Feedforward Control and Process Improvement for Some Time Series Disturbance Models

Lihui Shi

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2012

Program Authorized to Offer Degree: UW Industrial and Systems Engineering

University of Washington

Abstract

Feedforward Control and Process Improvement
for Some Time Series Disturbance Models

by Lihui Shi

Chair of Supervisory Committee:

Professor Kailash C. Kapur
Industrial and Systems Engineering

Process adjustment strategy is an important part of the process improvement methods. The feedback control technique is used to compensate the deviation of the output, and it has been intensively investigated. For continuous improvement and proactive strategies, feedback has a delay and is not the ideal solution. In this thesis we introduce the idea of feedforward control in our process adjustment procedure. This thesis focuses on studying the implementation of the feedforward control on many manufacturing processes and some common realistic disturbance models.

First, the basic concepts of statistical process control and engineering process control are considered, the basic ideas of the feedback and feedforward control are introduced, and some motivation examples of feedforward control are proposed. We introduce the philosophy of feedforward control from a new perspective, called disturbance decomposition viewpoint, and show the basic causations of the different factors in our control system. We illustrate this feedforward control idea with many real examples.

Then, some studies of feedforward control are investigated. We build new process adjustment models based on the time series disturbance models. We focus on complicated and synthesis disturbance models, which are more realistic in the real applications nowadays. Some common and appropriate time series models, both stationary and nonstationary, are

chosen as the background disturbances, such as the integrated moving average (IMA) model, autoregressive (AR) model, autoregressive moving average (ARMA) model.

Two types of disturbance models are investigated: periodic shift disturbance models and random step change disturbance models. We combine the feedforward control with the traditional feedback control in our adjustment system for maintaining the stability of the process and delivering products at target values. The performance evaluations for different control strategies, including feedback control, feedforward control and combined (feedback plus feedforward) control on those disturbance models are studied, and some numerical numerical results are obtained.

In the periodic shift disturbance models, we derive the feedforward control equations for each model and the closed forms for the long run output mean square error (MSEO). The detailed proofs of those formula are also provided. In the random step change disturbance model, we propose a quasi-feedforward feedback control strategy, and get the MSEO results through simulations.

In summary, a collection of process disturbance models for manufacturing and some others processes are proposed and studied systematically in this thesis. It is demonstrated that by implementing proper adjustment strategies, the stability of the process can be better maintained and the variation reduction can be efficiently achieved; thus, significant economic benefits obtained from the consistent quality of products will be achieved. This research contributes directly to quality improvement of the manufacturing and some other industries and to the field of applied statistics.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I wish to express sincere appreciation to the Department of Industrial and Systems Engineering at University of Washington, for their extended long-term support. Especially, I would like to express my deep gratitude to my PhD advisor, Professor Kailash C. Kapur, for his knowledge, patience, encouragement and support in the past five years. It is my great honor and fortune to be his student. I would like to thank the other committee members, Professor Christina Mastrangelo and Professor Santosh Devasia, for their invaluable feedbacks and suggestions throughout the course of this thesis development. Finally, I want to thank my parents, for their spiritual supports and deep love throughout my life. This thesis would never have been completed without the encouragement and devotion of my family and friends.

DEDICATION

to my dear parents, Shizhou Shi and Yulan Li

Chapter 1

INTRODUCTION

1.1 Motivation

Control is a continuous endeavor to keep measures of quality as close as possible to their target values for indefinite periods of time [9].

For any process, including industrial, chemical, economic, business, etc, in a general sense, our objective is to achieve the smallest variation around a fixed target value. There are two statistically based approaches for addressing this problem: statistical process control (SPC), which is a major tool in process monitoring, and engineering process control (EPC), which is used in process adjustment.

Process monitoring provides an ongoing check on the stability of the process. It uses control charts to identify variation which are due to special causes, so that its elimination can result in process improvement. Meanwhile, a stable process does not necessarily imply a desired result. With a stable process that does not trigger any of the detection rules for a control chart, a process capability analysis is also performed to evaluate the ability of the current process to produce conforming product, which are within the specification limits.

However, stability of the process is not always desired, since in many cases, the current process might not be a good one. For example, if the process mean is not on target, or if the process variation is too large, then the stability of this current process will definitely be hazardous. To achieve the ultimate goal of the process improvement, what we need is the change of the process, or more specifically, the adjustment of the process so that it can meet the customer needs.

1.2 Statistical Process Control

1.2.1 Basic Statistical Process Control Ideas

Statistical process control (SPC) is used to detect and remove the process disturbance due to the special causes in the process, mainly by using the control chart technique in process monitoring. SPC parallels the hypothesis test. For example, in the control chart for the process mean, we want to test if there is the process mean μ has a shift from the target value μ_0, then the hypothesis test is: $H_0 : \mu = \mu_0$ vs $H_a : \mu \neq \mu_0$. Similarly, in the control chart for the process variance, the corresponding hypothesis test is: $H_0 : \sigma = \sigma_0$ vs $H_a : \sigma \neq \sigma_0$. Or more generally, $H_0 : F = F_0$ vs $H_a : F \neq F_0$, where F_0 is a specific distribution, such as an independent and identical distributed (i.i.d.) normal $N(\mu_0, \sigma_0^2)$ process, or an autoregressive moving average (ARMA) (1,1) time series process, or a linear profile.

In SPC, when we are monitoring a process by using the control charts, there are control limits which usually includes the upper control limit (UCL) and the lower control limit (LCL). People usually use a 3-sigma limits as a rule of thumb, where sigma stands for the standard deviation for a reference distribution. The statistical basis of control chart is, as long as the points are plot within the control limits and did not trigger any detection rules, the process is assumed to be in control, and no action is necessary; however, a point plots outside of the control limits is interpreted as evidence that the process is out of control, and investigation and correction action are required to find and eliminate the assignable cause. For the basic principles of SPC, see [25].

1.2.2 Control Charts for i.i.d. Process

The most frequently used control charts include Shewhart chart, CUSUM chart and EWMA chart. The disturbance can be a shift on process mean or variance, or both.

The Shewhart chart is the most basic and straightforward one. On the Shewhart chart, we plot the measured or computed statistic directly [34].

CUSUM involves the calculation of a cumulative sum (which is what makes it "sequential") [28]. Samples from a process $\{x_n\}_{n\geq 1}$ are assigned weights K, and summed as follows:

$$\begin{cases} S_0^+ = 0 \\ S_{n+1}^+ = max(0, S_n^+ + x_n - K) \end{cases}$$

where $K = \frac{s}{2}\delta$ and s is the shift size that one wishes to detect. When the value of S exceeds a certain threshold value, a change in value has been found. The above formula only detects changes in the positive direction. When negative changes need to be found as well, the min operation should be used instead of the max operation, and this time a change has been found when the value of S is below the (negative) value of the threshold value.

$$\begin{cases} S_0^- = 0 \\ S_{n+1}^- = min(0, S_n^- - x_n - K) \end{cases}$$

An exponentially weighted moving average (EWMA), applies weighting factors which decrease exponentially. The weighting for each older data point decreases exponentially, giving much more importance to recent observations while still not discarding older observations entirely. The EWMA statistic is defined as

$$\begin{cases} EWMA_0 = \mu_0 \\ EWMA_n = \lambda EWMA_n + (1 - \lambda)x_{n-1}, 0 < \lambda < 1 \end{cases}$$

where λ is the smoothing parameter or weight, which accounts for how much of the past data should be discounted out at computing the current EWMA statistic. We will show later that the EWMA can be used for both monitoring and forecasting.

1.2.3 Control Charts for Autocorrelated Process

Usually the control charts are applied on i.i.d. processes, and independence between the observations is an important basic assumption for the control charts. However, in real applications, this independence assumption is not justified, since many processes exhibit autocorrelation. The efficiency of control charts will be substantially deteriorated with the presence of autocorrelation in the process, for example, the false alarm rate will be dramatically increased when we use the conventional control charts to monitor an autocorrelated process. Many authors, such as Alwan and Roberts, Montgomery and Mastrangelo, etc, have investigated in designing control charts for monitoring the autocorrelated process [1], [26].

These control charts for the autocorrelated process use the basic fact that, we check the series of the residuals obtained after fitting the process by a special time series model. If the model is adequate or correct, then the residuals should resemble a series of independent random noise. This is also called diagnostic checking in the iterative model building process [5].

For example, Montgomery and Mastrangelo proposed to use an integrated moving average (IMA) process to model a general nonstationary process so that the EWMA statistic of the process observations is the next step prediction of the process mean [26] .

1.3 Engineering Process Control

1.3.1 Three Types of Variability

Variation is the enemy of process improvement, if there is no variation, there will be no quality problems. Suppose we sample the products from a production line and measure some characteristic of interest, since all the measurements are subject to variation, so in order to find and apply the appropriate control methods, we need to be able to describe and understand variation.

There are different types of variabilities. Figure 1.1(a) presents the stationary and uncorrelated process, i.e., the white noise; Figure 1.1(b) presents the stationary and autocorrelated process; and Figure 1.1(c) presents the nonstationary process. Figures 1.1(a) and 1.1(b) illustrate stationary behavior, by which it means that the process data vary around a fixed mean in a stable and predictable manner.

The key property of the white noise is that the order of the data tells us nothing about the series, so consequently it provides no useful information in predicting the future values.

For the second process, positive deviations tend to follow and to be followed by positive deviations as well as the negative ones, so the adjacent observations from the process mean are not statistically independent of each other. Thus the lack of statistical independence implies some degree of predictability for a guessed value from previous ones.

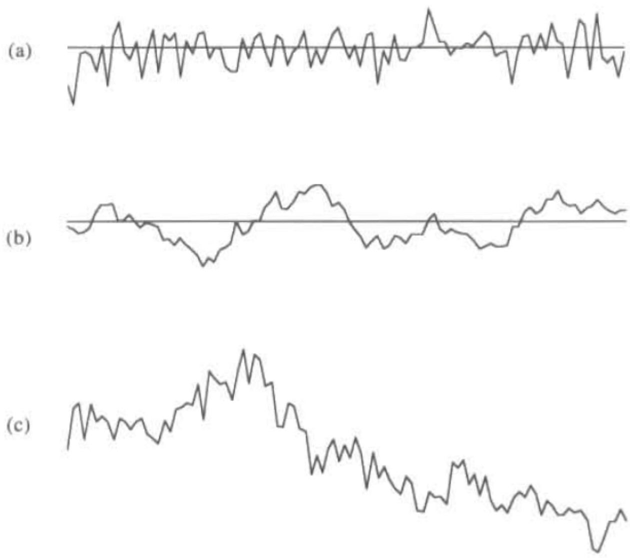

Figure 1.1: (a) A stationary white noise series; (b) A stationary autocorrelated series; (c) A nonstationary series.

Deming says, “no process, except in artificially demonstrations by use of random numbers, is steady and unwavering” [15], so a stable stationary state is an unnatural one, and it must be stabilized by using process adjustment. This is especially true for the process which is affected by factors that cannot be standardized.

To design an efficient adjustment methods we need to know something about the disturbance we are trying to control, i.e., the sequence of deviations that would have occurred if no control had been applied. Such a disturbance can frequently be represented by a nonstationary series. However, the output series we actually see is not the uncontrolled series but rather the series after the control had been applied. And if a control is effective, the controlled series will not look like a nonstationary one but will be like a stationary series such as Figures 1.1(a) and 1.1(b).

1.3.2 From SPC to EPC

SPC is traditionally applied to a process with a fixed mean, and usually successive observations are assumed to be statistically independent. It assumes that we can bring the process back into a state of statistical control, and once the process is in the in-control state, it will tend to stay there with stability for a relatively long period of time without continual ongoing adjustment.

However, sometimes, if no control were applied, the process will have a tendency to drift or wander away from the target, due to its nonstationarity property. These may due to some causes which are often known but cannot be economically removed. We stabilize this type of process by using EPC, such as feedback or feedforward control, which we will discuss later. Thus under these circumstances, only SPC is not enough, and EPC is required. For the control of such processes, a specific question is, if we reject $H_0 : \mu = \mu_0$, then do we need to make adjustment or not? If yes, how much adjustment should we make?

1.3.3 Basic Ideas of Feedback and Feedforward Control

As the two major statistical methods addressing the quality process, while SPC parallels hypothesis test, EPC parallels parameter estimation, through which it estimates the current level of the disturbance that is used to apply appropriate compensatory adjustment. Feedback and feedforward control are two important methods in EPC.

Figure 1.2 illustrates the flow diagrams of the typical open-loop system in Figure 1.2(a), feedforward system in Figure 1.2(b) and closed loop feedback system in Figure 1.2(c).

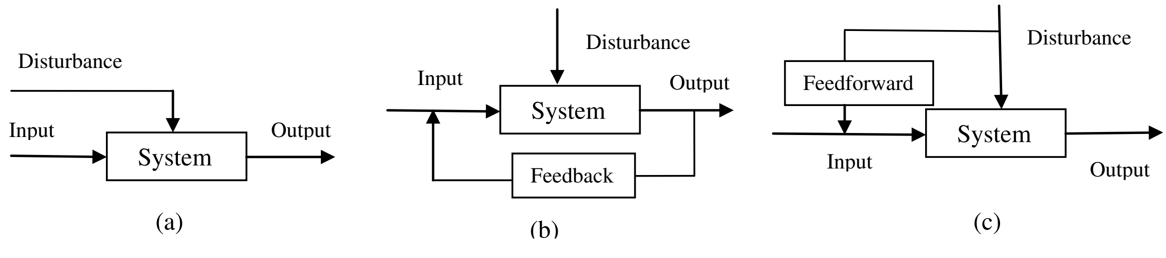

Figure 1.2: Feedforward and feedback control systems

In the open-loop system, there is always a disturbance term into the black-box system, which renders dispersion on the output value and makes it not necessarily equal to the target value that we desire. F eedback control is made in reaction to errors at the output. However, sometimes it is too late to wait until the process output is obtained and measured by using the feedback control only, so it is ideal if we can do some preaction to adjust the process before the output. Feedforward control is made to compensate those anticipated disturbances.

Generally, if we used the measurements of one or more of the known disturbing input factors to calculate the estimate of the disturbance, we would have an example of feedforward control. If the estimate of the disturbance directly or indirectly used are only the present and past values of the output disturbance, then we would have an example of feedback control [5].

While feedback control has been extensively investigated by many researchers [9], [12], etc, the existing research on feedforward control from a statistical perspective is quite limited. Therefore, further investigations and developments on feedforward control are both necessary and worthwhile.

1.3.4 Motivation of Feedforward Control

• Example 1: Basic grinding problem.

A grinding machine uses an abrasive wheel as the cutting tool. Each grain of abrasive on the wheel's surface cuts a small chip from the workpiece via shear deformation.

There are two wheels in our basic grinding problem, one is the grinding wheel working as a cutting tool, and the other is the workpiece wheel, which was cut by the grinding wheel to form a designed shape.

Suppose that the target value for the diameter of the workpiece wheel is 5.0 cm. The traditional control method is to use the feedback control, that is, to measure the diameter of the produced workpiece wheel, and adjust the position of the grinding wheel correspondingly. However, if we can apply the feedforward control idea, it might be preferable for producing better workpiece products. By using the feedforward control, we monitor the wear of the grinding wheel and measure its diameter after wearing. If it is too close to the workpiece wheel, then we have evidence to suspect that the diameter of the produced workpiece wheel is smaller than the target value 5.0 cm, so we adjustment the workpiece wheel with a bigger distance to the workpiece wheel, and vice versa.

• Example 2: Thermostat problem.

A thermostat will counteract a drop in temperature by switching on the heating. Feedforward control will suppress the disturbance before it has had the chance to affect the system's essential variables. This requires the capacity to anticipate the effect of perturbations on the system's goal. Otherwise the system would not know which external fluctuations to consider as perturbations, or how to effectively compensate their influence before it affects the system. This requires that the control system be able to gather early information about these fluctuations.

For example, feedforward control might be applied to the thermostatically controlled room by installing a temperature sensor outside of the room, which would warn the thermostat about a drop in the outside temperature, so that it could start heating before this would affect the inside temperature. In many cases, such advance warning is difficult to implement, or simply unreliable. For example, the thermostat might start heating the room, anticipating the effect of outside cooling, without being aware that at the same time someone in the room switched on the oven, producing more than enough heat to offset the drop in outside temperature. No sensor or anticipation can ever provide complete information about the future effects of an infinite variety of possible perturbations, and therefore feedforward control is bound to make mistakes.

• Example 3: Water heater problem.

Bela G. Liptak gave an example of water heater to illustrate the idea of feedforward control in Figure 1.3 [22]. The steam is put into cold water in a heat exchanger system to maintain a fixed temperature on the generated hot water. Instead of only using the thermostat on the hot water outlet, a skillful operator of a water heater system could use a simple feedforward strategy to compensate for changes in inlet water temperature by detecting a change in inlet water temperature and in response to that, increasing or decreasing the steam rate to counteract the change. This same compensation could be made automatically with an inlet temperature detector designed to initiate the corrective appropriate adjustment in the steam valve opening. Here the cold water temperature is our informative variable B.

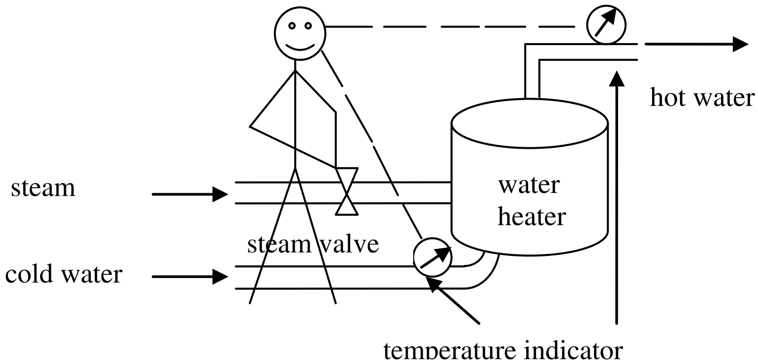

Figure 1.3: feedforward feedforward control implemented by an operator in a water heater system

Bela G. Liptak also pointed out that the concept of feedforward control is very powerful, but it is difficult to implement in most process control applications, since in many cases disturbances cannot be accurately measured. The main limitation of feedforward is due to our inability to prepare perfect process models or to make perfectly accurate measurements.

• Example 4: Drying process problem.

Myer Kutz considered a drying process with a long residence time [21]. Using feedback control, a moisture sensor would be placed at the output of the drying unit and its signal would be used to adjust the drying time and or drying temperature. However, if the residence time is 3 hours, then it will take close to 3 hours to determine if the right adjustment has been made. In the meantime, the product under production may not meet quality standards for moisture and require reprocessing or disposal. Both alternatives impose unacceptable costs during manufacturing.

A better solution is to measure the moisture of the incoming feed material and treat any deviation in feed moisture content from nominal as a disturbance variable and use this as a feedforward type of controller. By measuring the moisture content of the feed material entering the dryer, a mass balance calculation can be employed to determine the change in drying temperature or drying time required for that feedstock. The correct process is applied without feedback control.

• Example 5: Car cruise control

A feedforward system can be illustrated by comparing it with a familiar feedback system that of cruise control in a car. When in use, the cruise control enables a car to maintain a steady road speed. When an uphill stretch of road is encountered, the car slows down below the set speed; this speed error causes the engine throttle to be opened further, bringing the car back to its original speed.

A feedforward system on the other hand would in some way "predict" the slowing down of the car. For example it could measure the slope of the road and, upon encountering a hill, would open up the throttle by a certain amount, anticipating the extra load. The car does not have to slow down at all for the correction to come into play. Here the road slope is the informative variable B.

1.4 Basic Causations in Feedforward Feedback Control Model

In our feedback and feedforward control framework, all the variables in a system can be categorized into three groups: control factors X which values can be adjusted, uncontrollable noise factors e which are the source of variation, and the output Y. The relationship between those factors can be written as $Y = f(X, e)$, where f is the transfer function. Meanwhile, if e can be further decomposed into a new observable but uncontrollable factor B (possibly with observation errors) and the remaining noise e', then we can write a new transfer function $Y = f'(X, B, e')$ and apply feedforward control by using the information of B. We call B informative variable and e' noninformative noise.

See Figure 1.4 for illustrations, while the left graph is closed-loop feedback control, the right graph adds the feedforward control loop for the informative variable B. Apparently any process can be adjusted by feedback control, however, it is not true for feedforward control. We have illustrated the feasibility conditions for feedforward control from disturbance decomposition viewpoint: only if the disturbance e can be decomposed into a noninfor-

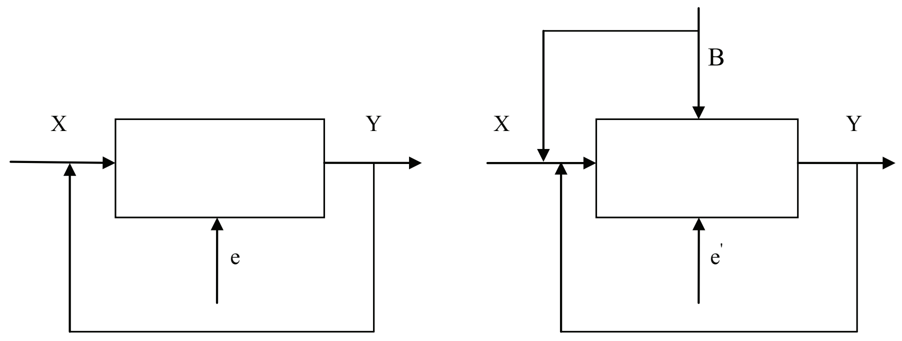

Figure 1.4: Disturbance Decomposition

matative noise (or background disturbance) e' and an informative variable B, which can be measured, forecasted, or estimated (with errors usually), but cannot be adjusted, then we can use the information of B to adjust X to make the output Y closer to the target value.

Since unmeasured background disturbances are always present in any control system, feedforward control is usually combined with feedback control to achieve the highest possible efficiency. How effective this feedforward control can be depends on the magnitude of the informative variable B that can be taken out from the noise e, and we will explain this later in detail.

1.5 *The Informative Variable B in Some Examples*

From those real examples with feedforward control, we can identify the noninformative variable B in each example. In the water heater example in which the steam is put into cold water in a heat exchanger system to maintain a fixed temperature on the generated hot water, since the operator would compensate for changes in inlet cold water temperature by monitoring it and in response to that, increasing or decreasing the steam rate to counteract the change, so the cold water temperature is our informative variable B.

In another example for the cruise control which enables a car to maintain a steady road speed, since feedforward control is applied by measuring the road slope, upon encountering a hill, then it would open up the throttle by a certain amount automatically and anticipate the extra load, so the road slope is the informative variable B.

1.6 Some More Comments on the Informative Variable B

From the previous illustration about the disturbance decomposition model and the basic causations between the different variables, we know that the key for the feedforward control application is the informative variable B. Here we will provide more detailed explanations on the informative variable B in different disturbance models we will investigate.

We mentioned that the basic requirement for the informative variable B is that it can be measured, forecasted, or estimated (with errors usually), but cannot be adjusted. Later we will show that for real disturbance models, feedforward control can be applied either when B can be measured, estimated, or even when its distribution is known.

1.7 Overview of Thesis

In Chapter 2, we will have a literature review on Box's feedback control by using the time series method. Both responsive system and process with inertia represented by the first-order dynamic model have been covered. We will also cover some preliminary knowledge on time series analysis and introduce some classical time series models. In Chapter 3 and Chapter 4, we will investigate two very common disturbance models in the manufacturing processes: the periodic shift disturbance models and random step change disturbance models. Specific control equations are derived for each control strategy under each disturbance model. We will evaluate the effect of feedback control, feedforward control and combined control for those disturbance models.

Chapter 2

LITERATURE REVIEW

In chapter 2, we first give a detailed review on the basic knowledge of time series models, and then introduce some common time series models, including autoregressive (AR) model, moving average (MA) model, autoregressive moving average (ARMA) model, integrated moving average (IMA) model, etc. Then we will make a literature review on the feedback control method proposed by Box and his collaborators, who studied the process adjustment problem by feedback control using time series models since 1960s. For those time series disturbance models, we will talk about their expressions in both the responsive system and system with inertia.

2.1 Review of Time Series Models

Before we introduce Box's feedback control in the next section, it is necessary to give a review for time series models as preliminary knowledge.

A time series is a set of observations generated sequentially in time. Here we do not consider the deterministic time series, in which future values are exactly determined by some mathematical function. Instead, we only consider time series as a stochastic process, in which the process evolves in time according to probabilistic laws. In general, a time series model is an equation that relates in some way or other a sequence $\{z_t\}_{t\geq 1}$ of values for the disturbance to a sequence $\{a_t\}_{t\geq 1}$ of a white noise series.

We know that as a stochastic process, the time series is strictly stationary if its properties are unaffected by a change of time origin, i.e., the joint distributions of two sequence $\{z_{t_1}, z_{t_2}, ..., z_{t_m}\}$ and $\{z_{t_1+k}, z_{t_2+k}, ..., z_{t_m+k}\}$ are the same, for any shift time k.

2.1.1 Autocorrelation Function

Usually, a stationary time series can be described by its mean, variance and autocorrelation function. Without loss of generality, we introduce time series in the context of modeling the disturbance $z_t = y_t - T$. When the series is stationary, we suppose that $T = \mu$ (Of course $T = \mu$ is not always true, and adjustment bias may exist, but we do not discuss this case now), so $z_t = y_t - \mu$ and $E(z_t) = E(y_t) - \mu = \mu - \mu = 0$. A measure of the dispersion of a stationary time series is its variance, i.e., its long-term average of the squares of deviations from the mean:

$$\sigma_y^2 = \sigma_z^2 = E[(z_t - E(z_t))^2] = E(z_t^2) = \frac{1}{n}(z_t^2 + z_{t-1}^2 + z_{t-2}^2 + ...), \tag{2.1}$$

where the number of terms n in the sum is supposed to be very large, and theoretically infinite.

Similarly, a measure of the series dependence between deviations k steps apart is the autocovariance function at lag k:

$$\begin{aligned} \gamma_k &= cov(z_t, z_{t+k}) \\ &= E[(z_t - E(z_t))(z_{t+k} - E(z_{t+k}))] \\ &= E(z_t z_{t+k}) = \tfrac{1}{n}(z_t z_{t-k} + z_{t+1} z_{t+1-k} + ...) \end{aligned} \tag{2.2}$$

So $\gamma_0 = \sigma_z^2$. The autocorrelation coefficients at lag k are defined to be

$$\rho_k = \gamma_k / \gamma_0. \tag{2.3}$$

Its series $\gamma_0, \gamma_1, \gamma_2, ...$ is called the autocorrelation function (ACF).

2.1.2 Variogram

Let $V_m = var(z_{t+m} - z_t)$ denote the variance of the difference between observations taken m steps apart, then a corresponding scaleless quantity measuring variance inflation is

$$V_m/V_1 = var(z_{t+m} - z_t)/var(z_{t+1} - z_t).$$

We call the series $\{1, V_2/V_1, V_3/V_1, ...\}$ the standardized variogram, or simply the variogram. We have the relationship between the variogram and the ACF:

$$V_m/V_1 = (1 - \rho_m)/(1 - \rho_1).$$

And since

$$\begin{aligned}
V_m/V_1 &= var(z_{t+m} - z_t)/var(z_{t+1} - z_t) \\
&= E[(z_{t+m} - z_t)^2]/E[(z_{t+1} - z_t)^2] \\
&= \frac{E(z_{t+m}^2) - 2E(z_{t+m}z_t) + E(z_t^2)}{E(z_{t+1}^2) - 2E(z_{t+1}z_t) + E(z_t^2)} \\
&= \frac{2E(z_{t+m}^2) - 2E(z_{t+m}z_t)}{2E(z_{t+1}^2) - 2E(z_{t+1}z_t)} \\
&= \frac{cov(z_{t+m}, z_t) - var(z_{t+m})}{cov(z_{t+1}, z_t) - var(z_{t+1})} \\
&= (\gamma_m - \gamma_0)/(\gamma_1 - \gamma_0) \\
&= (\frac{\gamma_m}{\gamma_0} - 1)/(\frac{\gamma_1}{\gamma_0} - 1) \\
&= (1 - \rho_m)/(1 - \rho_1)
\end{aligned}$$

Compared with the strict stationarity above, the time series is weakly stationary (second-order stationary) if $E(z_t) = \mu$ is a fixed constant for all t, and autocovariances $cov(z_t, z_{t+k}) = \gamma_k$ only depend on the time lag k for all t. It can be easily seen that weak stationarity plus normality is sufficient to produce strict stationarity.

2.1.3 General Linear Process with its Stationarity and Invertibility

Define $\mathbb{B}$ as the backshift-operator $\mathbb{B}y_t = y_{t-1}$. Similarly, define $\mathbb{F}$ as the forwardshift-operator $\mathbb{F}y_t = y_{t+1}$.

Let $\{a_t\}$ be a white noise series, i.e., a series of independent drawings from some distribution (usually supposed approximately normal) having mean zero and with fixed standard deviation σ_a, then the general linear process can be expressed as the form:

$$z_t = a_t + \sum_{i=1}^{\infty} \psi_i a_{t-i} = (1 + \sum_{i=1}^{\infty} \psi_i \mathbb{B}^i) a_t = \psi(\mathbb{B}) a_t, \tag{2.4}$$

where $\psi(\mathbb{B}) = 1 + \sum_{i=1}^{\infty} \psi_i \mathbb{B}^i = \sum_{i=0}^{\infty} \psi_i \mathbb{B}^i$, with $\psi_0 = 1$.

The general linear process can also be expressed as another form:

$$\begin{aligned}
&z_t = \sum_{i=1}^{\infty} \pi_i z_{t-i} + a_t \Longleftrightarrow \\
&(1 - \sum_{i=1}^{\infty} \pi_i \mathbb{B}^i) z_t = \pi(\mathbb{B}) z_t = a_t.
\end{aligned} \tag{2.5}$$

We can get the relationship $\pi(\mathbb{B}) = \psi^{-1}(\mathbb{B})$.

For a linear process, two very basic concepts are stationarity and invertibility. As we have already explained the meaning of stationarity, the basic idea of invertibility can be illustrated as follows: from the above expression $\pi(\mathbb{B})z_t = a_t$, we can write present event z_t as a linear combination of past happenings $z_{t-1}, z_{t-2}, ...$, as long as the weights π_j are absolutely summable, i.e., $\sum_{i=0}^{\infty} |\pi_i| < \infty$.

For a linear process above $z_t = \psi(\mathbb{B})a_t$, or $\pi(\mathbb{B})z_t = a_t$, it is stationary if $\sum_{i=0}^{\infty} |\psi_i| < \infty$, i.e., the series $\psi(\mathbb{B})$ converges for $|\mathbb{B}| \leq 1$, that is, on or within the unit circle. And it is invertible if $\sum_{i=0}^{\infty} |\pi_i| < \infty$. i.e., the series $\pi(\mathbb{B})$ converges for $|\mathbb{B}| \leq 1$, that is, on or within the unit circle. Invertibility condition is independent of the stationary condition.

2.1.4 Autoregressive Models

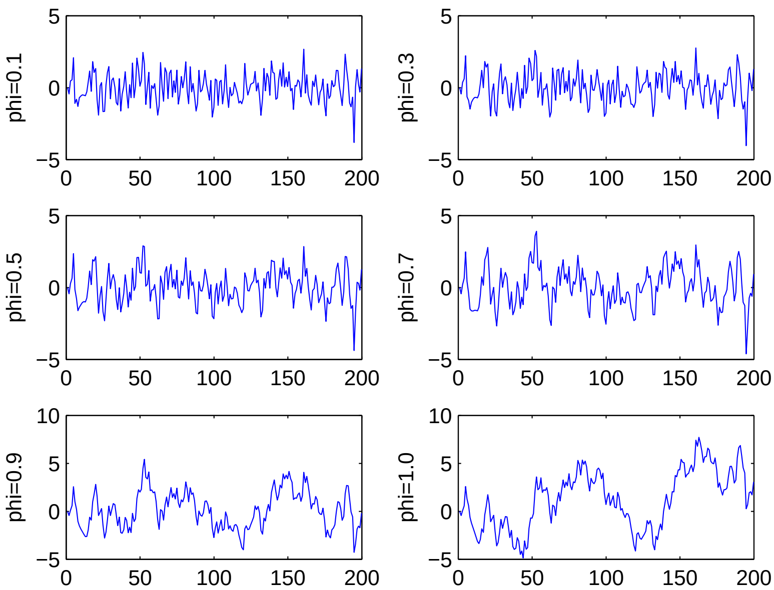

Figure 2.1: Six AR(1) series generated by the same white noise with ϕ=0.1,0.3,0.5,0.7,0.9,1.0.

The AR(p) model is given by $\phi(\mathbb{B})z_t = a_t$, where

$$\phi(\mathbb{B}) = 1 - \phi_1\mathbb{B} - \phi_2\mathbb{B}^2 - ... - \phi_p\mathbb{B}^p. \tag{2.6}$$

The equation $\phi(\mathbb{B}) = 0$ is called the characteristic equation for the process. The stationary condition is, the roots of $\phi(\mathbb{B}) = 0$ must be lie outside the unit circle.

For example, for the AR(1) process

$$\begin{aligned} & z_t = \phi_1 z_{t-1} + a_t, \\ & \Longrightarrow z_t = \phi_1 z_{t-1} + a_t \\ & = \phi_1(\phi_1 z_{t-2} + a_{t-1}) + a_t = ... \\ & = \sum_{j=0}^{\infty} \phi_1^j a_{t-j}, \end{aligned}$$

we must require $|\phi_1| < 1$ to ensure the process z_t has a finite variance so that it is stationary.

AR(1) and AR(2) processes are of considerable practical importance[5]. Sometimes the coefficient function ϕ_1^j in this expansion is called "Green's function" denoted by G_j[32]. So for AR(1), $G_j = \phi_1^j$.

2.1.5 Moving Average Models

The MA(q) model is given by $z_t = \theta(\mathbb{B}) a_t$, where

$$\theta(\mathbb{B}) = 1 - \theta_1 \mathbb{B} - \theta_2 \mathbb{B}^2 - ... - \theta_q \mathbb{B}^q. \tag{2.7}$$

MA(1) and MA(2) processes are of considerable practical importance[5] .

Time series that can be represented by stationary MA models are rather uncommon[9]. And compared with MA model, of greater interest for representing the stationary autocorrelated series are the AR models. But one example that can be modeled by the MA(1) process is the carry-over series[9, pg. 273].

2.1.6 Autoregressive Moving Average (ARMA) Models

The ARMA(p,q) model is given by $\phi(\mathbb{B}) z_t = \theta(\mathbb{B}) a_t$, where

$$\begin{aligned} \phi(\mathbb{B}) &= 1 - \phi_1 \mathbb{B} - \phi_2 \mathbb{B}^2 - ... - \phi_p \mathbb{B}^p, \\ \theta(\mathbb{B}) &= 1 - \theta_1 \mathbb{B} - \theta_2 \mathbb{B}^2 - ... - \theta_q \mathbb{B}^q, \end{aligned} \tag{2.8}$$

as we defined above.

The stationarity condition is, the characteristic function $\phi(\mathbb{B}) = 0$ has all its roots lying outside the unit circle. And the invertibility condition is, the function $\theta(\mathbb{B}) = 0$ has all its roots lying outside the unit circle.

Table 2.1 lists the properties of AR, MA and ARMA models.

Table 2.1: Summary of the properties of AR, MA and ARMA models

characteristics	AR(p) Process	MA(q) Process	ARMA(p,q) Process
In terms of previous $z_t^{\prime}$s	$\phi(\mathbb{B})z_t = a_t$	$\theta^{-1}(\mathbb{B})z_t = a_t$	$\theta^{-1}(\mathbb{B})\phi(\mathbb{B})z_t = a_t$
In terms of previous $a_t^{\prime}$s	$z_t = \phi^{-1}(\mathbb{B})a_t$	$z_t = \theta(\mathbb{B})a_t$	$z_t = \phi^{-1}(\mathbb{B})\theta(\mathbb{B})a_t$
π weights	Finite series	Infinite series	Infinite series
ψ weights	Infinite series	Finite series	Infinite series
Stationarity condition	Roots of $\phi(\mathbb{B}) = 0$ lie outside unit circle	Always stationary	Roots of $\phi(\mathbb{B})$=0 lie outside unit circle
Invertibility condition	Always invertible	Roots of $\theta(\mathbb{B}) = 0$ lie outside unit circle	Roots of $\theta(\mathbb{B})$=0 lie outside unit circle
Autocorrelation function	Infinite Tails off	Finite Cuts off after lag q	Infinite Tails off
Partial autocorrelation function	Finite Cuts off after lag p	Infinite Tails off	Infinite Tails off

2.1.7 Autoregressive Integrated Moving Average (ARIMA) Models

Most naturally occurring series, as well as the uncontrolled process disturbances cannot be adequately represented by stationary models. However, although the level $\{z_t\}$ is not stationary, it is frequent true that its rate of change, the first difference $\{z_t - z_{t-1}\}$ is stationary. Also, on very rare occasions, if the first difference is still nonstationary, it second difference, i.e., the difference of the differences,

$$\{(z_t - z_{t-1}) - (z_{t-1} - z_{t-2})\} = \{z_t - 2z_{t-1} + z_{t-2}\}$$

is stationary.

ARIMA(p,d,q) model can be expressed as:

$$\varphi(\mathbb{B})z_t = \phi(\mathbb{B})\nabla^d z_t = \theta(\mathbb{B})a_t,$$

where $\nabla = 1 - \mathbb{B}$, and $\nabla^d = \nabla^{d-1}\mathbb{B}$.

That is, after differencing d times, the ARIMA(p,d,q) model turns to be an ARMA(p,q) model. In practice, d is rarely great than 1[9, pg. 279].

2.1.8 Integrated Moving Average Model

Integrated moving average (IMA) model, i.e., ARIMA(0,1,1) model,

$$z_{t+1} - z_t = a_{t+1} - \theta a_t, \tag{2.9}$$

is considered to be the most important model for the nonstationary disturbance. Let $\lambda = 1 - \theta$, when $\lambda = 0(\theta = 1.0)$, the disturbance z_t is a white noise, which is typical for a process in a perfect state of control. Larger values of λ produce disturbances that display an increasing degree of instability, see Figures 2.2. They closely mimic the behavior of many industrial processes when no control is applied. When $\lambda = 1.0(\theta = 0)$, $z_{t+1} = z_t + a_{t+1}$, which is a random walk. The industrial time series usually have much smaller values of λ, usually $0.2 \leq \lambda \leq 0.4$ [9].

At times $t, t-1, ..., 1$, the IMA model is

$$\begin{aligned} z_t - z_{t-1} &= a_t - \theta a_{t-1} \\ z_{t-1} - z_{t-2} &= a_{t-1} - \theta a_{t-2} \\ &\dots \\ z_2 - z_1 &= a_2 - \theta a_1 \\ z_1 &= a_1 \end{aligned}$$

By adding those t equations with appropriate cancelation, we get

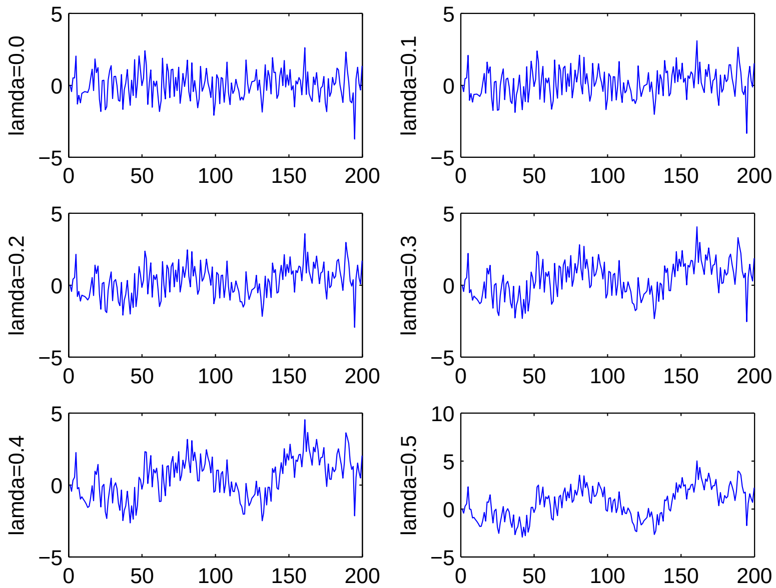

Figure 2.2: Six IMA series generated by the same white noise with θ=1.0,0.9,0.8,0.7,0.6,0.5.

$$
\begin{aligned}
z_t &= a_t + \lambda \sum_{i=1}^{t-1} a_i, \Longleftrightarrow \\
z_t &= \theta a_t + \lambda \sum_{i=1}^{t} a_i.
\end{aligned}
\tag{2.10}
$$

where $\lambda = 1 - \theta$, so z_t is a mixture of the current random shock a_t and the sum $\sum_{i=1}^{t-1} a_i$ of previous shocks. Thus it shows that the IMA model is equivalent to a random walk with added independent noise.

2.2 Box's Feedback Control

2.2.1 Basic Concepts and Notations

First we introduce some notations in Box's process adjustment strategies:

T: the target value.

y_t: the output value at time t.

a_t: white noise series, i.e., a series of independent and identically distributed (iid) random variables with mean 0 and standard deviation σ_a.

δ_t: periodical step change at time t.

z_t: disturbance at time t—the deviation from some target value T that would occur if no attempt at control were made.

$\hat{z}_t$: an estimate of z_t.

$\tilde{z}_t$: an EWMA forecast of z_t based on data up to and including time t.

e_t: forecast error at time t, $e_t = z_t - \hat{z}_t = z_t - \tilde{z}_{t-1}$ /output error $e_t = y_t - T$, the deviation from the target value T after the adjustment at time t. To be proved: forecast error=output error, i.e., $e_t = z_t - \tilde{z}_{t-1} = y_t - T$.

X: compensatory variable, or adjustment variable.

X_t: the level of the compensatory variable at time t.

$x_t = X_t - X_{t-1} = \nabla X_t$: adjustment—the change made in the level of the compensating variable X_t, at time t from the previous level X_{t-1}.

G: damping factor, its value can be changed by the operators.

g: system gain, the eventual change in the output that is induced by a unit adjustment at the input. Its value can be changed through system designs.

$\mathbb{B}$: common backshift-operator defined as $\mathbb{B}y_t = y_{t-1}$.

• Proof of e_t=output error=forecast error, i.e., $e_t = y_t - T = z_t - \tilde{z}_{t-1}$:

Proof: Since after adjustment $e_t = y_t - T$, and without control $z_t = y_t' - T$, where y_t' is the output value at time t if there is no control. Then we have

$e_t - z_t = y_t - y_t' \Longrightarrow$

$e_t = z_t + y_t - y_t' = z_t - (y_t' - y_t)$.

So to prove that e_t also satisfies $e_t = z_t - \tilde{z}_{t-1}$, it is equivalent to prove

$y_t' - y_t = \tilde{z}_{t-1}$.

Since $y_t' - y_t = \tilde{z}_{t-1}$

$\Longleftrightarrow (y_t' - T) - (y_t - T) = \tilde{z}_{t-1}$

$\Longleftrightarrow z_t - e_t = \tilde{z}_{t-1}$

$\Longleftrightarrow z_t - \tilde{z}_{t-1} = e_t$

The last equation is true from the definition of forecast error $e_t = z_t - \tilde{z}_{t-1}$.

Thus $e_t = y_t - T = z_t - \tilde{z}_{t-1}$ has been proved.

2.2.2 Minimum Mean Square Error

Like Taguchi's Beta coefficient method and Grubbs' harmonic rule, we assume that after adjustment the process mean and the target value coincides, i.e., $\mu_y = T$, so $E(e_t) = 0$, then we have:

$$\sigma_e^2 = E(e_t^2) - [E(e_t)]^2 = E[(y_t - T)^2] = E[(y_t - \mu_y)^2] = \sigma_y^2,$$

so we have $MSE \Longleftrightarrow \sigma_y^2$.

So one optimal criteria is minimum mean square error (MMSE): long-run average of squares of the e_t values is the smallest possible.

When there is only off-target cost, we only need to consider the output variance σ_y^2. However, in many cases there is also the adjustment cost, i.e., the larger the adjustment x_t we made, the larger the cost is, so in our adjustment, we need to balance σ_e^2 and σ_x^2. [9] proposed an optimal linear scheme that minimizes $\sigma_x^2 + \alpha\sigma_e^2$, called constrained adjustment.

2.2.3 Responsive System: Feedback Control Equation and PI Control

In this thesis we only consider the control of discrete systems, where measurements and (or) adjustments are made at equispaced intervals of time $t, t-1, t-2...$

Assume it is a responsive system, i.e., all the effect of a change x_t in the adjustment variable X_t will be realized at the output y_t in one time interval. The basic idea of feedback control is making an adjustment roughly proportional to the current output error $e_t = y_t - T$, so the control equation is $gx_t = -Ge_t$, where G is the proportion of the full adjustment that is put into effect.

The basic procedure for the feedback control is:

(1) At time t, we measure the output deviation e_t.

(2) Based on e_t, we use the feedback control equation $gx_t = -Ge_t$.

(3) Since it is a responsive system, so after we made the feedback control x_t at time t, its effect will be realized on y_{t+1} after one time interval. So we measure the output deviation e_{t+1}, and iterate the procedure (1) and (2) repeatedly.

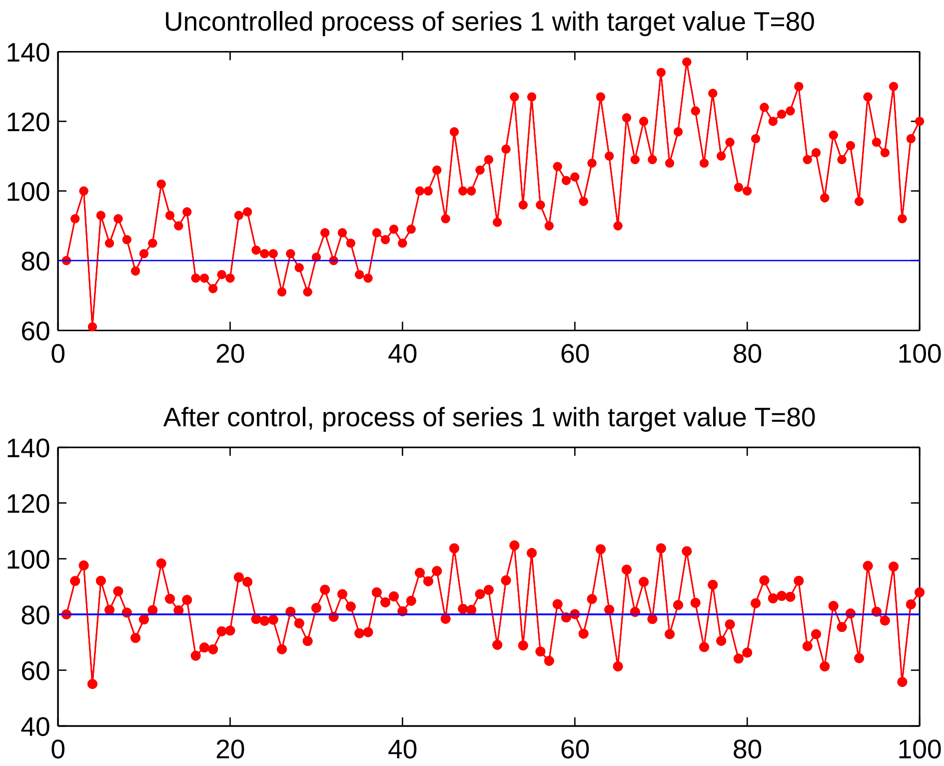

Figure 2.3: Series 1 without control and after control

The top part of Figure 2.3 shows 100 observations of a thickness measurement of a very thin metallic film, when no control was applied[9]. We desire to maintain this quality characteristic as closely as possible to the target value T=80. The low part shows the controlled series after feedback control was made. By comparing these two figures, the control is seen to be very effective.

Since the control equation can be written as

$$\begin{aligned} & gx_t = -Ge_t \\ \Longrightarrow\quad & g(X_t - X_{t-1}) = -Ge_t \\ \Longrightarrow\quad & gX_t = gX_0 - G\sum_{i=1}^{t} e_i \\ \Longrightarrow\quad & gX_t = k_0 + k_I \sum_{i=1}^{t} e_i, \end{aligned}$$

where $k_0 = gX_0, k_I = -G$, so it is a form of proportional integral (PI) control.

2.2.4 EWMA Forecast, IMA Model and Feedback Control

EWMA can be used for both process monitoring and forecasting. The use of exponential smoothing for forecasting is to give most weight to the last observation and less to the next-but-last and so on. Thus the general idea is that, given data up to and including time t, which is called the forecast origin, we can use the EWMA $\tilde{z}_t$ to provide an estimate $\hat{z}_{t+1}$ of the next value z_{t+1}:

$\tilde{z}_t = \hat{z}_{t+1} = (1-\theta)(z_t + \theta z_{t-1} + \theta^2 z_{t-2} + ...)$

$\Longleftrightarrow \tilde{z}_t = \lambda z_t + \theta \tilde{z}_{t-1}$, where $\lambda = 1 - \theta$,

$\Longleftrightarrow \hat{z}_{t+1} = \lambda z_t + \theta \hat{z}_t$.

So as soon as the actual value z_t becomes available at time t, we can interpolate between z_t and $\hat{z}_t$ to get $\hat{z}_{t+1}$. Figure 2.4 shows an example of EWMA forecast with θ=0.6.

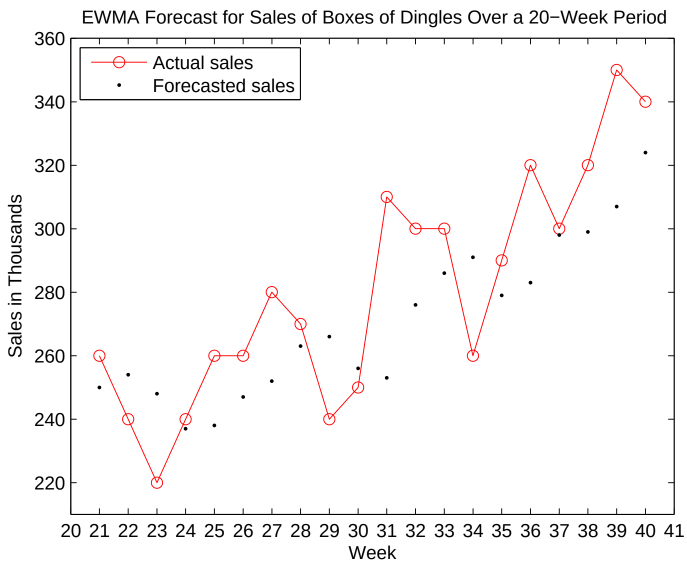

Figure 2.4: EWMA forecast for the sales of boxes from week 21 to week 41 using θ=0.6

There are three important relationships for EWMA forecast:

(1) $\tilde{z}_t = \lambda z_t + \theta \tilde{z}_{t-1}$ or $\hat{z}_{t+1} = \lambda z_t + \theta \hat{z}_t$.

(2) $\tilde{z}_t - \tilde{z}_{t-1} = \lambda(z_t - \tilde{z}_{t-1})$ or $\hat{z}_{t+1} - \hat{z}_t = \lambda e_t$, where $e_t = z_t - \tilde{z}_{t-1}$. Interpolate between z_t and $\tilde{z}_{t-1}$ to get $\tilde{z}_t$.

(3) $z_t - z_{t-1} = e_t - \theta e_{t-1}$.

From equation (3) we know that for IMA model, we have $e_t = a_t$, i.e., the optimal forecast, therefore EWMA is a good forecast for IMA model[27].

When the disturbance is not an IMA model, which forecast should we use? Box and Luceño showed the robustness of the EWMA forecast[9, pg. 122].

Now we show the equivalence of EWMA forecast and feedback control. Since at time t, if we can estimate the level of the disturbance one step ahead z_{t+1} by $\hat{z}_{t+1}$, then we could achieve perfect control by setting X_t so that at time t it cancels out the predicted value $\hat{z}_{t+1}$:

$gX_t = -\hat{z}_{t+1} \Longrightarrow gx_t = -(\hat{z}_{t+1} - \hat{z}_t)$,

And since $gx_t = -Ge_t = -G(z_t - \hat{z}_t)$, so we have

$$\begin{aligned} &\hat{z}_{t+1} - \hat{z}_t = G(z_t - \hat{z}_t) \Longrightarrow \\ &\hat{z}_{t+1} = Gz_t + (1-G)\hat{z}_t. \end{aligned}$$

This is a first-order difference equation. From equation (1) in the three important relationships for EWMA forecast, if we set $\theta = 1 - G$, it has the solution

$$\hat{z}_{t+1} = \tilde{z}_t = (1-\theta)(z_t + \theta z_{t-1} + \theta^2 z_{t-2} + ...). \tag{2.11}$$

Then the equation $gX_t = -\hat{z}_{t+1}$ becomes $gX_t = -\tilde{z}_t$. This proves the equivalence of $gx_t = -Ge_t$ and $gX_t = -\hat{z}_{t+1}$, i.e., the adjustment made which is proportional to the last error, turns out to be equivalent to setting gX_t at each stage so as to cancel an EWMA forecast of the disturbance with smoothing constant $\theta = 1 - G$.

From $gX_{t-1} = -\hat{z}_t$ and $e_t = z_t - \hat{z}_t$, we have

$$e_t = z_t + gX_{t-1}.$$

Since z_t is the disturbance, and gX_{t-1} is the total magnitude of feedback control, so the output error e_t is their sum. So we have three equivalent forms for the feedback control equations:

(1) $gx_t = -Ge_t$.

(2) $gX_t = -\hat{z}_{t+1}$.

(3) $e_t = z_t + gX_{t-1}$.

They are valid only for IMA disturbance model in a responsive system. For a general process with inertia, we have similar feedback control equations, which will be discussed later.

2.2.5 Inflation of the Output Variance

So far we assume that the disturbance z_t is an IMA model with smoothing constant θ_0:

$$z_t - z_{t-1} = a_t - \theta_0 a_{t-1}, |\theta_0| < 1.$$

However, in practice, we do not know the true value θ_0 of θ. Let $\lambda_0 = 1 - \theta_0$. Suppose the feedback control applied here uses $G \neq \lambda_0$, then the forecast errors $\{e_t\}_{t\geq 1}$ resulting from the feedback control using $\theta = 1 - G$ instead of θ_0 is

$$z_t - z_{t-1} = e_t - \theta e_{t-1}, |\theta| < 1.$$

From the above two equations we have:

$$e_t - \theta e_{t-1} = a_t - \theta_0 a_{t-1}.$$

This is an ARMA (1,1) model for e_t values, and we have the variance inflation

$$\frac{\sigma_e^2}{\sigma_a^2} = \frac{1 - 2\theta\theta_0 + \theta_0^2}{1 - \theta^2} = 1 + \frac{(\theta_0 - \theta)^2}{1 - \theta^2} = 1 + \frac{(G - \lambda_0)^2}{G(2 - G)}. \quad (2.12)$$

Without loss of generality, we usually assume $\sigma_a^2 = 1$.

We can reduce the value of G to a "suboptimal" value less than λ_0 to reduce the input variance while only mildly increase σ_e^2, see Figure 2.5 [9, pg. 142]. It can be seen that to reduce σ_x we need to cut back on the value of the damping factor G. When a small increase in σ_e is tolerated in order to achieve a large reduction in σ_x, the adjustment schemes are called constrained adjustment schemes.

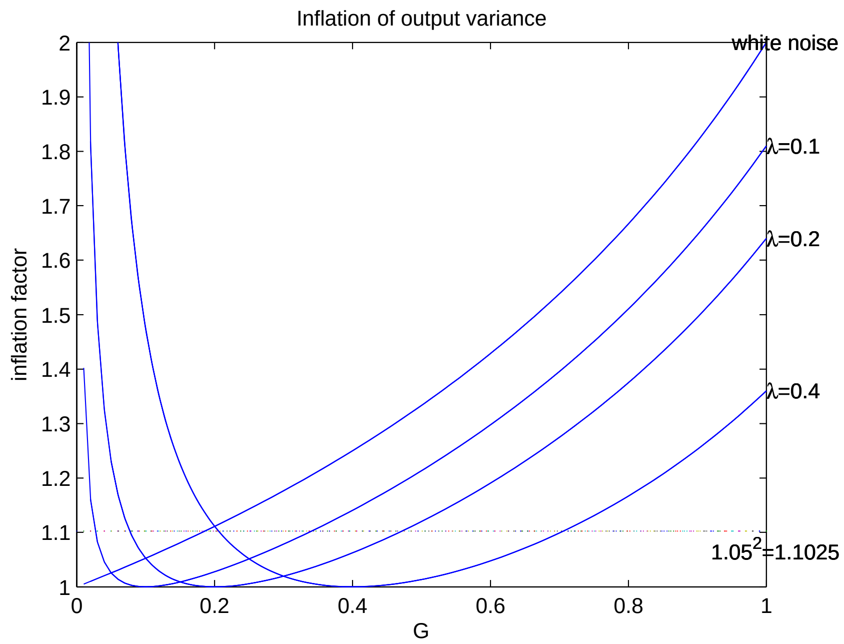

Figure 2.5: Inflation of process variance with different choice of G

2.3 Box's Feedback Control For Process With Inertia

2.3.1 First-order Dynamic System and EWMA

In some processes it may take some time before the full effect of adjustment is experienced at the process output, or there are one or two intervals of pure delay in the system response. We model it by a first-order dynamic system:

$$Y_{t+1} - Y_t = (1-\delta)(X_t - Y_t) \Longleftrightarrow$$
$$Y_{t+1} = (1-\delta)X_t + \delta Y_t, 0 \leq \delta < 1.$$

For a step change in the input X, the proportion of the total change that occurs in the first time interval is $1-\delta$. The above first-order dynamic system can be generalized to be

$$Y_t = \delta Y_{t-1} + g(1-\delta)X_{t-1} + C, 0 \leq \delta < 1. \tag{2.13}$$

where C is a constant.

When a unit step change is made in X, after t time periods, the change in Y will be $g(1-\delta^t)$. Thus for this dynamic model the output change asymptotically approaches g units, the system gain, as $t \to \infty$.

A further simplification of dynamics is that, all of the changed induced by a step change in X will occur in a single time interval, which corresponds to $\delta = 0$, so the dynamic model becomes

$$Y_t = gX_{t-1} + C.$$

where C is a constant. This is called the pure-gain model (responsive system), i.e., a process without inertia.

Process inertia can also be represented by EWMA. From $Y_{t+1} = (1-\delta)X_t + \delta Y_t$, we can get the EWMA form by iteration:

$$Y_{t+1} = (1-\delta)(X_t + \delta X_{t-1} + \delta^2 X_{t-2} + ...),$$

so the system equation can be written as the forecast form by $Y_{t+1} = \tilde{X}_t$, where $\tilde{X}_t$ is an EWMA with smoothing parameter δ, which is a measure of the process dynamics (inertia), with a small value of δ corresponds to a rapidly responding system, and a large value of δ corresponds to a slowly responding system. Figure 2.6 shows a process with $\delta = 0.5$.

Compared with $e_t = z_t + gX_{t-1}$ for the responsive system, and the output errors of the first-order dynamic system $Y_t - \delta Y_{t-1} = g(1-\delta)X_{t-1}$ satisfies[10]:

$$\begin{aligned} e_t - \delta e_{t-1} &= z_t - \delta z_{t-1} + g(1-\delta)X_{t-1} \\ \Longleftrightarrow e_t &= z_t + \tfrac{g(1-\delta)}{1-\delta\mathbb{B}} X_{t-1}. \end{aligned} \tag{2.14}$$

• Proof: Since

$$\begin{cases} e_t - \delta e_{t-1} = z_t - \delta z_{t-1} + g(1-\delta)X_{t-1} \\ Y_t - \delta Y_{t-1} = g(1-\delta)X_{t-1} \end{cases}$$

$$\begin{aligned} &\Longleftrightarrow \quad e_t - \delta e_{t-1} = z_t - \delta z_{t-1} + Y_t - \delta Y_{t-1}, \\ &\Longleftrightarrow \quad e_t = z_t + Y_t, \end{aligned}$$

which is always true[9, pg. 184].

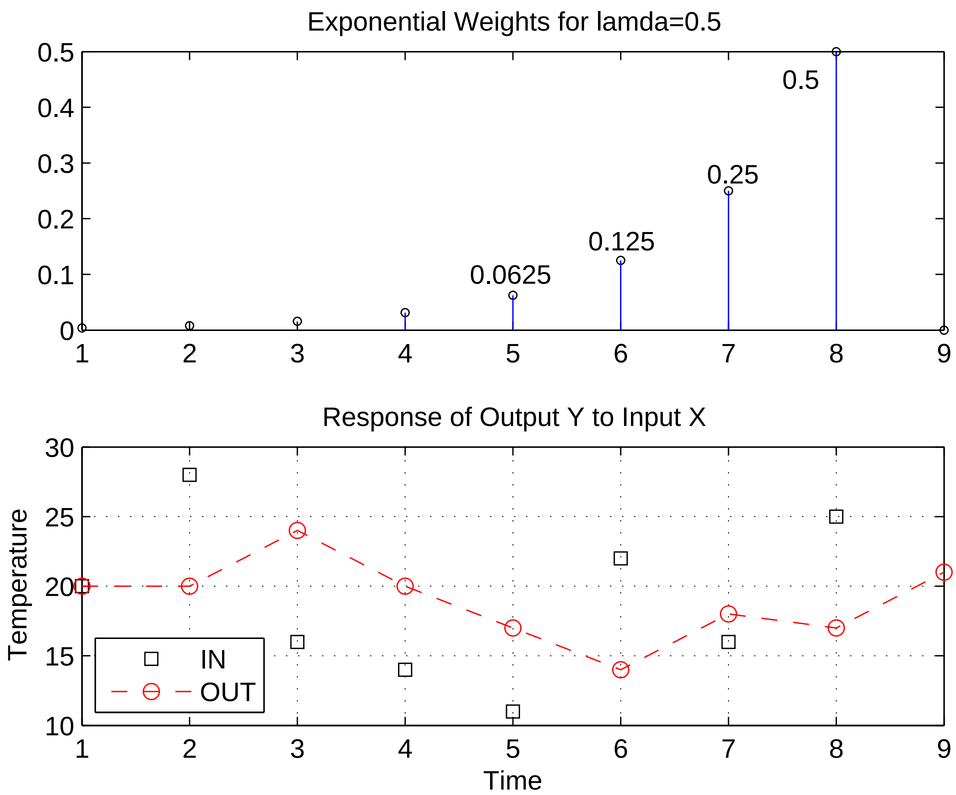

Figure 2.6: Input and output of a dynamic system

2.3.2 Feedback Control, PI Control and EWMA Forecast

As we point out, if a process is not a responsive system and there is process inertia, it takes more than one time interval for the full effect of the adjustment in the input realized at the output. And for most practical cases, $0 \leq \delta \leq 0.5$, so most effect of the the adjustment in the input can be realized at the output within two time intervals, and the effect of the the adjustment in the input after two time intervals is so small that it can be ignored. Thus Box and Luceño also considered "the slightly more adventurous possibility of making the adjustment x_t depend on the last two errors e_t and e_{t-1}", with the control equation [9]

$$gx_t = c_1 e_t + c_2 e_{t-1}. \tag{2.15}$$

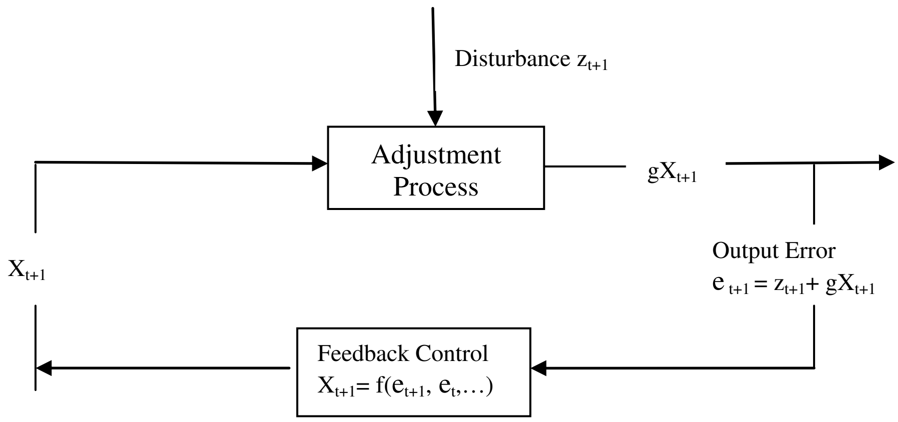

Figure 2.7: A flow diagram for a process subject to feedback control at time $t+1$

From the control equation $gx_t = c_1 e_t + c_2 e_{t-1}$, we can get its PI control form:

$$gX_t = k_0 + k_P + k_I \sum_{i=1}^{t} e_i,$$

where $k_0 = 0, k_p = -c_2, k_I = c_1 + c_2$.

Since

$$\begin{aligned} gx_t &= c_1 e_t + c_2 e_{t-1} \\ &= -[-(c_1 + c_2)][e_t - \tfrac{c_2}{c_1+c_2}(e_t - e_{t-1})] \\ &= -G[e_t + P(e_t - e_{t-1})], \end{aligned}$$

where $G = -(c_1 + c_2), P = -c_2/(c_1 + c_2)$, so the PI control equation can be written as

$$gx_t = -G[e_t + P(e_t - e_{t-1})], \tag{2.16}$$

We have $c_1 \leq 0, G \geq 0, P \in R$. It is obvious that when $P > 0$ we extrapolate a line joining the last two points of e_t and e_{t-1} by P time intervals, and when $P < 0$ we interpolate between e_t and e_{t-1}. We will show later that when we interpolate by using $P < 0$, the adjustment process is more stable, and it requires less adjustment in the input variable.

We can also write the control equation in another form. Set $\delta = -c_2/c_1$, then since $G = -(c_1 + c_2)$, we have $c_1 = (c_1 + c_2)/(1 + c_2/c_1) = -G/(1 - \delta)$. Then we have

$$\begin{aligned} gx_t &= c_1 e_t + c_2 e_{t-1} = c_1(e_t + \tfrac{c_2}{c_1} e_{t-1}) = -\tfrac{G}{1-\delta}(e_t - \delta e_{t-1}), \\ \Longleftrightarrow gX_t &= gX_0 - \tfrac{\delta G}{1-\delta} - G \sum_{i=1}^{t} e_i. \end{aligned}$$

This PI control is equivalent to setting $g\tilde{X}_t = -\tilde{z}_t$, where $\tilde{X}_t$ is an EWMA of X_t, X_{t-1},... with smoothing constant $\delta = -c_2/c_1$ and $\tilde{z}_t$ is an EWMA of z_t, z_{t-1},..., with smoothing constant $1 - G$.

- Proof of the equivalence of $gx_t = -\frac{G}{1-\delta}(e_t - \delta e_{t-1})$ and $g\tilde{X}_t = -\tilde{z}_t$:

From $g\tilde{X}_t = -\tilde{z}_t$ we have:

$$g\tilde{x}_t = g(\tilde{X}_t - \tilde{X}_{t-1}) = -(\tilde{z}_t - \tilde{z}_{t-1}),$$

where $\tilde{x}_t = (1 - \delta)(x_t + \delta x_t + \delta^2 x_t + ...)$, i.e., it is an EWMA of current and previous adjustments x_t, x_t,... with smoothing constant δ.

From $\tilde{z}_t - \tilde{z}_{t-1} = \lambda e_t$, i.e., equation (2) of the three equivalent relationships for the EWMA forecast in section 2.4.4, we have

$$g\tilde{x}_t = -(\tilde{z}_t - \tilde{z}_{t-1}) = -\lambda e_t.$$

By setting $G = \lambda$, we have $g\tilde{x}_t = -Ge_t$. Then we have

$$g(\tilde{x}_t - \delta\tilde{x}_{t-1}) = -G(e_t - \delta e_{t-1}).$$

Since $\tilde{X}_t$ is an EWMA of X_t, X_{t-1},... with smoothing constant δ, i.e.,

$$\tilde{X}_t = (1 - \delta)X_t + \delta\tilde{X}_{t-1},$$

we have

$$\tilde{x}_t = (1 - \delta)x_t + \delta\tilde{x}_{t-1},$$

so we have

$$\tilde{x}_t - \delta\tilde{x}_{t-1} = (1 - \delta)x_t,$$

Combining it with $g(\tilde{x}_t - \delta\tilde{x}_{t-1}) = -G(e_t - \delta e_{t-1})$, we get

$$g(1 - \delta)\tilde{x}_t = -G(e_t - \delta e_{t-1}),$$

so that

$$gx_t = -\frac{G}{1 - \delta}(e_t - \delta e_{t-1}).$$

2.3.3 MMSE Control and PI Control with Constrained Adjustment

We have the assumption on the process inertia and the disturbance model:

1. The adjustment process inertia is represented by

$$X_{t+1} = \tilde{X}_t = (1-\delta_0)(X_t + \delta_0 X_{t-1} + \delta_0^2 X_{t-2} + ...).$$

2. The process disturbance is an IMA model

$$z_t - z_{t-1} = a_t - \theta_0 a_{t-1},$$

and $\lambda_0 = 1 - \theta_0$.

Then it is natural that the MMSE control is $g\tilde{X}_t = -\tilde{z}_t$, where $\tilde{X}_t$ is an EWMA of X_t, X_{t-1},... with smoothing constant δ_0 and $\tilde{z}_t$ is an EWMA of z_t, z_{t-1},..., with smoothing constant θ_0.

The three control equations are

(1) $gx_t = c_1 e_t + c_2 e_{t-1}$, with $c_1 = -\frac{\lambda_0}{1-\delta_0}$.

(2) $gX_t = k_0 + k_P + k_I \sum_{i=1}^{t} e_i$, with $k_p = -\frac{\delta_0 \lambda_0}{1-\delta_0}$ and $k_I = -\lambda_0$.

(3) $gx_t = -G[e_t + P(e_t - e_{t-1})]$, with $G = \lambda_0$ and $P = \frac{\delta_0}{1-\delta_0}$.

It is well known that MMSE control is not necessarily the best choice, since it often requires excessive manipulation [2], [9], [13]. MMSE control may not be of much practical interest unless the process is only slightly nonstationary (λ_0 is small) and has rapid dynamic response (δ_0 is not far from 0) [9].

Since P is the degree of extrapolation along a line joining the last two points e_t and e_{t-1}, so when δ_0 is larger, i.e., for slower dynamics, $P = \frac{\delta_0}{1-\delta_0}$ is larger, which requires greater extrapolation, i.e., greater manipulation on the adjustment variable.

To avoid the excessive manipulation on the adjustment variable in the MMSE control, we can use the PI control with constrained adjustment. From numerical results, its performance on the output and input variances are almost as good as the optimal control which usually is not easy to derive and has the complicated form [9].

The PI control with constrained adjustment $gx_t = -G[e_t + P(e_t - e_{t-1})]$ usually has $P < 0$, i.e., we interpolate between e_t and e_{t-1}, so it is a more stable procedure and of course has smaller input variance. In practice, the particularly useful range of P for constrained PI scheme is $-0.25 \leq P \leq 0$[9, pg. 180]. From $\tilde{x}_t = (1-\delta)(x_t + \delta x_{t-1} + \delta^2 x_{t-2} + ...)$, the weights of $x_{t-2}, x_{t-3}...$ are becoming so small that they can be ignored, thus

$$\begin{aligned} & g\tilde{x}_t = g(1-\delta)(x_t + \delta x_{t-1} + \delta^2 x_{t-2} + ...) = -Ge_t \\ \Longrightarrow \quad & g(1-\delta)(x_t + \delta x_{t-1}) = -Ge_t \\ \Longleftrightarrow \quad & x_t = -\delta x_{t-1} - \tfrac{G}{g(1-\delta)} e_t. \end{aligned}$$

For the detailed discussions of optimal choices of G and P, see[9].

While the MMSE forecast for IMA is EWMA forecast $\hat{z}_{t+1} = (1-\theta)z_t + \theta \hat{z}_t$, the MMSE forecast for ARMA(1,1) is $\hat{z}_{t+1} = (\phi - \theta)z_t + \theta \hat{z}_t$. But since

$$\hat{z}_{t+1} = (\phi - \theta)(z_t + \theta z_{t-1} + \theta^2 z_{t-2} + ...),$$

so the sum of weights is $(\phi - \theta)/(1-\theta) < 1$ when $\phi < 1$, thus it is not an average of the past data[23].

2.3.4 *Output Error Relationship for First-Order Dynamics Process Under Feedback Control*

If the disturbance is IMA model, and the process is first-order dynamic system represented by Equation (2.2), $Y_t = \delta Y_{t-1} + g(1-\delta)X_{t-1}$, where $0 \leq \delta < 1$, under the feedback control equation $gx_t = -Ge_t$, the output error e_t satisfying[9, pg. 184]

$$e_t - [1 + \delta - G(1-\delta)]e_{t-1} + \delta e_{t-2} = a_t - (\delta + \theta)a_{t-1} + \delta\theta a_{t-2}. \tag{2.17}$$

• Proof:

$$\begin{aligned} & e_t - \delta e_{t-1} = z_t - \delta z_{t-1} + g(1-\delta)X_{t-1} (e.q.(2.3)) \\ \Longrightarrow \quad & (e_t - e_{t-1}) - \delta(e_{t-1} - e_{t-2}) = (z_t - z_{t-1}) - \delta(z_{t-1} - z_{t-2}) + g(1-\delta)x_{t-1} \\ \Longrightarrow \quad & (e_t - e_{t-1}) - \delta(e_{t-1} - e_{t-2}) = (z_t - z_{t-1}) - \delta(z_{t-1} - z_{t-2}) - G(1-\delta)e_{t-1} \\ & z_t - z_{t-1} = a_t - \theta a_{t-1}, \\ \Longrightarrow \quad & e_t - [1 + \delta - G(1-\delta)]e_{t-1} + \delta e_{t-2} = (a_t - \theta a_{t-1}) - \delta(a_{t-1} - \theta a_{t-2}) \\ \Longrightarrow \quad & e_t - [1 + \delta - G(1-\delta)]e_{t-1} + \delta e_{t-2} = a_t - (\delta + \theta)a_{t-1} + \delta\theta a_{t-2}. \end{aligned}$$

Similarly, under the feedback control equation $gx_t = -G[e_t + P(e_t - e_{t-1})]$, we can prove the output error e_t satisfying

$$e_t - [1 + \delta - G(1 + P)(1 - \delta)]e_{t-1} + [\delta - (1 - \delta)GP]e_{t-2} = a_t - (\delta + \theta)a_{t-1} + \delta\theta a_{t-2}.$$

2.4 Box's Feedforward Control by Transfer Function Models

Box *et.al.* discussed the feedforward control for a system with inertia that minimizes the mean square error at the output, by using the transfer function model[5]. Suppose y_t is the output, u_t is the observed but uncontrollable input disturbance, and X_t is the input adjustment variable, and N_t is the total effect in the output of all other sources of disturbance, assume that

$$\begin{aligned} y_{1t} &= \delta^{-1}(\mathbb{B})\omega(\mathbb{B})\mathbb{B}^b u_t, \\ y_{2t} &= L_1^{-1}(\mathbb{B})L_2(\mathbb{B})\mathbb{B}^{f+1} X_t. \end{aligned}$$

Then we have, the total effect of the input disturbance u is

$$\delta^{-1}(\mathbb{B})\omega(\mathbb{B})u_{t-b},$$

and the the total effect of the adjustment variable X is

$$L_1^{-1}(\mathbb{B})L_2(\mathbb{B})X_{t-f-1}.$$

Also assume that the effects of the input variables u and X on the output y are addictive, then the effect of the observed input disturbance u will be canceled if we set

$$L_1^{-1}(\mathbb{B})L_2(\mathbb{B})X_{t-f-1} = -\delta^{-1}(\mathbb{B})\omega(\mathbb{B})u_{t-b},$$

thus the control should be such that

$$L_1^{-1}(\mathbb{B})L_2(\mathbb{B})X_t = -\delta^{-1}(\mathbb{B})\omega(\mathbb{B})u_{t-(b-f-1)}.$$

Case 1: $b \geq f + 1$. Let $x_t = X_t - X_{t-1}$, then MMSE feedforward control is

$$\begin{aligned} X_t &= -\tfrac{L_1(\mathbb{B})\omega(\mathbb{B})}{L_2(\mathbb{B})\delta(\mathbb{B})} u_{t-(b-f-1)} \Longleftrightarrow \\ x_t &= -\tfrac{L_1(\mathbb{B})\omega(\mathbb{B})}{L_2(\mathbb{B})\delta(\mathbb{B})}(u_{t-(b-f-1)} - u_{t-1-(b-f-1)}). \end{aligned}$$

Case 2: $b < f + 1$. The MMSE feedforward control is

$$X_t = -\frac{L_1(\mathbb{B})}{L_2(\mathbb{B})}\hat{u}'_t(f+1-b) \Longleftrightarrow$$
$$x_t = -\frac{L_1(\mathbb{B})}{L_2(\mathbb{B})}(\hat{u}'_t(f+1-b) - \hat{u}'_t(f+1-b)),$$

where the needed forecasts $\hat{u}'_t(f+1-b)$ can be obtained by using the forecasting methods in Chapter 5 of [5].

2.5 Model Identification for Disturbance and Process Dynamics

A generic model building procedure is Box-Jenkin iterative approach[5]:

Identification

$\Longrightarrow$Fitting (Estimation)

$\Longrightarrow$Diagnostic Checking

$\Longrightarrow$Then goes back to the identification phase again to iterate.

Here we only consider some useful graphical tools for model identification, which is the first stage in model building. In this stage, no precise formulation of the problem is available, and graphical methods are particular useful. For detailed information in other stages, please refer to Chapters 6,7,8 of [5].

2.5.1 Autocorrelation Function for Model Identification

Table 2.3 summarizes how sample ACF can be used for model identification of stationary processes.

sometimes we only have the data, but do not know which time series model it follows, so we need some tool for the model identification. One very important tool for model identification of the time series is the sample ACF.

2.5.2 Variograms for Different Models

Variogram is a very important and efficient tool for identifying different models, both stationary and nonstationary, as well as the ACF and PACF. Figure 2.8 shows the variograms of both stationary and nonstationary time series models, including white noise, AR(1), and IMA models with different values of λ.

Table 2.2: Shape of autocorrelation functions

Shape	Indicated model
Exponential decay to zero	AR—use the PACF plot to identify the order p
Alternating positive, negative decay to zero	AR—use the PACF plot to identify the order p
One or more spikes, rest are essentially zero	MA—order identified by where ACF becomes zero
Decay, starting after a few lags	ARMA model
All zero or close to zero	Data is essentially random
High values at fixed intervals	Include seasonal autoregressive term
No decay to zero	Series is not stationary

From Figure 2.8 we can see that:

(1) For any stationary model, the variogram eventually flattens out (in mathematical language, it approaches an asymptote)[9].

(2) For all these models whose first difference is stationary, the variogram may show some curvature initially but eventually increases linearly without limit.

(3) For ARIMA (p,2,q) whose first difference is not stationary but second difference is a stationary ARMA(p,q) model, the variogram increases as the square of m irrespective of the values of the parameters, and also the rate of increase of the variogram increases without limit. So such models do not appear particular useful for representing a process disturbance because they imply that the effect of the “sticky innovations” gets larger and larger as m increases[6], [9, pg. 117].

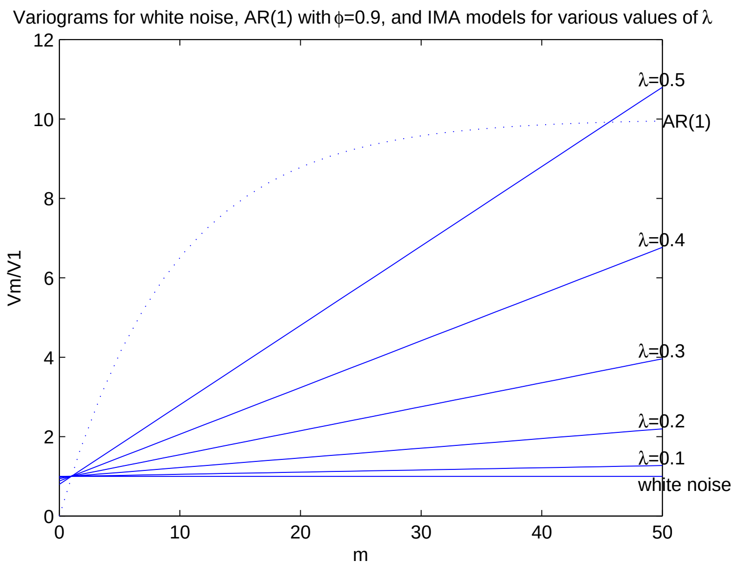

Figure 2.8: Variograms of some time series models

Chapter 3

PERIODIC SHIFT DISTURBANCE MODELS

In chapter 3, I will introduce the first type of disturbance models that are under study by this thesis: the periodic shift disturbance model. First I will give the motivations for such periodic shift disturbance model, then I will derive the control equations for each model, after that I will derive the closed-form MSEO for each disturbance model under different control strategies, including feedback control, feedforward control and combined control. Following parts are the performance analysis, sensitivity analysis and robust analysis, and those issues are evaluated in detail.

3.1 Motivations of the Periodic Shift Disturbance Models

In many real applications, the true disturbance is not an IMA process itself (also true for AR(1) or ARMA(1,1) process), instead, it often includes the IMA process as the background disturbance and some other additional stochastic or deterministic part, like a spike, a mean sustain shift, a ramp, an exponential rise to new levels, etc, see Vander Wiel (1996), Box and Luceño (1997), Box and Luceño (2002), Tsiamyrtzis and Hawkins (2010), etc.

In this chapter our proposed model is motivated by a feedstock change problem discussed by Box and Luceño (2002), and we find that it fits well for a real manufacturing scenario in which a production line works for producing very thin metallic films.

In this process, feedstock material is frequently fed into the production line to produce the metallic films. After a certain number of metallic films are produced, a lot is formed and then a new lot is beginning to be formed by subsequent metallic films. When each time a new batch of feedstock material is introduced at the beginning of each lot, there is a variability caused by different stock material put into each batch.

For such a production process, it is obvious that both within-lots variability and between-lots variability exist. Within each lot, the feedstock material is relatively homogeneous and a lack of uniformity appears between different lots, which effects the thickness of the metallic films produced. The amount of feedstock material in each lot can be easily measured by a precise instrument, so that the changes between lots can be anticipated, thus these collected data can be used for feedforward control application. Meanwhile, the within-lot variability is modeled by an IMA process, which is mentioned earlier as the most common nonstationary disturbance model. As soon as enough batches of metallic film was produced, reliable parameter estimates can be made on θ for feedback control application. It should be noted that this type of periodic cycles commonly exists in many real applications, especially in some high-value discrete-part manufacturing processes.

The model investigated by Box and Luceño (2002) is an additive model of IMA process and periodic shifts:

$$z_t - z_{t-1} = a_t - \theta a_{t-1} + \delta_t, |\theta| \leq 1, \tag{3.1}$$

where δ_t is the shift that occurs periodically at times $t = T, 2T, 3T...$, and $m_t = \delta_t + \varepsilon_t$ is the one step ahead estimator of δ_t at time $t-1$, with estimate error ε_t. Assume

$$E(\delta_t) = E(\varepsilon_t) = 0, var(\delta_t) = \sigma_\delta^2, var(\varepsilon_t) = \sigma_\varepsilon^2. \tag{3.2}$$

Usually we are concerned with variance inflation $\sigma_\delta^2 > \sigma_a^2$, with a small estimate error $\sigma_\varepsilon^2 < \sigma_\delta^2$. We call it model (1).

From our earlier statements, since the stationary AR(1) and ARMA (1,1) processes are common disturbance models as well, it is reasonable to consider them as background disturbances and add periodic shifts to get two alternative disturbance models:

$$z_t - \phi z_{t-1} = a_t + \delta_t, 0 < \phi < 1, \tag{3.3}$$

$$z_t - \phi z_{t-1} = a_t - \theta a_{t-1} + \delta_t, 0 < \phi < 1, 0 < |\theta| < 1, \phi \neq \theta. \tag{3.4}$$

We call them model (2) and model (3), and they are typical periodic shifts disturbance models with a stationary background disturbance. In real applications, due to the nature of the industrial and manufacturing processes and the relatively short sampling intervals, the processes are usually positively correlated with $\phi > 0$ and $\theta > 0$. Also, since we mentioned earlier that the AR(1) model is an important disturbance model in many cases, we assume $\theta \neq 0$ in model (3) to avoid model (2) as its special case.

Box and Luceño (2002) studied the characteristics of different control methods on model (1), including feedback control, feedforward control and combined control. However, they did not evaluate their sensitivity and robustness. We investigate the efficiencies, sensitivity and robustness of these control strategies for the nonstationary disturbance model (1), as well as the stationary disturbance model (2) and model (3). Through all of these evaluations we can gain some valuable insights on these disturbance models and different control strategies.

3.2 Control Equations

To evaluate the efficiencies of these control strategies, the feedback as well as feedforward control equations for each model are needed. In this section we assume the true model parameter values are all known, and the unknown parameter scenario will be discussed in the later robustness analysis section.

The major source of variation for the uncontrolled process of each disturbance model is caused by the background disturbances e', i.e., IMA, AR(1), ARMA(1,1) process in model (1), model (2) and model (3) respectively. This is the reason why we call e' noninformative "noise" in this thesis, and it must be adjusted by feedback control directly for variation reduction purpose. Since the variation caused by each periodic shift δ_t is alleviated throughout a period T, so usually it is just the minor source of variation in the uncontrolled process, and it can be adjusted by feedforward control for further improvement.

First we focus on the feedback control equations. Since feedback control is applied to compensate the effect of e', so it should achieve the MMSE control for these background disturbances, i.e., producing white noise output errors $e_t = a_t$. We showed the feedback control equation for IMA process in the introduction section, Box and Luceño (1997, pp. 151) provided the feedback control equation for the AR(1) process, and Tsung et al. (1998) provided the feedback control equation for the ARMA(1,1) process.

Compared with feedback control, feedforward control is applied to compensate m_t, i.e., the periodic shift δ_t estimated at time $t-1$. For a responsive system, the feedforward control equation can be written in the form of

$$gx_{t-1} = -l(m_t, m_{t-1}, ...), \tag{3.5}$$

where l is a suitable model based function of the periodic shifts estimated at current and previous time, i.e., l has different forms based on different disturbance models. Box and Luceño (2002) propose the feedforward control equation $gx_{t-1} = -m_t$ for model (1), which does not depend on the model parameter θ. Since δ_t occur at $t = T, 2T, 3T...$ periodically, so the nonnull feedforward control actions are made periodically at $t = T-1, 2T-1, 3T-1...$

Now we derive the feedforward control equations for model (2) and model (3). Notice that under feedback control, we have $e_t = z_t + gX_{t-1}$, which shows the basic relationship between the output error e_t under control, the disturbance z_t without control, and the cumulated amount of adjustments $gX_{t-1} = \sum_{i=1}^{t-1} gx_i + gX_0$ made before time t. After some simple algebra manipulations of this relationship with the disturbance model equation in term of z_t and a_t, δ_t, as well as the feedback control equation in term of gX_t and e_t, we can derive the output error relationship in terms of e_t and a_t, δ_t under feedback control.

The output error relationship is very important, since they can be used for the derivation of feedforward control equations and the output variance σ_e^2 under feedback control. Since feedforward control is applied to compensate the effect of δ_t, so the output error relationship we want to achieve after adding feedforward control is just replacing δ_t by ε_t in the output error relationship under feedback control.

For model (2), from $e_t = z_t + gX_{t-1}$ and $z_t - \phi z_{t-1} = a_t + \delta_t$, we have

$$e_t - \phi e_{t-1} - gX_{t-1} + \phi gX_{t-2} = a_t + \delta_t.$$

Plug in the feedback control equation $gX_t = -\frac{\phi}{1-\phi\mathbb{B}} e_t$, we get the output error relationship $e_t = a_t + \delta_t$ under feedback control. Since we want to achieve $e_t = a_t + \varepsilon_t$ by adding feedforward control gX^{ff}, so it should satisfy

$$gX^{ff}_{t-1} - \phi gX^{ff}_{t-2} = -m_t.$$

Then we have

$$gx^{ff}_{t-1} = (1-\mathbb{B})gX^{ff}_{t-1} = -\frac{1-\mathbb{B}}{1-\phi\mathbb{B}} m_t = -(m_t - \tilde{m}_{t-1}),$$

where $\tilde{m}_{t-1}$ is an EWMA with smoothing constant ϕ.

For model (3), similarly from $e_t = z_t + gX_{t-1}$ and $z_t - \phi z_{t-1} = a_t - \theta a_{t-1} + \delta_t$, we have the output error relationship

$$e_t - \phi e_{t-1} - gX_{t-1} + \phi g X_{t-2} = a_t - \theta a_{t-1} + \delta_t.$$

Plug in the feedback control equation

$$gX_t = -\frac{\phi - \theta}{1 - \phi \mathbb{B}} e_t,$$

we get

$$e_t - \theta e_{t-1} = a_t - \theta a_{t-1} + \delta_t$$

under feedback control. Since we want to achieve $e_t - \theta e_{t-1} = a_t - \theta a_{t-1} + \varepsilon_t$ by adding feedforward control gX^{ff}, so it should satisfy

$$gX^{ff}_{t-1} - \phi g X^{ff}_{t-2} = -m_t,$$

which is the same as model (2).

Table 3.1: Feedback and Feedforward Control Equations for Disturbance Models

	Feedforward control equation	Feedback control equation
Model (1)	$gx_{t-1} = -m_t$	$gx_t = -Ge_t$
	$t = T, 2T, 3T...$	where $G = 1 - \theta$
Model (2)	$gx_{t-1} = -(m_t - \tilde{m}_{t-1})$	$gX_t = -\frac{\phi}{1-\phi\mathbb{B}} e_t = -\phi \sum_{i=1}^{t} \phi^{t-i} e_i$
	$t = 1, 2, 3, ...$	$gx_t = -\phi(e_t - \tilde{e}_{t-1})$
Model (3)	$gx_{t-1} = -(m_t - \tilde{m}_{t-1})$	$gX_t = -\frac{\phi-\theta}{1-\phi\mathbb{B}} e_t = -(\phi - \theta) \sum_{i=1}^{t} \phi^{t-i} e_i$
	$t = 1, 2, 3, ...$	$gx_t = -(\phi - \theta)(e_t - \tilde{e}_{t-1})$

All the control equations for each model are listed in Table 3.1, in which $\tilde{e}_{t-1}$ and $\tilde{m}_{t-1}$ are both EWMA forecasts with smoothing constant ϕ, and all the control equations have fairly simple structures. We demonstrate that while the model (1) requires periodic adjustments for feedforward control, the feedforward control for model (2) and model (3) are repeated adjustments with the same control equation, and interestingly the feedforward control equation for model (3) only depends on ϕ but not on θ. While the feedback control for model (1) has an integral control form as we mentioned earlier, the feedback control equations for model (2) and model (3), with equivalent forms in terms of gX_t and gx_t, have a weighted integral form which is also easy for implementation.

3.3 Output Mean Square Error Comparisons

EPC had been largely developed in process industry in which usually we have the tacit assumption that the only cost is the off target cost (Box and Kramer (1992), Box and Luceño (1997), pp. 191). The long-run output mean square error (MSEO), defined as

$$MSEO = \lim_{t\to\infty} \frac{1}{t} \sum_{i=1}^{t} var(e_i), \tag{3.6}$$

is widely applied as the comparison criterion. The MSEOs formulae of M_1, M_2, M_3 for model (1) have been provided by Box and Luceño (2002). Now we derive the MSEOs formulea for model (2) and model (3).

In the previous section we mentioned that the output error relationship under combined control is just replacing δ_t using ε_t in the output error relationship under feedback control, so for the MSEOs formulae under combined control, we only need to replace σ_δ^2 by σ_ε^2 in the MSEOs formulae under feedback control. For feedforward control, since it only compensates the periodic shifts δ_t instead of the background disturbance e', so its output error relationship is just by replacing δ_t by ε_t in the disturbance model equation. And since they are all additive models, thus their MSEOs formulae under feedforward control is just by adding the term $\frac{\sigma_\varepsilon^2}{T}$ to the variance of e' directly.

For model (2) we have the output error relationship $e_t = a_t + \delta_t$ under feedback control, so we can easily have $M_1' = \sigma_a^2 + \frac{\sigma_\delta^2}{T}$. For model (3), from $e_t - \theta e_{t-1} = a_t - \theta a_{t-1} + \delta_t$ under feedback control, we can derive $M_1'' = \sigma_a^2 + \frac{\sigma_\delta^2}{T(1-\theta^2)}$. Then the closed form MSEOs formulae for each disturbance model under different control strategies can be derived in strict mathematical proofs and easily verified by simulations, as listed in Table 2.

Table 3.2: MSEOs Formulae for Disturbance Models (1), (2), (3)

	Model (1)	Model (2)	Model (3)
FB	$M_1 = \sigma_a^2 + \frac{\sigma_\delta^2}{T(1-\theta^2)}$	$M_1' = \sigma_a^2 + \frac{\sigma_\delta^2}{T}$	$M_1'' = \sigma_a^2 + \frac{\sigma_\delta^2}{T(1-\theta^2)}$
FF	$M_2 = \infty$	$M_2' = \frac{1}{1-\phi^2}\sigma_a^2 + \frac{\sigma_\varepsilon^2}{T}$	$M_2'' = \frac{1-2\phi\theta+\theta^2}{1-\phi^2}\sigma_a^2 + \frac{\sigma_\varepsilon^2}{T}$
FB +FF	$M_3 = \sigma_a^2 + \frac{\sigma_\varepsilon^2}{T(1-\theta^2)}$	$M_3' = \sigma_a^2 + \frac{\sigma_\varepsilon^2}{T}$	$M_3'' = \sigma_a^2 + \frac{\sigma_\varepsilon^2}{T(1-\theta^2)}$

For nonstationary model (1), its MSEO tends to inflate to infinity in the long run under feedforward control since it only compensates δ_t. Box and Luceño (2002) applied an alternative formula

$$MSEO = \frac{1}{T}\sum_{i=1}^{T} var(e_i)$$

for M_2 and got $M_2 = [1 + \frac{T-1}{2}(1-\theta)^2]\sigma_a^2 + \sigma_\varepsilon^2$. MSEO can be substantially reduced by feedback control which compensates the process nonstationarity. For model (2) and model (3), since an AR (1) process has variance $\frac{1}{1-\phi^2}\sigma_a^2$ and an ARMA (1,1) process has variance $[1 + \frac{(\phi-\theta)^2}{1-\phi^2}]\sigma_a^2$, so their MSEOs can be substantially reduced by feedback control, especially when ϕ is large. In each model, the MSEO achieved by combined control is slightly larger than the white noise process variance σ_a^2, which is the minimum possible value that can be possibly achieved, thus the combined control is always the best control strategy for these disturbance models.

It is worth noticing that in model (2), under feedback control and combined control, the MSEOs M_1' and M_3' do not depend on parameter ϕ. Instead, the MSEOs M_1'' and M_3'' in model (3) only depend on parameter θ but not on ϕ. Also, it is interesting that M_1 and M_1'', or M_3 and M_3'' for model (1) and model (3) are the same, i.e., for the processes with same value of θ, regardless of the value of ϕ in model (3), their MSEOs under feedback (or combined) control are the same.

Evidently only feedforward control is not sufficient, since it compensates δ_t instead of the background disturbance e' that produces the major part of the variation, thus feedforward control should be combined with feedback control whenever possible.

A stationary process such as AR(1) or ARMA(1,1) can be considered as generated by passing white noise through a filter, it is worth pointing out that an undesirable initial transient period often exists in the simulation, and it is especially long when the filter has poles near the unit circle, i.e., $|\phi| \to 1$, resulting in a discrepancy between the theoretical MSEOs formulae results and the simulation results of M_2' and M_2''. We found that in model (2) this discrepancy is consistently small, however, it becomes non negligible when $|\phi| \to 1$ in model (3), which makes our M_2'' formulation result not very accurate.

Usually this transition period problem can be alleviated by discarding the observations until the transient period dies out and it reaches a stationary state. Also, Kay (1981) proposed an initial conditions of the filter which can guarantee a stationary output process by using the Levinson-Durbin algorithm. Fortunately, this transition period problem does not arise in the simulation of the processes under feedback control and combined control, due to the fact that feedback control may conceal the true nature of the disturbance affecting the process that might have been seen in the original uncontrolled data, see Box and Kramer (1992), Box and Luceño (1997), Luceño (1998), etc. While this had been argued as a flaw for feedback control under the process monitoring scenario, it turns to be desirable for feedback control under process adjustment scenario. Therefore we will only evaluate the efficiencies of feedback control and combined control in this thesis.

While MSEO corresponds to the off target cost, sometimes the adjustment cost is also taken into account, which is often proportional to the amount of adjustment x_t, see Box and Kramer (1992), Box and Luceño (1997), Luceño (1998), Del Castillo (2002b). The long-run input mean square error (MSEI), defined as

$$MSEI = \lim_{t\to\infty} \frac{1}{t}\sum_{i=1}^{t}(gx_i)^2,$$

is often used as the relevant criterion. It is obvious that compared with feedforward or feedback control, the combined control always has a larger MSEI, nevertheless, it is still the preferable control strategy, justified by its much smaller MSEO. Traditional method involving both MSEO and MSEI is to using their linear combination as an objective function and formulate it into a linear optimization problem, but it usually takes very complicated forms, see Box and Luceño (1997, pp. 164), Del Castillo (2002b), etc. Therefore we will only investigate MSEOs to choose the best control strategy.

3.4 Sensitivity Analysis

We present some sensitivity analysis on MSEOs to evaluate the effect of parameter values in each disturbance model under these control strategies. Since there are many parameters such as T, ϕ, θ, σ_a, σ_δ and σ_ε, so we will investigate the sensitivity on two groups of parameters: first we fix σ_δ, σ_ε and change the values of θ, ϕ, then we fix θ, ϕ and change the values of σ_δ, σ_ε. For simplicity, we set $\sigma_a = 1$ and $T = 10$ in this section.

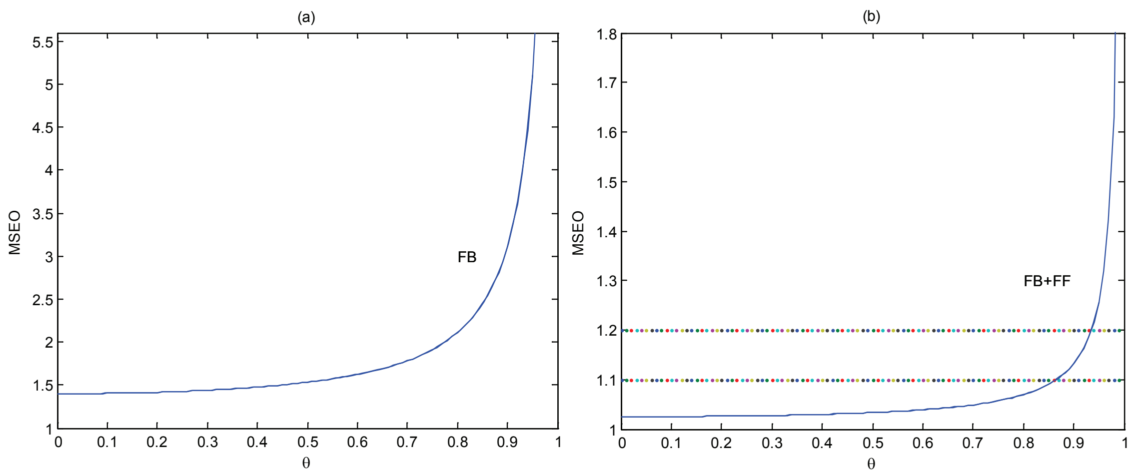

Figure 3.1: MSEOs of model (1) and model (3) with T=10, σ_δ=2, σ_ε=0.5, after feedback control (a) and combined control (b), for different values of θ

First we investigate model (1) and model (3). Table 3.2 shows they have the same MSEOs under feedback control (FB) and combined control (FB+FF). In Figure 3.1, we draw the MSEO curves under FB (a) and FB+FF (b), with different values of θ. Here we follow Box and Luceño (2002) by setting $\sigma_\delta = 2, \sigma_\varepsilon = 0.5$. It shows that their performance deteriorates as θ increases, especially when θ is close to 1. Fortunately, the combined control is very efficient in a very wide range of θ, with an inflation of MSEO less than 10% when $\theta \leq 0.85$, less than 20% when $\theta \leq 0.95$, and it only performs poorly when θ is very close to 1, with the background disturbance IMA process reduced to an approximate white noise process so that feedback control will "temper" the process, as shown in the Deming's funnel experiment, see Del Castillo (2002a).

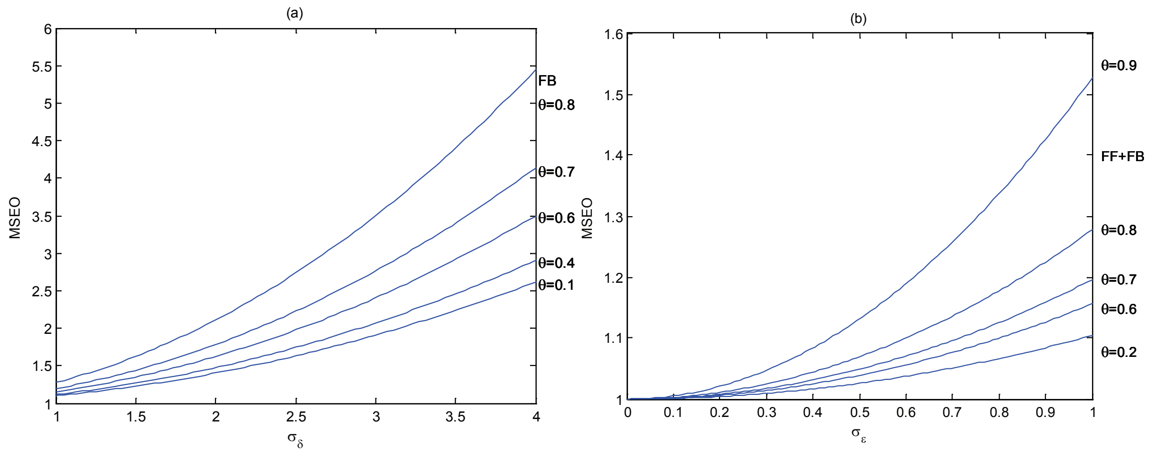

Figure 3.2: MSEOs of model (1) and model (3) with T=10, after feedback control for different values of σ_δ in (a), and after combined control for different values of σ_ε in (b), both with different θ curves

Table 3.2 also shows the MSEOs formulae M_1, M_1'' under FB only depend on σ_δ, and M_3, M_3'' under FB+FF only depend on σ_ε. In Figure 3.2, we draw the MSEO curves as a function of σ_δ under FB (a), and under FB+FF as a function of σ_ε (b), with different values of θ. As we expected, MSEO increases as σ_δ and σ_ε increases.

Then we investigate model (2). As shown in Table 3.2, M_1' under FB and M_3' under FB+FF are independent of ϕ and only depend on σ_δ and σ_ε, so in Figure 3.3, we draw the MSEO curves as a function of σ_δ under FB (a), and as a function of σ_ε under FB+FF (b), and it also shows that MSEO increases as σ_δ and σ_ε increases. By comparing Figure 3.3(a) with Figure 3.2(a) for model (1) and model (3), it can be noticed that within the same range of σ_δ, the performance of FB for model (2) is much better than that for model (1)

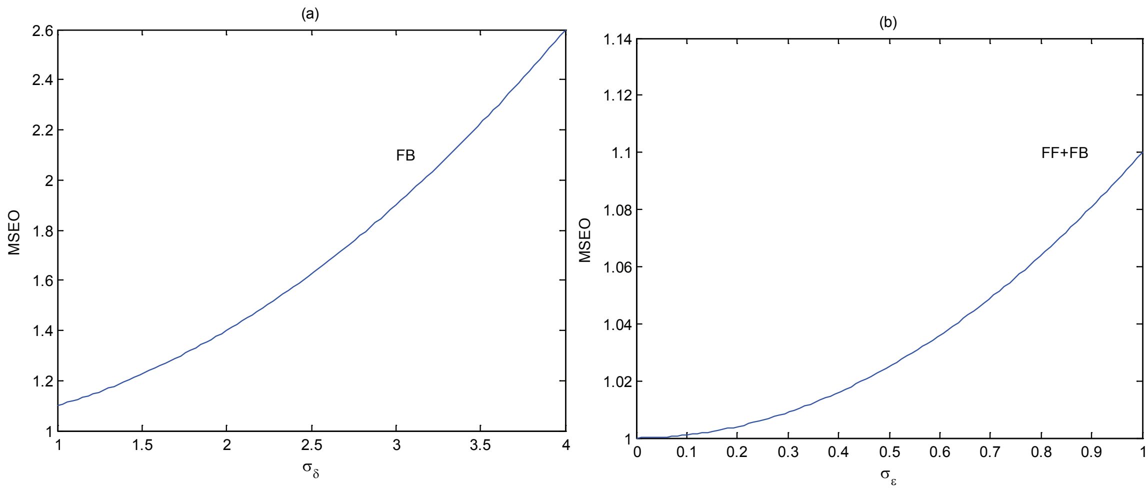

Figure 3.3: MSEOs of model (2) with T=10, after feedback control for different values of σ_δ (a), after combined control for different values of σ_ε with different ϕ (b)

and model (3), since the MSEO inflation range is only (1, 2.6) instead of (1, 6) on Figure 3.3(a). More specifically, under FB for model (2) is as good as under FB for model (1) and model (3) with a very small θ=0.1, with the same MSEO inflation 2.6. Fortunately the combined control is always very efficient, with MSEO inflation less than 10% invariably, as good as under FB+FF for model (1) and model (3) with a very small θ=0.2.

3.5 Robustness Analysis

All the above formulae derivations and performance comparisons are based on the ideal assumption that both the disturbance model structure and parameters are known, however, in many real applications, this assumption is unlikely to be true. In this section we present some robustness analysis on MSEO for these control strategies on the three disturbance models, under two types of model misspecifications: model parameter misspecification and model structure misspecification.

3.5.1 Robustness on Model Parameter Misspecification

Assume we know the true disturbance structure but do not know its true parameter values, so we use the control equations in Table 3.1 with estimated parameters $\hat{\phi}$, $\hat{\theta}$, and the EWMA forecasts with smoothing constant $\hat{\phi}$. Let $\sigma^2_{e'}$ be the output variance of the background disturbance under feedback control, then the MSEO formula can be expressed as a function of $\sigma^2_{e'}$. Now we derive the $\sigma^2_{e'}$ formula and the MSEOs formulae for each model under feedback control and combined control.

In model (1), under feedback control, the background disturbance IMA process has the output error relationship

$$e_t - \hat{\theta} e_{t-1} = a_t - \theta a_{t-1},$$

which is an ARMA(1,1) process for $\{e_t\}$, so we have

$$\sigma^2_{e'} = \frac{1 - 2\theta\hat{\theta} + \theta^2}{1 - \hat{\theta}^2} \sigma^2_a,$$

see Box and Luceño (1997, pp. 152).

In model (2), from $e_t = z_t + gX_{t-1}$ and feedback control equation $gX_t = -\frac{\phi}{1-\hat{\phi}\mathbb{B}} e_t$, we get $z_t = \frac{e_t}{1-\hat{\phi}\mathbb{B}}$. Plug it into the AR(1) equation, we get the output error relationship

$$(1 - \phi\mathbb{B}) e_t = (1 - \hat{\phi}\mathbb{B}) a_t,$$

which is an ARMA(1,1) process for $\{e_t\}$, so we have $\sigma^2_{e'} = \frac{1-2\phi\hat{\phi}+\hat{\phi}^2}{1-\phi^2} \sigma^2_a$. Similarly under feedback control, we have

$$(1 - \phi\mathbb{B}) e_t = (1 - \hat{\phi}\mathbb{B})(a_t + \delta_t),$$

and under combined control, we have

$$(1 - \phi\mathbb{B}) e_t = (1 - \hat{\phi}\mathbb{B}) a_t + (\phi - \hat{\phi})\mathbb{B}\delta_t - (1 - \phi\mathbb{B})\varepsilon_t.$$

In model (3), from $e_t = z_t + gX_{t-1}$ and feedback control equation $gX_t = -\frac{\hat{\phi}-\hat{\theta}}{1-\hat{\phi}\mathbb{B}} e_t$, we get $z_t = \frac{1-\hat{\theta}\mathbb{B}}{1-\hat{\phi}\mathbb{B}} e_t$. Plug it into the ARMA(1,1) equation, we get

$$(1 - \phi\mathbb{B})(1 - \hat{\theta}\mathbb{B}) e_t = (1 - \theta\mathbb{B})(1 - \hat{\phi}\mathbb{B}) a_t,$$

which is an ARMA(2,2) process for $\{e_t\}$, then we can solve $\sigma^2_{e'}$ through a routine procedure, see Box et al. (1994). We can derive the output error relationship

$$(1-\phi\mathbb{B})(1-\hat{\theta}\mathbb{B})e_t = (1-\theta\mathbb{B})(1-\hat{\phi}\mathbb{B})a_t + (1-\hat{\phi}\mathbb{B})\delta_t$$

under feedback control, and

$$(1-\phi\mathbb{B})(1-\hat{\theta}\mathbb{B})e_t = (1-\theta\mathbb{B})(1-\hat{\phi}\mathbb{B})a_t + (\phi-\hat{\phi})\mathbb{B}\delta_t - (1-\phi\mathbb{B})\varepsilon_t$$

under combined control.

Table 3.3: MSEOs Formulae Under Model Parameter Misspecification

	Model (1)	Model (2)	Model (3)
FB	$M_1 = \sigma^2_{e'} + \frac{\sigma^2_\delta}{T(1-\hat{\theta}^2)}$	$M'_1 = \sigma^2_{e'} + \frac{1-2\phi\hat{\phi}+\hat{\phi}^2}{1-\phi^2}\frac{\sigma^2_\delta}{T}$	$M''_1 = \sigma^2_{e'} + \frac{\phi_1(1+\phi_2)(\theta_1+\theta_2\psi_1)+(1-\phi_2)(\phi_2\theta_2+\theta_1\psi_1+\theta_2\psi_2-1)}{T[(1+\phi_2)(\phi_1^2+2\phi_2-\phi_2^2-1)]}\sigma^2_\delta$
FB +FF	$M_2 = \sigma^2_{e'} + \frac{\sigma^2_\varepsilon}{T(1-\hat{\theta}^2)}$	$M'_2 = \sigma^2_{e'} + \frac{(\phi-\hat{\phi})^2}{1-\phi^2}\frac{\sigma^2_\delta}{T} + \frac{\sigma^2_\varepsilon}{T}$	$M''_2 = \sigma^2_{e'} + \frac{-K_1\sigma^2_\delta+K_2\sigma^2_\varepsilon}{TK_3}$
$\sigma^2_{e'}$	$\frac{1-2\theta\hat{\theta}+\theta^2}{1-\hat{\theta}^2}\sigma^2_a$	$\frac{1-2\phi\hat{\phi}+\hat{\phi}^2}{1-\phi^2}\sigma^2_a$	$\frac{\phi_1(1+\phi_2)(\theta_1+\theta_2\psi_1)+(1-\phi_2)(\phi_2\theta_2+\theta_1\psi_1+\theta_2\psi_2-1)}{(1+\phi_2)(\phi_1^2+2\phi_2-\phi_2^2-1)}\sigma^2_a$

We derive the MSEOs formulae of each model under feedback control and combined control in Table 3.3. In model (3),

$$K_1 = (1-\phi_2)(\phi-\hat{\phi})^2,$$
$$K_2 = \phi\phi_1(1+\phi_2) - (1-\phi_2)(1+\phi^2),$$
$$K_3 = (1+\phi_2)(\phi_1^2+2\phi_2-\phi_2^2-1),$$

with $\psi_1 = \phi_1 - \theta_1$, $\psi_2 = \phi_1\psi_1 + \phi_2 - \theta_2$, in which we have

$$\phi_1 = \phi + \hat{\theta}, \phi_2 = -\phi\hat{\theta}, \theta_1 = \theta + \hat{\phi}, \theta_2 = -\theta\hat{\phi}.$$

It is interesting that unlike Table 3.2, here the MSEOs M'_2 and M''_2 under combined control not only depend on σ^2_ε but also on σ^2_δ. The reason is due to the parameter uncertainty in feedforward control equations of model (2) and model (3) as shown in Table 1, it cannot exactly compensate the δ_t term.

For model (1), based on the $\sigma_{e'}^2$ formula, Box and Luceño (1997, pp. 143) recommended using $G = 1 - \hat{\theta} = 0.2$, i.e., equivalently $\hat{\theta} = 0.8$, and they showed that it will only result in at most 10% inflation in the minimum MSEO achievable if the true value of the model parameter θ were exactly known, for the processes with $0.6 \leq \theta < 1$.

For model (2), we found that $\hat{\phi} = 0.3$ performs well for the weakly correlated AR(1) process with $\phi \leq 0.5$, and $\hat{\phi} = 0.8$ performs well for the strongly correlated AR(1) process with $0.5 < \phi \leq 0.9$, with at most 10% inflation of in the minimum MSEO achievable with known ϕ. And the process with $\phi > 0.9$ requires a high value $\hat{\phi} = 0.9$, which can be seen from its $\sigma_{e'}^2$ formula. Therefore we suggest that first find a rough estimation for the range of ϕ, and then choose the corresponding recommended value of $\hat{\phi}$.

The circumstance for model (3) becomes more complicated, since there are uncertainties for two parameters ϕ and θ simultaneously. We found that for any ARMA(1,1) process with (ϕ, θ), any control equation with $(\hat{\phi}, \hat{\theta})$ satisfying $\hat{\phi} - \hat{\theta} \approx \phi - \theta$ will perform very well, with at most 10% inflation of MSEO with known (ϕ, θ). Moreover, as long as $\hat{\phi} - \hat{\theta}$ and $\phi - \theta$ has the same sign, the MSEO result remains to be acceptably good, while the opposite sign usually results in undesirable large MSEOs, due to the reason that the sign of $\phi - \theta$ determines the direction of our feedback control for model (3), as seen from the feedback control equation $gx_t = -(\phi - \theta)(e_t - \tilde{e}_{t-1})$ in Table 3.1.

3.5.2 Robustness on Model Structure Misspecification

Compared with the model parameter misspecification, a more interesting and valuable topic is model structure misspecification, i.e., if the true model is not the same as our assumed one, then how about the performance of different control strategies? This problem has been discussed by many researchers such as Box and Luceño (1997), Apley and Kim (2004), Tsiamyrtzis and Howkins (2010), etc. Luceño (1998) demonstrated that under feedback control scenario, if we are adjusting a stationary disturbance process, then we will lose very little efficiency by misidentifying it as nonstationary; but in contrast, for a nonstationary disturbance process, our adjustment will be very inefficient if we misidentify it as stationary. When we misspecify the nonstationary model (1) to be the stationary model (2) or model (3), similar results can be obtained as well through simulations. The reason is when we control a nonstationary disturbance by using the adjustment equations for the stationary disturbance, it would produce a nonstationary series of output errors. We consider four cases of model structure misspecifications in Table 3.4.

Table 3.4: True Models and Assumed Models Under Model Structure Misspecifications

Case	1	2	3	4
True Model	Model (2)	Model (3)	Model (2)	Model (3)
Assumed Model	Model (1)	Model (1)	Model (3)	Model (2)

Case 1: Misspecify model (2) to be model (1)

When an AR(1) background disturbance is misspecified to be an IMA process, we apply the control equation with estimated parameters for model (1) by setting $G = 1 - \hat{\theta}$. We chose $\phi = 0.9, 0.8, 0.7, 0.6, 0.3$ for the true AR(1) model and $G = 0.95, 0.9, 0.8, 0.7, 0.6, 0.5$ in the control equation. For every ϕ, we find the optimal value of G (G_{opt}) in the control equation, and its corresponding MSEO under combined control ($MSEO_{opt}$) in Table 3.5. We still set $\sigma_\delta = 2$ and $\sigma_\varepsilon = 0.5$ as we did in previous sections. The results show that G_{opt} is always larger than the true value of ϕ. When the true model structure and the true model parameter are both known, the MSEO under combined control ($MSEO_0$) is $\sigma_a^2 + \frac{\sigma_\varepsilon^2}{T} = 1.025$ for any θ, as seen in Table 3.2. The results in Table 3.5 show that the percentage increase in standard deviation (ISD) is acceptably small, especially for the highly correlated AR(1) process with $\phi \geq 0.7$.

Case 2: Misspecify model (3) to be model (1)

When an ARMA(1,1) background disturbance is misspecified to be an IMA process, we apply the control equation with estimated parameters for model (1) by setting $G = 1 - \hat{\theta}$. For the sake of simplicity, we fix $\phi = 0.8$ and $\phi = 0.5$, and choose $\theta = 0.9, 0.8, 0.6, 0.3, 0.1$ for the true ARMA(1,1) model, and $G = 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3$ in the control equation.

Table 3.5: MSEOs Under Combined Control for Different ϕ and G for Case 1

$G \backslash \phi$	0.9	0.8	0.7	0.6	0.3
0.95	1.10	1.17	1.26	1.35	1.73
0.9	1.10	1.17	1.25	1.34	1.69
0.8	1.13	1.19	1.26	1.34	1.64
0.7	1.18	1.23	1.29	1.35	1.61
0.6	1.25	1.30	1.34	1.39	1.60
0.5	1.38	1.40	1.43	1.47	1.62
G_{opt}	0.95-0.97	0.9-0.95	0.9	0.8	0.6
$MSEO_{opt}$	1.096	1.174	1.254	1.339	1.598
$MSEO_0$	1.025	1.025	1.025	1.025	1.025
$ISD(\%)$	3.4	7.0	10.6	14.3	24.9

For every combination of (ϕ, θ), we find G_{opt} and $MSEO_{opt}$ under combined control in Table 3.6. It shows that the G_{opt} is always slightly larger than $1 - \theta$. This conclusion is the same as the recommendation given by Luceño (1998) under the feedback control scenario, if misspecifying an ARMA(1,1) disturbance as an IMA process. From Table 3.2 we have $MSEO_0 = \sigma_a^2 + \frac{\sigma_\varepsilon^2}{T(1-\theta^2)}$ only depends on parameter θ but not on ϕ, and the results in Table 3.6 show that the model (3) with the same θ and a larger ϕ has a smaller ISD, and the one with the same ϕ and a smaller θ has a smaller ISD.

It is worth noticing that for both case 1 and case 2, we can easily verify that when we use a G which is not necessarily G_{opt}, the MSEO inflation is usually very small compared with the $MSEO_{opt}$ achieved by using G_{opt}, i.e., the MSEO performance is very robust on the choice of G. It can also be seen from Table 3.5 and Table 3.6 that sometimes the G_{opt} is chosen as a range instead of an exact number, which means that all the G in this range achieves the same $MSEO_{opt}$. This robustness property is very desirable in real applications, since we usually do not know the true model structures and their true parameter values, so

Table 3.6: MSEOs Under Combined Control for Different θ, G for Case 2 (ϕ=0.8/0.5)

$G \backslash \theta$	0.9	0.8	0.6	0.3	0.1
0.9	2.08/2.88	1.89/2.62	1.58/2.17	1.28/1.70	1.19/1.50
0.8	1.92/2.64	1.75/2.41	1.48/2.02	1.24/1.61	1.18/1.46
0.7	1.79/2.46	1.64/2.25	1.40/1.90	1.216/1.55	1.20/1.444
0.6	1.70/2.31	1.56/2.12	1.35/1.81	1.220/1.523	1.25/1.452
0.5	1.62/2.20	1.50/2.03	1.33/1.757	1.25/1.526	1.33/1.49
0.4	1.59/2.14	1.48/1.99	1.34/1.751	1.34/1.57	1.46/1.57
0.3	1.62/2.17	1.53/2.03	1.43/1.82	1.50/1.70	1.69/1.73
G_{opt}	0.4/0.4	0.4/0.4	0.5/0.4-0.5	0.6-0.7/0.5-0.6	0.8/0.7
MSE_{opt}	1.59/2.14	1.48/1.99	1.33/1.75	1.22/1.52	1.18/1.44
$MSEO_0$	1.132	1.069	1.039	1.028	1.025
$ISD(\%)$	18.5/37.5	17.7/36.4	13.1/29.8	8.9/21.6	7.3/18.5

the G_{opt} cannot be obtained.

Case 3: Misspecify model (2) to be model (3)

Compared with case 1 and case 2 in which we misspecify a stationary disturbance to be a nonstationary disturbance, it is more interesting to consider the case that we misspecify a stationary disturbance to be another stationary one.

When an AR(1) disturbance is misspecified to be an ARMA(1,1), from $e_t = z_t + gX_{t-1}$ and the feedback control equation $gX_t = -\frac{\hat{\phi}-\hat{\theta}}{1-\hat{\phi}\mathbb{B}}e_t$, we have the output error relationship

$$(1-\phi\mathbb{B})(1-\hat{\theta}\mathbb{B})e_t = (1-\hat{\phi}\mathbb{B})a_t,$$

which is an ARMA(2,1) process for $\{e_t\}$. By adding the period shift term δ_t, we have

$$(1-\phi\mathbb{B})(1-\hat{\theta}\mathbb{B})e_t = (1-\hat{\phi}\mathbb{B})a_t + (1-\hat{\phi}\mathbb{B})\delta_t$$

under feedback control. And under combined control, from $e_t = z_t + gX_{t-1}$, feedback control equation $gX_t = -\frac{\hat{\phi}-\hat{\theta}}{1-\hat{\phi}\mathbb{B}}e_t$, feedforward control equation $gX_{t-1} = -\frac{m_t}{1-\hat{\phi}\mathbb{B}}$, and model

(2) equation, we can get

$$(1-\phi\mathbb{B})(1-\hat{\theta}\mathbb{B})e_t = (1-\hat{\phi}\mathbb{B})a_t + (\phi-\hat{\phi})\mathbb{B}\delta_t - (1-\phi\mathbb{B})\varepsilon_t.$$

So due to the parameter uncertainty in feedforward control equations, the feedforward control cannot exactly compensate the periodic shift, thus the δ_t term is included in the output error relationship under combined control.

Case 4: Misspecify model (3) to be model (2)

When an ARMA(1,1) disturbance is misspecified to be an AR(1), from $e_t = z_t + gX_{t-1}$ and control equation $gX_t = -\frac{\hat{\phi}}{1-\hat{\phi}\mathbb{B}}e_t$, we have the output error relationship

$$(1-\phi\mathbb{B})e_t = (1-\hat{\phi}\mathbb{B})(1-\theta\mathbb{B})a_t,$$

which is an ARMA(1,2) process for $\{e_t\}$. By adding the period shift term δ_t, we have

$$(1-\phi\mathbb{B})e_t = (1-\hat{\phi}\mathbb{B})(1-\theta\mathbb{B})a_t + (1-\hat{\phi}\mathbb{B})\delta_t$$

under feedback control. Under combined control, from $e_t = z_t + gX_{t-1}$, feedback control equation $gX_t = -\frac{\hat{\phi}}{1-\hat{\phi}}\tilde{e}_t$, feedforward control equation $gX_{t-1} = -\frac{m_t}{1-\hat{\phi}\mathbb{B}}$, and model (3) equation, we can get the output error relationship

$$(1-\phi\mathbb{B})e_t = (1-\hat{\phi}\mathbb{B})(1-\theta\mathbb{B})a_t + (\phi-\hat{\phi})\mathbb{B}\delta_t - (1-\phi\mathbb{B})\varepsilon_t.$$

From the output error relationship under each scenario, we can derive the MSEOs formulae for case 3 and case 4 in Table 3.7. In case 3, we have

$$K_1 = (1-\phi_2)(\phi-\hat{\phi})^2,$$
$$K_2 = (1-\phi_2)(1+\phi^2-\phi\phi_1) - \phi\phi_1(1+\phi_2),$$
$$K_3 = 1-\phi_2-\phi_1^2-\phi_1^2\phi_2-\phi_2^2(1-\phi_2).$$

Similar to Table 3.3 for model parameter misspecification, the MSEOs M'_{23} and M'_{32} under combined control for model structure misspecification include σ_δ^2 term as well.

In case 4, since there is only one parameter $\hat{\phi}$ in the assumed model (2), we can investigate the optimal values of $\hat{\phi}$ for $\sigma_{e'}^2$ under background disturbance. When an ARMA(1,1)

Table 3.7: Case 3 and Case 4 Comparisons

	Case 3: Model (2) to be Model (3)	Case 4: Model (3) to be Model (2)
FB	$M_{23} = \sigma^2_{e'} + \frac{1-2\phi_1\hat{\phi}+\hat{\phi}^2}{T[1-\phi_1^2-\phi_1^2\phi_2-\phi_2-\phi_2^2+\phi_2^3]}\sigma^2_\delta$	$M_{32} = \sigma^2_{e'} + \frac{1-2\phi\hat{\phi}+\hat{\phi}^2}{T(1-\phi^2)}\sigma^2_\delta$
FB +FF	$M'_{23} = \sigma^2_{e'} + \frac{K_1\sigma^2_\delta+K_2\sigma^2_\varepsilon}{TK_3}$	$M'_{32} = \sigma^2_{e'} + \frac{(\phi-\hat{\phi})^2}{T(1-\phi^2)}\sigma^2_\delta + \frac{\sigma^2_\varepsilon}{T}$
$\sigma^2_{e'}$	$\frac{1+\theta_1^2-\phi_2-2\phi_1\theta_1-\phi_2\theta_1^2}{1-\phi_1^2-\phi_1^2\phi_2-\phi_2-\phi_2^2+\phi_2^3}\sigma^2_a$	$\frac{1-\phi_1\theta_1-(\phi_1\theta_2+\theta_1)\psi_1-\theta_2\psi_2}{1-\phi_1^2}\sigma^2_a$
	$\phi_1 = \phi + \hat{\theta}$, $\phi_2 = -\phi\hat{\theta}$, $\theta_1 = \hat{\phi}$	$\phi_1 = \phi$, $\theta_1 = \theta + \hat{\phi}$, $\theta_2 = -\theta\hat{\phi}$ $\psi_1 = \phi_1 - \theta_1$, $\psi_2 = \phi_1\psi_1 - \theta_2$

background disturbance is misspecified to be an AR(1), we have the $\sigma^2_{e'}$ formula from case 4 in Table 3.7. Using the original parameters ϕ, θ and $\hat{\phi}$, we get

$$\sigma^2_{e'} = \frac{(\theta^2 - 2\phi\theta + 1)\hat{\phi}^2 + (2\phi^2\theta + 2\theta - 2\phi - 2\phi\theta^2)\hat{\phi} + (\theta^2 - 2\phi\theta + 1)}{1-\phi^2}\sigma^2_a. \tag{3.7}$$

Since the minimization of $\sigma^2_{e'}$ is equivalent to the minimization of the numerator, which is a quadratic function of $\hat{\phi}$, so we can get its optimal value

$$\hat{\phi}_{opt} = \frac{\phi\theta^2 - \phi^2\theta + \phi - \theta}{\theta^2 - 2\phi\theta + 1} = \frac{(\phi-\theta)(1-\phi\theta)}{(\phi-\theta)^2 + (1-\phi^2)}, \tag{3.8}$$

and it is easy to see that $\hat{\phi}_{opt}$ has the same sign with $\phi - \theta$, thus we have $\hat{\phi}_{opt} < 0$ if $\phi < \theta$, and $\hat{\phi}_{opt} \geq 0$ otherwise. At first glance, it seems counterintuitive to use $\hat{\phi} < 0$ in feedback control, which appears to be deteriorating the process instead of improving it. However, this is not surprising, since $\hat{\phi} < 0$ is only detrimental for the assumed model (2), but not for the true model (3).

In Table 3.8, we list the $\hat{\phi}_{opt}$ and the corresponding $MSEO_{opt}$ for the true ARMA(1,1) process with different (ϕ, θ) in case 4. It can be seen that although the MSE_{opt} achieved by using $\hat{\phi}_{opt}$ is not as good as the MMSE control which can achieve $\sigma^2_{e'} = 1$ with known model structure ARMA(1,1) and parameters (ϕ, θ), it is still very good, with a MSEO inflation less than 5% except for the cases of $\theta - \phi \geq 0.4$.

Table 3.8: $\hat{\phi}_{opt}$/MSEO$_{opt}$ for Different (ϕ, θ) of True ARMA(1,1) Process in Case 4

$\theta \backslash \phi$	0.8	0.6	0.4	0.2
0.8	0.0/1.000	-0.15/1.038	-0.27/1.102	-0.38/1.175
0.6	0.26/1.036	0.0/1.000	-0.17/1.016	-0.31/1.051
0.4	0.52/1.049	0.22/1.009	0.0/1.000	-0.18/1.006
0.2	0.70/1.020	0.44/1.008	0.21/1.002	0.0/1.000

It is worth noticing that in case 4, we do not know the true model structure is an ARMA(1,1) process, and do not know the value of (ϕ, θ) either, so the $\hat{\phi}_{opt}$ cannot be obtained in real applications. However, we can easily verify that when we use a $\hat{\phi}$ which is not necessarily $\hat{\phi}_{opt}$, the MSEO inflation is usually very small compared with the $MSEO_{opt}$ achieved by using $\hat{\phi}_{opt}$, i.e., the MSEO performance is very robust on the choice of $\hat{\phi}$. Especially, as long as we $\hat{\phi}$ has the same sign as $\hat{\phi}_{opt}$, the MSEO result remains to be acceptably good, while the opposite sign usually results in undesirable large MSEOs. The reason is that the sign of $\hat{\phi}$ determines the direction of our feedback control for our assumed model (2), as seen from the feedback control equation $gx_t = -\hat{\phi}(e_t - \tilde{e}_{t-1})$ in Table 3.1 with estimated parameter $\hat{\phi}$. Since we mentioned earlier that $\hat{\phi}_{opt}$ has the same sign with $\phi - \theta$, and the $\phi - \theta$ determines the direction of our feedback control for our true model (3), as seen from the feedback control equation $gx_t = -(\phi - \theta)(e_t - \tilde{e}_{t-1})$ in Table 3.1, so it is reasonable to believe that when the directions of the feedback control equations of our assumed model (2) and true model (3) are the same, i.e., when $\phi - \theta$, $\hat{\phi}_{opt}$ and $\hat{\phi}$ has the same sign, the MSEO performance will be consistently good.

3.6 Simulated Examples

In this section, we illustrate some of our main results and conclusions through two simulated examples, called example (1) and example (2). Each example includes an evolving disturbance process of 500 observations, consists of 5 pieces of 100 observations, with different model structures and parameters evolving over time. It is similar to the simulated evolving process given by Tsung et al. (1998, pp. 216, Figure 1 (a), (b)), which consists of 1000 observations, with the first 250 observations generated from an ARMA(1,1) model with $(\phi, \theta) = (0.8, 0.3)$, and the next 250 observations from an IMA model with $\theta = 0.25$, then the next 250 observations from an ARMA(1,1) model with $(\phi, \theta) = (0.8, 0.2)$, and the last 250 observations from an ARMA(1,1) model with $(\phi, \theta) = (0.6, 0.3)$.

For convenience, we take model (1) with $\theta = 0.25$ as process (I), model (2) with $\phi = 0.5$ as process (II), model (3) with $(\phi, \theta) = (0.8, 0.3)$ as process (III), model (3) with $(\phi, \theta) = (0.8, 0.2)$ as process (IV) and model (3) with $(\phi, \theta) = (0.6, 0.3)$ as process (V). We use the same process from (I) to (V) in our two simulated examples, just by varying their orders. In our simulations, we still use $\sigma_a = 1$, $\sigma_\delta = 2$ and $\sigma_\varepsilon = 0.5$.

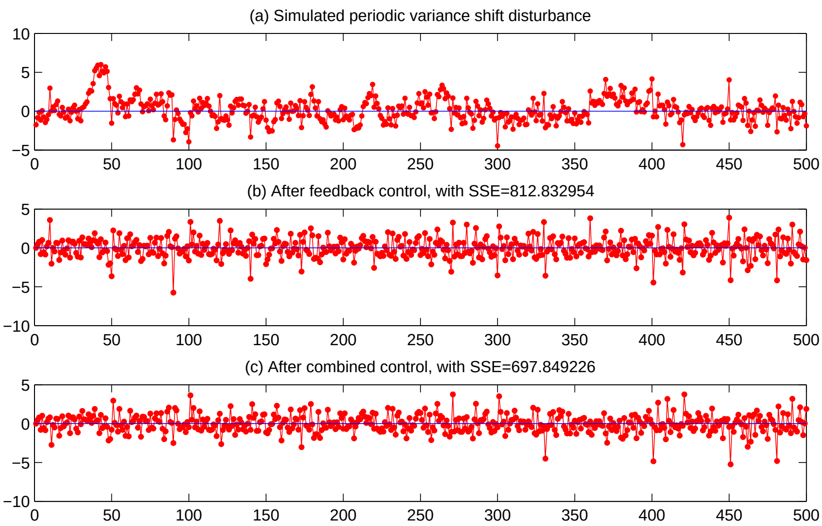

Figure 3.4: Example (1), Simulated Data and Processes Under Control for Model (1)

In our example (1), we consider the evolving disturbance consists of processes (I), (II), (III), (IV), (V) in order. Assume we only know the initial model structure and parameter but do not know how it changes over time, so we adjust the whole process by using the feedback and feedforward control equations for process (I). In Figure 3.4 we plot three processes, in which (a) is the uncontrolled disturbance, (b) is the process under feedback control, and (c) is the process under combined control. The sum of squares error (SSE) for each process in (b) and (c) is calculated. The three plots show that after feedback control, the process improves substantially since the nonstationarity of the first 100 observations disappears, and combined control further improves the process with a much smaller SSE.

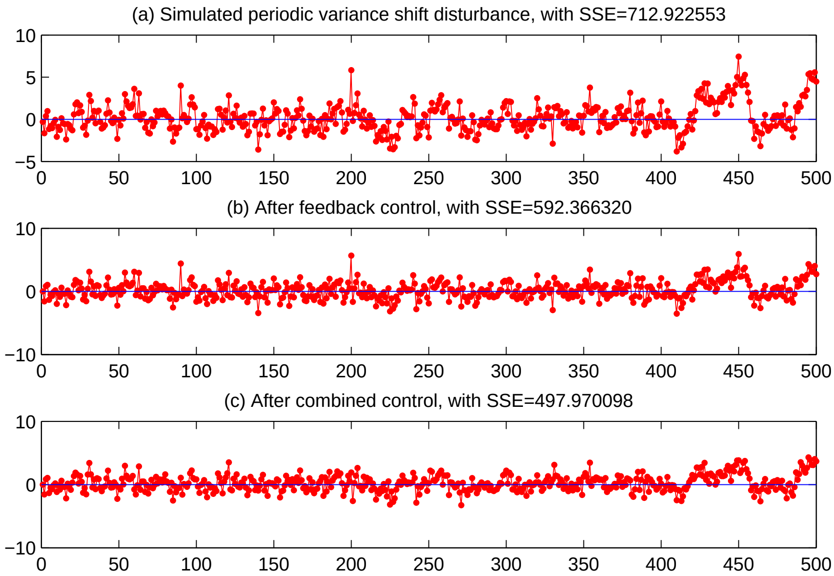

Figure 3.5: Example (2), Simulated Data and the Processes Under Control for Model (3)

In our example (2), we consider the evolving disturbance consists of processes (III), (II), (IV), (V), (I) in order. Similarly we adjust the whole process by using the feedback and feedforward control equations for the initial process (III). In Figure 3.5 we get three processes corresponding to Figure 3.4, and one conspicuous difference between these two figures is shown from the last 100 observations in Figure 3.5(b) and Figure 3.5(c) which still exhibit nonstationarity, implying that the adjustment is inefficient. Under this case, the SSE for each process is calculated only for its stationary part, i.e, the first 400 observations.

These simulated examples serve as a synthesis flavor for the conclusions illustrated in the model structure and parameter misspecifications in the robustness analysis section. However, since the plots in Figure 3.4 and Figure 3.5 are both generated by simulation with only one replication, so each figure only represents one possible result. Without loss of generality, now we repeat each example by additional simulations with 10,000 replications and calculate the MSEOs in Table 3.9.

Table 3.9: MSEOs for Two Simulation Examples

	Example (1) (last 400)	Example (1) (all 500)	Example (2) (first 400)	Example (2) (all 500)
(a) Disturbance	1.98	7.35-7.55	1.93	7.45-7.55 (100%)
(b) Feedback	1.89	1.83	1.59	4.06-4.11 (54.5%)
(c) Combined	1.70	1.64	1.30	3.76-3.82 (50.5%)

In each example, we calculate MSEOs under two cases, for only stationary process part as well as the whole process involving the nonstationary part. For only the stationary part, we can get relatively small MSEOs. For the processes including the nonstationary part, their MSEOs are quite large with large variations, so their ranges are given instead of the fixed numbers. Fortunately, their MSEOs ratios are fixed as given in the last column. The results in column 2 show that when we are adjusting a stationary disturbance part with model (2) and model (3) by using the control equations for nonstationary model (1), the MSEO can be reduced from 1.98 to 1.70 under combined control. In column 4, when we are adjusting the same stationary part by using the control equations for model (3) with initial model parameters, their MSEOs can be reduced substantially from 1.93 to 1.30 under combined control. Notice that here the amount of MSEO reduction is much greater than column 2, since it has much less degree of model misspecifications. Similarly, the MSEOs results in columns 3 are slightly better than column 2 for example (1), due to the same reason of less degree of model misspecifications. In the last column, although the unacceptably large MSEOs under control still exist due to the last 100 observations of nonstationary part, they can be reduced by half amount by combined control, taking the MSEO of the uncontrolled disturbance as the 100% baseline.

3.7 Proofs of Some Formulas for Some Disturbance Models

3.7.1 Proof of Formulas for Periodic Shift Disturbance Model (1)

• Proof of $M_2 = [1 + \frac{T-1}{2}(1-\theta)^2]\sigma_a^2 + \sigma_\varepsilon^2$:

We first prove that without control, $MSEO = [1 + \frac{T-1}{2}(1-\theta)^2]\sigma_a^2 + \sigma_\delta^2$.

If we do not consider the step change term δ_t, then only with IMA disturbance model $z_t - z_{t-1} = a_t - \theta a_{t-1}$, from Equation (2.1) in the last chapter, $z_t = a_t + \lambda \sum_{i=1}^{t-1} a_i$, where $\lambda = 1-\theta$, by taking the variance, we have $var(z_t) = \sigma_a^2 + \lambda^2 \sum_{i=1}^{t-1} \sigma_a^2 = \sigma_a^2(1 + (t-1)\lambda^2)$, for $t \neq 0, T, 2T, 3T...$ Then at times $t = 1, 2, ..., T$, we have

$$\begin{aligned} var(z_1) &= \sigma_a^2 \\ var(z_2) &= \sigma_a^2(1+\lambda^2) \\ &\ldots \\ var(z_{T-1}) &= \sigma_a^2(1+(T-2)\lambda^2) \\ var(z_T) &= \sigma_a^2(1+(T-1)\lambda^2) \end{aligned}$$

Then we can get

$$\sum_{i=1}^{T} var(z_i) = \sigma_a^2(T + \frac{T(T-1)}{2}\lambda^2).$$

Adding the step change term δ_t, since there is only one feedforward control being made during $t = 1, ..., T$, then the MSEO without control is

$$\frac{1}{T}\sum_{i=1}^{T} var(z_i) = [1 + \frac{T-1}{2}(1-\theta)^2]\sigma_a^2 + \sigma_\delta^2.$$

Since the effect of the feedforward control will be to reduce the step change term δ_t to be a white noise term ε_t, so it will be to reduce this MSEO to be

$$M_2 = [1 + \frac{T-1}{2}(1-\theta)^2]\sigma_a^2 + \sigma_\varepsilon^2.$$

• Proof of $e_t - \theta e_{t-1} = a_t - \theta a_{t-1} + \delta_t$:

$$\begin{cases} z_t = e_t - gX_{t-1}, \\ z_t - z_{t-1} = a_t - \theta a_{t-1} + \delta_t, \end{cases}$$

$$\begin{aligned}
&\Longrightarrow \quad (e_t - gX_{t-1}) - (e_{t-1} - gX_{t-2}) = a_t - \theta a_{t-1} + \delta_t, \\
&\Longleftrightarrow \quad e_t - e_{t-1} - gx_{t-1} = a_t - \theta a_{t-1} + \delta_t, \\
&\Longleftrightarrow \quad e_t - e_{t-1} - (-Ge_{t-1}) = a_t - \theta a_{t-1} + \delta_t, (gx_t = -Ge_t) \\
&\Longleftrightarrow \quad e_t - (1-G)e_{t-1} = a_t - \theta a_{t-1} + \delta_t, \\
&\Longleftrightarrow \quad e_t - \theta e_{t-1} = a_t - \theta a_{t-1} + \delta_t.
\end{aligned}$$

• Proof of $M_1 = \sigma_a^2 + \frac{\sigma_\delta^2}{T(1-\theta^2)}$.

First we calculate $var(e_T)$. From $e_T - \theta e_{T-1} = a_T - \theta a_{T-1} + \delta_T$, multiply e_T and e_{T-1} on both sides respectively, we get

$$\begin{cases} e_T e_T - He_T e_{T-1} = a_T e_T - \theta a_{T-1} e_T + \delta_T e_T \\ e_{T-1} e_T - He_{T-1} e_{T-1} = a_T e_{T-1} - \theta a_{T-1} e_{T-1} + \delta_T e_{T-1} \end{cases}$$

Taking expectations of these two equations, we get

$$\begin{cases} \gamma_0 - \theta\gamma_1 = \sigma_a^2 + \sigma_\delta^2 \\ \gamma_1 - \theta\gamma_0 = -\theta\sigma_a^2 \end{cases}$$

where $\gamma_0 = \sigma_e^2$. From the second equation we get $\gamma_1 = \theta(\sigma_e^2 - \sigma_a^2)$, plug it into the first equation we get $\sigma_e^2 = \sigma_a^2 + \frac{\sigma_\delta^2}{1-\theta^2}$.

Then it is apparent that $var(e_1) = var(e_2) = ... = var(e_{T-1}) = \sigma_e^2$, so

$$M_1 = \frac{1}{T}\sum_{i=1}^{T} var(e_i) = \frac{1}{T}(T\sigma_a^2 + \frac{\sigma_\delta^2}{1-\theta^2}) = \sigma_a^2 + \frac{\sigma_\delta^2}{T(1-\theta^2)}.$$

• Proof of $M_3 = \sigma_a^2 + \frac{\sigma_\varepsilon^2}{T(1-\theta^2)}$.

First we prove that for feedforward plus feedback control, we have $e_t - \theta e_{t-1} = a_t - \theta a_{t-1} + \varepsilon_t$.

Similar to the proof of $e_t - He_{t-1} = a_t - \theta a_{t-1} + \delta_t$ for the feedback control. When we get $e_t - e_{t-1} - gx_{t-1} = a_t - \theta a_{t-1} + \delta_t$. Plug $gx_{t-1} = -Ge_{t-1} - m_t$ into it, we have

$$\begin{aligned}
&\Longrightarrow \quad e_t - e_{t-1} - (-Ge_{t-1} - m_t) = a_t - \theta a_{t-1} + \delta_t, \\
&\Longleftrightarrow \quad e_t - (1-G)e_{t-1} + (\delta_t + \varepsilon_t) = a_t - \theta a_{t-1} + \delta_t, \\
&\Longleftrightarrow \quad e_t - \theta e_{t-1} = a_t - \theta a_{t-1} - \varepsilon_t, \\
&\Longleftrightarrow \quad e_t - \theta e_{t-1} = a_t - \theta a_{t-1} + \varepsilon_t.
\end{aligned}$$

So similarly we have $M_3 = \sigma_a^2 + \frac{\sigma_\varepsilon^2}{T(1-\theta^2)}$.

3.7.2 Proof of Formulas for Equation (3.2)

• Proof of after feedforward control, $M_2' = [\frac{1}{1-\phi^2} - \frac{1-\phi^{2T}}{T(1-\phi^2)^2}]\sigma_a^2 + \sigma_\varepsilon^2$:

For Equation (3.2), from

$$\begin{aligned}
z_t &= \phi z_{t-1} + a_t \\
\phi z_{t-1} &= \phi^2 z_{t-2} + \phi a_{t-1} \\
\phi^2 z_{t-1} &= \phi^3 z_{t-2} + \phi^2 a_{t-1} \\
&\quad ...
\end{aligned}$$

By adding those t equations together, we have

$$z_t = a_t + \phi a_{t-1} + \phi^2 a_{t-1} ... = \frac{\tilde{a}_t}{1-\phi},$$

where $\tilde{a}_t$ is an EWMA of a_t, a_{t-1},... with smoothing constant ϕ. Then we have

$$\begin{aligned}
var(z_t) &= var(a_t + \phi a_{t-1} + \phi^2 a_{t-1} ...) \\
&= var(a_t) + var(\phi a_{t-1}) + var(\phi^2 a_{t-1} ...) \\
&= (1 + \phi^2 + \phi^4 + \phi^6 + ... + \phi^{t-1})\sigma_a^2 \\
&= \tfrac{1-(\phi^2)^t}{1-\phi^2}\sigma_a^2
\end{aligned}$$

Then at times $t = 1, 2, ..., T$, we have

$$\begin{aligned}
var(z_1) &= \tfrac{1}{1-\phi^2}\sigma_a^2 \\
var(z_2) &= \tfrac{1-\phi^2}{1-\phi^2}\sigma_a^2 = \sigma_a^2 \\
&\quad ... \\
var(z_{T-1}) &= \tfrac{1-\phi^{2(T-1)}}{1-\phi^2}\sigma_a^2 \\
var(z_T) &= \tfrac{1-\phi^{2T}}{1-\phi^2}\sigma_a^2
\end{aligned}$$

And by adding those T equations, we get

$$\sum_{i=1}^{T} var(z_i) = (\frac{T}{1-\phi^2} - \frac{1+\phi^2+\phi^4+\phi^6+...\phi^{2T}}{1-\phi^2})\sigma_a^2 = (\frac{T}{1-\phi^2} - \frac{1-\phi^{2(T+1)}}{(1-\phi^2)^2})\sigma_a^2,$$

Adding the step change term δ_t, since there is only one feedforward control being made during $t = 1, ..., T$, then the MSEO without control is

$$\frac{1}{T}\sum_{i=1}^{T} var(z_i) = [\frac{1}{1-\phi^2} - \frac{1-\phi^{2T+2}}{T(1-\phi^2)^2}]\sigma_a^2 + \sigma_\delta^2.$$

Since the effect of the feedforward control will be to reduce the step change term δ_t to be a white noise term ε_t, so it will reduce this MSEO to be

$$M_2' = [\frac{1}{1-\phi^2} - \frac{1-\phi^{2T+2}}{T(1-\phi^2)^2}]\sigma_a^2 + \sigma_\varepsilon^2.$$

3.7.3 Proof of Formulas for Periodic Shift Disturbance Model (2)

• Proof of $z_t = a_t + \frac{\phi-\theta}{1-\phi}\tilde{a}_{t-1}$:

From $z_{t+1} - \phi z_t = a_{t+1} - \theta a_t$, we have

$$\begin{aligned} z_t &= \phi z_{t-1} + a_t - \theta a_{t-1} \\ \phi z_{t-1} &= \phi^2 z_{t-2} + \phi a_{t-1} - \theta\phi a_{t-2} \\ \phi^2 z_{t-2} &= \phi^3 z_{t-3} + \phi^2 a_{t-2} - \theta\phi^2 a_{t-3} \\ &\ ... \end{aligned}$$

Adding them together, we get

$$\begin{aligned} z_t &= a_t + (\phi-\theta)a_{t-1} + \phi(\phi-\theta)a_{t-2} + \phi^2(\phi-\theta)a_{t-3} + ... \\ &= a_t + (\phi-\theta)(a_{t-1} + \phi a_{t-2} + \phi^2 a_{t-3} + ...) \\ &= a_t + \tfrac{\phi-\theta}{1-\phi}\tilde{a}_{t-1} \end{aligned}$$

• Proof of Equation (3.4), the feedback control equation $gX_t = -\frac{\phi-\theta}{1-\phi\mathbb{B}}e_t$:

From $gX_t = -\hat{z}_{t+1} = -\phi z_t + \theta a_t$, and $z_t = a_t + \frac{\phi-\theta}{1-\phi}\tilde{a}_{t-1}$, we have

$$\begin{aligned} gX_t &= -\phi z_t + \theta a_t \\ &= -\tfrac{\phi(1-\theta\mathbb{B})}{1-\phi\mathbb{B}}a_t + \theta a_t \\ &= -\tfrac{\phi(1-\theta\mathbb{B})}{1-\phi\mathbb{B}}a_t + \tfrac{\theta(1-\phi\mathbb{B})}{1-\phi\mathbb{B}}a_t \\ &= -\tfrac{\phi-\theta}{1-\phi\mathbb{B}}a_t \end{aligned}$$

Since when the feedback control scheme is in operation, the observed output at time t is the output error e_t, so the controlled output series would then be such that $e_t = a_t$ and $\sigma_e^2 = \sigma_a^2$, and the feedback control equation is

$$gX_t = -\frac{\phi - \theta}{1 - \phi\mathbb{B}} e_t.$$

3.7.4 Alternative Simpler Proofs of Formula for Periodic Shift Disturbance Model (3)

(1) $z_t = a_t + \frac{\phi-\theta}{1-\phi}\tilde{a}_{t-1}$

(2) $gX_t = -\frac{\phi-\theta}{1-\phi\mathbb{B}} e_t$

(3) $e_t = a_t + \delta_t$

• Proof of (1): It is equivalent to ARMA(1,1): $z_t = \frac{1-\theta\mathbb{B}}{1-\phi\mathbb{B}} a_t$

• Proof of (2): In the dynamic system $Y_t - \delta Y_{t-1} = g(1-\delta)X_{t-1}$, assume that $\delta = 0$, so we have the pure-gain model $Y_t = gX_{t-1}$. Since $gX_t = -\hat{z}_{t+1} = -\phi z_t + \theta a_t$, and from

$$\begin{cases} e_t = Y_t + z_t \\ Y_t = gX_{t-1} \\ z_t = \dfrac{1-\theta\mathbb{B}}{1-\phi\mathbb{B}} a_t \end{cases}$$

we have

$$gX_t = -\phi z_t + \theta a_t = -\frac{\phi(1-\theta\mathbb{B})}{1-\phi\mathbb{B}} a_t + \theta a_t = -\frac{\phi-\theta}{1-\phi\mathbb{B}} a_t.$$

Since when the feedback control scheme is in operation, the observed output at time t is the output error e_t, so the controlled output series would then be such that $e_t = a_t$, thus

$$gX_t = -\frac{\phi-\theta}{1-\phi\mathbb{B}} e_t.$$

• Proof of (3): From $e_t = Y_t + z_t = gX_{t-1} + z_t = -\frac{\phi-\theta}{1-\phi\mathbb{B}} e_{t-1} + \frac{1-\theta\mathbb{B}}{1-\phi\mathbb{B}} a_t$, we have

$$\begin{aligned} e_t + \tfrac{\phi-\theta}{1-\phi\mathbb{B}}\mathbb{B} e_t &= \tfrac{1-\theta\mathbb{B}}{1-\phi\mathbb{B}} a_t \\ \Longrightarrow \tfrac{1-\theta\mathbb{B}}{1-\phi\mathbb{B}} e_t &= \tfrac{1-\theta\mathbb{B}}{1-\phi\mathbb{B}} a_t \\ \Longrightarrow e_t &= a_t, \end{aligned}$$

so for Equation (3.3) we have $e_t = a_t + \delta_t$.

3.8 Feedback and Feedforward Control Under First-Order Dynamics

Finally we give some results for the feedback and feedforward control under first-order dynamic system, given by Box and Luceño[10].

For the first-order dynamic model of Equation (2.2),

$$Y_t = \delta Y_{t-1} + g(1-\delta)X_{t-1},$$

where $0 \leq \delta < 1$, together with the PI control of Equation (2.4)

$$gx_t = -G[e_t + P(e_t - e_{t-1})],$$

with output error in Equation (2.3)

$$e_t = z_t + \frac{g(1-\delta)}{1-\delta\mathbb{B}}X_{t-1},$$

and feedforward disturbance model in Equation (3.1)

$$z_t - z_{t-1} = a_t - \theta a_{t-1} + \delta_t, |\theta| \leq 1,$$

Box and Luceño derived the very complicated forms of the long-run MSEO and MSEI for feedback control, and feedback plus feedforward control, in the appendix of [10].

For feedback control, we have

$$MSEO = \zeta_0 + \tfrac{\sigma_\delta^2}{T}(\tfrac{1-\phi_2}{1+\phi_2})\tfrac{1}{(1-\phi_2)^2-\phi_1^2},$$
$$MSEI = G^2\zeta_0[1 + 2P(1+P)(1-\tfrac{\zeta_1}{\zeta_0})] + \tfrac{\sigma_\delta^2}{T}(\tfrac{G}{1+\phi_2})\tfrac{1}{(1-\phi_2)^2-(\phi_1)^2},$$

where $\zeta_0 = var(e_t)$,

$\zeta_1 = Cov(e_t, e_{t-1})$,

$\phi_1 = 1 + \delta - G(1+P)(1-\delta)$,

$\phi_2 = -\delta + GP(1-\delta)$.

For feedback plus feedforward control, we have

$MSEO = \zeta_0 + \frac{\sigma_\varepsilon^2}{T}(\frac{1-\phi_2}{1+\phi_2})\frac{1}{(1-\phi_2)^2-\phi_1^2}$.

$MSEI = \frac{\sigma_m^2}{T}(1+\delta^2)(1-\delta)^2 + G^2\zeta_0[1+2P(1+P)(1-\frac{\zeta_1}{\zeta_0})] + \frac{\sigma_\varepsilon^2}{T}(\frac{G}{1+\phi_2})\frac{1}{(1-\phi_2)^2-(\phi_1)^2}\{(1-\phi_2)[(1+P)^2+P^2] - 2\phi_1 P(1+P)\}$.

Chapter 4

RANDOM STEP CHANGE DISTURBANCE MODELS

In chapter 4 we will introduce and investigate anther type of disturbance model, called random step change disturbance model. Step change (or called step shift) models have been studied by many researchers, either sustained step shift, or random step shift. They have been studied and discussed in mostly SPC area, and many control charts have been proposed for monitoring such kind of disturbance models.

In this chapter, first we will introduce the motivation for the random step change disturbance models, in a way from sustained step shift model to random step change model, and from SPC to EPC scenario.

4.1 Motivation of Random Step Change Disturbance Models

Process monitoring of sustained step shift disturbance models are discussed in some SPC literature. Vander Wiel [6] investigated the process monitoring problem of IMA process with a sustained level shift,

$$z_t - z_{t-1} = a_t - \theta a_{t-1} + D_t, \tag{4.1}$$

where

$$D_t = \begin{cases} 0, 0 \le t < t_s \\ s, t \ge t_s \end{cases}$$

and t_s is the unknown change-point. Nembhard and Valverde-Ventura [16] proposed the cumulative score (Cuscore) statistics

$$Q_t = Q_{t-1} + s\theta^{t-1} e_t$$

in their Cuscore control chart to detect a sustained step shift in the IMA disturbance.

Other researchers focused on the more challenging disturbance models subject to random step changes instead of the sustained shift, which are usually caused by variations in the physical conditions, such as the environmental temperature and raw material qualities. Chen and Elsayed [17] studied using an EWMA estimator to monitor the i.i.d normally distributed process with random step changes. In their model, the random step-change occurrence has a constant probability p which is independent of the prior history of the process, with size

$$r = \frac{\tau}{\sigma}, \tag{4.2}$$

where σ is the standard deviation of the background normal process, and τ is the standard deviation of the process mean. They proposed an EWMA estimator with closed-form expression for the optimal value of the weighted variable λ, as a function of the estimates $\hat{p}$ and $\hat{r}$ derived from historical data.

Tsiamyrtzis and Hawkins [18] investigated the process monitoring of a mean drift model of AR(1) process subject to random step changes, in a Bayesian framework. They supposed that the process mean has a jump of size δ that occurs with probability p, and assumed that the prior information about the process mean is available. Then at each time when the new data comes, they get the posterior distribution for the process mean through Bayes theorem to check if the mean has drifted or not. If there is no significant change, they use this posterior as the prior for the next stage. Their model is suitable for some practical problems, for example, tool wear problems in which the wear incorporates a random step change (due, e.g., to tool chipping) as well as drift. They also generalized the model with assigning a prior distribution to the size of the jumps δ, with AR(1) process subject to random step changes [13].

Since "a step change is more difficult to see when buried in an IMA than when buried in iid noise" [6] or a stationary background noise, now we further generalize the background disturbance to be the nonstationary IMA model subject to random step changes,

$$z_t - z_{t-1} = a_t - \theta a_{t-1} + D_t, \tag{4.3}$$

$$D_t \sim \begin{cases} N(0, \sigma_a^2), & \text{with probability} 1 - p, t = 1, 2, ... \\ N(\delta, \sigma_a^2), & \text{with probability} p \end{cases}$$

and σ_a is the standard deviation of the white noise $\{a_t\}_{t\geq 1}$. Since we do not know the actual size of the step shift δ, we assume it is random with a certain prior distribution:

$$\pi(\delta) \sim N(\mu_\delta, \sigma_\delta^2), \tag{4.4}$$

where μ_δ and σ_δ^2 are the expected mean and variation of the step shift size δ respectively. Both μ_δ and σ_δ^2 are assumed to be known, considering that they can be determined from previous engineering knowledge or experiences about the process. Small values of σ_δ^2 lead to informative settings while large values indicate a priori ignorance for the size of the step shift. The size of the step shift can be modeled by δ and the uncertainty of the step shift can be modeled by the standard deviation ratio $r = \sigma_\delta/\sigma_a$. In this thesis, we call this disturbance model to be model (4).

In the process monitoring problem, the proposal for the sizes of μ_δ is that they need to be "big enough" to be of practical interest, but at the same time "small enough" not to be obvious visually, followed by the suggestion of Tsiamyrtzis and Hawkins [13]. However, in our process adjustment scenario, we do not need to have such restrictions, and all sizes of the step shift can be under our investigation.

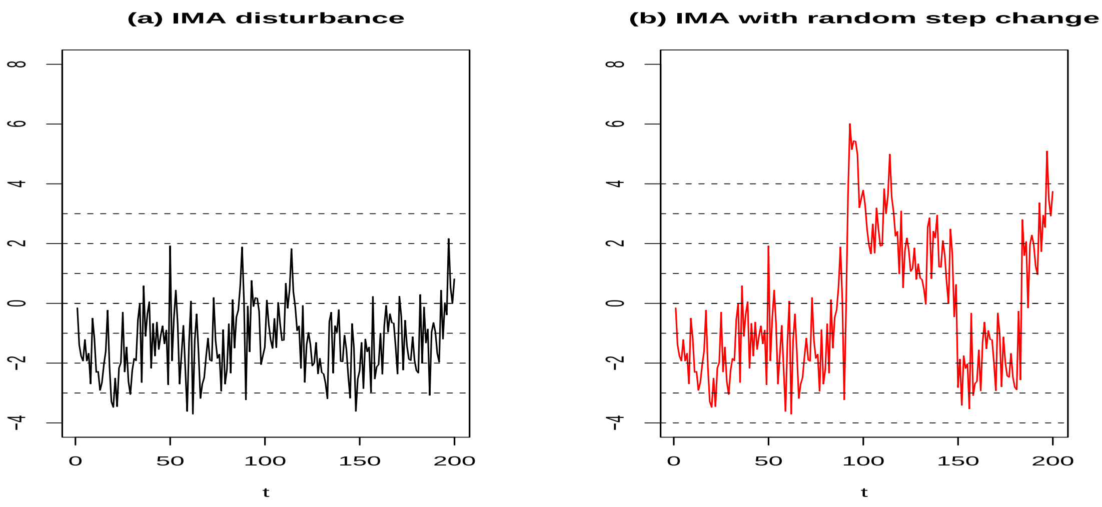

Figure 4.1: IMA disturbance without and with random step change

Figure 4.1(a) shows an IMA disturbance process of 200 data with $\theta = 0.8$, and Figure

4.1(b) shows the same IMA disturbance subject to random step change with the occurrence probability $p = 0.02$. The size of the step shift δ has the prior distribution $\pi(\delta) \sim N(\mu_\delta, \sigma_\delta^2)$ with $\mu_\delta = 2$ and $\sigma_\delta = 3$, so the standard deviation ratio $r = \sigma_\delta/\sigma_a = 3$. There are 4 random step changes: a 5.25 upward step change in the 92th value of the IMA disturbance z_t, a 2.08 downward step change in the 101th value, a 3.72 downward step change in the 150th value, and a 3.49 upward step change in the 184th value.

4.2 Control Equations for Random Step Change Models

For model (4), we introduce a process adjustment procedure based on feedback control and an added adjustment based on output errors monitoring. They proposed feedback control $gx_t = -Ge_t$, plus added adjustments as soon as a possible random step change is detected, i.e., when an output error e_t falls outside of the 3σ control limits, with control equation

$$gx_t = \begin{cases} -(e_t - 3\sigma_a), & \text{if} e_t > 3\sigma_a \\ -(e_t + 3\sigma_a), & \text{if} e_t < -3\sigma_a \end{cases}$$

This added adjustment is applied to compensate the possible random step change, and can be considered as *quasi-feedforward control*, and the overall combined control was called as *quasi-feedback feedforward control*. The rationale for this quasi-feedforward control is due to the reason that the feedback control equation is only for the IMA disturbance, and it is not sufficient when other disturbances are present (random step change here). The amount of the quasi-feedforward control is chosen to be the distance between output error e_t and the 3σ control limit on the same side of the control chart.

It is called quasi-feedforward control, since it is based on neither pure output error, nor the direct measure of the random step change; instead, it is based on the outliers in the sequence $\{e_t\}_{t\geq 1}$ which are outside of the 3σ control limits under feedback control, and these outliers call for further remedial actions. Thus such added adjustment goes beyond feedback control and does not meet the requirement for feedforward control directly, so in some sense, it is in between the traditional feedback control and feedforward control categories, which justifies such a name. Box and Kramer [2] discussed a similar process subjected to feedback control to show that the feedback control does not necessarily conceal the nature of the disturbance that is being compensated. Here we go a further step beyond them by applying an added adjustment to continue improving the process.

4.3 MSEO Results for Random Step Change Models

Now we present some numerical results on the MSEO under combined control (quasi-feedback feedforward control) in our random step change model (4).

Most industrial time series of IMA disturbances have $0.6 \leq \theta \leq 0.8$ [1], so we choose $\theta = 0.8$ and $\sigma_a = 1$ in model (4). We consider two occurrence probabilities $p = 0.02$ and $p = 0.05$ for the random step change, and choose some different combinations for the prior distribution of the step change $\pi(\delta) \sim N(\mu_\delta, \sigma_\delta^2)$. We investigate the expected size of the step shift μ_δ from the integer values between -3 and 3, for both upward change and downward change, and the uncertainty of the step shift $r = \sigma_\delta/\sigma_a$ from the integer values between 1 and 3. Due to symmetry, we only need to include $0 \leq \mu_\delta \leq 3$ in our numerical results summary.

We choose the sample size $n = 200$ in each simulated disturbance process and make $10,000$ iterations for each simulated disturbance process in our simulation. We apply two control methods: feedback control and combined control, for the same disturbance process data, and compare their MSEOs results. In our numerical results summary, we show both the MSEO under combined control and the improvement percentage for the quasi-feedforward control, which is calculated by

$$\frac{MSEO_{FB} - MSEO_{comb}}{MSEO_{FB}}, \tag{4.5}$$

where $MSEO_{FB}$ is the MSEO under feedback control and $MSEO_{comb}$ is the MSEO under combined control.

The numerical results are shown in Table 4.1, notice that the improvement percentage is given in the parenthesis followed by the MSEO under combined control. Since there are three parameters μ_δ, σ_δ and p, we will fix each two out of the three parameter values, and compare the MSEO and the improvement percentage for different values of the other parameter. Some useful conclusions can be drawn from Table 4.1:

Table 4.1: MSEOs under combined control for different p, μ_δ and σ_δ in model (4) ($\theta = 0.8$)

μ_δ	σ_δ	p=0.02	p=0.05
3	3	1.51 (20.6%)	2.31 (36.7%)
3	2	1.42 (16.1%)	2.14 (32.0%)
3	1	1.40 (11.5%)	2.06 (25.2%)
2	3	1.40 (15.5%)	2.01 (29.2%)
2	2	1.30 (9.8%)	1.78 (21.5%)
2	1	1.23 (4.6%)	1.63 (12.2%)
1	3	1.32 (11.7%)	1.81 (22.0%)
1	2	1.20 (5.2%)	1.50 (12.0%)
1	1	1.10 (1.0%)	1.27 (3.3%)
0	3	1.30 (10.2%)	1.73 (19.3%)
0	2	1.16 (3.7%)	1.42 (7.8%)
0	1	1.05 (0.2%)	1.13 (0.6%)

(1) For the disturbance model with the same prior distribution of the random step change $N(\mu_\delta, \sigma_\delta^2)$, the higher the occurrence probability p, the larger improvement percentage for the quasi-feedforward control can be achieved. However, the disturbance with a larger p results in a larger MSEO under combined control.

(2) For the disturbance model with the same expected size of the random step change μ_δ and the occurrence probability p, the larger the uncertainty r (equivalently σ_δ) of the random step change, the larger improvement percentage for the quasi-feedforward control can be achieved. However, the disturbance with a larger r results in a larger MSEO under combined control.

(3) For the disturbance model with the same uncertainty of the random step change r (equivalently σ_δ) and the occurrence probability p, the larger the expected size of the random step change μ_δ, the larger improvement percentage for the quasi-feedforward control can

be achieved. However, the disturbance with a larger μ_δ results in a larger MSEO under combined control.

All the conclusion are consistent with our intuition: the larger the μ_δ, r or p, i.e., the more information on the random step change, which is the informative variable, can be decomposed from the total disturbance, the more improvement can be achieved from feedforward control application. However, the larger and more frequent random step change always potentially increases the process variability, even after adjusted by feedback and feedforward control, resulting in a larger MSEO. This explains the reason for the interesting point that a lager improvement percentage always corresponds to a larger MSEO under combined control.

4.4 Type I and Type II Errors in Process Monitoring and Adjustment

As we mentioned at the beginning that in SPC, the process monitoring problem can be considered as a hypothesis test in statistical area. It is known that for any hypothesis test, the type I and type II errors always exist. Therefore, the type I and type II errors also exist in control charts, with the following explanations:

• Type I error: Ascribe a variation or a mistake to a special cause (assignable cause) when in fact the cause belongs to the system (common cause).

• Type II error: Ascribe a variation or a mistake to the system (common causes) when in fact the cause was a special cause.

Table 4.2: Type I and Type II Error Probabilities in Process Monitoring

	True Model Common Cause	True Model Special Cause
Detect Common Cause	$1-\alpha$	β
Detect Special Cause	α	$1-\beta$

Table 4.2 illustrates the four possible cases for true model and detected model in SPC with their probabilities expressed by using type I error α and type II error β. We can see

that for a certain true model, either with common cause or special cause, control charts can detect either correctly or incorrectly, with total probability 1. In SPC, a rule of thumb for different control charts comparison is that with the same in-control average run length (ARL), they try to achieve the minimum out-of-control ARL. The control limits are adjusted by controlling the type I error to be $\alpha = 0.05$.

In SPC, the type I error α is also called the false alarm rate, meaning that the plotted point falls outside of the control limits when the process is actually in control. People always try to achieve the smallest type II error (equivalent to the minimum out-of-control ARL), while controlling the type I error to be a fixed value like $\alpha = 0.05$.

In process adjustment scenario, type I and type II errors have rarely been discussed in the EPC literature. Notice that here the type I and type II errors only related to the hypothesis test problem in process monitoring, in which we try to detect a special cause, like a step change, and it is not related to the case of Deming's funnel experiment, in which the "leave it alone" no adjustment rule is better than the active adjustment strategy. Therefore these two types of errors related to the process monitoring problem that exists in some process adjustment procedures.

The lack of discussion on type I and type II errors in EPC might due to the fact that most disturbance models under investigation belong to a particular kind of time series model, like a nonstationary IMA model, or a stationary AR(1) or ARMA (1,1) model, etc. For such kind of disturbances, the issue of type I and type II errors does not exist at all, since they all require continuous adjustment other than the periodic adjustment. Actually, as we mentioned at the beginning of Chapter 3, the true disturbance in many real applications is not a particular time series model itself, instead, it often includes a particular time series process as the background disturbance, together with some other additional part, like a spike, a mean sustain shift, a ramp, an exponential rise to new levels, etc. For example, we investigated periodic shift disturbance models in Chapter 3, with IMA, AR(1) and ARMA(1,1) as the background disturbances, and in Chapter 4 we introduced the random step change disturbance model with the IMA background disturbance.

Another reason that type I and type II errors have not been well investigated under process adjustment scenario might be, due to the functionality of process adjustment. And since we can only see the process after adjustment other than before adjustment, which is the output error series $\{e_t\}_{t\geq 1}$ other than the disturbance $\{z_t\}_{t\geq 1}$, so we can only observe the process under control and the disturbance cannot be directly seen. When the size of the step change is small, or when there are very few change-points, the signal (i.e., the nature of the disturbance) is more likely to be hidden in the the output error series $\{e_t\}_{t\geq 1}$ after adjustment.

However, as we mentioned before, the only aim of process adjustment is to reduce the process variation through adjusting the process, i.e., to achieve the minimum MSEO, so no matter how many observations will result in type I and type II errors, as long as we can have small MSEO, it will be enough for our process adjustment.

Also, we are not certain on the relationship between the type I, II errors and MSEO under process adjustment, which might be a very complicated issue. Intuitively, less type I and type II errors might achieve smaller MSEO, but besides these two types of errors, the effects of the type I and type II errors on MSEO also relate to the magnitude of the step change in the random step change model (4). If a large step change is not detected (type II error), then it might substantially increase the MSEO, while a smaller step change might not have much influence.

4.5 Type I and Type II Errors in Random Step Change Disturbance Models

In the control equations we proposed for the random step change disturbance model, the quasi-feedforward control is applied through monitoring the output error series $\{e_t\}_{t\geq 1}$ using the 3σ control limits. Here the process monitoring serves as a complementary tool for the feedforward control application on this disturbance model, therefore the issue of type I and type II error exists in such process adjustment procedures.

All the possibilities for the type I and type II errors in process adjustment scenario are listed in Table 4.3. If the random step change occurs, the output error e_t may still fall inside of the 3σ control limit, resulting in a type II error with probability β. Similarly, the output error e_t may fall outside of the 3σ control limit, even though there is no change occurs at all, resulting in a type I error with probability α. Intuitively, we should try to maximize the change of correctly detecting a step change when it occurs, which has the probability $1-\beta$. Meanwhile, it would be very desirable that we do not make the feedforward control whenever the random step change does not occur.

Table 4.3: Type I and Type II Errors in Process Adjustment

	True Model No Change	True Model Change
Detect No Change	$1-\alpha$	β
Detect Change	α	$1-\beta$

4.6 Generalized Type II Errors in Random Step Change Disturbance Models

Recall that in SPC, the average run length (ARL), defined as the number of points that, on average, will be plotted on a control chart before an out of control condition is indicated, is usually applied as the comparison criterion for different control charts.

By applying the feedback and quasi-feedforward control strategies in our random step change disturbance model, it might take a while to detect the random step change after it happens, so the feedforward control might be applied a few observations later after the true step change location. In our simulations, we found this phenomenon is very frequent.

We relax the definition of type II error a little bit, by choosing five observations interval after the true step change location, and do not consider a feedforward control made during this time interval as a type II error. We define the **generalized** type II error if we fail to detect a random step change (not make a feedforward control) within the next five observations after it actually happens.

Notice that there is no generalized type I error in the similar sense, since once it gives a false alarm of a step change, it will trigger a feedforward control.

This newly defined generalized type II error is very helpful for us to evaluate the influence of type II errors on the control performance, i.e., output mean square error, since it successfully separates the two cases that it fails to detect a random step change completely, or fails to detect a step change only when it happens but successfully track it later.

4.7 *Evaluation of Type I and Generalized Type II Errors in Random Step Change Disturbance Models*

Now we investigate the issue between our proposed process adjustment efficiency and type I, (generalized) type II errors in the random step change disturbance models. We investigate the random step change model (4) with $\theta = 0.8$ in the IMA background disturbance, with two random step change occurrence probabilities $p = 0.02$ and $p = 0.05$, and the magnitude of the random step change has a prior distribution $\pi(\delta) \sim N(\mu_\delta, \sigma_\delta^2)$. We consider the expected size of the step shift μ_δ between 1 and 3 for upward change, and the uncertainty of the step shift $r = \sigma_\delta/\sigma_a$ between integer values between 1 and 3.

Table 4.4: The Number of Type I, Generalized Type II Errors in Model (4) ($p = 0.02$)

μ_δ	σ_δ	α error No.	β error No.	$1-\beta$	β_G error No.	$1-\beta_G$	MSEO
3	3	1.49	1.85	54.6%	1.49	62.8%	1.51
2	3	1.39	2.25	44.1%	1.84	54.0%	1.40
1	3	1.31	2.51	37.9%	2.14	46.4%	1.32
3	2	1.57	1.91	52.8%	1.46	63.5%	1.42
2	2	1.35	2.58	36.5%	2.14	46.5%	1.30
1	2	1.10	3.06	24.4%	2.70	32.5%	1.20
3	1	1.80	1.84	53.6%	1.24	69.0%	1.40
2	1	1.28	2.90	27.9%	2.36	41.0%	1.23
1	1	0.84	3.65	10.0%	3.33	16.8%	1.10

We choose the sample size $n = 200$ in each simulated disturbance process and make $10,000$ iterations for each simulated disturbance process. We record the average number of type I error observations (α error No.), the average number of type II error observations (β

error No.), the average number of generalized type II error observations (β_G error No.), and correct detection percentage of the step change observations, which are $1-\beta$ and $1-\beta_G$. The MSEOs are also added into the last columns of Table 4.4 and Table 4.5. All the numerical results for model (4) with random step change occurrence probabilities $p = 0.02$ and $p = 0.05$ are shown in Table 4.4 and Table 4.5 respectively.

Table 4.5: The Number of Type I, Generalized Type II Errors in Model (4) ($p = 0.05$)

μ_δ	σ_δ	α error No.	β error No.	$1-\beta$	β_G error No.	$1-\beta_G$	MSEO
3	3	2.77	4.51	54.6%	3.40	66.0%	2.31
2	3	2.58	5.53	44.6%	4.40	56.0%	2.01
1	3	2.37	6.26	37.3%	5.06	49.4%	1.81
3	2	3.04	5.49	45.1%	3.24	67.6%	2.14
2	2	2.52	6.28	37.3%	4.98	50.2%	1.78
1	2	1.91	7.58	24.0%	6.44	35.6%	1.50
3	1	3.55	4.14	58.6%	2.45	75.4%	2.06
2	1	2.45	6.87	31.7%	5.16	48.4%	1.63
1	1	1.29	8.95	10.5%	8.03	19.7%	1.27

We can find that for the same prior distribution of the random step change $\pi(\delta) \sim N(\mu_\delta, \sigma_\delta^2)$, with our new definition of the generalized type II error, its number reduces substantially compared with the traditional type II error.

Also, in the prior distribution of the random step change $\pi(\delta) \sim N(\mu_\delta, \sigma_\delta^2)$, for the same σ_δ, when μ_δ increases, type I error α increases but (generalized) type II error β decreases. It is easy to understand, since when the step change has bigger magnitude μ_δ, it is easier to be detected, so that the probability of failed to detect the step change when it occurs is becoming smaller, i.e., (generalized) type II error is becoming smaller. Instead, if the step change has smaller magnitude μ_δ, the probability that they fail to detect the step change when it occurs is becoming larger, in other words, the step change is more easily to be hidden, i.e., (generalized) type II error is becoming larger.

Similarly, the numerical results for occurrence probability $p = 0.05$ are given in Table 4.5. All the other parameters are the same as those in Table 4.4.

In Table 4.4 and Table 4.5, for type I and (generalized) type II errors, it looks like larger type I error usually results in a larger MSEO, which means that if we mistakenly detect a random change when it actually does not happen, and make the unnecessary quasi-feedforward control, then it will increase the MSEO. Instead, for the (generalized) type II error, i.e., if we fail to detect a random change when it actually occurs and do not make adjustment, since the MSEO decreases (a sign of improvement) when it increase, so it seems that the (generalized) type II error does not influence the MSEO substantially.

However, we should be very cautious about this conclusion, since there are many other factors that might influence the control performance and the type I, (generalized) type II errors. For example, in our random step change model, since the magnitude of the step change has a prior normal distribution, so its size is not a fixed value. Sometimes the step change can be large and sometimes it can be small, so on any two observations when the same type of error happens, their effects on MSEO are usually different. For instance, a larger random change must have more influence on the MSEO compared to s smaller one. In order to evaluate it more carefully, we need to make more investigations on this issue, we need to fix the size of the random step change.

4.8 Evaluation of Type I and Generalized Type II Errors for Fixed Size Random Step Change

Now we investigate the effects of Type I and generalized Type II errors on MSEO, by using the fixed size step change with IMA background disturbance model, with the step change size ranges from $\delta = 1$ to $\delta = 4$. It is obvious that this is just the degenerated model of disturbance model (4) with $\sigma_\delta = 0$. In Table 4.6 we provide the numerical results for model (4) with random step change occurrence probabilities $p = 0.02$, with the same parameters settings in our simulations, i.e., $\theta = 0.8$ in the IMA background disturbance, sample size $n = 200$ in each simulated disturbance process and make $10,000$ iterations for each simulated disturbance process.

Table 4.6: Type I and Generalized Type II Errors for Random Step Change Model with Fixed Change Size ($p = 0.02$)

α error No.	$\delta = 1$	$\delta = 2$	$\delta = 3$	$\delta = 4$
0	1.03	1.13	1.22	1.32
1	1.08	1.21	1.33	1.44
2	1.14	1.28	1.42	1.56
3	1.19	1.36	1.51	1.66
4	1.24	1.41	1.61	1.78
5	1.25	1.48	1.67	1.88
6	NA	1.55	1.75	1.99
0	1.022	1.12	1.35	1.57
1	1.017	1.15	1.42	1.66
2	1.042	1.20	1.48	1.83
3	1.051	1.25	1.54	2.19
4	1.063	1.29	1.63	NA
5	1.078	1.34	1.687	NA
6	1.092	1.40	1.694	NA

The numerical results in Table 4.6 show that for big step changes, the generalized type II error has more effects on MSEO, with a large MSEO range from 1.57 to 2.10 when $\delta = 4$ for 0 to 3 generalized type II errors, while the MSEO range for 0 to 3 type I errors is only from 1.32 to 1.66, which means that we should not ignore the possible big step change without applying feedforward control. For small step changes, the type I error has more effects on MSEO while the MSEO range of different generalized type II is very small (from 1.022 to 1.092 when $\delta = 1$), which suggests that if there is no obvious evidence of a step change, we should not applied feedforward control to temper the relatively stable process under feedback control, which is also consistent with the conclusion we get from the Deming's funnel experiment.

Notice that a few NA values exist in Table 4.6, which means that such type I or generalized type II errors never happens in our simulations with $10,000$ iterations. For example, for the small size random change $\delta = 1$, there is no case for more than 6 type I errors among all the $10,000$ iterations in our simulation, and there is no case for more than 3 generalized type II errors among all the $10,000$ iterations in our simulation. This is easy to be explained by our intuition, since for small size random step changes, it will be less likely that it falls outside the 3σ control limits many times, which causes a type I error; similarly, for large size random step changes, it will be less likely that it completely fails to detect such step changes many times, which causes a generalized type II error.

Chapter 5

SOME COMMENTS ON PROCESS MONITORING AND FEEDBACK, FEEDFORWARD CONTROL

There are some relevant issues on SPC and EPC that might cause misunderstanding, confusion and controversies from researchers and practitioners. Now we discuss these topics in greater detail and further illustrate the rationales behind our new perspectives on the process monitoring, feedback and feedforward control framework.

5.1 When SPC is Not Enough

As we know, SPC provides an ongoing check on the stability of a process by using control charts to identify variation which are due to special causes. However, the process stability is not the only thing we need to care for, since a stable process might not be a capable process, for example, if the process mean is not on target, or if the process variation is too large. Under such cases, we need to change the process so it can meet the customer requirements, instead of maintaining the stability of this noncapable process. This calls for process adjustment strategies like feedback control and feedforward control to make the process on target and reduce variation.

5.2 When Feedforward Control Application is Possible?

Koontz and Bradspies [3] claimed that “even the most enthusiastic proponents of feedforward control admit that, if input variables are not known or unmeasurable, the system will not work.” Notice that their input variables just correspond to our informative variables B. However, the ideal assumption of absolute certainty on the informative variables have practical limitations, and sometimes parameter uncertainty should be treated as additional source of variability, possibly due to a poor understanding of the process behavior.

From our disturbance decomposition viewpoint, the above statement of Koontz and Bradspies [3] can be relaxed in some sense: as long as we have some knowledge on the informative variable B, either complete knowledge or partial knowledge, feedforward control or at least quasi-feedforward control is possible. For example, in model (4), without knowing the exact size of the random step change each time, we made the realistic assumption that with some empirical knowledge on the random step change, its occurrence probability p is known, and its prior distribution is normally distributed with some known mean and variance, following the assumption of Tsiamyrtzis and Hawkins [13].

5.3 Disturbance Nature and Change-point Identification in EPC

Box and Kramer [2] argued that the feedback control does not necessarily conceal the nature of the disturbance. However, sometimes feedback control does conceal it. This is due to the objective of EPC, and we will explain it as follows.

In SPC, when we monitor a stationary or nonstationary process, we can either monitor the forecast errors through fitting a time series model, or monitor the original observations (disturbance) directly. For different disturbance models, no method is uniformly better than the other. When there is no advantage to using forecast errors, directly monitoring the raw data is much more convenient.

However, when we need to adjust such a process in EPC, usually we cannot monitor a process without making any adjustment and wait until a signal appears; instead, we need to adjust the process from the beginning (usually by feedback control) and try to make it on target with minimum variation as much as possible, so we can only observe the process under control and the disturbance cannot be directly seen. When the size of the step change is small, or when there is only one change-point, the signal is more likely to be hidden in the process after adjustment. Therefore sometimes the nature of the disturbance might be concealed by feedback control.

Fortunately, process adjustment usually does not require the identification of the change-point, for example, the EWMA forecast (feedback control) filters out the noise and gives a clear picture of how the true mean level varies [17]. In other words, as long as the process under control has small variation (i.e., MSEO), then it is a good control, and usually we do not need to worry about how the process looks like without control.

However, in some particular circumstances we can benefit substantially from the change-point identification in EPC, and this leads us to the next discussion topic.

5.4 The Role of Process Monitoring in the Integration of SPC and EPC

Process adjustment itself usually is insufficient if we have high uncertainty on the disturbance model, for example, when the disturbance has *multiple* change-points, the size of the step change is unknown or under high uncertainty. Under such circumstances, process adjustment needs to be combined with process monitoring to achieve higher efficiency, i.e., smaller MSEO.

In model (4), we only know the occurrence probability of the random step change and its prior distribution, but we do not know the more detailed information about the informative variable, i.e., the random step change, including the exact change-points locations and the magnitude of every random step change. For such case, control chart for process monitoring can be used as a supplemental tool for locating the possible change-point, serving for the further process improvement function. This idea of "process monitoring serves for better process adjustment" was advocated by Tucker long time ago when feedback control was first introduced by Box and his collaborators in early 1960s [20], by arguing argued that "the development of a proper monitoring function in the presence of feedforward and/or feedback control is (should be) a central research issue in the quest for continuous quality improvement". This is exactly our motivation for proposing the quasi-feedforward control strategy in our random step change disturbance model (4).

Since in SPC, false alarms and failure of detection always exist due to the type I and type II errors, so similar problems exist in the integration of SPC and EPC as well. Intuitively, with larger p, μ_δ and r in our random step change disturbance model (4), our quasi-feedback feedforward control tends to have potentially higher efficiency.

5.5 Disturbance Decomposition and Cuscore Charts

While this disturbance decomposition viewpoint is relatively new in EPC for feedforward control application, similar ideas can be found on some disturbance models in SPC. Sometimes people have experience on how the process will change so that the signal patterns are anticipated, and Cuscore chart is devised to detect such anticipated systematic *signals* hidden in certain *noise* for process monitoring function. Box and Luceño [1] discussed a variety of different signal and noise combinations for Cuscore chart monitoring. Very naturally, the signal and the noise monitored by the Cuscore chart in SPC just parallel the informative variable B and the background disturbance e' in EPC from our disturbance decomposition viewpoint.

5.6 Disturbance Decomposition and Robust Parameter Design

One of our motivations on disturbance decomposition viewpoint in EPC comes from robust parameter design. Now we explain how the rationale behind our feedforward control idea can be unified in both robust parameter design and EPC framework.

It is well known that robust parameter design can be successful only if the control factors X are interacting with the noise factors e, which is often called "signal-by-noise interaction". This implies that after disturbance decomposition, the control factors X should also interact with the informative variable B as well as the noninformative noise e', while the B and e' are independent of each other.

It can be proved that these conditions are satisfied in our disturbance decomposition model as well. We only investigate model (1) here, since model (2) and model (3) can be similarly justified.

We show that X_t interacts with δ_t and the IMA noise. Since the feedforward control is applied to compensate the effect of δ_t, so different levels of δ_t call for different levels of the adjustment amount x_t (equivalently X_t), with control equation

$$gx_t = -m_t = -(\delta_t + \epsilon_t),$$

thus X_t definitely interacts with δ_t. Similarly the feedback control is intended to compensate the IMA process, with control equation

$$gX_t = -\tilde{z}_t = -(1-\theta)(z_t + \theta z_{t-1} + \theta^2 z_{t-2} + ...),$$

where the EWMA forecast $\tilde{z}_t$ is a function of the current and all the previous $\{z_t\}$'s, so x_t (equivalently X_t) and z_t are interacting to each other as well. Finally, since model (1) is an additive model, so the decomposed informative variable δ_t and the IMA noninformative noise are independent of each other.

Chapter 6

RESEARCH CONTRIBUTIONS

This chapter summarizes the research contributions of this thesis. The contributions of this thesis are summarized as follows:

- A new point of view for the feedforward control problem was presented based on a disturbance decomposition approach, which reveals the basic causations among different variables in our feedforward feedback control disturbance model with a more clear illustration. This new idea of feedforward control is also illustrated by some real application examples.

- Periodic shift models with different (statinoary and nonstationary) background disturbances have been investigated. The closed-form feedback and feedforward control equations for each disturbance model have been proposed, and the closed-form MSEO expressions are derived. Performance analysis, sensitivity analysis and robust analysis have been made.

- Random step change disturbance models have been investigated, with nonstationary IMA background disturbance. Both feedback and quasi-feedforward control equations have been proposed, and the MSEO results are derived through simulations. Type I and type II errors in process adjustment procedure for this random step change disturbance models have been investigated, and their relationship with the control performance have been evaluated.

- Some critical issues on process monitoring and process adjustment have been discussed, and some insights on both SPC and EPC, as well as other related topics have been gained through those discussions.

BIBLIOGRAPHY

[1] L.C. Alwan and H.V. Roberts. Time-series modeling for statistical process control. *Journal of Business and Economic Statistics*, 6:87–95, 1988.

[2] K. J. Astrom. *Introduction to stochastic control.* Mathematics in Science and Engineering Series, Academic, New Work, 1970.

[3] G. A. Barnard. Control charts and stochastic processes. *Journal of the Royal Statistical Society, Series B*, 21:239–271, 1959.

[4] G. E. P. Box and G.M. Jenkins. Some statistical aspects of adaptive optimization and control. *Journal of the Royal Statistical Society, Series B*, 24:297–343, 1962.

[5] G. E. P. Box, G.M. Jenkins, and G.C. Reinsel. *Time series analysis: forecasting and control.* Wiley, 4th edition, 2008.

[6] G. E. P. Box and Tim. Kramer. Statistical process monitoring and feedback adjustment-a discussion. *Technometrics*, 34:251–285, 1992.

[7] G. E. P. Box, A. Luceno, and Maria del Carmen Paniagua-Quinones. *Statistical Control by Monitoring and Adjustment.* Wiley, 2nd edition, 2009.

[8] G. E. P. Box and Alberto Luceno. Selection of sampling interval and action limit for discrete feedback adjustment. *Technometrics*, 36:369–378, 1994.

[9] G. E. P. Box and Alberto Luceno. *Statistical Control by Monitoring and Feedback Adjustment.* Wiley, 1997.

[10] G. E. P. Box and Alberto Luceno. Feedforward as a supplement to feedback adjustment in allowing for feedstock changes. *Journal of Applied Statistics*, 29:1241–1254, 2002.

[11] G. E. P. Box and S. Narasimhan. Rethinking statistics for quality control. *Quality Engineering*, 22:60–72, 2010.

[12] E. D Castillo. *Statistical Process Adjustment for Quality Control*. Wiley, 2002.

[13] E. D. Castillo, R. Pan, and B.M. Colosimo. A unifying view of some process adjusmtnet methods. *Journal of Quality Technology*, 35:286–293, 2003.

[14] S. V. Crowder. An spc model for short production runs: minimizing expected cost. *Technometrics*, 34:64–73, 1992.

[15] W.E. Deming. *Out of Crisis*. Cambridge, MA, 1986.

[16] Bernard Friedland. *Control System Design: An Introduction to State-Space Methods*. McGraw-Hill, 2005.

[17] F.E. Grubbs. An optimum procedure for setting machines or adjusting processes. *Journal of Quality Technology*, 15:186–189, 1983.

[18] T. J. Harris. Optimal controllers for nonsymmetric and nonquadratic loss functions. *Technometrics*, 34:298–306, 1992.

[19] K. L. Jensen and S. B. Vardeman. Optimal adjustment in the presence of deterministic process drift and random adjustment error. *Technometrics*, 35:376–389, 1993.

[20] R.E. Kalman. A new approach to linear filtering and prediction problems. *Transaction ASME, Journal of Basic Engineering*, 82:35–45, 1960.

[21] Myer Kutz. *Handbook of Farm, Dairy, and Food Machinery*. William Andrwe Publishing, 2008.

[22] Bela G. Liptak. *Instrument Engineers' Handbook, Fourth Edition, Volume Two: Process Control and Optimization*. Taylor and Francis, 2nd edition, 2005.

[23] A. Luceno. Performance of discrete feedback adjustment schemes with dead band, under stationary versus nonstationary stochastic disturbance. *Technometrics*, 40:223–233, 1998.

[24] J.F. Macgregor. A different view of the funnel experiment. *Journal of Quality Technology*, 22:255–259, 1990.

[25] D. C Montgomery. *Introduction to Statistical Quality Control*. Wiley, 6th edition, 2009.

[26] D.C. Montgomery and C.M. Mastrangelo. Some statistical process control methods for autocorrelated data (with discussion). *Journal of Quality Technology*, 23:179–204, 1991.

[27] J.F. Muth. Optimal properties of exponential weighted forecasts of time series with permanent and transitory component. *Journal of the American Statistical Association*, 55:299–306, 1960.

[28] E. S. Page. Continuous inspection schemes. *Biometrika*, 41:100–115, 1954.

[29] E. S. Page. On problems in which a change in a parameter occurs at an unknownpoint. *Biometrika*, 44:248–252, 1957.

[30] E. S. Page. Cumulative sum charts. *Technometrics*, 3:1–9, 1961.

[31] R. Pan. *Statistical process adjustment methods for quality control in short-run manufacturing*. Unpublished PhD Thesis, Department of Industrial and Manufacturing Engineering, Penn State University, 2002.

[32] S.M. Pandit and S.M. Wu. *Time series and system analysis with applications*. Wiley, 1983.

[33] R.L. Plackett. Some theorems in least squares. *Biometrika*, 37:149–157, 1950.

[34] W.A. Shewart. *Economic control of quality of manufacturing product.* Princeton, NJ, 1931.

[35] J. D. Spurrier and L.A. Thombs. Control charts for detecting cyclical behavior. *Technometrics*, 32:163–171, 1990.

[36] G. Taguchi. *System of experimental design.* Unipub, Kraus International Publications, White Plains, NY, 1987.

[37] L. G. Tatum. Control charts for the detection of a periodic component. *Technometrics*, 38:152–160, 1996.

[38] D. Trietsch. The harmonic rule for process setup adjustment with quadratic loss. *Journal of Quality Technology*, 30:75–84, 1998.

[39] F. Tsung, H. Wu, and V. N. Nair. On the efficiency and robustness of discrete proportional-integral control schemes. *Technometrics*, 40:214–222, 1998.

[40] S.A. Vander Wiel. Monitoring processes that wander using integrated moving average models. *Technometrics*, 38:139–151, 1996.

[41] S.A. Vander Wiel, F.W. Faltin, G.J. Hahn, and W.T. Tucker. Algorithmic statistical process control: concepts and an application. *Technometrics*, 34:286–297, 1992.

[42] N.F. Zhang. A statistical control chart for stationary process data. *Technometrics*, 40:24–38, 1998.

[43] N.F. Zhang and J.F. Pollard. Analysis of autocorrelations in dynamic processes. *Technometrics*, 36:354–368, 1994.

VITA

Lihui Shi was born in Handan, Hebei, China. He earned a Bachelor of Science degree in Information and Computational Science in 2004 at Hebei University of Technology, Tianjin, China, and a Master of Science in 2007 in Statistics from Nankai University, Tianjin, China. He went to the United States to continue his education in University of Washington at Seattle since fall 2007. In 2009 he earned a Master in Industrial and Systems Engineering at University of Washington. In 2012 he earned a Doctor of Philosophy at the University of Washington in Industrial and Systems Engineering, and a Master of Science in Statistics from University of Washington.

Lihui Shi has been worked in Webtrends Inc. at Seattle as a research analytic intern for one year, and he started his full time position there as a research scientist since June 2012.

Lihui Shi's main research areas are quality, reliability and applied statistics, including engineering process control, statistical process control, design of experiments, time series analysis, etc.