Univerzita Karlova v Praze

Matematicko-fyzikální fakulta

Diplomová práce

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

Charles University in Prague
Faculty of Mathematics and Physics

Diploma thesis

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

Název práce: Predikce transformavaných 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 proporcionálního 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

Chapter 1

Introduction

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 $\widehat{X}_{t+h}$ for the original series. The obvious method of simply transforming this prediction to obtain prediction $\widehat{Y}_{t+h} = T(\widehat{X}_{t+h})$ 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 μ, variance σ^2 and autocorrelation function* $\text{corr}(X_t, X_{t+\tau}) = \rho_\tau$*. Define*

$$Z_t = \frac{X_t - \mu}{\sigma} \textit{ for all } t.$$

Then Z_t and $Z_{t+\tau}$ have bivariate normal distribution with zero means, unit variance and correlation ρ_τ. 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

$$\mathsf{E}(Y_t) = \alpha_0, \tag{2.1}$$

the covariance between the original and transformed series is

$$\mathsf{cov}(X_t, Y_{t+\tau}) = \alpha_1 \rho_\tau \sigma \tag{2.2}$$

and finally the autocovariance function of the transformed series is

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

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

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

For the formula (2.1) we write

$$\begin{aligned} \mathsf{E}(Y_t) &= \mathsf{E}[H_0(Z_t)Y_t] = \mathsf{E}\left[H_0(Z_t)\sum_{j=0}^{M}\alpha_j H_j(Z_t)\right] \\ &= \sum_{j=0}^{M}\alpha_j \mathsf{E}\left[H_0(Z_t)H_j(Z_t)\right] = \alpha_0 \mathsf{E}\left[H_0(Z_t)H_0(Z_t)\right] = \alpha_0. \end{aligned}$$

Now we have

$$\begin{aligned} \mathsf{cov}(X_t, Y_{t+\tau}) &= \mathsf{E}(X_t Y_{t+\tau}) - \mathsf{E}(X_t)\mathsf{E}(Y_{t+\tau}) \\ &= \mathsf{E}\left[(\sigma Z_t + \mu)Y_{t+\tau}\right] - \mu\alpha_0 \\ &= \sigma\mathsf{E}(Z_t Y_{t+\tau}) + \mathsf{E}(\mu Y_{t+\tau}) - \mu\alpha_0 \\ &= \sigma\mathsf{E}\left[H_1(Z_t)\sum_{j=0}^{M}\alpha_j H_j(Z_{t+\tau})\right] + \mathsf{E}\left[\mu\sum_{j=0}^{M}\alpha_j H_j(Z_{t+\tau})\right] - \mu\alpha_0 \\ &= \sigma\alpha_1\rho_\tau + \mu\alpha_0 - \mu\alpha_0 = \alpha_1\rho_\tau\sigma \end{aligned}$$

and finally

$$\begin{aligned} \mathsf{cov}(Y_t, Y_{t+\tau}) &= \mathsf{E}\left[\sum_{j=0}^{M}\alpha_j H_j(Z_t)\sum_{i=0}^{M}\alpha_i H_i(Z_{t+\tau})\right] \\ &= \sum_{j=1}^{M}\sum_{i=1}^{M}\alpha_j\alpha_i \mathsf{E}\left[H_j(Z_t)H_i(Z_{t+\tau})\right] \\ &= \sum_{j=1}^{M} j!\alpha_j^2\rho_\tau^j. \end{aligned}$$

□

Let us now apply this theorem to three transformations: exponential, quadratic and cubic. First it is necessary to find the Hermite expansion for each of these transformations.

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

$$\begin{aligned} Y_t = \exp(X_t) &= \exp\left(\sigma Z_t + \mu - \frac{1}{2}\sigma^2 + \frac{1}{2}\sigma^2\right) = \exp\left(\mu + \frac{1}{2}\sigma^2\right) \exp\left(\sigma Z_t - \frac{1}{2}\sigma^2\right) \\ &= \exp\left(\mu + \frac{1}{2}\sigma^2\right) \sum_{j=o}^{\infty} \frac{\sigma^j}{j!} H_j(Z_t). \end{aligned}$$

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

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

the covariance between the original and the transformed series is

$$\mathsf{cov}(X_t, Y_{t+\tau}) = \alpha_1 \rho_\tau \sigma = \exp\left(\mu + \frac{1}{2}\sigma^2\right) \sigma^2 \rho_\tau,$$

and finally the autocovariance function of the transformed series is

$$\mathsf{cov}(Y_t, Y_{t+\tau}) = \sum_{j=1}^{M} j! \alpha_j^2 \rho_\tau^j = \exp\left(2\mu + \sigma^2\right) \sum_{j=1}^{\infty} \frac{(\sigma^2 \rho_\tau)^j}{j!} = \exp\left(2\mu + \sigma^2\right)\left(\exp(\sigma^2 \rho_\tau) - 1\right). \tag{2.4}$$

Corollary 2.3. (Quadratic transformation) Recall that

$$\begin{aligned} H_0(x) &= 1, \\ H_1(x) &= x, \\ H_2(x) &= x^2 - 1, \\ H_3(x) &= x^3 - 3x. \end{aligned}$$

Consider the transformation

$$\begin{aligned} Y_t &= X_t^2 = (\sigma Z_t + \mu)^2 = \mu^2 + 2\mu\sigma Z_t + \sigma^2 Z_t^2 \\ &= \mu^2 + \sigma^2 + 2\sigma\mu Z_t + \sigma^2 Z_t^2 - \sigma^2. \end{aligned}$$

Denoting

$$\begin{aligned}\alpha_0 &= \mu^2+\sigma^2,\\ \alpha_1 &= 2\sigma\mu,\\ \alpha_2 &= \sigma^2\end{aligned}$$

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

$$\mathsf{E}(Y_t) = \alpha_0 = \mu^2+\sigma^2,$$

the covariance between the original and the transformed series

$$\mathsf{cov}(X_t, Y_{t+\tau}) = \alpha_1\rho_\tau\sigma = 2\sigma^2\mu\rho_\tau$$

and finally the autocovariance function of the transformed series

$$\mathsf{cov}(Y_t, Y_{t+\tau}) = \sum_{j=1}^{M} j!\alpha_j^2\rho_\tau^j = \alpha_1^2\rho_\tau + 2\alpha_2^2\rho_\tau^2 = 4\sigma^2\mu^2\rho_\tau + 2\sigma^4\rho_\tau^2. \tag{2.5}$$

Corollary 2.4. (Cubic transformation) Consider the transformation

$$\begin{aligned}Y_t &= X_t^3 = (\sigma Z_t+\mu)^3\\ &= \mu^3 + 3\sigma\mu^2 Z_t + 3\sigma^2\mu Z_t^2 + 3\sigma^3 Z_t^3\\ &= \mu^3 + 3\sigma^2\mu + \left(3\sigma\mu^2+3\sigma^3\right)Z_t + 3\sigma^2\mu Z_t^2 - 3\sigma^2\mu + 3\sigma^3 Z_t^3 - 3\sigma^3 Z_t.\end{aligned}$$

Denoting

$$\begin{aligned}\alpha_0 &= \mu^3+3\sigma^2\mu,\\ \alpha_1 &= 3\sigma\mu^2+3\sigma^3,\\ \alpha_2 &= 3\sigma^2\mu,\\ \alpha_3 &= \sigma^3\end{aligned}$$

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

$$\mathsf{E}(Y_t) = \alpha_0 = \mu^3 + 3\sigma^2\mu,$$

the covariance between the original and the transformed series is

$$\mathsf{cov}(X_t, Y_{t+\tau}) = \alpha_1\rho_\tau\sigma = 3\sigma^2\mu^2\rho_\tau + 3\sigma^4\rho_\tau$$

and finally the autocovariance function of the transformed series is

$$\begin{aligned}\mathsf{cov}(Y_t, Y_{t+\tau}) &= \sum_{j=1}^{M} j!\alpha_j^2\rho_\tau^j = \alpha_1^2\rho_\tau + 2\alpha_2^2\rho_\tau^2 + 6\alpha_3^2\rho\tau^3 &(2.6)\\ &= \left(3\sigma\mu^2+3\sigma^3\right)^2\rho_\tau + 2\left(3\sigma^2\mu\right)^2\rho_\tau^2 + 6(\sigma^3)^2\rho_\tau^3. &(2.7)\end{aligned}$$

2.1.2 Application to ARMA processes

In order to apply general formulas to a given process we need to know its mean μ and variance σ^2, as well as its autocorrelation function ρ_τ. This is very easy to calculate for MA(∞) process.

Example 2.5. Consider a MA(∞) 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 $\epsilon_t \sim \mathsf{N}(0, \sigma_\epsilon^2)$. The mean of this process is obviously zero, its variance is

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

the autocovariance function is

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

and the autocorrelation function is

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

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

$$\begin{aligned}
\mathsf{E}(Y_t) &= \exp\left(\frac{1}{2}\sigma_\epsilon^2 \sum_{i=0}^{\infty} \theta_i^2\right), \\
\mathsf{cov}(X_t, Y_{t+\tau}) &= \exp\left(\frac{1}{2}\sigma_\epsilon^2 \sum_{i=0}^{\infty} \theta_i^2\right)\left(\sigma_\epsilon^2 \sum_{t=0}^{\infty} \theta_t \theta_{t+\tau}\right), \\
\mathsf{cov}(Y_t, Y_{t+\tau}) &= \exp\left(\sigma_\epsilon^2 \sum_{i=0}^{\infty} \theta_i^2\right)\left[\exp\left(\sigma_\epsilon^2 \sum_{t=0}^{\infty} \theta_t \theta_{t+\tau}\right) - 1\right],
\end{aligned}$$

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

$$\begin{aligned}
\mathsf{E}(Y_t) &= \sigma_\epsilon^2 \sum_{i=0}^{\infty} \theta_i^2, \\
\mathsf{cov}(X_t, Y_{t+\tau}) &= 0, \\
\mathsf{cov}(Y_t, Y_{t+\tau}) &= 2\sigma_\epsilon^4 \left(\sum_{t=0}^{\infty} \theta_t \theta_{t+\tau}\right)^2
\end{aligned}$$

and finally for $Y_t = X_t^3$:

$$\begin{aligned}
\mathsf{E}(Y_t) &= 0, \\
\mathsf{cov}(X_t, Y_{t+\tau}) &= 3\sigma_\epsilon^4 \left(\sum_{i=0}^{\infty} \theta_i^2\right)\left(\sum_{t=0}^{\infty} \theta_t\theta_{t+\tau}\right), \\
\mathsf{cov}(Y_t, Y_{t+\tau}) &= 9\sigma_\epsilon^6 \left(\sum_{i=0}^{\infty} \theta_i^2\right)^2 \left(\sum_{t=0}^{\infty} \theta_t\theta_{t+\tau}\right) + 6\sigma_\epsilon^6 \left(\sum_{t=0}^{\infty} \theta_t\theta_{t+\tau}\right)^3.
\end{aligned}$$

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 $\epsilon_t \sim \mathsf{N}(0,1)$ 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

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

taking into consideration that ϵ_t are i.i.d. The autocovariance function of this process is

$$\begin{aligned}
R_X(\tau) &= \mathsf{cov}(X_t, X_{t+\tau}) = \mathsf{cov}\left(\sum_{j=0}^{\infty} \varphi^j \epsilon_{t-j}, \sum_{j=0}^{\infty} \varphi^j \epsilon_{t+\tau-j}\right) \\
&= \sum_{j=0}^{\infty} \varphi^{j+\tau}\varphi^j \mathsf{var}(\epsilon_{t-j+\tau}\epsilon_{t+\tau-j}) = \sum_{j=0}^{\infty} \varphi^{2j+\tau} = \frac{\varphi^\tau}{1-\varphi^2}.
\end{aligned}$$

The autocorrelation function is then

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

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

$$\begin{aligned}
R_Y^{(1)}(\tau) &= \mathsf{cov}(Y_t, Y_{t+\tau}) = \exp\left(2\mu + \sigma^2\right)\left(\exp(\sigma^2\rho_\tau) - 1\right) \\
&= \exp\left(\frac{1}{1-\varphi^2}\right)\left[\exp\left(\frac{\varphi^\tau}{1-\varphi^2}\right) - 1\right].
\end{aligned}$$

Using formula (2.5) we can find the autocovariance function of the transformed series $Y_t = X_t^2$

$$
\begin{aligned}
R_Y^{(2)}(\tau) &= \mathsf{cov}(Y_t, Y_{t+\tau}) = 4\sigma^2\mu^2\rho_\tau + 2\sigma^4\rho_\tau^2 \\
&= 2\varphi^\tau \left(\frac{1}{1-\varphi^2}\right)^2.
\end{aligned}
$$

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

$$
\begin{aligned}
R_Y^{(3)}(\tau) &= \mathsf{cov}(Y_t, Y_{t+\tau}) = \left(3\sigma\mu^3 + 3\sigma^3\right)^2 \rho_\tau + 2\left(3\sigma^2\mu\right)^2 \rho_\tau^2 + 6\sigma^6\rho_\tau^3 \\
&= 9\varphi^\tau \left(\frac{1}{1-\varphi^2}\right)^3 + 6\left(\frac{\varphi^\tau}{1-\varphi^2}\right)^3 = (9\varphi^\tau + 6\varphi^{3\tau})\left(\frac{1}{1-\varphi^2}\right)^3.
\end{aligned}
$$

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

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

The autocovariance function of this process is

$$
\begin{aligned}
R_X(\tau) &= \mathsf{cov}(X_t, X_{t+\tau}) = \mathsf{cov}(\epsilon_t + \theta\epsilon_{t-1}, \epsilon_{t+\tau} + \theta\epsilon_{t+\tau-1}) \\
&= \begin{cases} 1+\theta^2, & \text{if } \tau = 0, \\ \theta, & \text{if } \tau = 1 \text{ or } \tau = -1, \\ 0, & \text{otherwise} \end{cases}
\end{aligned}
$$

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

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

and so

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

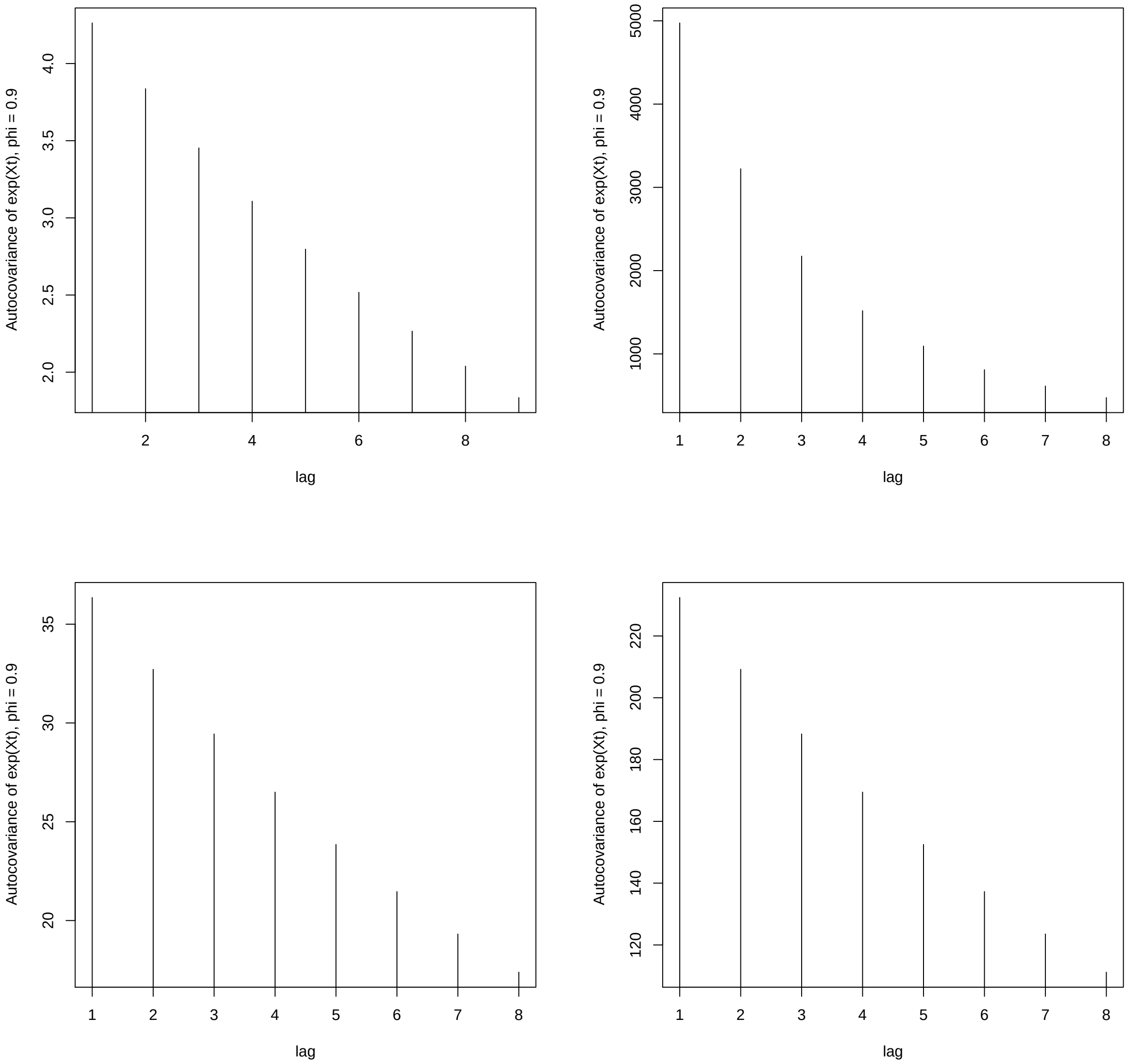

Figure 2.1: Autocovariance functions of the original process AR(1) and its transforms.

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.

$\theta = 0.25$	$\tau = 0$	$\tau = 1$	$\theta = 0.75$	$\tau = 0$	$\tau = 1$
X_t	1.063	0.250	X_t	1.562	0.750
$\exp(X_t)$	5.479	0.723	$\exp(X_t)$	17.989	3.578
X_t^2	2.258	0.265	X_t^2	4.882	1.172
X_t^3	10.795	2.540	X_t^3	34.332	16.479

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

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

ϵ_t has finite kurtosis η and B is the lag operator. In the following we will assume that ϵ_t are uncorrelated, but not necessarily independent.

Theorem 2.8. *Let us assume that ϵ_t are i.i.d. random variables with zero mean, finite kurtosis, $\eta = \mathsf{E}(\epsilon_t^4)/\left[\mathsf{E}(\epsilon_t^2)\right]^2$. Let us assume that*

$$\mathsf{E}(\epsilon_u\epsilon_v) = \begin{cases} \sigma_\epsilon^2, & \text{if } u = v, \\ 0, & \text{otherwise,} \end{cases}$$

autocorrelation function function of ϵ_t^2 is $\rho_{\epsilon^2}(\tau)$ and

$$\mathsf{E}(\epsilon_s\epsilon_t\epsilon_u\epsilon_v) = \begin{cases} \left[1 + (\eta-1)\rho_{\epsilon^2}(s-v)\right]\sigma_\epsilon^4, & \text{if } s = t,\ u = v \text{ or } s = u,\ t = v, \\ \left[1 + (\eta-1)\rho_{\epsilon^2}(s-t)\right]\sigma_\epsilon^4, & \text{if } s = v,\ t = u, \\ 0, & \text{otherwise.} \end{cases}$$

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

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

Let us define

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

$$\Delta(\tau) = \frac{\sum_{s=0}^{\infty}\sum_{t=0}^{\infty} \theta_t\theta_s\theta_{t+\tau}\theta_{s+\tau}\rho_{\epsilon^2}(t-s)}{\left(\sum_{i=0}^{\infty} \theta_i^2\right)^2},$$

$$\phi(\tau) = \frac{\sum_{s=0}^{\infty}\sum_{t=0}^{\infty} \theta_s^2\theta_t^2\rho_{\epsilon^2}(\tau+t-s)}{\left(\sum_{i=0}^{\infty} \theta_i^2\right)^2},$$

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

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

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

Proof. See Palma and Zevallos [10]. □

Corollary 2.9. ***(Linear process)***
Let ϵ_t be i.i.d. random variables with zero mean and finite kurtosis η, i.e. ϵ_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 κ is the kurtosis of Y_t given by

$$\kappa = (\eta-3)\frac{\sum_{i=0}^{\infty} \theta_i^4}{\left(\sum_{i=0}^{\infty} \theta_i^2\right)^2} + 3$$

Proof. In this case we can write

$$\phi(\tau) = \Delta(\tau) = \Delta(0)\alpha(\tau) = \frac{\sum_{t=0}^{\infty} \theta_t^2 \theta_{t+\tau}^2}{\left(\sum_{i=0}^{\infty} \theta_i^2\right)^2}$$

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

$$\Delta(0) = \frac{\sum_{i=0}^{\infty} \theta_i^4}{\left(\sum_{i=0}^{\infty} \theta_i^2\right)^2}.$$

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(∞) 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 $\epsilon_t \sim \mathsf{N}(0, \sigma_\epsilon^2)$. Autocorrelation sequence of ϵ_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 ϵ_t is $\eta = 3$. Hence

$$\rho_Y(\tau) = \rho_X^2(\tau),$$

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

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

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

$$\rho_Y(\tau) = \frac{\mathsf{cov}(Y_t, Y_{t+\tau})}{\mathsf{var}(Y_t)} = \frac{2\sigma_\epsilon^4 \left(\sum_{t=0}^{\infty} \theta_t \theta_{t+\tau}\right)^2}{2\sigma_\epsilon^4 \left(\sum_{t=0}^{\infty} \theta^2\right)^2} = \rho_X^2(\tau).$$

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

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

and

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

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

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

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

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

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

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 $\epsilon_t \sim \mathsf{N}(0,1)$ 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

$$f(\lambda) = \frac{1}{2\pi}|1 + \theta\mathrm{e}^{-\mathrm{i}\lambda}|^2 \text{, where } \lambda \in [-\pi, \pi].$$

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

$$f_Y(\lambda) = \alpha_0^2 f^{*0} + 2!\alpha_2^2 f^{*2} = \delta(\lambda) + 2f^{*2}$$

and

$$\begin{aligned}
f^{*2} &= \int_{-\pi}^{\pi} f(\lambda - a) f(a)\,\mathrm{d}a = \int_{-\pi}^{\pi} \frac{1}{2\pi}|1 + \theta \mathrm{e}^{-\mathrm{i}(\lambda - a)}|^2 \frac{1}{2\pi}|1 + \theta \mathrm{e}^{-\mathrm{i}a}|^2\,\mathrm{d}a \\
&= \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \left(1 + \theta \mathrm{e}^{-\mathrm{i}(\lambda - a)}\right)\left(1 + \theta \mathrm{e}^{\mathrm{i}(\lambda - a)}\right)\left(1 + \theta \mathrm{e}^{-\mathrm{i}a}\right)\left(1 + \theta \mathrm{e}^{\mathrm{i}a}\right)\,\mathrm{d}a \\
&= \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \left(1 + \theta \mathrm{e}^{-\mathrm{i}(\lambda - a)} + \theta \mathrm{e}^{\mathrm{i}(\lambda - a)} + \theta^2\right)\left(1 + \theta \mathrm{e}^{-\mathrm{i}a} + \theta \mathrm{e}^{\mathrm{i}a} + \theta^2\right)\,\mathrm{d}a \\
&= \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \left(1 + \theta^2 + 2\theta\cos(\lambda - a)\right)\left(1 + \theta^2 + 2\theta\cos(a)\right)\,\mathrm{d}a \\
&= \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \left[(1 + \theta^2)^2 + 4\theta^2 \cos(a)\cos(\lambda - a)\right]\,\mathrm{d}a \\
&= \frac{1}{4\pi^2}\left[2\pi(1 + \theta^2)^2 + 4\pi\theta^2\cos(\lambda)\right] \\
&= \frac{1}{2\pi}(1 + \theta^2)^2 + \frac{1}{\pi}\theta^2\cos(\lambda).
\end{aligned}$$

Hence the spectral density of the transformed process is

$$f_Y(\lambda) = \delta(\lambda) + \frac{1}{\pi}\left[(1 + \theta^2)^2 + 2\theta^2\cos(\lambda)\right].$$

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).

$$R(t) = \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}t\lambda} f(\lambda)\,\mathrm{d}\lambda.$$

In our case

$$R_Y^{(2)}(\tau) = \frac{1}{\pi}\int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}t\lambda}\left[\pi\delta(\lambda) + (1 + \theta^2)^2 + 2\theta^2\cos(\lambda)\right]\,\mathrm{d}\lambda.$$

Calculating this integral for concrete values of τ we have

$$\begin{aligned}
R_Y^{(2)}(0) &= 2(1 + \theta^2)^2, \\
R_Y^{(2)}(1) &= R_Y^{(2)}(-1) = 2\theta^2, \\
R_Y^{(2)}(\tau) &= 0, \text{ for } \tau \geq 2,
\end{aligned}$$

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 μ and variance σ^2, and $\widehat{X}_{t+h}$ 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. $\widehat{X}_{t+h} = \mathsf{E}(X_{t+h}|I_t)$. Assume that the error of prediction $e_{t+h} = X_{t+h} - \widehat{X}_{t+h}$ given I_t has conditionally normal distribution with zero mean and variance $S^2(h)$. Define $Z_t = \frac{X_t - \mu}{\sigma} \sim \mathsf{N}(0,1)$ 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

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

Then constants γ_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

$$\widehat{Y}^{(1)}_{t+h} = \gamma_0 \tag{3.1}$$

and the unconditional mean square error of this prediction is

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

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 $\widehat{X}_{t+h}$ 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 $C = \frac{\widehat{X}_{t+h} - \mu}{\sqrt{\sigma^2 - S^2(h)}}$. We have

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

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

$$\begin{aligned} H_j(Z_{t+h}) &= H_j(BW_{t+h} + AC) \\ &= \sum_{k=0}^{j} \binom{j}{k} B^k A^{j-k} H_{j-k}(C) H_k(W_{t+h}). \end{aligned}$$

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

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

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

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

To calculate the unconditional mean square error we first write

$$V^{(1)}(h) = \mathsf{E}\left[Y_{t+h} - \widehat{Y}_{t+h}^{(1)}\right]^2 = \mathsf{E}\left[Y_{t+h}\right]^2 - 2\mathsf{E}\left[Y_{t+h}\widehat{Y}_{t+h}^{(1)}\right] + E\left[\widehat{Y}_{t+h}^{(1)}\right]^2. \quad (3.2)$$

Recall from the properties of Hermite polynomials, formula (7.3), that

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

Therefore

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

Similarly,

$$\begin{aligned} \mathsf{E}\left[\widehat{Y}_{t+h}^{(1)}\right]^2 &= \mathsf{E}\left[\sum_{j=0}^{M} \alpha_j A^j H_j(C)\right]^2 \\ &= \sum_{i=0}^{M}\sum_{j=0}^{M} \alpha_i\alpha_j A^{i+j}\mathsf{E}\left[H_i(C)H_j(C)\right] = \sum_{i=0}^{M} j!\alpha_j^2 A^{2j}, \end{aligned}$$

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

$$\begin{aligned} \mathsf{cov}(X_{n+h}, \widehat{X}_{t+h}) &= \mathsf{cov}(X_{t+h}, X_{t+h} - e_{t+h}) = \mathsf{cov}(X_{t+h}, X_{t+h}) - \mathsf{cov}(X_{t+h}, e_{t+h}) \\ &= \sigma^2 - \mathsf{cov}(\widehat{X}_{t+h} + e_{t+h}, e_{t+h}) = \sigma^2 - \mathsf{cov}(\widehat{X}_{t+h}, e_{t+h}) - \mathsf{cov}(e_{t+h}, e_{t+h}) \\ &= \sigma^2 - S^2(h) \end{aligned}$$

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

$$\begin{aligned} \rho &= \mathsf{corr}(Z_{t+h}, C) = \frac{\mathsf{cov}(Z_{t+h}, C)}{\sqrt{\mathsf{var}(Z_{t+h})\mathsf{var}(C)}} = \mathsf{cov}(Z_{t+h}, C) \\ &= \mathsf{cov}\left[\frac{X_{t+h} - \mu}{\sigma}, \frac{\widehat{X}_{t+h} - \mu}{\sqrt{\sigma^2 - S^2(h)}}\right] \\ &= \frac{1}{\sigma\sqrt{\sigma^2 - S^2(h)}}\mathsf{cov}(X_{n+h}, \widehat{X}_{t+h}) \\ &= \frac{1}{\sigma\sqrt{\sigma^2 - S^2(h)}}(\sigma^2 - S^2(h)) = \sqrt{\frac{\sigma^2 - S^2(h)}{\sigma^2}} = A. \end{aligned}$$

Now we can write

$$\begin{aligned} 2\mathsf{E}\left[\widehat{Y}_{t+h}^{(1)} Y_{t+h}\right] &= 2\mathsf{E}\left[\sum_{i=0}^{M} \alpha_i H_i(Z_{t+h})\right]\left[\sum_{j=0}^{M} \alpha_j A^j H_j(P)\right] \\ &= 2\mathsf{E}\left[\sum_{i=0}^{M}\sum_{j=0}^{M} \alpha_i\alpha_j A^j H_i(Z_{t+h})H_j(C)\right] \\ &= 2\sum_{j=0}^{M} j!\alpha_j^2 A^j \rho^j = 2\sum_{j=0}^{M} j!\alpha_j^2 A^{2j}. \end{aligned}$$

Returning back to formula (3.2) we derived the expected expression for the forecast error

$$\begin{aligned} V^{(1)}(h) &= \sum_{j=0}^{M} \alpha_j^2 j! - 2\sum_{j=0}^{M} \alpha_j^2 A^{2j} j! + \sum_{j=0}^{M} \alpha_j^2 A^{2j} j! \\ &= \sum_{j=0}^{M} \alpha_j^2 j! \left(1 - A^{2j}\right). \end{aligned}$$

□

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}

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

Then the mean square error of this prediction is

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

Proof. We have

$$\begin{aligned} V^{(2)}(h) &= \mathsf{E}\left[Y_{t+h} - \widehat{Y}_{t+h}^{(2)}\right]^2 = \mathsf{E}\left[\left(Y_{t+h} - \widehat{Y}^{(1)}(h)\right) - \left(Y_{t+h}^{(2)} - Y_{t+h}^{(1)}\right)\right]^2 \\ &= \mathsf{E}\left(Y_{t+h} - \widehat{Y}^{(1)}(h)\right)^2 - 2\mathsf{E}\left(Y_{t+h} - \widehat{Y}^{(1)}(h)\right)\left(Y_{t+h}^{(2)} - Y_{t+h}^{(1)}\right) + \mathsf{E}\left(Y_{t+h}^{(2)} - Y_{t+h}^{(1)}\right)^2. \end{aligned}$$

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:*

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

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

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

The mean square prediction error is

$$V^{(3)}(h) = \mathsf{E}\left(Y_{t+h} - \widehat{Y}_{t+h}^{(3)}\right)^2 = \mathsf{E}\left(Y_{t+h}\right)^2 - 2\mathsf{E}\left(Y_{t+h}\widehat{Y}_{t+h}^{(3)}\right) + \mathsf{E}\left(\widehat{Y}_{t+h}^{(3)}\right)^2$$

where

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

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

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

see Theorem 3.1. Hence we have

$$\mathsf{E}\left(Y_{t+h}\widehat{Y}_{t+h}^{(3)}\right)^2 = \mathsf{E}\left(\widehat{Y}_{t+h}^{(3)}\right)^2 = \sum_{j=0}^{1} j!\alpha_j^2 A^{2j}.$$

And so the mean square prediction error is

$$\begin{aligned} V^{(3)}(h) &= \sum_{j=0}^{M} j!\alpha_j^2 - \sum_{j=0}^{1} j!\alpha_j^2 A^{2j} = \sum_{j=0}^{M} j!\alpha_j^2 - \alpha_0^2 - \alpha_1^2 A^2 = \sum_{j=1}^{M} j!\alpha_j^2 - \alpha_1^2 A^2 = \\ &= \sum_{j=2}^{M} j!\alpha_j^2 + \alpha_1^2\left(1 - A^2\right) = \sum_{j=2}^{M} j!\alpha_j^2 + \alpha_1^2 B^2 = \sum_{j=2}^{M} j!\alpha_j^2 + \frac{\alpha_1^2 S^2(h)}{\sigma^2}. \end{aligned}$$

□

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

$$G^{(2)}(h) = \frac{V^{(2)}(h) - V^{(1)}(h)}{V^{(1)}(h)}, \tag{3.4}$$

$$G^{(3)}(h) = \frac{V^{(3)}(h) - V^{(1)}(h)}{V^{(1)}(h)}. \tag{3.5}$$

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

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

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

Then the predictions for the transformed series are

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

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

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

their mean square prediction errors

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

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

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

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

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

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

Using Theorem 3.1 we have

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

and

$$\begin{aligned} V^{(1)}(h) &= \sum_{j=1}^{\infty} j!\alpha_j^2\left(1 - A^{2j}\right) \\ &= \exp\left(2\mu + \sigma^2\right)\sum_{j=1}^{\infty} j!\frac{\sigma^{2j}}{(j!)^2}\left(1 - A^{2j}\right) \\ &= \exp\left(2\mu + \sigma^2\right)\left[\sum_{j=1}^{\infty}\frac{\sigma^{2j}}{j!} - \sum_{j=1}^{\infty}\frac{\sigma^{2j}}{j!}A^{2j}\right] \\ &= \exp\left(2\mu + \sigma^2\right)\left[\exp(\sigma^2) - 1 - \exp(\sigma^2A^2) + 1\right] \\ &= \exp\left[2(\mu + \sigma^2)\right]\left[1 - \exp(\sigma^2A^2 - \sigma^2)\right] \\ &= \exp\left[2(\mu + \sigma^2)\right]\left[1 - \exp(-S^2(h))\right], \end{aligned}$$

Using Theorem 3.2 we have

$$\widehat{Y}^{(2)}_{t+h} = \exp(X_{t+h})$$

and to evaluate $V^{(1)}(h)$ we need first to calculate the second term in (3.3)

$$\begin{aligned}
&\lim_{M\to\infty}\sum_{j=0}^{M}A^{2j}(j!)^{-1}\left[\sum_{k=1}^{\lfloor\frac{1}{2}(M-j)\rfloor}\alpha_{j+2k}\frac{(j+2k)!}{k!}\left(-\frac{1}{2}B^2\right)^k\right]^2\\
= &\lim_{M\to\infty}\exp(2\mu+\sigma^2)\sum_{j=0}^{M}\frac{(A^2\sigma^2)^j}{j!}\left[\sum_{k=1}^{\lfloor\frac{1}{2}(M-j)\rfloor}\frac{1}{k!}\left(-\frac{1}{2}B^2\sigma^2\right)^k\right]^2\\
= &\exp(2\mu+\sigma^2)\exp(\sigma^2A^2)\left[\exp\left(-\frac{1}{2}B^2\sigma^2\right)-1\right]^2\\
= &\exp(2\mu+\sigma^2)\exp(\sigma^2-S^2(h))\left[1-\exp\left(-\frac{1}{2}S^2(h)\right)\right]^2\\
= &\exp\left[2(\mu+\sigma^2)\right]\left[\exp\left(-\frac{1}{2}S^2(h)\right)-\exp(-S^2(h))\right]^2.
\end{aligned}$$

Now we can write

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

Using Theorem 3.3 we have

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

and

$$\begin{aligned}
V^{(3)}(h) &= \lim_{M\to\infty}\sum_{j=2}^{M}j!\alpha_j^2+\frac{\alpha_1^2S^2(h)}{\sigma^2}\\
&= \exp\left(\mu+\frac{1}{2}\sigma^2\right)^2\left[\sum_{j=0}^{\infty}\frac{\sigma^{2j}}{(j!)^2}j!-1-\sigma^1+S^2(h)\right]\\
&= \exp(2\mu+\sigma^2)\left[\exp(\sigma^2)-1-\sigma^2+S^2(h)\right].
\end{aligned}$$

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 $V^{(3)}(h)$. □

3.2.2 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 (α_0, α_1 and α_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

$$\begin{aligned} \alpha_0 &= \mu^2 + \sigma^2, \\ \alpha_1 &= 2\sigma\mu, \\ \alpha_2 &= \sigma^2, \\ \gamma_0 &= \widehat{X}_{t+h}^2 + S^2(h), \\ \gamma_1 &= 2\widehat{X}_{t+h} S(h), \\ \gamma_2 &= S^2(h), \end{aligned}$$

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

$$\begin{aligned} \widehat{Y}_{t+h}^{(1)} &= \widehat{X}_{t+h}^2 + S^2(h), \\ \widehat{Y}_{t+h}^{(2)} &= \widehat{X}_{t+h}^2, \\ \widehat{Y}_{t+h}^{(3)} &= \sigma^2 - \mu^2 + 2\mu\widehat{X}_{t+h}, \end{aligned}$$

their mean square expected errors

$$\begin{aligned} V^{(1)}(h) &= 4(\mu^2 + \sigma^2)S^2(h) - 2S^4(h), \\ V^{(2)}(h) &= 4(\mu^2 + \sigma^2)S^2(h) - S^4(h), \\ V^{(3)}(h) &= 2\sigma^4 + 4\mu^2 S^2(h), \end{aligned}$$

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

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

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

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

and

$$\begin{aligned}V^{(1)}(h) &= \alpha_1^2(1-A^2)+\alpha_2^2(1-A^4)\\ &= 4\sigma^2\mu^2\left(\frac{S^2(h)}{\sigma^2}\right)+2\sigma^4\left[1-\left(1-\frac{S^2(h)}{\sigma^2}\right)^2\right]\\ &= 4\mu^2S^2(h)+2\left[\sigma^4-(\sigma^2-S^2(h))^2\right]\\ &= 4(\mu^2+\sigma^2)S^2(h)-2S^4(h).\end{aligned}$$

Using Theorem 3.2 we write

$$\widehat{Y}^{(2)}_{t+h}=\widehat{X}^2_{t+h}$$

and the second term in (3.3) (only the term with $k=1$ is greater than zero)

$$\sum_{j=1}^{M}\alpha_j^2 j!(1-A^{2j})+\sum_{j=0}^{M}A^{2j}(j!)^{-1}\left[\sum_{k=1}^{\lfloor\frac{1}{2}(M-j)\rfloor}\alpha_{j+2k}\frac{(j+2k)!}{k!}\left(-\frac{1}{2}B^2\right)^k\right]^2$$
$$=\left[2\alpha_2\left(-\frac{1}{2}B^2\right)\right]^2=\left[\sigma^2\frac{S^2(h)}{\sigma^2}\right]^2=S^4(h),$$

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

$$\begin{aligned}\widehat{Y}^{(3)}_{t+h} &= \alpha_0+\alpha_1\frac{\widehat{X}_{t+h}-\mu}{\sigma}\\ &= \mu^2+\sigma^2+2\sigma\mu\frac{\widehat{X}_{t+h}-\mu}{\sigma}=\sigma^2-\mu^2+2\mu X_{t+h}\end{aligned}$$

and

$$V^{(3)}(h)=2\alpha_2^2+\frac{\alpha_1^2S^2(h)}{\sigma^2}=2\sigma^2+4\mu^2S^2(h).$$

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 the cubic transformation. From Corollary 2.4 we know that only the first four coefficients (α_0, α_1, α_2 and α_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*

$$Y_t = X_t^3 = \sum_{i=0}^{3} \alpha_i H_i(Z_t) = \sum_{i=0}^{3} \gamma_i H_i(W_t),$$

where

$$\begin{aligned}
\alpha_0 &= \mu^3 + 3\sigma^2\mu, \\
\alpha_1 &= 3\sigma\mu^2 + 3\sigma^3, \\
\alpha_2 &= 3\sigma^2\mu, \\
\alpha_3 &= \sigma^3, \\
\gamma_0 &= \widehat{X}_{t+h}^3 + 3S^2(h)\widehat{X}_{t+h}, \\
\gamma_1 &= 3\widehat{X}_{t+h}^2 S(h) + 3S^3(h), \\
\gamma_2 &= 3S^2(h)\widehat{X}_{t+h},\ \gamma_3 = S^3(h),
\end{aligned}$$

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

$$\begin{aligned}
\widehat{Y}_{t+h}^{(1)} &= \widehat{X}_{t+h}^3 + 3S^2(h)\widehat{X}_{t+h}, \\
\widehat{Y}_{t+h}^{(2)} &= \widehat{X}_{t+h}^3, \\
\widehat{Y}_{t+h}^{(3)} &= 3(\mu^3 + \sigma^2)\widehat{X}_{t+h} - 2\mu^3
\end{aligned}$$

their mean square errors

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

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

$$\begin{aligned}
G^{(2)}(h) &= \frac{3S^2(h)(\mu^2+\sigma^2-S^2(h))}{3(\mu^2+\sigma^2)^2 + 2S^4(h) - 6\mu^2 S^2(h) + 6\sigma^2(2\mu^2+\sigma^2-S^2(h))}, \\
G^{(3)}(h) &= \frac{6\sigma^4(3\mu^2+\sigma^2) - S^2(h)(6S^4(h) - 18\mu^2 S^2(h) + 18\sigma^2(2\mu^2+\sigma^2-S^2(h)))}{S^2(h)\left[9(\mu^2+\sigma^2)^2 + 6S^4(h) - 18\mu^2 S^2(h) + 18\sigma^2(2\mu^2+\sigma^2-S^2(h))\right]}.
\end{aligned}$$

Proof. To see that the coefficients α_j are calculated correctly we write:

$$\begin{aligned}
Y_t &= \sum_{j=0}^{4} \alpha_j H_j(Z_t) \\
&= \mu^3 + 3\sigma^2\mu + (3\sigma\mu^2 + 3\sigma^3)\frac{X_t-\mu}{\sigma} + \\
&\quad +3\sigma^2\mu\left[\left(\frac{X_t-\mu}{\sigma}\right)^2 - 1\right] + \sigma^3\left[\left(\frac{X_t-\mu}{\sigma}\right)^3 - 3\frac{X_t-\mu}{\sigma}\right] \\
&= X_t^3.
\end{aligned}$$

The coefficients γ_j are obtained in similar way as in the case previous transformations; with $\widehat{X}_{t+h}$ instead of μ and $S^2(h)$ instead of σ^2. Using Theorem 3.1 we write

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

and

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

Using Theorem 3.2 we write

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

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

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

Hence

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

Finally, using Theorem 3.3 we have

$$\begin{aligned}
\widehat{Y}^{(3)}_{t+h} &= \alpha_0 + \alpha_1\left(\frac{\widehat{X}_{t+h} - \mu}{\sigma}\right) = \mu^3 + 3\sigma^2\mu + (3\sigma\mu^2 + 3\sigma^3)\left(\frac{\widehat{X}_{t+h} - \mu}{\sigma}\right) \\
&= 3(\mu^3 + \sigma^2)\widehat{X}_{t+h} - 2\mu^3
\end{aligned}$$

and

$$\begin{aligned} V^{(3)}(h) &= \sum_{j=2}^{3} j!\alpha_j^2 + \alpha_1^2 \frac{S^2(h)}{\sigma^2} \\ &= 18\sigma^4\mu^2 + 6\sigma^6 + (3\sigma\mu^2 + 3\sigma^3)^2 \frac{S^2(h)}{\sigma^2} \\ &= 6\sigma^4(3\mu^2 + \sigma^2) + 9S^2(h)(\mu^2 + \sigma^2)^2, \end{aligned}$$

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 $V^{(3)}(h)$. □

3.3 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 σ^2 and spectral density*

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

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

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

where $\widehat{X}_{t+1}$ 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

$$X_t = \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}t\lambda}\, \mathrm{d}z(\lambda).$$

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

$$\widehat{X}_{t+1} = \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}(t+1)\lambda} \frac{c(\mathrm{e}^{-\mathrm{i}\lambda}) - c(0)}{c(\mathrm{e}^{-\mathrm{i}\lambda})}\, \mathrm{d}z(\lambda).$$

Define $e_{t+1} = X_{t+1} - \widehat{X}_{t+1}$. Then $\mathsf{E}(e_{t+1}) = 0$ and

$$\begin{aligned} \epsilon_{t+1} &= \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}(t+1)\lambda}\,\mathrm{d}z(\lambda) - \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}(t+1)\lambda} \frac{c(\mathrm{e}^{-\mathrm{i}\lambda}) - c(0)}{c(\mathrm{e}^{-\mathrm{i}\lambda})}\,\mathrm{d}z(\lambda) = \\ &= c(0) \int_{-\pi}^{\pi} \frac{\mathrm{e}^{\mathrm{i}(t+1)\lambda}}{c(\mathrm{e}^{-\mathrm{i}\lambda})}\,\mathrm{d}z(\lambda) \end{aligned}$$

and so

$$\mathsf{E}(e_{t+1})^2 = \sigma_e^2.$$

Because we assume that X_t is Gaussian also ϵ_{t+1} is Gaussian and from the basic properties of the normal distribution (see e.g. Anděl [4]) we have $\mathsf{E}\epsilon_{t+1}^3 = 0$ and $\mathsf{E}\epsilon_{t+1}^4 = 3\sigma_e^4$. Hence we can write

$$\mathsf{E}X_{t+1}^2 = \mathsf{E}(\epsilon_{t+1} + \widehat{X}_{t+1})^2 = \mathsf{E}\epsilon_{t+1}^2 + 2\mathsf{E}\epsilon_{t+1}\widehat{X}_{t+1} + \mathsf{E}\widehat{X}_{t+1}^2 = \mathsf{E}\widehat{X}_{t+1}^2 + \sigma_e^2.$$

The mean square prediction error is then

$$\begin{aligned} \mathsf{E}[X_{t+1}^2 - \widehat{X}_{t+1}^2]^2 &= \mathsf{E}[(\epsilon_{t+1} + \widehat{X}_{t+1})^2 - \widehat{X}_{t+1}^2]^2 = \mathsf{E}[\epsilon_{t+1}^2 + 2\epsilon_{t+1}\widehat{X}_{t+1}]^2 \\ &= \mathsf{E}[\epsilon_{t+1}^4 + 4\epsilon_{t+1}^3\widehat{X}_{t+1} + 4\epsilon_{t+1}^2\widehat{X}_{t+1}^2] = \mathsf{E}\epsilon_{t+1}^4 + 4\mathsf{E}(\epsilon_{t+1}^3\widehat{X}_{t+1}) + 4\mathsf{E}(\epsilon_{t+1}^2\widehat{X}_{t+1}^2) \\ &= 3\sigma_e^4 + 4\mathsf{E}\epsilon_{t+1}^3\mathsf{E}\widehat{X}_{t+1} + 4\mathsf{E}\epsilon_{t+1}^2\widehat{X}_{t+1}^2 = 3\sigma_e^4 + 4\sigma_e^2\mathsf{E}\widehat{X}_{t+1}^2 \\ &= 3\sigma_e^4 + 4\sigma_e^2(\mathsf{E}X_{t+1}^2 - \sigma_e^2) = 3\sigma_e^4 + 4\sigma_e^2\mathsf{E}X_{t+1}^2 - 4\sigma_e^4 = 4\sigma_e^2\mathsf{E}X_{t+1}^2 - \sigma_e^4 \\ &= 4\sigma_e^2\sigma^2 - \sigma_e^4 = V^{(2)}(1). \end{aligned}$$

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* $V_t = Y_t - \mathsf{E}Y_t = X_t^2 - \mathsf{E}X_t^2$ *and bias-adjusted prediction* $\widehat{V}_{t+1} = \widehat{Y}_{t+1} - \mathsf{E}\widehat{Y}_{t+j} = \widehat{X}_{t+1}^2 - \mathsf{E}\widehat{X}_{t+1}^2$. *The expected error of the bias-adjusted prediction is*

$$\mathsf{E}\left(V_{t+1} - \widehat{V}_{t+1}\right)^2 = V^{(2)}(1) - \left(\sigma^2 - \sigma_e^2\right)^2.$$

Proof. We have

$$\begin{aligned} \mathsf{E}\left(V_{t+1} - \widehat{V}_{t+1}\right)^2 &= \mathsf{E}\left[\left(X_{t+1}^2 - \mathsf{E}X_{t+1}^2\right) - \left(\widehat{X}_{t+1}^2 - \mathsf{E}\widehat{X}_{t+1}^2\right)\right]^2 = \\ &= \mathsf{E}\left[\left(X_{t+1}^2 - \widehat{X}_{t+1}^2\right) - \left(\mathsf{E}X_{t+1}^2 - \mathsf{E}\widehat{X}_{t+1}^2\right)\right]^2 = \\ &= \mathsf{E}\left(U_{t+1} - \mathsf{E}U_{t+1}\right)^2 = \mathsf{E}U_{t+1}^2 - \left(\mathsf{E}U_{t+1}\right)^2, \end{aligned}$$

where $U_t = X_t^2 - \widehat{X}_t^2$. Now we have

$$\begin{aligned} \mathsf{E}U_{t+1}^2 &= \mathsf{E}\left(X_{t+1}^2 - \widehat{X}_{t+1}^2\right)^2 = V^{(2)}(1) \\ \mathsf{E}U_{t+1} &= \mathsf{E}X_{t+1}^2 - \mathsf{E}\widehat{X}_{t+1}^2 = \sigma^2 - \sigma_e^2. \end{aligned}$$

□

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 μ and variance σ^2 of the original series X_t, the optimal prediction in the original series $\widehat{X}_{t+h}$ 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 $\epsilon_t \sim \mathsf{N}(0, \sigma_\epsilon^2)$ are i.i.d. Obviously $\mu = 0$. Variance of the process is

$$\sigma^2 = \mathsf{var}(X_t) = \sigma_\epsilon^2(1 + \theta^2)$$

The optimal prediction is

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

and the error of the 1-step ahead prediction is

$$S^2(1) = \sigma_\epsilon^2.$$

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

$$\begin{aligned}
\widehat{Y}_{t+1}^{(1)} &= \exp\left(-\sum_{j=1}^{\infty}(-\theta)^j X_{t+1-j} + \frac{1}{2}\sigma_\epsilon^2\right), \\
\widehat{Y}_{t+1}^{(2)} &= \exp\left[-\sum_{j=1}^{\infty}(-\theta)^j X_{t+1-j}\right], \\
\widehat{Y}_{t+1}^{(3)} &= \exp\left[\frac{1}{2}\sigma_\epsilon^2(1+\theta^2)\right]\left[1-\sum_{j=1}^{\infty}(-\theta)^j X_{t+1-j}\right]
\end{aligned}$$

and the mean square errors of these predictions are

$$\begin{aligned}
V^{(1)}(1) &= \exp[2\sigma_\epsilon^2(1+\theta^2)]\left[1-\exp(\sigma_\epsilon^2)\right], \\
V^{(2)}(1) &= \exp[2\sigma_\epsilon^2(1+\theta^2)]\left[1-2\exp\left(-\frac{3}{2}\sigma_\epsilon^2\right)+\exp(-2\sigma_\epsilon^2)\right], \\
V^{(3)}(1) &= \exp\left[\sigma_\epsilon^2(1+\theta^2)\right]\left[\exp[\sigma_\epsilon^2(1+\theta^2)]-1-\sigma_\epsilon^2\theta^2\right].
\end{aligned}$$

4.1.2 Quadratic transformation

Consider the transformation

$$Y_t = X_t^2.$$

From corollary 3.5 the predictions for this transformation are

$$\begin{aligned}
\widehat{Y}_{t+1}^{(1)} &= \left[-\sum_{j=1}^{\infty}(-\theta)^j X_{t+1-j}\right]^2 + \sigma_\epsilon^2, \\
\widehat{Y}_{t+1}^{(1)} &= \left[-\sum_{j=1}^{\infty}(-\theta)^j X_{t+1-j}\right]^2, \\
\widehat{Y}_{t+1}^{(1)} &= \sigma_\epsilon^2(1+\theta^2)
\end{aligned}$$

and the mean square errors of these predictions are

$$\begin{aligned}
V^{(1)}(1) &= 4\left[\sigma_\epsilon^2(1+\theta^2)\right]\sigma_\epsilon^2 - 2(\sigma_\epsilon^2)^2, \\
V^{(2)}(1) &= 4\left[\sigma_\epsilon^2(1+\theta^2)\right]\sigma_\epsilon^2 - (\sigma_\epsilon^2)^2, \\
V^{(3)}(1) &= 2\left[\sigma_\epsilon^2(1+\theta^2)\right]^2.
\end{aligned}$$

4.1.3 Cubic transformation

Consider the transformation

$$Y_t = X_t^3.$$

From corollary 3.6 the predictions for this transformations are

$$\begin{aligned}
\widehat{Y}^{(1)}_{t+1} &= \left[-\sum_{j=1}^{\infty}(-\theta)^j X_{t+1-j}\right]^3 + 3\sigma_\epsilon^2\left(-\sum_{j=1}^{\infty}(-\theta)^j X_{t+1-j}\right),\\
\widehat{Y}^{(2)}_{t+1} &= \left[-\sum_{j=1}^{\infty}(-\theta)^j X_{t+1-j}\right]^3,\\
\widehat{Y}^{(3)}_{t+1} &= 3\sigma_\epsilon^2(1+\theta^2)\left(-\sum_{j=1}^{\infty}(-\theta)^j X_{t+1-j}\right)
\end{aligned}$$

and the mean square prediction errors

$$\begin{aligned}
V^{(1)}(1) &= 6\sigma_\epsilon^6\left[4+6\theta^2+3\theta^4\right]\\
V^{(2)}(1) &= V^{(1)}(h) + 9\theta^2\sigma_\epsilon^6,\\
V^{(3)}(1) &= 3\sigma_\epsilon^6\left(1+\theta^2\right)^3\left(5+3\theta^2\right).
\end{aligned}$$

$Y_t = \exp(X_t)$		$h=1$	$h=2$	$h=3$
$\theta = 0.25$	$G^{(2)}(h)$	0.015	0.015	0.015
$\theta = 0.25$	$G^{(3)}(h)$	4.097	4.097	4.097
$\theta = 0.75$	$G^{(2)}(h)$	0.080	0.080	0.080
$\theta = 0.75$	$G^{(3)}(h)$	0.350	0.350	0.350

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

$Y_t = X_t^2$		$h=1$	$h=2$	$h=3$
$\theta = 0.25$	$G^{(2)}(h)$	0.015	0.015	0.015
$\theta = 0.25$	$G^{(3)}(h)$	7.758	7.758	7.758
$\theta = 0.75$	$G^{(2)}(h)$	0.110	0.110	0.110
$\theta = 0.75$	$G^{(3)}(h)$	0.694	0.694	0.694

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

$Y_t = X_t^3$		$h = 1$	$h = 2$	$h = 3$
$\theta = 0.25$	$G^{(2)}(h)$	0.019	0.019	0.019
$\theta = 0.25$	$G^{(3)}(h)$	3.275	3.275	3.275
$\theta = 0.75$	$G^{(2)}(h)$	0.097	0.097	0.097
$\theta = 0.75$	$G^{(3)}(h)$	0.205	0.205	0.205

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

4.2 MA(∞) process

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

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

where again $\epsilon_t \sim \mathsf{N}(0, \sigma_\epsilon^2)$ are i.i.d. The variance of this process is

$$\sigma^2 = \mathsf{var}(X_t) = \sigma_\epsilon^2 \sum_{j=0}^{\infty} \theta_j^2,$$

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

$$S^2(1) = \sigma_\epsilon^2.$$

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

4.2.1 Exponential transformation

Consider the transformation

$$Y_t = \exp(X_t).$$

From corollary 3.4 the predictions for this transformation are

$$\begin{aligned} \widehat{Y}_{t+1}^{(1)} &= \exp\left(\widehat{X}_{t+1} + \frac{1}{2}\sigma_\epsilon^2\right), \\ \widehat{Y}_{t+1}^{(2)} &= \exp\left[\widehat{X}_{t+1}\right], \\ \widehat{Y}_{t+1}^{(3)} &= \exp\left[\frac{1}{2}\sigma_\epsilon^2 \sum_{j=0}^{\infty} \theta_j^2\right]\left[1 + \widehat{X}_{t+1}\right] \end{aligned}$$

and the mean square errors of these predictions are

$$
\begin{aligned}
V^{(1)}(1) &= \exp\left(2\sigma_\epsilon^2 \sum_{j=0}^{\infty} \theta_j^2\right)\left[1-\exp(\sigma_\epsilon^2)\right],\\
V^{(2)}(1) &= \exp\left(2\sigma_\epsilon^2 \sum_{j=0}^{\infty} \theta_j^2\right)\left[1-2\exp\left(-\frac{3}{2}\sigma_\epsilon^2\right)+\exp(-2\sigma_\epsilon^2)\right],\\
V^{(3)}(1) &= \exp\left[\sigma_\epsilon^2 \sum_{j=0}^{\infty} \theta_j^2\right]\left[\exp\left(\sigma_\epsilon^2 \sum_{j=0}^{\infty} \theta_j^2\right)-1-\sigma_\epsilon^2 \sum_{j=0}^{\infty} \theta_j^2+\sigma_\epsilon^2\right].
\end{aligned}
$$

4.2.2 Quadratic transformation

Consider the transformation

$$Y_t = \exp(X_t).$$

From corollary 3.5 the predictions for this transformation are

$$
\begin{aligned}
\widehat{Y}_{t+1}^{(1)} &= \widehat{X}_{t+1}^2 + \sigma_\epsilon^2,\\
\widehat{Y}_{t+1}^{(2)} &= \widehat{X}_{t+1}^2,\\
\widehat{Y}_{t+1}^{(3)} &= \sigma_\epsilon^2 \sum_{j=0}^{\infty} \theta_j^2
\end{aligned}
$$

and the mean square errors of these predictions are

$$
\begin{aligned}
V^{(1)}(1) &= 4\left[\sigma_\epsilon^2 \sum_{j=0}^{\infty} \theta_j^2\right]\sigma_\epsilon^2 - 2\sigma_\epsilon^4,\\
V^{(2)}(1) &= 4\left[\sigma_\epsilon^2 \sum_{j=0}^{\infty} \theta_j^2\right]\sigma_\epsilon^2 - \sigma_\epsilon^4,\\
V^{(3)}(1) &= 2\left[\sigma_\epsilon^2 \sum_{j=0}^{\infty} \theta_j^2\right]^2.
\end{aligned}
$$

4.2.3 Cubic transformation

Consider the transformation

$$Y_t = X_t^3.$$

From corollary 3.6 the predictions for this transformation are

$$
\begin{aligned}
\widehat{Y}_{t+1}^{(1)} &= \widehat{X}_{t+1}^3 + 3\sigma_\epsilon^2 \widehat{X}_{t+1},\\
\widehat{Y}_{t+1}^{(2)} &= \widehat{X}_{t+1}^3,\\
\widehat{Y}_{t+1}^{(3)} &= 3\sigma_\epsilon^2 \widehat{X}_{t+1}
\end{aligned}
$$

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 $\epsilon_t \sim \mathsf{N}(0, \sigma_\epsilon^2)$ 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

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

The optimal prediction is

$$\widehat{X}_{t+1} = \varphi X_t$$

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

$$\begin{aligned}
\widehat{Y}_{t+1}^{(1)} &= \exp\left[\varphi X_t + \frac{1}{2}\sigma_\epsilon^2\right], \\
\widehat{Y}_{t+1}^{(2)} &= \exp\left[\varphi X_t\right], \\
\widehat{Y}_{t+1}^{(3)} &= \exp\left[\frac{\sigma_\epsilon^2}{2(1-\varphi^2)}\right](1+\varphi X_t)
\end{aligned}$$

and the mean square errors of these predictions are

$$\begin{aligned}
V^{(1)}(1) &= \exp\left(2\frac{\sigma_\epsilon^2}{1-\varphi^2}\right)\left[1-\exp(-\sigma_\epsilon^2)\right], \\
V^{(2)}(1) &= \exp\left(2\frac{\sigma_\epsilon^2}{1-\varphi^2}\right)\left[1-2\exp\left(-\frac{3}{2}\sigma_\epsilon^2\right)+\exp\left(-2\sigma_\epsilon^2\right)\right], \\
V^{(3)}(1) &= \exp\left(\frac{\sigma_\epsilon^2}{1-\varphi^2}\right)\left[\exp\left(\frac{\sigma_\epsilon^2}{1-\varphi^2}\right)-1-\frac{\sigma_\epsilon^2}{1-\varphi^2}+\sigma_\epsilon^2\right].
\end{aligned}$$

4.3.2 Quadratic transformation

Consider the transformation

$$Y_t = X_t^2$$

The predictions for this transformation are

$$\begin{aligned}
\widehat{Y}_{t+1}^{(1)} &= \varphi X_t^2 + \sigma_\epsilon^2, \\
\widehat{Y}_{t+1}^{(1)} &= \varphi X_t^2, \\
\widehat{Y}_{t+1}^{(1)} &= \frac{\sigma_\epsilon^2}{1-\varphi^2}
\end{aligned}$$

and the mean square errors of these predictions are

$$\begin{aligned}
V^{(1)}(1) &= 4\left(\frac{\sigma_\epsilon^4}{1-\varphi^2}\right) - 2\sigma_\epsilon^4, \\
V^{(2)}(1) &= 4\left(\frac{\sigma_\epsilon^4}{1-\varphi^2}\right) - \sigma_\epsilon^4, \\
V^{(3)}(1) &= 2\left(\frac{\sigma_\epsilon^2}{1-\varphi^2}\right)^2.
\end{aligned}$$

4.3.3 Cubic transformation

Consider the transformation

$$Y_t = X_t^3$$

The predictions for this transformation are

$$\begin{aligned}
\widehat{Y}_{t+1}^{(1)} &= \varphi X_t^3 + 3\sigma_\epsilon^2 \widehat{X}_{t+h}, \\
\widehat{Y}_{t+1}^{(2)} &= \varphi X_t^3, \\
\widehat{Y}_{t+1}^{(3)} &= 3\left(\sigma_\epsilon^2 \sum_{j=0}^{\infty} \varphi^{2j}\right) \widehat{X}_{t+1}
\end{aligned}$$

and the mean square errors of these predictions are

$$\begin{aligned}
V^{(1)}(1) &= 3\sigma_\epsilon^6 \left[9\left(\frac{1}{1-\varphi^2}\right)^2 + 6\left(\frac{1}{1-\varphi^2}\right) + 2\right], \\
V^{(2)}(1) &= V^{(1)} + 9\sigma_\epsilon^6 \left(\frac{\varphi^2}{1-\varphi^2}\right), \\
V^{(3)}(1) &= 3\sigma_\epsilon^6 \left(\frac{1}{1-\varphi^2}\right)^2 \left[\frac{2}{1-\varphi^2} + 3\right].
\end{aligned}$$

$Y_t = \exp(X_t)$		$h=1$	$h=2$	$h=3$
$\varphi = 0.25$	$G^{(2)}(h)$	0.015	0.016	0.016
$\varphi = 0.25$	$G^{(3)}(h)$	4.120	3.850	3.834
$\varphi = 0.75$	$G^{(2)}(h)$	0.080	0.090	0.090
$\varphi = 0.75$	$G^{(3)}(h)$	0.681	0.292	0.185

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

$Y_t = X_t^2$		$h=1$	$h=2$	$h=3$
$\varphi = 0.25$	$G^{(2)}(h)$	0.015	0.016	0.016
$\varphi = 0.25$	$G^{(3)}(h)$	7.791	7.289	7.260
$\varphi = 0.75$	$G^{(2)}(h)$	0.070	0.119	0.150
$\varphi = 0.75$	$G^{(3)}(h)$	1.317	0.610	0.407

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

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

$Y_t = X_t^3$		$h = 1$	$h = 2$	$h = 3$
$\varphi = 0.25$	$G^{(2)}(h)$	0.019	0.020	0.020
$\varphi = 0.25$	$G^{(3)}(h)$	3.290	3.068	3.055
$\varphi = 0.75$	$G^{(2)}(h)$	0.073	0.102	0.112
$\varphi = 0.75$	$G^{(3)}(h)$	0.456	0.174	0.101

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

where $\epsilon_t \sim \mathsf{N}(0, \sigma_\epsilon^2)$ 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

$$\sigma^2 = \sigma_\epsilon^2 \frac{1 + 2\varphi\theta + \theta^2}{1 - \varphi^2}$$

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

$$S^2(1) = \sigma_\epsilon^2$$

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

$$\begin{aligned}
\widehat{Y}_{t+1}^{(1)} &= \exp\left(\widehat{X}_{t+1} + \frac{1}{2}\sigma_\epsilon^2\right), \\
\widehat{Y}_{t+1}^{(2)} &= \exp\left(\widehat{X}_{t+1}\right), \\
\widehat{Y}_{t+1}^{(3)} &= \exp\left(\frac{1}{2}\sigma_\epsilon^2 \frac{1 + \varphi\theta + \theta^2}{1 - \varphi^2}\right)(1 + \widehat{X}_{t+1})
\end{aligned}$$

and the mean square errors of these predictions are

$$
\begin{aligned}
V^{(1)}(1) &= \exp\left(2\sigma_\epsilon^2 \frac{1+\varphi\theta+\theta^2}{1-\varphi^2}\right)\left[1-\exp(-\sigma_\epsilon^2)\right], \\
V^{(2)}(1) &= \exp\left(2\sigma_\epsilon^2 \frac{1+\varphi\theta+\theta^2}{1-\varphi^2}\right)\left[1-2\exp\left(-\frac{3}{2}\sigma_\epsilon^2\right)+\exp\left(-2\sigma_\epsilon^2\right)\right], \\
V^{(3)}(1) &= \exp\left[\sigma_\epsilon^2 \frac{1+2\varphi\theta+\theta^2}{1-\varphi^2}\right]\left[\exp\left(\sigma_\epsilon^2 \frac{1+2\varphi\theta+\theta^2}{1-\varphi^2}\right)-1-\sigma_\epsilon^2\frac{1+2\varphi\theta+\theta^2}{1-\varphi^2}+\sigma_\epsilon^2\right].
\end{aligned}
$$

4.4.2 Quadratic transformation

Consider the transformation

$$Y_t = X_t^2$$

The predictions for this transformation are

$$
\begin{aligned}
\widehat{Y}_{t+1}^{(1)} &= \varphi X_t^2 + \sigma_\epsilon^2, \\
\widehat{Y}_{t+1}^{(2)} &= \varphi X_t^2, \\
\widehat{Y}_{t+1}^{(3)} &= \sigma_\epsilon^2 \frac{1+2\varphi\theta+\theta^2}{1-\varphi^2}
\end{aligned}
$$

and the mean square errors of these predictions are

$$
\begin{aligned}
V^{(1)}(1) &= 4\sigma_\epsilon^4 \frac{1+2\varphi\theta+\theta^2}{1-\varphi^2} - 2\sigma_\epsilon^2, \\
V^{(2)}(1) &= 4\sigma_\epsilon^4 \frac{1+2\varphi\theta+\theta^2}{1-\varphi^2} - \sigma_\epsilon^2, \\
V^{(3)}(1) &= 2\sigma^4 \left(\frac{1+2\varphi\theta+\theta^2}{1-\varphi^2}\right)^2.
\end{aligned}
$$

4.4.3 Cubic transformation

Consider the transformation

$$Y_t = X_t^3$$

The predictions for this transformation are

$$
\begin{aligned}
\widehat{Y}_{t+1}^{(1)} &= \varphi X_t^3 + 3\sigma_\epsilon^2 \widehat{X}_{t+1}, \\
\widehat{Y}_{t+1}^{(2)} &= \varphi X_t^3, \\
\widehat{Y}_{t+1}^{(3)} &= 3\left(\sigma_\epsilon^2 \frac{1+2\varphi\theta+\theta^2}{1-\varphi^2}\right)\widehat{X}_{t+1}
\end{aligned}
$$

and the mean square errors of these predictions are

$$
\begin{aligned}
V^{(1)}(1) &= 3\sigma_\epsilon^6\left[9\left(\frac{1+2\varphi\theta+\theta^2}{1-\varphi^2}\right)^2+6\left(\frac{1+2\varphi\theta+\theta^2}{1-\varphi^2}\right)+2\right],\\
V^{(2)}(1) &= V^{(1)}+9\sigma_\epsilon^6\left(\frac{1+2\varphi\theta+\theta^2}{1-\varphi^2}-1\right),\\
V^{(3)}(1) &= 3\sigma_\epsilon^6\left(\frac{1+2\varphi\theta+\theta^2}{1-\varphi^2}\right)^2\left[2\frac{1+2\varphi\theta+\theta^2}{1-\varphi^2}+3\right].
\end{aligned}
$$

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 $\mathsf{var}(\epsilon) = \sigma_\epsilon^2 = 1$. The observed average squared errors for each prediction

$$\begin{aligned}
\overline{V}^{(1)}(1) &= \frac{\sum_{i=1}^{n}(\widehat{Y}_{t+1}^{(1)} - Y_{t+1}^{(1)})^2}{n} \\
\overline{V}^{(2)}(1) &= \frac{\sum_{i=1}^{n}(\widehat{Y}_{t+1}^{(2)} - Y_{t+1}^{(2)})^2}{n} \\
\overline{V}^{(3)}(1) &= \frac{\sum_{i=1}^{n}(\widehat{Y}_{t+1}^{(3)} - Y_{t+1}^{(3)})^2}{n}
\end{aligned}$$

were compared with the theoretical values $V^{(1)}(1)$, $V^{(2)}(1)$ and $V^{(3)}(1)$. The variance of innovations is in all cases $\sigma_\epsilon^2 = 1$ 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 θ 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 (φ 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 φ. 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 φ 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.

θ	σ^2	$\overline{V}^{(1)}(1)$	$V^{(1)}(1)$	$\overline{V}^{(2)}(1)$	$V^{(2)}(1)$	$\overline{V}^{(3)}(1)$	$V^{(3)}(1)$
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

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

θ	σ^2	$\overline{V}^{(1)}(1)$	$V^{(1)}(1)$	$\overline{V}^{(2)}(1)$	$V^{(2)}(1)$	$\overline{V}^{(3)}(1)$	$V^{(3)}(1)$
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

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

θ	σ^2	$\overline{V}^{(1)}(1)$	$V^{(1)}(1)$	$\overline{V}^{(2)}(1)$	$V^{(2)}(1)$	$\overline{V}^{(3)}(1)$	$V^{(3)}(1)$
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

Table 5.3: MA(1) process, $Y_t = X_t^3$

φ	σ^2	$\overline{V}^{(1)}(1)$	$V^{(1)}(1)$	$\overline{V}^{(2)}(1)$	$V^{(2)}(1)$	$\overline{V}^{(3)}(1)$	$V^{(3)}(1)$
0.1	1.010	4.518	4.766	4.935	5.196	4.518	4.766
0.5	1.333	9.293	9.097	10.155	9.917	9.543	9.334
0.9	5.263	8 203	23 567	8 811	25 691	13 629	36 267

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

φ	σ^2	$\overline{V}^{(1)}(1)$	$V^{(1)}(1)$	$\overline{V}^{(2)}(1)$	$V^{(2)}(1)$	$\overline{V}^{(3)}(1)$	$V^{(3)}(1)$
0.1	1.010	2.019	2.040	3.011	3.040	2.020	2.041
0.5	1.333	3.389	3.333	4.408	4.333	3.611	3.55600
0.9	5.263	19.634	19.053	20.669	20.053	57.215	55.402

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

φ	σ^2	$\overline{V}^{(1)}(1)$	$V^{(1)}(1)$	$\overline{V}^{(2)}(1)$	$V^{(2)}(1)$	$\overline{V}^{(3)}(1)$	$V^{(3)}(1)$
0.1	1.010	15.875	15.366	15.997	15.457	15.875	15.366
0.5	1.333	34.177	30.000	37.795	33.000	34.482	30.222
0.9	5.263	771.326	659.186	819.446	697.554	1471.947	1124.071

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

(φ, θ)	σ^2	$\overline{V}^{(1)}(1)$	$V^{(1)}(1)$	$\overline{V}^{(2)}(1)$	$V^{(2)}(1)$	$\overline{V}^{(3)}(1)$	$V^{(3)}(1)$
(0.25,0.25)	1.267	9.220	7.962	10.033	8.679	9.383	8.100
(0.75,0.25)	3.286	388.534	451.580	418.022	492.268	534.034	626.569
(0.25,0.75)	2.067	49.236	39.435	53.018	42.988	55.565	46.062
(0.75,0.75)	6.142	56 338	136 904	32 703	149 239	40 816	213 721

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

(φ, θ)	σ^2	$\overline{V}^{(1)}(1)$	$V^{(1)}(1)$	$\overline{V}^{(2)}(1)$	$V^{(2)}(1)$	$\overline{V}^{(3)}(1)$	$V^{(3)}(1)$
(0.25,0.25)	1.267	3.211	3.067	4.234	4.067	3.375	3.209
(0.75,0.25)	3.286	12.152	11.143	13.233	12.143	24.935	21.592
(0.25,0.75)	2.067	6.637	6.267	7.668	7.267	9.080	8.542
(0.75,0.75)	6.142	22.976	22.571	23.978	23.571	77.575	75.469

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

(φ,θ)	σ^2	$\overline{V}^{(1)}(1)$	$V^{(1)}(1)$	$\overline{V}^{(2)}(1)$	$V^{(2)}(1)$	$\overline{V}^{(3)}(1)$	$V^{(3)}(1)$
(0.25,0.25)	1.267	29.205	26.520	31.974	28.920	29.353	26.634
(0.75,0.25)	3.286	290.404	238.347	317.687	258.918	405.426	309.997
(0.25,0.75)	2.067	94.258	84.120	105.163	93.720	102.832	91.402
(0.75,0.75)	6.142	882.115	914.265	924.160	960.551	1637.966	1730.405

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

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*

$$\begin{aligned} H_n(x) &= (-1)^n \exp\left(\frac{x^2}{2}\right) \frac{\mathrm{d}^n}{\mathrm{d}x^n} \exp\left(\frac{x^2}{2}\right), \\ H_n(x) &= (-1)^n \frac{\phi^{(n)}(x)}{\phi(x)} \end{aligned}$$

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

$$\begin{aligned} H_0(x) &= 1, \\ H_1(x) &= x, \\ H_2(x) &= x^2 - 1, \\ H_3(x) &= x^3 - 3x, \\ H_4(x) &= x^4 - 6x^2 + 3, \\ H_5(x) &= x^5 - 10x^3 + 15x \end{aligned}$$

and so on. Define operators E_0 and E by

$$\begin{aligned}\mathsf{E}_0\left\{\psi(x)\right\} &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\psi(x)\exp\left(\frac{-x^2}{2}\right)\mathrm{d}x,\\ \mathsf{E}\left\{\psi(x)\right\} &= \frac{1}{\sqrt{2\pi\sigma^2}}\int_{-\infty}^{\infty}\psi(x)\exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)\mathrm{d}x.\end{aligned}$$

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

$$\mathsf{E}_0\{H_n(x)H_k(x)\} = \begin{cases} 0, & n \neq k, \\ n!, & n = k, \end{cases} \tag{7.1}$$

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

$$\mathsf{E}_0\{H_n(x)\} = 0 \text{ for } n > 0. \tag{7.2}$$

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

$$\mathsf{E}\{H_n(X)|Y=y\} = \rho^n H_n(y)$$

and

$$\mathsf{E}\{H_n(X)H_k(Y)\} = \begin{cases} 0, & n \neq k, \\ n!\rho^n, & n = k. \end{cases} \tag{7.3}$$

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:

$$\exp\left(tx - \frac{t^2}{2}\right) = \sum_{n=0}^{\infty} H_n(x)\frac{t^n}{n!}. \tag{7.4}$$

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

$$H_n(Ax+By) = \sum_{k=0}^{n}\binom{n}{k}A^k B^{n-k}H_k(x)H_{n-k}(y). \tag{7.5}$$

In general, the coefficients α_j for Hermite polynomial expansion of a function

$$T(x) = \sum_{j=0}^{\infty}\alpha_j H_j(x)$$

are given by

$$\alpha_n = \mathsf{E}_0 \left\{ \frac{\mathrm{d}^n}{\mathrm{d}x^n} \frac{T(x)}{n!} \right\}. \tag{7.6}$$

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

$$x^{2p} = \frac{(2p)!}{2^{2p}} \sum_{j=0}^{p} \frac{H_{2j}(x)}{(2j)!(p-j)!}, \qquad p = 0, 1, 2, \ldots \tag{7.7}$$

and

$$x^{2p+1} = \frac{(2p+1)!}{2^{2p+1}} \sum_{j=0}^{p} \frac{H_{2j+1}(x)}{(2j+1)!(p-j)!}, \qquad p = 0, 1, 2, \ldots. \tag{7.8}$$

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*

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

where

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

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

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

Then the best linear h-step ahead prediction is

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

The one-step ahead prediction is given by

$$\widehat{X}_{t+1} = \int_{-\pi}^{\pi} \mathrm{e}^{\mathrm{i}(t+1)\lambda} \frac{c(\mathrm{e}^{-\mathrm{i}\lambda}) - c(0)}{c(\mathrm{e}^{-\mathrm{i}\lambda})} \mathrm{d}z(\lambda)$$

and the mean square error of this prediction is

$$\mathsf{E}(X_{t+1} - \widehat{X}_{t+1})^2 = 2\pi \exp\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} \log f(\lambda) \mathrm{d}\lambda\right) > 0.$$

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

Bibliography

[1] Abadir K.M., Talmain G. (2004): Autocovariance functions of series and of their transforms. Journal of Econometrics 124, pp. 227-252.

[2] Abramowitz M., Stegun I.A. (1964): Handbook of Mathematical Functions with Formulas, Graphs and Mathematical tables, National Bureau of Standards, Washington. Available at http://www.convertit.com/Go/ConvertIt/Reference/AMS55.ASP.

[3] Anděl, J. (1976): Analýza časových řad, SNTL, Praha.

[4] Anděl, J. (2005): Základy matematické statistiky. Matfyzpress, Praha.

[5] Choi I-B., Taniguchi M. (2003): Prediction problems for square-transformed stationary processes. Statist. Inference for Stoch. Processes 6, 43-64.

[6] Granger C.W.J., Newbold P. (1976): Forecasting transformed series. J. R. Statist. Soc. B 38, No. 2, 189-203

[7] Grenander U., Rosenblatt M. (1957): Statistical Analysis of Stationary Time Series, Wiley, New York.

[8] Hannan E.J. (1970): Multiple Time Series, Wiley, USA.

[9] Lebedev N.N. (1972): Special Functions and their Applications, Dover Publications, New York.

[10] Palma W., Zevallos M. (2004): Analysis of the correlation structure of square time series. Journal of Time Series Analysis, Vol. 25, No. 4., pp. 529-550.

[11] Prášková Z. (2004): Základy náhodných procesů 2, Karolinum, Praha.