z^j$$

is such that

$$\frac{1}{2\pi} |c(e^{-i\lambda})|^2 = f(\lambda).$$

Then the best linear $h$-step ahead prediction is

$$\hat{X}_{t+h} = \int_{-\pi}^{\pi} e^{i(t+h)\lambda} \frac{\sum_{j=h}^{\infty} c_j e^{-ij\lambda}}{c(e^{-i\lambda})} \, dz(\lambda).$$

The one-step ahead prediction is given by

$$\hat{X}_{t+1} = \int_{-\pi}^{\pi} e^{i(t+1)\lambda} \frac{c(e^{-i\lambda}) - c(0)}{c(e^{-i\lambda})} \, dz(\lambda).$$
and the mean square error of this prediction is
\[
E(X_{t+1} - \hat{X}_{t+1})^2 = 2\pi \exp \left( \frac{1}{2\pi} \int_{-\pi}^{\pi} \log f(\lambda) d\lambda \right) > 0.
\]

Proof: See Grenander and Rosenblatt [7], p. 69.
Bibliography


