

## Chapter 2

# Assessment Based on Minimum-Variance Principles

An important goal of quality improvement in manufacturing is the reduction of variability in product attributes. Producing more consistent output improves product performance and may reduce manufacturing costs. Therefore, the standard control performance assessment methods are based on the MV principle or modifications of it. The key point is that the MV benchmark (as a reference performance bound) can be estimated from routine operating data without additional experiments, provided that the system delay is known or can be estimated with sufficient accuracy.

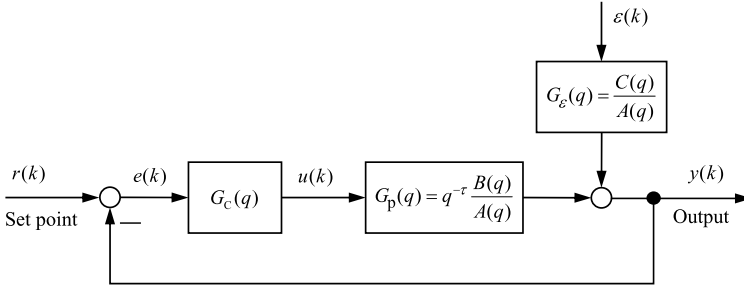
This chapter provides an introduction to the theory of MV performance assessment. Some basic notations and concepts are given in Sect. 2.1. The derivation of minimum variance control is recalled in Sect. 2.2. In Sect. 2.3, the autocorrelation test to check minimum variance is considered. Section 2.4 presents the celebrated MVC-based performance index, known as the Harris index, and how to estimate it using different algorithms. In Sect. 2.5, the extension of MV assessment to feedback-plus-feedback loops is described. Its extension to the assessment of set-point tracking and cascade control will be provided in Sect. 2.6. All methods presented are illustrated using many examples.

### 2.1 System Descriptions and Basics

For the description of the methods in this chapter, we assume generic feedback control systems shown in Fig. 2.1, where  $r(k)$  is the set point,  $u(k)$  the controller output,  $e(k)$  the control error,  $y(k)$  the process output, and  $\varepsilon(k)$  is the unmeasured disturbance.  $G_c$ ,  $G_p$  and  $G_\varepsilon$  denote the transfer functions of the feedback controller, the process and disturbance dynamics, respectively. The set point is set to zero by convenience, and the disturbances are assumed to be zero mean. If the reference value and/or the mean of the disturbances are not zero, they can be made mean-free by a simple transformation.

Let the system under consideration be described by an ARMAX model (see Fig. 2.1)

$$A(q)y(k) = q^{-\tau}B(q)u(k) + C(q)\varepsilon(k), \quad (2.1)$$



**Fig. 2.1** Generic feedback control system structure

where  $\varepsilon(k)$  is a zero-mean white noise with the variance  $\sigma_\varepsilon^2$ , also referred to as chocks.  $A(q)$ ,  $B(q)$  and  $C(q)$  are polynomials<sup>1</sup> in  $q^{-1}$  of order  $n$ ,  $m$  and  $p$ , respectively:

$$\begin{aligned} A(q) &= 1 + a_1 q^{-1} + a_2 q^{-2} + \cdots + a_n q^{-n}, \\ B(q) &= b_0 + b_1 q^{-1} + b_2 q^{-2} + \cdots + b_m q^{-m}, \\ C(q) &= 1 + c_1 q^{-1} + c_2 q^{-2} + \cdots + c_p q^{-p}. \end{aligned} \quad (2.2)$$

$\tau$  is an integer number of sampling periods<sup>2</sup> (i.e. the dynamics contain a delay of  $\tau$  samples), so that the leading term of  $B$  is non-zero constant. This means that  $B$  is strictly rational or that the input  $u$  does not affect the output  $y$  immediately, i.e., there is at least one-sample delay ( $\tau \geq 1$ ). Also note that the polynomials  $A$  and  $C$  are monic, because their leading term is unity.

The noise model of an ARMAX model includes only random steps. For a generalisation of the treatment, an ARIMAX model of the form

$$A(q)y(k) = q^{-\tau} B(q)u(k) + \frac{C(q)}{\Delta} \varepsilon(k) \quad (2.3)$$

may be needed to describe drifting (non-stationary) disturbances. As before,  $u$  is the input,  $y$  is the output, and  $\varepsilon$  is the white noise.  $\Delta$  is the backward difference operator, i.e.  $\Delta = 1 - q^{-1}$ . ARIMAX models are typically used for the design of model predictive controllers, particularly DMC and GPC.

If the process is assumed to be stable, it can be expressed as the *infinite impulse response* (IIR)

$$G_p(q) = \sum_{i=\tau}^{\infty} h_i q^{-i}, \quad \lim_{i \rightarrow \infty} h_i = 0. \quad (2.4)$$

<sup>1</sup>Following Ljung (1999),  $q$  is chosen as an argument of the polynomials rather than  $q^{-1}$  (which perhaps would be more natural in view of the right side) in order to be in formal agreement with  $z$ -transform and Fourier-transform expressions.

<sup>2</sup>For discrete systems with no time delay, there is a minimum one-sample delay because the output depends on the previous input, i.e.  $\tau = 1$ .

In practice, the impulse response is truncated at time  $n_p$ , called “time-to-steady-state”:

$$G_p(q) \approx \sum_{i=\tau}^{n_p} h_i q^{-i}, \quad \lim_{i \rightarrow \infty} h_i = 0. \quad (2.5)$$

When considering an ARIMAX model, the disturbance transfer function is marginally stable due to the pole at  $q = 1$ , so it can be shown that

$$G_\varepsilon(q) = \sum_{i=0}^{\infty} e_i q^{-i}, \quad (2.6)$$

where the coefficients  $e_i$  converge to  $C(1)/A(1)$ . Defining  $n_\varepsilon$  as the settling time of the disturbance dynamics enables Eq. 2.6 to be expressed as

$$G_\varepsilon(q) = \sum_{i=0}^{n_\varepsilon} e_i q^{-i} + e_{n_\varepsilon} \sum_{i=n_\varepsilon+1}^{\infty} q^{-i}. \quad (2.7)$$

As the “differenced” load transfer function, i.e.  $\Delta G_\varepsilon$ , is stable, it can be written as

$$\Delta G_\varepsilon(q) = \frac{C(q)}{A(q)} = \sum_{i=0}^{\infty} d_i q^{-i} \approx \sum_{i=0}^{n_\varepsilon} d_i q^{-i} \quad (2.8)$$

with  $d_0 = e_0 = 1$ ,  $d_i = e_i - e_{i-1}$  for  $i = 1, 2, \dots, n_\varepsilon$ .

### 2.1.1 Note on Input–Output Models

Following the system identification and control performance monitoring literature, the *polynomial operator form* is used throughout this book for the description of input–output models. This makes use of the backwards-shift (or unit delay) operator  $q^{-1}$  defined as

$$q^{-1}f(k) = f(k-1).$$

For instance, the difference equation of a linear system

$$y(k) + a_1 y(k-1) + \dots + a_n y(k-n) = b_0 u(k) + b_1 u(k-1) + \dots + b_m u(k-m) \quad (2.9)$$

thus becomes

$$(1 + a_1 q^{-1} + \dots + a_n q^{-n})y(k) = (b_0 + b_1 q^{-1} + \dots + b_m q^{-m})u(k). \quad (2.10)$$

This will simply be denoted by

$$A(q)y(k) = B(q)u(k), \quad (2.11)$$

where  $A(q)$  and  $B(q)$  are the following polynomials, in fact, depending on  $q^{-1}$  in the form

$$A(q) = 1 + a_1 q^{-1} + \cdots + a_n q^{-n}, \quad (2.12)$$

$$B(q) = b_0 + b_1 q^{-1} + \cdots + b_m q^{-m}. \quad (2.13)$$

The ratio of both polynomials

$$G(q) = \frac{B(q)}{A(q)} \quad (2.14)$$

is considered as the discrete *transfer operator* of the discrete transfer function (strictly speaking,  $G(z)$  should be used in the latter case) of the system. For *time-invariant linear systems*, the forward-shift operator  $q$  and the complex variable  $z$  defining the  $z$ -transform are equivalent. In this case, one can use either one ( $q$  is just replaced with  $z$ ), and the appropriate signification will result from the context; see Ratjen and Jelali (2006).

However, the shift operator  $q$  and thus the transfer operator  $G(q)$  can be applied for any discrete-time system, thus as well to linear systems with time-varying coefficients (e.g. in the context of adaptive control) or nonlinear systems, where the  $z$ -transform and thus the concept of transfer function *do not* apply.

Note that the variable  $z$  is analytical: we speak of numerical values  $z_i$  of the poles of a transfer function  $G(z)$ . The operator  $q$  does not possess any numerical values; it gives the transfer function  $G(q)$ , whose mathematical expression is *strictly* identical to  $G(z)$ .

## 2.2 Minimum-Variance Control (MVC)

The minimum-variance control (MVC), also referred to as optimal  $H_2$  control and first derived by Åström (1979), is the best possible feedback control for linear systems in the sense that it achieves the smallest possible closed-loop output variance. More specifically, the MVC task is formulated as minimisation of the variance of the error between the set point and the actual output at  $k + \tau$ , given all the information up to time  $k$ :

$$J = E \{ [r - y(k + \tau)]^2 \} \quad (2.15)$$

or

$$J = E \{ y^2(k + \tau) \}, \quad (2.16)$$

when the set point is assumed zero (without loss of generality), i.e., the case of regulation or disturbance rejection is considered. The discrete time delay  $\tau$  is defined as the number of whole periods of delay in the process, i.e. (Harris 1989)

$$\tau = 1 + f = 1 + \text{int}(T_d/T_s), \quad (2.17)$$

where  $T_d$  is the (continuous) process delay arising from true process dead time or analysis delay, and  $T_s$  denotes the sampling time.  $f$  is the number of integer periods of delay.

The design of minimum-variance controller requires a perfect system model and a perfect disturbance model and will result in a complete cancellation of the error (other than measurement noise) one sample time after the system time delay  $\tau$ . The test for detecting MVC follows immediately (see Sect. 2.3): if the sample auto-correlations of the system output are zero beyond  $\tau$ , then MVC is being achieved.<sup>3</sup> Further, if there is no process noise, i.e.  $\varepsilon(k) = 0$ , then MVC is equivalent to a dead-beat controller.

To enable minimisation of Eq. 2.15 with respect to the control input  $u$ , first we need to relate the controlled output  $y$  to  $u$ . When both sides of Eq. 2.1 are multiplied by  $E_\tau$  and the left side is substituted using the *Diophantine equation*, also known as the *polynomial division identity*,

$$E_\tau(q)A(q) = -q^{-\tau}F_\tau(q) + C(q), \quad (2.18)$$

where

$$E_\tau(q) = e_0 + e_1q^{-1} + e_2q^{-2} + \dots + e_{\tau-1}q^{-(\tau-1)}, \quad (2.19)$$

$$F_\tau(q) = f_0 + f_1q^{-1} + f_2q^{-2} + \dots + f_{n-1}q^{-(n-1)}, \quad (2.20)$$

we get the prediction of the output  $\tau$  steps ahead as

$$y(k + \tau) = \frac{F_\tau(q)}{C(q)}y(k) + \frac{E_\tau(q)B(q)}{C(q)}u(k) + E_\tau(q)\varepsilon(k + \tau). \quad (2.21)$$

The right-hand side of this equation contains the three terms: present and past output signals, present and past control signals and future error signals, respectively. As future terms are not available at time  $k$ , only the realisable terms of the optimal output prediction are then given by

$$\hat{y}(k + \tau) = \frac{F_\tau(q)}{C(q)}y(k) + \frac{E_\tau(q)B(q)}{C(q)}u(k). \quad (2.22)$$

Now, the control action is selected to optimise the variance of the output ( $\tau$  steps ahead), i.e.

$$\begin{aligned} \min_{u(k)} J(k) &= \min_{u(k)} E\{y^2(k + \tau)\} \\ &= \min_{u(k)} E\left\{\frac{F_\tau(q)}{C(q)}y(k) + \frac{E_\tau(q)B(q)}{C(q)}u(k) + E_\tau(q)\varepsilon(k + \tau)\right\}. \end{aligned} \quad (2.23)$$

---

<sup>3</sup>One should here remember the linear correlation test used for the validation of identified linear models; see Sect. 2.3.

(The set point is first assumed to be zero.)<sup>4</sup> This equation contains past inputs, past outputs and future disturbances. As the disturbance is assumed to be white noise, its future values cannot be correlated with past signals. Therefore, the minimum will be achieved when the sum of the first two components is set to zero:

$$\frac{F_\tau(q)}{C(q)}y(k) + \frac{E_\tau(q)B(q)}{C(q)}u(k) = 0, \quad (2.24)$$

which gives the MVC law

$$u(k) = -\frac{F_\tau(q)}{E_\tau(q)B(q)}y(k). \quad (2.25)$$

The same procedure applied to ARIMAX models (Eq. 2.3) leads to

$$\Delta u(k) = -\frac{F_\tau(q)}{E_\tau(q)B(q)}y(k). \quad (2.26)$$

These control laws imply that, no matter what the system dynamics is, all system poles (included in  $A(q)$  and thus  $F(q)$ ) and zeros, included in  $B(q)$ , are cancelled by MVC. Consequently, the basic MVC design is restricted for stable and minimum-phase systems. In practice, cancelling of system dynamics means to exert aggressive control effort, which may not be tolerated from the operational point of view. Another limitation is the sensitivity against system changes, i.e. the lack of robustness to modelling errors.

For non-minimum-phase systems, i.e. with unstable  $B(q)$ , MVC can be designed with some (minor) modifications. The unstable zeros are not inverted, similar to the treatment in the IMC design (Morari and Zafiriou 1989). The control law for non-minimum-phase (ARMAX) processes is given by

$$\Delta u(k) = -\frac{S(q)}{R(q)}y(k), \quad (2.27)$$

where  $S$  and  $R$  are the solution of the Diophantine equation

$$A(q)R(q) = -q^{-\tau}B(q)S(q) + C(q)B_-(q)B_+^{-1}(q). \quad (2.28)$$

The polynomial  $B$  is decomposed into a minimum phase part  $B_-$  and non-minimum phase part  $B_+$ .

From the MVC laws given above it is clear that the main vehicle for calculating minimum-variance controllers is the solution of the Diophantine Eqs. 2.18 and 2.28. For simple cases, it is possible to get solutions; see Example 2.1. However, constructing solutions for Diophantine equations usually requires the use of a software

---

<sup>4</sup>The basic MVC is designed to solve regulation problems, where the objective is to compensate for stochastic disturbances and not to follow a reference trajectory. However, MVC can be extended to include variations in the reference, as described below.

package. A standard one for this purpose is available from Kwakernaak and Sebek (2000). Another solver is provided by Moudgalya (2007) in form of a MATLAB function `xdync`.

Using the MVC, the minimum value of the output variance, shortly denoted minimum variance, is achieved:

$$\begin{aligned} J_{\min}(k) &= \min_{u(k)} E \{ y^2(k + \tau) \} = E \{ E_{\tau}(q) \varepsilon(k + \tau) \} \\ &= \left( \sum_{i=0}^{\tau-1} e_i^2 \right) \sigma_{\varepsilon}^2 \equiv \sigma_{\text{MV}}^2, \end{aligned} \quad (2.29)$$

where  $\sigma_{\varepsilon}^2$  is the (disturbance) noise variance. Note that  $\sigma_{\text{MV}}^2$  is the same as the variance of the prediction error  $y - \hat{y}$ . The achieved output of the closed-loop system under MVC is

$$y(k) = E_{\tau}(q) \varepsilon(k). \quad (2.30)$$

Note that whereas the controller itself may require the specification of the system model and disturbance model, both are not needed for MVC-based performance assessment, as described below (Sect. 2.4). It is important to stress that the adoption of MVC as a benchmark does not imply that it should be the goal towards which the existing control should be driven, or that it is always practical, desirable, or even possible to implement. Nevertheless, the performance bound set by the MVC is exceeded by all other (linear) controllers; hence, it serves as an appropriate benchmark against which the performance of other controllers may be compared.

The reader is encouraged to consult the textbook by Moudgalya (2007: Chap. 11), including many examples and MATLAB functions (`mv` for minimum phase systems, `mv_nm` for non-minimum phase systems). Minimum variance control (placed in a conventional feedback structure) can be viewed in an IMC structure or an SPC structure; see Sect. 3.2. The equivalence between MVC and IMC was revealed by Bergh and MacGregor (1987) to analyse the robustness of MVC. Refer also to Qin (1998), who derived the MVC using the IMC structure.

*Example 2.1* Consider the first-order system described by the transfer function

$$y(k) = \frac{q^{-\tau}}{1 + a_1 q^{-1}} u(k) + \frac{1}{1 + a_1 q^{-1}} \varepsilon(k) \quad (2.31)$$

with  $a_1 = -0.9$  and the time delay  $\tau = 3$ . This is an ARMAX model with

$$\begin{aligned} A(q^{-1}) &= 1 + a_1 q^{-1}, & n &= 1, \\ B(q^{-1}) &= 1, & m &= 0, \\ C(q^{-1}) &= 1, & p &= 0. \end{aligned}$$

The Diophantine Eq. 2.18 takes the form

$$E_3(q^{-1})A(q^{-1}) + q^{-\tau}F_3(q^{-1}) = C(q^{-1}),$$

$$(1 + e_1 q^{-1} + e_2 q^{-2})(1 + a_1 q^{-1}) + f_0 q^{-3} = 1.$$

Comparing the same powers of  $q^{-1}$  gives

$$\begin{aligned} q^0: \quad & 1 = 1, \\ q^{-1}: \quad & e_1 + a_1 = 0 \Rightarrow e_1 = -a_1, \\ q^{-2}: \quad & e_2 + a_1 e_1 = 0 \Rightarrow e_2 = a_1^2, \\ q^{-3}: \quad & a_1 e_2 + f_0 = 0 \Rightarrow f_0 = -a_1^3. \end{aligned}$$

The closed loop is then given by (Eq. 2.30)

$$y(k) = E_\tau(q^{-1})\varepsilon(k) = (1 - a_1 q^{-1} + a_1^2 q^{-2} + \dots)\varepsilon(k).$$

In fact, the first three terms will be the same irrespective of the (linear) controller used. The MVC law has the form (Eq. 2.25):

$$u(k) = -\frac{-a_1^3}{1 - a_1 q^{-1} + a_1^2 q^{-2}} y(k).$$

This gives for  $a_1 = -0.9$ :

$$u(k) = -\frac{0.729}{1 + 0.9q^{-1} + 0.81q^{-2}} y(k). \quad (2.32)$$

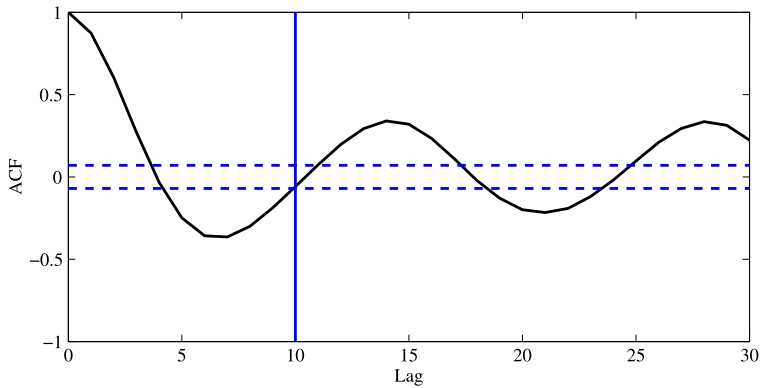
This control law can also be determined using the function `mv` from Moudgalya's MATLAB software (Moudgalya 2007: Sect. 11.4).

## 2.3 Auto-Correlation Test for Minimum Variance

Auto-correlation is a method that is used to determine how data in a time series are related. Auto-correlation-based analysis provides to discover the nature of disturbances acting on the process and how they affect the system by comparing current process measurements patterns with those exhibited in the past during “normal” operation.

A fundamental test for assessing the performance of control loops is to check the auto-correlation of the output samples: the autocorrelation should die out beyond the time delay  $\tau$ . Using a representative sample of measured output data, the sample auto-correlation can be computed (i.e. estimated). Statistically significant values of the estimated auto-correlations existing beyond the delay provide evidence that the current controller is not minimum variance. Furthermore, if there exist many large auto-correlation values that persist beyond  $\tau$ , the control performance deviates substantially from the MV performance bound. If only few slightly significant values exist beyond  $\tau$ , the performance is close to that of MVC. Since the auto-correlations are statistical estimates based on a finite sample of data, they will never be truly zero.





**Fig. 2.2** Example of auto-correlation test

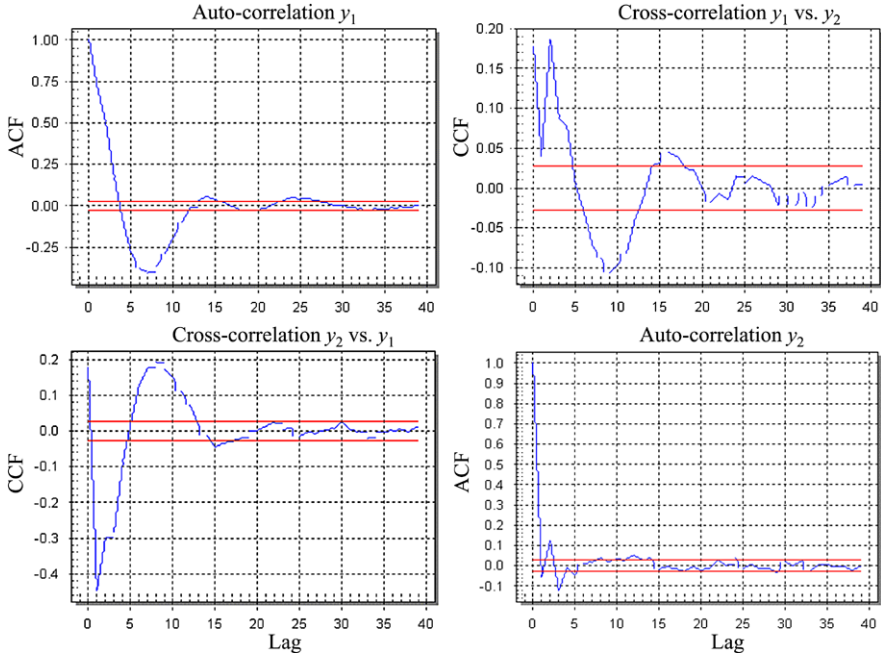
Therefore, to assess whether the true auto-correlations  $\rho_{yy}(j)$  might be zero or not, their estimated values must be compared to their statistical confidence intervals, e.g. 95 % or  $2\sigma$ . Box and Jenkins (1970) showed that, if  $\rho_{yy}(j)$  is zero for  $j \geq \tau$ , then the variance is

$$\sigma^2 = \text{var}\{\rho_{yy}(j)\} \approx \frac{1}{N} \left[ 1 + 2 \sum_{i=1}^{\tau-1} \rho_{yy}^2(i) \right], \quad j \geq \tau. \quad (2.33)$$

Therefore, the 95 % confidence interval for  $\rho_{yy}(j)$  is  $[-2\sigma, 2\sigma]$ . If most auto-correlation coefficients  $\rho_{yy}(j)$  are inside this interval for  $j \geq \tau$ , the control is roughly achieving minimum variance; otherwise, it is not.

*Example 2.2* Figure 2.2 depicts the auto-correlation estimates for a gauge control loop and their 95 % confidence levels. It is observed that the auto-correlation functions are far outside the confidence limits after the time delay of 10. Therefore, we conclude that the control is not achieving minimum variance. Furthermore, the auto-correlation function is oscillatory, indicating that oscillation exists in the original data.

The motivation behind the use of auto-correlation function (ACF) is that it can be easily estimated from plant response data. Moreover, the dynamic response characteristics for data trends can be inferred without having to resort to the more complicated tasks associated with the identification and interpretation of time-series models. For example, a slowly decaying auto-correlation function implies an under-tuned loop, and an oscillatory ACF typically implies an over-tuned loop. For multi-variable systems, off-diagonal plots can be used to trace the source of disturbance or the interaction between each process variables. Figure 2.3 shows an example taken from Huang et al. (1999), clearly indicating that the first loop has relatively poor performance while the second loop has very fast decay dynamics and thus good



**Fig. 2.3** Correlation functions of a multivariate process (Huang et al. 1999)

performance. The off-diagonal subplots indicate interaction between the two loops. Note that the ACF plot of the multivariate system is not necessarily symmetric.

## 2.4 Minimum-Variance Index/Harris Index

In the following, we present algorithms that will use routine (closed-loop) operating data to assess the performance of control loops against MVC as benchmark. MVC-based assessment first described by Harris (1989) compares the actual system-output variance  $\sigma_y$  to the output variance  $\sigma_{MV}^2$  as obtained using minimum-variance controller applied to an estimated time-series model from measured output data.

The Harris index is defined as

$$\eta_{MV} = \frac{\sigma_{MV}^2}{\sigma_y^2}. \quad (2.34)$$

This index will of course be always within the interval  $[0, 1]$ , where values close to unity indicate good control with respect to the theoretically achievable output variance. “0” means the worst performance, including unstable control. No matter what the current controller is, we need only the following information about the system:

- Appropriately collected closed-loop data for the controlled variable.
- Known or estimated system time delay ( $\tau$ ).

Moreover, there are two advantages for using this index over a simple error variance metric:

1. Taking the ratio of the two variances results in a metric that is (supposedly) independent of the underlying disturbances—a key feature in an industrial situation, where the disturbances can vary widely.
2. The metric is scale independent, bounded between 0 and 1. This is an important consideration for a plant user, who might be faced with evaluating hundreds or even thousands of control loops.

### 2.4.1 Estimation from Time-Series Analysis

From the measured (closed-loop) output data, a time-series model, typically of AR/ARMA type, is estimated:

$$y(k) = \frac{\hat{C}(q)}{\hat{A}(q)} \varepsilon(k). \quad (2.35)$$

A series expansion, i.e. impulse-response, of this model gives

$$\begin{aligned} y(k) &= \left( \sum_{i=0}^{\infty} e_i q^{-i} \right) \varepsilon(k) \\ &= \underbrace{(e_0 + e_1 q^{-1} + e_2 q^{-2} + \dots + e_{\tau-1} q^{-(\tau-1)})}_{\text{feedback-invariant}} \varepsilon(k) \\ &\quad + \underbrace{(e_{\tau} q^{-\tau} + e_{\tau+1} q^{-(\tau+1)} + \dots)}_{\text{feedback-varying}} \varepsilon(k). \end{aligned} \quad (2.36)$$

The first  $\tau$  impulse response coefficients can be estimated through  $\tau$ -term polynomial long division, or equivalently via resolution of the Diophantine identity:

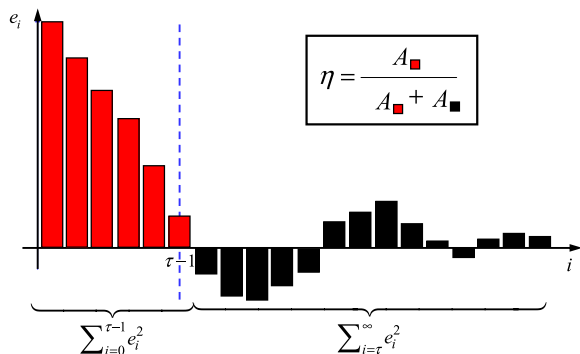
$$\hat{C}(q) = \hat{E}_{\tau}(q) \hat{A}(q) + q^{-\tau} \hat{F}_{\tau}(q), \quad (2.37)$$

where  $\hat{E}_{\tau}$  is an estimate of  $E_{\tau}$  in Eq. 2.18. The *feedback-invariant terms* are not functions of the process model or the controller; they depend only on the characteristics of the disturbance acting on the process.

Since the first  $\tau$  terms are invariant irrespective of the controller (Fig. 2.4), the minimum-variance estimate corresponding to the feedback-invariant part is given by

$$\sigma_{\text{MV}}^2 = \sum_{i=0}^{\tau-1} e_i^2 \sigma_{\varepsilon}^2. \quad (2.38)$$

**Fig. 2.4** An impulse response showing the contributions to the Harris index



The first coefficient of the impulse response,  $e_0$ , is often normalised to be equal to unity.

The estimate of the actual output variance can be directly estimated from the collected output samples using the standard relation in Eq. 1.1. However, it is suggested to use the (already) estimated time-series model also for evaluating the current variance. From the series expansion of the time-series model (Eq. 2.36), we obtain

$$\sigma_y^2 = \sum_{i=0}^{\infty} e_i^2 \sigma_\varepsilon^2. \quad (2.39)$$

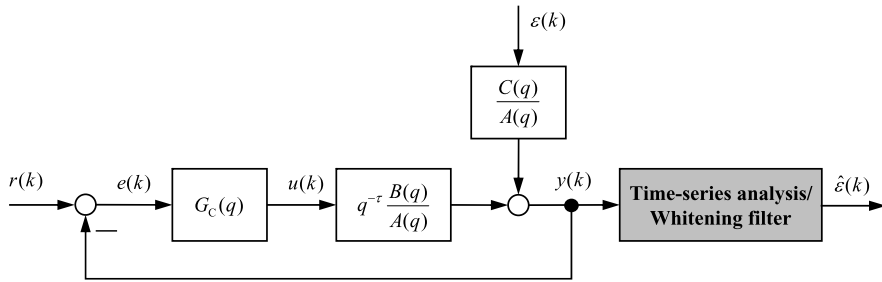
Since the noise variance will be cancelled in Eq. 2.34, it is neither needed nor has an effect on the performance index. This compares the sum of the  $\tau$  first impulse-response coefficients squared to the total sum; see Fig. 2.4.

The performance index  $\eta_{MV}$  corresponds to the ratio of the variance, which could theoretically be achieved under minimum variance control, to the actual variance.  $\eta_{MV}$  is a number between 0 (far from minimum-variance performance) and 1 (minimum-variance performance) that reflects the inflation of the output variance over the theoretical minimum variance bound. As indicated in Desborough and Harris (1992), it is more useful to replace  $\sigma_y^2$  by the mean-squares error of  $y$  to account for offset

$$\eta_{MV} = \frac{\hat{\sigma}_{MV}^2}{\text{MSE}} = \frac{\hat{\sigma}_{MV}^2}{\hat{\sigma}_y^2 + \bar{y}^2}. \quad (2.40)$$

See also Sect. 2.4.2. If  $\eta_{MV}$  is considerably less than 1, re-tuning the controller will yield benefits. If  $\eta_{MV}$  is close to 1, the performance cannot be improved by re-tuning the existing controller; only process or plant changes, such as changes in the location of sensors and actuators, inspection of valves, other control loop components, or even alterations to the control structure can lead to better performance.

Although  $\varepsilon(k)$  is unknown, it can be replaced by the estimated innovations sequence. This can be obtained by pre-whitening the system output variable  $y(k)$  via time series analysis based on an AR or ARMA model (alternatively a Kalman filter-based innovation model in state-space form); see Box and MacGregor (1974),



**Fig. 2.5** Schematic representation of the white noise or innovation sequence estimation

Söderström and Stoica (1989) and Goodwin and Sin (1984). An estimate for the random chocks is then found, e.g. by inverting the estimated ARMA model

$$\varepsilon = \hat{C}^{-1}(q) \hat{A}(q) y, \quad (2.41)$$

where  $y$  is the vector of the output data.

The aim of pre-whitening (or simply whitening) is at tracking back the source of variations in a regulatory closed-loop system to white noise excitation (“the driving force”), as shown in Fig. 2.5. This means reversing the relationship between  $y(k)$  and  $\varepsilon(k)$ . The process of obtaining a “whitening” filter is analogous to time-series modelling, where the final test of the adequacy of the model, i.e. validation, consists of checking if the residuals are “white”. These residuals are the estimated white noise sequence. In contrast to time-series modelling, where the estimation of the model is of core interest, the residual or innovation sequence is the main item of interest in the “whitening” process and thus in control performance assessment.

To summarise, the complete algorithm to evaluate the MVC-based (Harris) index and to assess feedback controls contains the steps described in Procedure 2.1.

**Procedure 2.1** Performance assessment based the Harris index.

1. Select the time-series-model type and orders.
2. Determine/estimate the system time delay  $\tau$ .
3. Identify the closed-loop model from collected output samples [ar/arma(x)] (Eq. 2.35).
4. Calculate the series expansion (impulse response) for the estimated model (Eq. 2.36) [dimpulse].
5. Estimate the minimum variance from Eq. 2.38.
6. Estimate the actual output variance from Eqs. 1.1 or 2.39.
7. Compute the (Harris) performance index (Eq. 2.34).

### 2.4.2 Estimation Algorithms

In this section, some different algorithms are described for the estimation of the Harris index from normal operating data, irrespective of the controller installed on the process. These algorithms do not necessitate the solution of Diophantine equation.

#### 2.4.2.1 Direct Least-Squares Estimation

A simple way to estimate the Harris index  $\eta_{MV}$  from closed-loop routine data is to use linear regression methods, without the necessity of solving any Diophantine equation or performing polynomial long divisions. From Eq. 2.21 the process output under any installed feedback controller  $G_c(q)$  can be expressed as

$$y(k) = E_\tau(q)\varepsilon(k + \tau) + q^{-\tau} \frac{F_\tau(q) - B(q)E_\tau(q)G_c(q)}{C(q)} y(k). \quad (2.42)$$

Under the assumption of closed-loop stability, the second term in the previous equation can be approximated by a finite-length ( $n$ ) AR model:

$$y(k) = \sum_{i=1}^n \Theta_i y(k - \tau - i + 1) + E_\tau(q)\varepsilon(k) \quad (2.43)$$

with the unknown model parameters  $\Theta_i$ .

Running  $k$  over a range of values and stacking up similar terms yields

$$\mathbf{y} = \mathbf{X}\boldsymbol{\Theta} + E_\tau(q)\boldsymbol{\varepsilon}(k) \quad (2.44)$$

with

$$\mathbf{y} = \begin{bmatrix} y(N) \\ y(N-1) \\ \vdots \\ y(n+\tau) \end{bmatrix}, \quad \boldsymbol{\Theta} = \begin{bmatrix} \Theta_1 \\ \Theta_2 \\ \vdots \\ \Theta_n \end{bmatrix},$$

$$\mathbf{X} = \begin{bmatrix} y(N-\tau) & y(N-\tau-1) & \cdots & y(N-\tau-n+1) \\ y(N-\tau-1) & y(N-\tau-2) & \cdots & y(N-\tau-n) \\ \vdots & \vdots & \cdots & \vdots \\ y(n) & y(n-1) & \cdots & y(1) \end{bmatrix}.$$

The parameter vector  $\boldsymbol{\Theta}$  can be estimated with LS method, i.e. by fitting the recorded closed-loop data  $\{y_1, y_2, \dots, y_N\}$  to the model Eq. 2.43. The LS solution follows as

$$\hat{\boldsymbol{\Theta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}. \quad (2.45)$$

An estimate of the minimum variance can be determined as the residual mean square error

$$\hat{\sigma}_{\text{MV}}^2 = \frac{1}{N - \tau - 2n + 1} (\mathbf{y} - \mathbf{X}\boldsymbol{\Theta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\Theta}), \quad (2.46)$$

while the actual variance results as

$$\hat{\sigma}_y^2 = \frac{1}{N - \tau - n + 1} \mathbf{y}^T \mathbf{y}. \quad (2.47)$$

The Harris index can then be formed as

$$\hat{\eta}_{\text{MV}}(\tau) = \frac{\hat{\sigma}_{\text{MV}}^2}{\hat{\sigma}_y^2} = \frac{N - \tau - n + 1}{N - \tau - 2n + 1} \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\Theta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\Theta})}{\mathbf{y}^T \mathbf{y}} \quad (2.48)$$

or

$$\hat{\eta}_{\text{MV}}(\tau) = \frac{\hat{\sigma}_{\text{MV}}^2}{\text{MSE}} = \frac{N - \tau - n + 1}{N - \tau - 2n + 1} \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\Theta})^T (\mathbf{y} - \mathbf{X}\boldsymbol{\Theta})}{\mathbf{y}^T \mathbf{y} + (N - \tau - n + 1) \bar{y}^2}, \quad (2.49)$$

when the mean square error is used rather than the variance to penalise non-zero steady-state errors; see Eq. 2.40. It is important to note that the signal  $y(k)$  has always to be made free from the set point value prior to the index calculation.

Exact distributional properties of the estimated performance indices are complicated and not amenable to a closed-form solution. Desborough and Harris (1992) approximated first and second moments for the estimated performance indices and resorted to a normal theory to develop approximate confidence intervals. Asymptotically, the performance indices are ratios of correlated quadratic forms, and as such the distributions of the performance indices are non-symmetric. Refinements to the confidence intervals developed in Desborough and Harris (1992) can be obtained with little extra computational effort, by resorting to the extensive statistical literature on the distributional properties of quadratic forms (Harris 2004).

### 2.4.2.2 Online/Recursive Least-Squares Estimation

One advantage of the LS approach is that recursive algorithms to find  $\hat{\eta}_{\text{MV}}(k)$  are readily available. An online estimation of the index becomes possible. This is useful to detect change points in control monitoring. Also, if the process is nonlinear and the dynamics are slow enough that the process can be considered locally linear, recursive estimation of the performance index provides a local estimate of the controller performance. Alternatively,  $\hat{\eta}_{\text{MV}}(k)$  can be used online as a tuning tool to immediately show whether the tuning changes have improved or degraded control performance (Desborough and Harris 1992). This assumes that the disturbance model does not change significantly.

Typically, recursive LS (RLS) algorithms minimise a cost function of the form

$$V = (\mathbf{y} - \mathbf{X}\boldsymbol{\Theta})^T \boldsymbol{\Lambda} (\mathbf{y} - \mathbf{X}\boldsymbol{\Theta}), \quad (2.50)$$

where  $\mathbf{A}$  is a diagonal matrix with elements  $(\lambda, \lambda^2, \dots, \lambda^N)$ .  $\lambda$  is the so-called forgetting factor used to place more emphasis on recent data. An estimate of the MV at time  $k$  is given by

$$\sigma_{\text{MV}}^2(k) = \lambda \sigma_{\text{MV}}^2(k-1) + \varepsilon^2(k). \quad (2.51)$$

An estimate of the performance index is computed as

$$\hat{\eta}_{\text{MV}}(k) = \frac{\sigma_{\text{MV}}^2(k)}{\sigma_y^2(k)}, \quad (2.52)$$

where  $\sigma_y^2(k)$  is the exponentially weighted moving mean square error

$$\sigma_y^2(k) = \lambda \sigma_y^2(k-1) + y^2(k). \quad (2.53)$$

Instead of an RLS method, a stochastic gradient algorithm, which does not need matrix computations, can be used as well. This has been proposed by Ingimundarson (2002, 2003) for performance assessment of  $\lambda$ -tuned PI controllers.

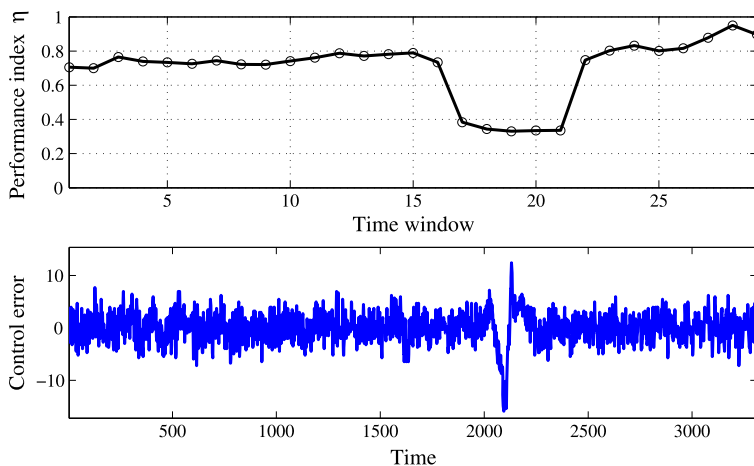
As stated by Tyler and Morari (1995), the recursive estimation described works well as long as the closed loop is accurately represented by an AR(MA) model. This does not apply for closed-loop models with moving average parameters. An alternative approach to the recursive index estimation is therefore to use a hierarchical method based on data windowing: the data are first broken into segments with similar dynamic properties. Efficient algorithms, such as those proposed by Basseville (1988), can be applied to rapidly detect changes in the closed-loop dynamics. Once a change has been detected, the Harris index can be computed for the largest data segment available with similar dynamics.

In practice, it is often sufficient to use moving windows to study the change in performance of the process over time. Drops or drifts in the performance index can be easily observed in such performance pictures like those shown in Fig. 2.6 ( $N = 500$ ). However, care has to be taken to not use too small data windows; see the guidelines in Sects. 7.1.2 and 7.3.2. In this example, a performance deterioration caused by a big (non-stationary) disturbance appearing at time  $k = 2025$  can be observed.

### 2.4.2.3 Filtering and Correlation Analysis (FCOR) Method

Huang et al. (1997a, 1997b, 1997c, 2000) have developed a method to derive the MVC-based performance index by filtering (i.e. pre-whitening) and subsequent correlation analysis (thus called FCOR) between of the delay-free output and estimated random shocks obtained by a pre-whitening filter. Calculation of the system correlation eliminates the need to determine the impulse response coefficients from the estimated closed-loop transfer function. The FCOR algorithm is presented in this section following Huang and Shah (1999).





**Fig. 2.6** An example of the Harris index trend computed for moving data windows (gauge control loop)

Consider the (stable) closed-loop system described by the infinite-order moving average process in Eq. 2.36. Multiplying this equation by the error terms  $\varepsilon(k)$ ,  $\varepsilon(k-1)$ ,  $\dots$ ,  $\varepsilon(k-\tau+1)$  respectively and then taking the expectation of both sides of the equation yields

$$\begin{aligned}
 r_{y\varepsilon}(0) &= E\{y(k)\varepsilon(k)\} = e_0\sigma_\varepsilon^2, \\
 r_{y\varepsilon}(1) &= E\{y(k)\varepsilon(k-1)\} = e_1\sigma_\varepsilon^2, \\
 r_{y\varepsilon}(2) &= E\{y(k)\varepsilon(k-2)\} = e_2\sigma_\varepsilon^2, \\
 &\vdots \\
 r_{y\varepsilon}(\tau-1) &= E\{y(k)\varepsilon(k-\tau+1)\} = e_{\tau-1}\sigma_\varepsilon^2.
 \end{aligned} \tag{2.54}$$

Therefore, the minimum variance is

$$\sigma_{MV}^2 = \sum_{i=0}^{\tau-1} e_i^2 \sigma_\varepsilon^2 = \sum_{i=0}^{\tau-1} \left( \frac{r_{y\varepsilon}(i)}{\sigma_\varepsilon^2} \right)^2 \sigma_\varepsilon^2 = \sum_{i=0}^{\tau-1} r_{y\varepsilon}^2(i) / \sigma_\varepsilon^2. \tag{2.55}$$

Substituting Eq. 2.55 into Eq. 2.34 leads to the performance index

$$\eta_{MV, \text{cor}} = \sum_{i=0}^{\tau-1} r_{y\varepsilon}^2(i) / (\sigma_y^2 \sigma_\varepsilon^2) = \sum_{i=0}^{\tau-1} \rho_{y\varepsilon}^2(i) = \mathbf{Z}^T \mathbf{Z}, \tag{2.56}$$

where  $\mathbf{Z}$  is the *cross-correlation coefficient* vector between  $y(k)$  and  $\varepsilon(k)$  for lags 0 to  $\tau-1$  and is denoted

$$\mathbf{Z} := [\rho_{y\varepsilon}(0), \rho_{y\varepsilon}(1), \rho_{y\varepsilon}(2), \dots, \rho_{y\varepsilon}(\tau-1)]^T. \tag{2.57}$$

The corresponding sampled version of the performance index is therefore given by

$$\hat{\eta}_{\text{MV,cor}} = \sum_{i=0}^{\tau-1} \hat{\rho}_{y\varepsilon}^2(i) = \hat{\mathbf{Z}}^T \hat{\mathbf{Z}}, \quad (2.58)$$

where

$$\hat{\rho}_{y\varepsilon}(l) = \frac{\sum_{k=1}^M y(k)\varepsilon(k-l)}{\sum_{k=1}^M y^2(k) \sum_{k=1}^M \varepsilon^2(k)}.$$

$\varepsilon(k)$  can be determined from pre-whitening of  $y(k)$  via time series, as explained in Sect. 2.4.1. The complete FCOR algorithm is described in Procedure 2.2.

**Procedure 2.2** Filtering and correlation-based (FCOR) algorithm.

1. Select the time-series-model type and orders.
2. Determine/estimate the system time delay  $\tau$ .
3. Identify an appropriate closed-loop model from collected output samples  $y(k)$ .
4. Filter the system output data  $y(k)$  from the model to obtain an estimate for the whitened sequence (Eq. 2.41).
5. Calculate the cross-correlation coefficients between  $y(k)$  and  $\varepsilon(k)$  for lags 0 to  $\tau - 1$  from Eq. 2.54.
6. Use Eq. 2.58 to compute the performance index.

Note that FCOR is a general methodology rather than a specific algorithm. It consists of two steps: filtering/whitening and correlation analysis. The advantage of FCOR is its flexibility in the form of the filter and its suitability for MIMO systems due to its simplicity in computation, namely computation of the Markov matrices can be avoided. The filtering step in FCOR can use any algorithm, such as AR, ARMA, state space, or even a nonlinear time series modelling algorithm.

#### 2.4.2.4 Examples

The following examples illustrate the performance assessment results in terms of the Harris index obtained using Procedure 2.2. Some controller tuning rules will be evaluated using the MVC-based assessment introduced above.

*Example 2.3* Consider the first-order system from Example 2.1. The MVC is used here just to simulate the process under this ideal controller and to show that the Harris index will take the value of 1 in this case. This is confirmed by the Harris index value given in Table 2.1 (fourth row). In this simulation,  $\varepsilon(k)$  was a normally distributed noise with the variance  $\sigma_\varepsilon^2 = 0.01$ . The Harris index values have been determined from using  $N = 1500$  simulated data points ( $T_s = 0.5$ ) and modelling the closed loop by an AR model of order  $n = 30$ .

**Table 2.1** Harris index value for the different controllers

Controller	$K_c$	$\hat{\eta}$
P1	0.05	0.62
P2	0.25	0.76
P3	0.50	0.37
MVC	–	0.99

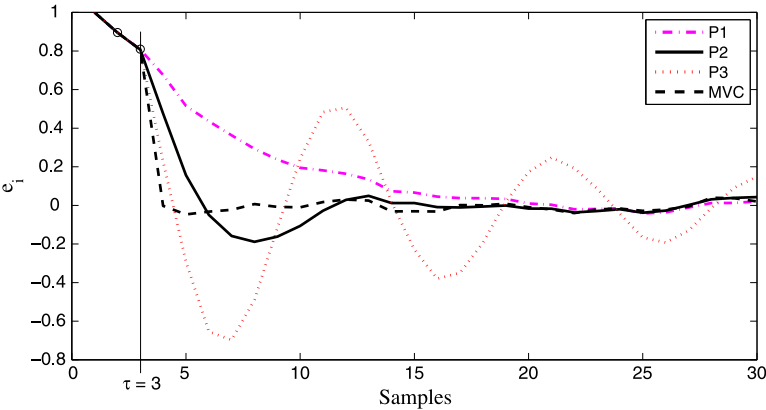
The impulse responses for a P-only controller with different gains and for the MVC are illustrated in Fig. 2.7. It can be seen that P1 is a sluggish controller, P2 a well-tuned controller and P3 an aggressive controller. The figure also shows how the impulse response for the MVC dies beyond  $\tau = 3$ . The first three (controller-invariant) coefficients are marked with circles in the figure. From this example, it can be learned that the Harris index for a well-tuned controller ( $\eta = 0.76$ ) does not always achieve that of MVC ( $\eta = 0.99 \approx 1$ ).

In the following, we consider two simulated processes used by Seborg et al. (2004: Chap. 12) to compare different controller tuning rules in terms of (deterministic) set-point tracking and disturbance rejection. Here we evaluate the stochastic control performance using the Harris index. For all processes, we assume that the process model (without time delay) and disturbance model are identical and the disturbance is normally distributed noise with the variance  $\sigma_\varepsilon^2 = 0.01$ .

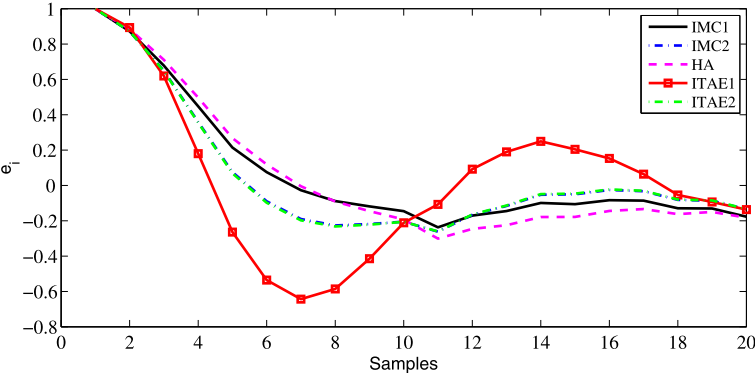
*Example 2.4* A blending system with a measurement time delay modelled by

$$G_p(s) = \frac{1.54}{5.93s + 1} e^{-1.07s} \tag{2.59}$$

and controlled by a PI controller is considered. Figure 2.8 illustrates the results gained from modelling the closed loop by an AR model of order  $n = 20$  using



**Fig. 2.7** Impulse responses for different controllers



**Fig. 2.8** Harris index value for the blending process and different controllers

**Table 2.2** Impulse responses for the blending process and different controllers

Controller/tuning rule	Acronym	$K_c$	$T_I$	$\eta$
IMC ( $\lambda = T/3$ )	IMC1	1.27	5.93	0.76
IMC ( $\lambda = T_d$ )	IMC2	1.80	5.93	0.81
Hägglund and Åström	HA	1.10	2.95	0.65
ITAE (disturbance)	ITAE1	2.97	2.75	0.59
ITAE (set point)	ITAE2	1.83	5.93	0.81

**Table 2.3** Harris index value for the lag-dominant process and different controllers

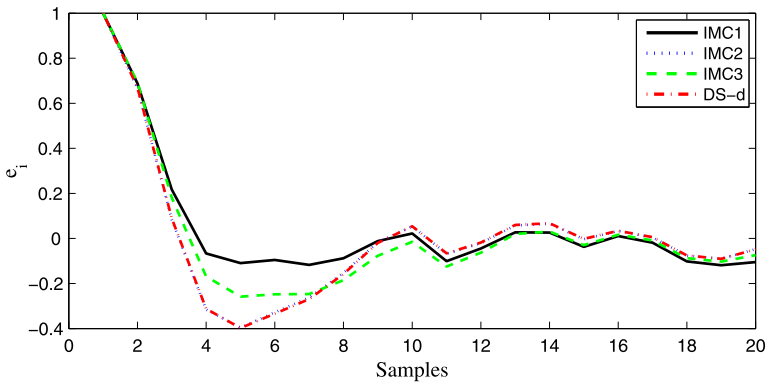
Controller/tuning rule	Acronym	$K_c$	$T_I$	$\hat{\eta}$
IMC ( $\lambda = 1.0$ )	IMC1	0.5	100	0.63
IMC ( $\lambda = 2.0$ ) based on integrator approximation	IMC2	0.556	5	0.52
IMC ( $\lambda = 1.0$ ) based on Skogestad's modification	IMC3	0.5	8	0.56
Direct synthesis (disturbance)	DS-d	0.551	4.91	0.52

1500 output samples. IMC and ITAE (set point) yield the best performance, and ITAE (disturbance) the least performance, as a consequence of the most aggressive settings; see Table 2.2. Note that IMC2 and ITAE (set point) have almost identical impulse responses for this example, but this is not true in general.

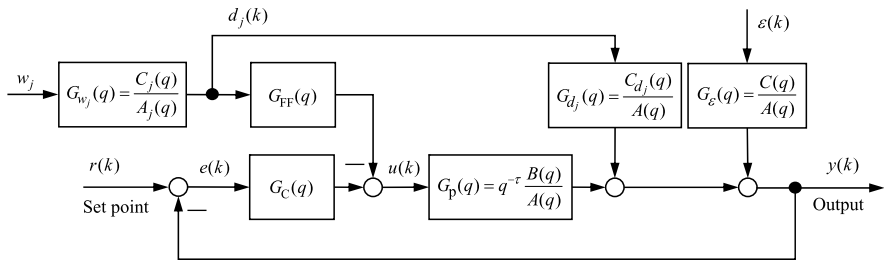
*Example 2.5* This example is a lag-dominant model with  $T_d/T = 0.01$ :

$$G_P(s) = \frac{100}{100s + 1} e^{-s}. \tag{2.60}$$

Table 2.3 contains the results gained from modelling the closed loop by an AR model of order  $n = 20$  using 1500 output samples. IMC1 leads to the best per-



**Fig. 2.9** Impulse responses for the lag-dominant process and different controllers



**Fig. 2.10** Generic feedback plus feedforward control system structure

formance and both IMC2 and DS-d to the least performance, which have almost identical controller settings and thus almost identical impulse responses, as can be seen in Fig. 2.9.

## 2.5 Assessment of Feedback/Feedforward Controls

Feedforward control (FFC) should always be introduced to reduce the process variability due to disturbances. An ineffective FFC will contribute to a large variance due to the measured disturbances. Therefore, one additional task in assessing feedback/feedforward control loops is to diagnose whether a poor performance is due to feedback control or feedforward control. This section focuses on the analysis of variance (ANOVA), to quantify major contributions to system-output variance (Desborough and Harris 1993; Huang and Shah 1999). A detailed derivation of the algorithm can also be found by Ratjen and Jelali (2006).

The MV is calculated differently in feedback/feedforward control (Fig. 2.10) than feedback control alone. The major difference is in the estimation of the variance of

**Table 2.4** Analysis of variance for feedback plus feedforward control

Disturbance	MV	FB	FF	FB/FF	Total
$\varepsilon(k)$	$\sigma_{\text{MV},\varepsilon}^2$	$\sigma_{\text{FB},\varepsilon}^2$	—	—	$\sigma_{y,\varepsilon}^2$
$w_1(k)$	$\sigma_{\text{MV},w_1}^2$	—	$\sigma_{\text{FF},w_1}^2$	$\sigma_{\text{FB/FF},w_1}^2$	$\sigma_{y,w_1}^2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$w_p(k)$	$\sigma_{\text{MV},w_p}^2$	—	$\sigma_{\text{FF},w_p}^2$	$\sigma_{\text{FB/FF},w_p}^2$	$\sigma_{y,w_p}^2$
Total	$\sigma_{\text{MV}}^2$		$\sigma_y^2 - \sigma_{\text{MV}}^2$		$\sigma_y^2$

the unmeasured disturbance  $\varepsilon(k)$ . An ARMAX model of the MISO form

$$y(k) = \frac{C(q^{-1})}{A(q^{-1})} \varepsilon(k) + \sum_{j=1}^p \frac{B_j(q^{-1})}{A(q^{-1})} d_j(k - \tau_j) \quad (2.61)$$

should be used for identifying the closed-loop model to include the effect of  $p$  measured disturbances  $d_j$ , as opposed to an AR(MA) model only. Then each measured disturbance model is identified as an AR(I)MA time-series model, i.e.  $A_j(q)d_j(k) = C_j(q)w_j(k)$ , leading to

$$y(k) = \frac{C(q^{-1})}{A(q^{-1})} \varepsilon(k) + \sum_{j=1}^p \frac{B_j(q^{-1})C_j(q^{-1})}{A(q^{-1})A_j(q^{-1})} w_j(k - \tau_j). \quad (2.62)$$

Time delays ( $\tau$  and  $\tau_j$ ) are required and the model orders also need to be determined. The identified closed-loop model is then used to carry out an ANOVA table for the output  $y$  based on the time delays in the feedforward and feedback paths. This method can yield valuable information about the sources of variability, provided that all measured disturbances are mutually independent.

The cross-correlation between the measured disturbances (as potential feedforward variables) and the output can be used to determine which of them could be used for FFC. The analysis of variance (Desborough and Harris, 1993) highlights the contribution of the disturbances to the overall variance (Table 2.4):

$$\sigma_y^2 = \sigma_{\text{MV},\varepsilon}^2 + \sigma_{\text{FB},\varepsilon}^2 + \sum_{j=1}^{n_w} (\sigma_{\text{MV},w_j}^2 + \sigma_{\text{FF},w_j}^2 + \sigma_{\text{FB/FF},w_j}^2), \quad (2.63)$$

where

- $\sigma_{\text{MV},\varepsilon}^2$ : the minimum variance of the FBC, arising from unmeasured disturbance  $\varepsilon$
- $\sigma_{\text{FB},\varepsilon}^2$ : the variance due to the non-optimality of the FBC
- $\sum_{j=1}^p \sigma_{\text{MV},w_j}^2$ : the minimum variance of the FFC, coming from  $r$  measured disturbances  $w_i$
- $\sum_{j=1}^p \sigma_{\text{FF},w_j}^2$ : the variance due to the non-optimality of the FFC

- $\sum_{j=1}^p \sigma_{\text{FB/FF}, w_j}^2$ : the variance due to the non-optimality of the combination FBC/FFC.

The bottom row in Table 2.4 consists of, from left to right, the summation of the minimum variances, the sum of all the variance due to non-optimality of the controller components and the total variance. It is important to note that it is not possible to unambiguously attribute variance inflation to either the feedback controller alone or the feedforward controller alone, hence the column labelled “FF/FB” in the table. Note that if the process is invertible, it is always possible to eliminate the variance inflation due to both this component and the feedforward component using a feedforward controller, regardless of the feedback controller (Desborough and Harris 1993). If one row contains a considerable portion of the total variance in the columns FF and FB + FF, this implies that re-tuning is needed. If only the term FB + FF is large, it can be expected that the feedback controller may handle the disturbance satisfactory. The analysis of variance helps quantify how much the performance of the control loop can be improved, which can be translated in terms of increased product quality and/or material/energy consumption; see Sect. 13.2.3.

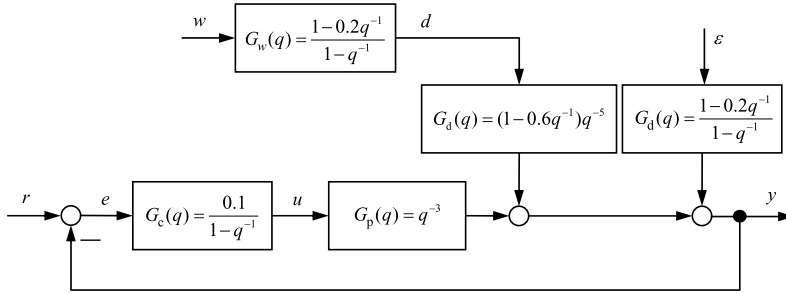
The procedure for variance estimation for feedforward/feedback control loops can be outlined as follows. It is demonstrated in simulation studies discussed below. An industrial application of this algorithm is presented in Sect. 16.3.1.

### Procedure 2.3 Variance estimation for feedforward/feedback control.

1. Determine or estimate the time delays.
2. Fit an ARMAX model to the closed-loop output samples  $y(k)$  and measured disturbance samples  $d_i(k)$  as inputs.
3. Fit individual AR(IMA) models to each of the feedforward variables  $d_i(k)$ .
4. Calculate the series expansions (impulse responses) for the estimated models.
5. Compute the variances as in Table 2.4.

*Example 2.6* The system consists of a pure time-delay process affected by output noise and a measurable disturbance. This linear system, adopted from Desborough and Harris (1993), has the structure and parameters illustrated in Fig. 2.11. For the simulation study, the driving noises were Gaussian random signals with the variances  $\sigma_w^2$  and  $\sigma_\varepsilon^2$ . A simple integral feedback controller was used as the initial controller.

Five cases will be studied. The first case considers the effect of weak disturbances, i.e. with low disturbance variance. In the second case, we increase the disturbance variance, compared to the first case. In the third case, the disturbance dynamics will be altered so that its average residence time will be significantly shorter. In the first three cases, the system was operated under feedback-only control, so the assessment method will give hints whether the loop should be extended with feedforward control. In the fourth case, a feedforward component will be added to the controller. Finally, the feedback controller will be re-tuned and evaluated again.



**Fig. 2.11** Structure and transfer functions of the considered control loop

**Table 2.5** Analysis of variance table [% of total variance] for Case 1

Disturbance	MV	FB	FF	FB/FF	Total
$\varepsilon(k)$	53.4	42.8	—	—	96.3
$w(k)$	0	—	3.0	0.7	3.7
Total	53.4		46.5		100

*Case 1. Weak Disturbances.* The variances of the driving noises were  $\sigma_w^2 = 0.1$  and  $\sigma_\varepsilon^2 = 1$ . A simulation of the system was carried out, and equidistant data were collected at a sampling time  $T_s = 0.1$  s. Steady-state operating data with 1000 samples were selected for calculating the ANOVA table; see Table 2.5. From this it can be deduced that the feedback controller ( $\sigma_{y,\varepsilon}^2 = 96.3$  %) is far from the minimum achievable variance ( $\sigma_{MV,\varepsilon}^2 = 53.4$  %) for the unmeasured noise. There is a major portion of the variance (42.8 %) which can be handled by a feedback tuning. However, the contribution (3.7 %) to the variance from the measured disturbance is small. 3 % can be reduced if an optimal feedforward controller is implemented. There is also a negligible portion of the variance (0.7 %) which can be handled by a combination of feedforward/feedback tuning. The conclusion is that the assessment method suggests that the feedback controller would benefit from improved tuning. However, an implementation of a feedforward controller is not recommended.

*Case 2. Strong Disturbances.* Here the transfer functions of the system are left unchanged, but the noise variance of the disturbance has been increased, i.e.  $\sigma_w^2 = \sigma_\varepsilon^2 = 1$ . Performing an ANOVA-analysis shows that the measured disturbance is now responsible for 27.7 % of the total variance, from which 22.6 % can be handled by feedforward control, see Table 2.6. This implies that the control loop will benefit from implementing a feedforward controller in this case.

*Case 3. Speed-up of Disturbance Dynamics.* In this case, the disturbance dynamics have changed such that there is no time delay, i.e.

$$G_d(s) = \frac{0.8}{1 + 4s}. \quad (2.64)$$



**Table 2.6** Analysis of variance table [% of total variance] for Case 2

Disturbance	MV	FB	FF	FB/FF	Total
$\varepsilon(k)$	40.5	31.9	–	–	72.3
$w(k)$	0	–	22.6	5.0	27.7
Total	40.5		59.5		100

**Table 2.7** Analysis of variance table [% of total variance] for Case 3

Disturbance	MV	FB	FF	FB/FF	Total
$\varepsilon(k)$	61.3	38.1	–	–	99.4
$w(k)$	0	–	0.4	0.2	0.6
Total	61.3		38.7		100

**Table 2.8** Analysis of variance table [% of total variance] for Case 4 with static feedforward control

Disturbance	MV	FB	FF	FB/FF	Total
$\varepsilon(k)$	46.1	38.1	–	–	84.2
$w(k)$	0	–	15.4	0.4	15.8
Total	46.1		53.9		100

**Table 2.9** Analysis of variance table [% of total variance] for Case 4 with dynamic feedforward control

Disturbance	MV	FB	FF	FB/FF	Total
$\varepsilon(k)$	49.8	41.7	–	–	91.6
$w(k)$	0	–	6.0	2.4	8.4
Total	49.8		50.1		100

The variances of the driving noises are the same as in Case 2. From performing similar simulations and looking at the ANOVA table (Table 2.7), the same conclusions can be drawn as in Case 1.

**Case 4. Effect of Feedforward Control.** The implementation of a simple proportional feedforward controller  $G_{FF} = 0.5$  leads to a 7 % decrease of the feedforward portion of variance, as shown in the ANOVA Table 2.8, compared to the performance in Case 2 (Table 2.6). This is due to the static feedforward which attempts to compensate for changes in the feedforward variable before these changes appear in the output. When now a two-sample delay is included in the feedforward controller, i.e.  $G_{FF} = 0.5q^{-2}$ , the feedforward portion of variance decreases up to 6 %, as shown in the ANOVA Table 2.9. From this table it can also be deduced that re-tuning or redesigning the feedback controller will yield the largest return, since it is still far from the minimum-variance performance.

**Table 2.10** Analysis of variance table [% of total variance] for Case 5 with dynamic feedforward control and re-tuned feedback control

Disturbance	MV	FB	FF	FB/FF	Total
$\varepsilon(k)$	71.4	14.1	–	–	85.6
$w(k)$	0	–	8.5	6.0	14.4
Total	71.4		28.6		100

*Case 5. Re-tuning of the Feedback Controller.* Just an increase of the proportional controller gain to  $K_c = 0.31$  yields substantial performance improvement for the feedback controller, as can be seen in Table 2.10. A further decrease of variance may be only achieved by redesigning the controller.

## 2.6 Assessment of Set-Point Tracking and Cascade Control

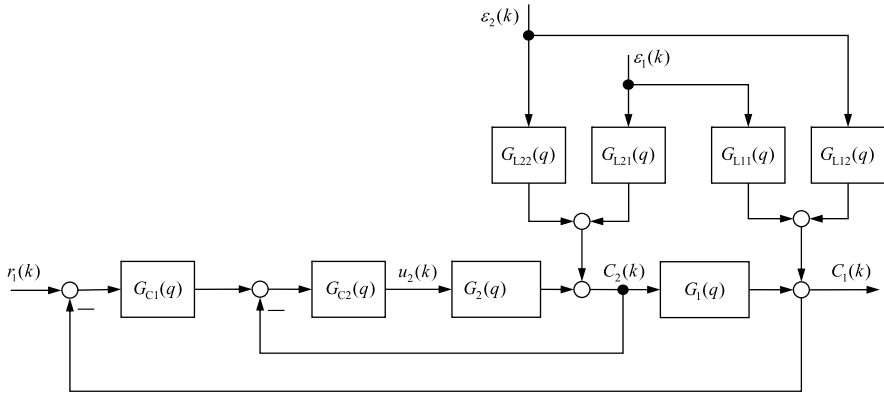
Most of the techniques presented above can be applied to single control loops operating in a regulatory mode, i.e. with constant set point. However, when set-point variations occur frequently, neglecting them will lead to under-estimation of the regulatory performance improvement. Equation 2.30 has to be extended to include the transfer function relating the control error to the set-point changes. The superposition principle gives the resulting closed-loop relation

$$y(k) = \frac{F_\tau(q)}{E_\tau(q)A(q)} q^{-\tau} r(k) + E_\tau(q)\varepsilon(k). \quad (2.65)$$

The estimation procedure of Sect. 2.4 thus has to be modified by estimating an ARMAX model with  $r(k)$  as the input signal, similar to the case of feedback plus feedforward control; see Sect. 2.5.

### 2.6.1 Performance Assessment of Cascade Control Systems

In process control applications, the rejection of load disturbances is often of main concern. To improve the control performance for this task, the implementation of a cascade control system is a good option. Indeed, cascade control is widely used in the process industries and is particularly useful when the disturbances are associated with the manipulated variable or when the final control element exhibits nonlinear behaviour (Shinskey 1996). Therefore, the main criterion to assess cascade control loops is its capability to reject load disturbances. Typical examples of cascade control from the metal processing industry are strip thickness control and flatness control systems; see Chap. 15.



**Fig. 2.12** Block diagram of a cascade control system

The minimum achievable variance with cascade control is generally lower than that from single-loop feedback control and can provide useful information on potential performance improvement. The above techniques can directly be used to analyse the performance of the primary loop under the assumption of constant set point, but require modifications for the analysis of the secondary loop. The relationships for the minimum-variance assessment of cascade control systems (Fig. 2.12) are derived following Ko and Edgar (2000).

Subscript 1 in this figure refers to the primary control loop, while subscript 2 refers to the secondary control loop.  $C_1(k)$  and  $C_2(k)$  are the process outputs of primary loop and the secondary loop, respectively.  $C_1(k)$  is the deviation variable from its set point, and  $C_2(k)$  the deviation of the secondary output from its steady-state value, which is required to keep the primary output at its set point.  $G_1(q) \equiv G_1^*(q)q^{-\tau_1}$  is the process transfer function in the primary loop with time delay equal to  $\tau_1$ , and  $G_1^*(q)$  is the primary process model without any time delay. It is assumed that  $G_1(q)$  has a stable inverse, i.e. all zeros lie inside the unit circle. The disturbance filters  $G_{L11}(q)$  and  $G_{L12}(q)$  are assumed to be rational functions of  $q^{-1}$ , and they are driven by zero-mean white noise sequences  $\varepsilon_1(k)$  and  $\varepsilon_2(k)$ , respectively. Similarly, for the secondary loop, we have  $G_2(q) \equiv G_2^*(q)q^{-\tau_2}$  as the process transfer function in the secondary loop with time delay equal to  $\tau_2$  and  $G_2^*(q)$  is the secondary process model without any time delay.  $G_2(q)$  is also assumed to be minimum-phase. The combined effect of all unmeasured disturbances to the secondary output is represented as a superposition of disturbance filters  $G_{L21}(q)$  and  $G_{L22}(q)$  driven by zero-mean white noise sequences  $\varepsilon_1(k)$  and  $\varepsilon_2(k)$ , respectively.

Using block-diagram algebra, from Table 2.11 it can simply be seen that:

$$\begin{aligned} C_1(k) &= G_1(q)C_2(k) + G_{L11}(q)\varepsilon_1(k) + G_{L12}(q)\varepsilon_2(k), \\ C_2(k) &= G_2(q)u_2(k) + G_{L21}(q)\varepsilon_1(k) + G_{L22}(q)\varepsilon_2(k), \end{aligned} \quad (2.66)$$

where  $u_2(k)$  is the manipulated variable in the secondary control loop. The MVC algorithm for the equation system 2.66 is given by

- *Primary Controller*

$$G_{c1,MV} = \frac{G_1^*(Q_{22}R_{21} - Q_{21}R_{22}) + (R_{11} + T_1)G_{L22} - (R_{12} + T_2)G_{L21}}{(Q_{11} + S_1q^{-\tau_1})(R_{12} + T_2 + G_1^*R_{22}) - (Q_{12} + S_2q^{-\tau_1})(R_{11} + T_1 + G_1^*R_{21})}. \quad (2.67)$$

- *Secondary Controller*

$$G_{c2,MV} = \frac{(Q_{11} + S_1q^{-\tau_1})(R_{12} + T_2 + G_1^*R_{22}) - (Q_{12} + S_2q^{-\tau_1})(R_{11} + T_1 + G_1^*R_{21})}{G_1^*[G_{L11}S_2 - G_{L12}S_1 + (R_{11}Q_{12} - R_{12}Q_{11})q^{-\tau_2}]}, \quad (2.68)$$

where  $Q_{11}(q)$  and  $Q_{12}(q)$  are polynomials in  $q^{-1}$  of order  $\tau_1 + \tau_2 - 1$ ,  $Q_{21}(q)$ ,  $Q_{22}(q)$ ,  $S_1(q)$  and  $S_2(q)$  are polynomials in  $q^{-1}$  of order  $\tau_2 - 1$ , and  $R_{ij}(q)$  ( $i, j = 1, 2$ ) are proper transfer functions that satisfy the following Diophantine identities:

$$\begin{aligned} G_{L11} &= Q_{11} + R_{11}q^{-\tau_1-\tau_2}, \\ G_{L12} &= Q_{12} + R_{12}q^{-\tau_1-\tau_2}, \\ G_{L21} &= Q_{21} + R_{21}q^{-\tau_2}, \\ G_{L22} &= Q_{22} + R_{22}q^{-\tau_2}, \\ G_1^*Q_{21} &= S_1 + T_1q^{-\tau_2}, \\ G_1^*Q_{22} &= S_2 + T_2q^{-\tau_2}. \end{aligned} \quad (2.69)$$

The primary output  $C_1(k)$  under this optimal control algorithm is an MA process of order  $\tau_1 + \tau_2 - 1$

$$C_1(k) = [Q_{11}(q) + S_1(q)q^{-\tau_1}]\varepsilon_1(k) + [Q_{12}(q) + S_2(q)q^{-\tau_1}]\varepsilon_2(k), \quad (2.70)$$

and the MV of  $C_1(k)$  is

$$\sigma_{C_1,MV}^2 = \text{trace} \left\{ \left( \sum_{i=0}^{\tau_1+\tau_2-1} N_i^T N_i \right) \Sigma_\varepsilon \right\}, \quad (2.71)$$

where  $N_i$  ( $i = 0, 1, \dots, \tau_1 + \tau_2 - 1$ ) are defined as the coefficient matrices of the matrix polynomial  $[(Q_{11} + S_1q^{-\tau_1})(Q_{12} + S_2q^{-\tau_1})]$ , and  $\Sigma_\varepsilon$  is the variance-covariance matrix of the white noise vector  $[\varepsilon_1(k) \varepsilon_2(k)]^T$ . The derivation of the above relationships can be found by Ko and Edgar (2000). Since the polynomials  $Q_{11}$ ,  $Q_{12}$ ,  $S_1$ ,  $S_2$  in Eq. 2.70 are all feedback-invariant, the expression for the primary output under MV cascade control can be estimated from the first  $\tau_1 + \tau_2 - 1$  MA coefficients of the closed-loop transfer functions relating  $\varepsilon_1(k)$  to  $C_1(k)$  and  $\varepsilon_2(k)$  to  $C_1(k)$ . No joint identification of the process dynamics and the disturbance model is needed.

The closed-loop transfer functions in this case can be obtained from the first row of the transfer function matrix estimated via *multivariate time-series analysis* of

$[C_1(k), C_2(k)]^T$ . For this analysis, an AR model [arx setting an empty input] can be used efficiently with its computational speed. Alternatively, a state space model can be estimated via the prediction error method [pem] or a subspace identification method [n4sid]; see Sect. 7.2.6. The sample variance-covariance matrix of the residual vectors thus provides an estimate of the variance and the covariance elements of the innovation sequences. The closed-loop impulse-response coefficients can then be determined via simple correlation analysis between the output variables and the estimated innovations sequences, or by solving a suitable Diophantine identity concerning the estimated parameter matrix polynomial,

$$\begin{aligned}\sigma_{C_1, MV}^2 &= \text{var}\{(h_{10} + h_{11}q^{-1} + \cdots + h_{1, \tau_1 + \tau_1 - 1}q^{-(\tau_1 + \tau_1 - 1)})\varepsilon_1 \\ &\quad + (h_{20} + h_{21}q^{-1} + \cdots + h_{2, \tau_1 + \tau_1 - 1}q^{-(\tau_1 + \tau_1 - 1)})\varepsilon_2\} \\ &= \text{trace}\left\{\left(\sum_{i=0}^{\tau_1 + \tau_1 - 1} N_i^T N_i\right) \Sigma_\varepsilon\right\}.\end{aligned}\quad (2.72)$$

The MV performance index for the cascade control system is defined as

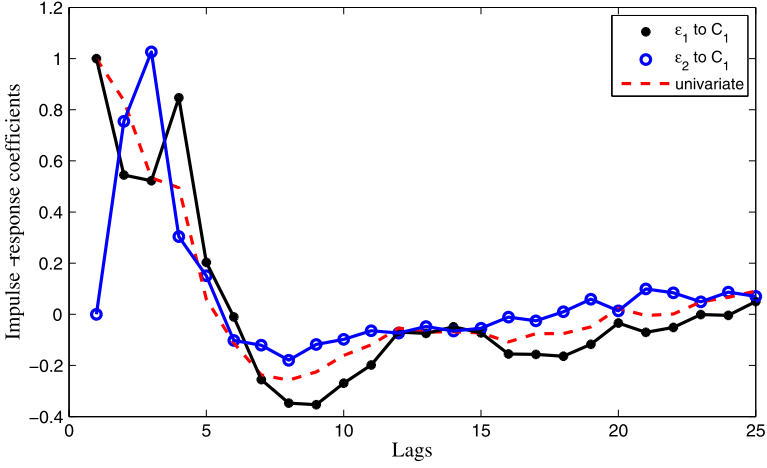
$$\eta = \frac{\sigma_{C_1, MV}^2}{\sigma_{C_1}^2}.\quad (2.73)$$

Naturally, the question arises if the MV can be calculated by just applying univariate analysis on  $C_1(k)$ . Indeed, this would lead to similar results, but only in the case where the net disturbance effect driven by  $\varepsilon_2(k)$  is negligible. Otherwise, the estimated performance index from univariate analysis is higher than the estimated performance index value obtained through multivariate analysis. Thus, univariate analysis of cascade control loops would yield an over-estimate of the control performance.

*Example 2.7* We consider the example of a process described by Eqs. 2.66 with (Ko and Edgar 2000)

$$\begin{aligned}G_1(q) &= \frac{q^{-2}}{1 - 0.9q^{-1}}, & G_{L11}(q) &= \frac{1}{1 - 0.8q^{-1}}, & G_{L12}(q) &= \frac{q^{-1}}{1 - 0.1q^{-1}}, \\ G_2(q) &= \frac{q^{-1}}{1 - 0.5q^{-1}}, & G_{L21}(q) &= \frac{q^{-1}}{1 - 0.2q^{-1}}, & G_{L22}(q) &= \frac{1}{1 - 0.3q^{-1}}\end{aligned}\quad (2.74)$$

to illustrate how the assessment procedure of cascade control systems works in detail. For this system, the primary process time delay is two samples ( $\tau_1 = 2$ ), and the secondary process time delay is one sample ( $\tau_2 = 1$ ). The process is subjected to disturbances in the form of white noise sequences  $\{\varepsilon_1(k), \varepsilon_2(k)\}$  with unity variance-covariance matrix  $\Sigma_\varepsilon = \begin{bmatrix} 1.0 & 0.1 \\ 0.1 & 1.0 \end{bmatrix}$ . A closed-loop simulation was performed using a



**Fig. 2.13** Closed-loop impulse responses from simulated data

PI controller for the primary loop and P-only controller for the secondary loop. The transfer functions for the controllers used are

$$G_{c1}(q) = \frac{0.48 - 0.4q^{-1}}{1 - q^{-1}}; \quad G_{c2}(q) = 0.7.$$

A data set of 2000 samples for the primary and secondary outputs was collected and a multivariate AR model of 25th order was fitted to the gathered data. The estimated variance-covariance matrix of the white noise sequence is  $\hat{\Sigma}_\varepsilon = \begin{bmatrix} 1.36 & 0.43 \\ 0.43 & 1.30 \end{bmatrix}$ . From this model the estimated closed-loop impulse responses have been obtained, as shown in Fig. 2.13.

The estimated minimum variance by multivariate analysis is (Eq. 2.72)

$$\begin{aligned} \sigma_{C_1, MV}^2 = & \text{trace} \left\{ \left( \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} + \begin{bmatrix} 0.48 \\ 0.81 \end{bmatrix} \begin{bmatrix} 0.48 & 0.81 \end{bmatrix}, \right. \\ & \left. + \begin{bmatrix} 0.49 \\ 1.03 \end{bmatrix} \begin{bmatrix} 0.49 & 1.03 \end{bmatrix} \right) \begin{bmatrix} 1.36 & 0.43 \\ 0.43 & 1.30 \end{bmatrix} \right\} \approx 5.01. \end{aligned} \quad (2.75)$$

This gives the estimated performance index (Eq. 2.73)

$$\eta_{C_1, \text{multi}} = \frac{\sigma_{C_1, MV}^2}{\sigma_{C_1}^2} = \frac{5.01}{7.28} \approx 0.69. \quad (2.76)$$

For a comparison, univariate analysis using an AR model of 25th order fitted to the primary response data  $C_1(k)$  has also been carried out. The obtained univariate closed-loop impulse response is also illustrated in Fig. 2.13. The estimated mini-

minimum variance by the univariate analysis is (Eq. 2.38)

$$\sigma_{C_1, MV}^2 = (1.0^2 + 0.836^2 + 0.534^2)(2.822) = 5.596.$$

This yields the estimated performance index as

$$\eta_{C_1, uni} = \frac{\sigma_{C_1, MV}^2}{\sigma_{C_1}^2} = \frac{5.596}{7.28} \approx 0.79. \quad (2.77)$$

The estimated performance index by univariate analysis is (13 %) higher than the estimated performance index value obtained through a multivariate analysis. Thus, univariate analysis of cascade control loops would yield an over-estimate of the controller's performance. Note that analysis of the inner loop yields an estimated performance index of

$$\eta_{C_2, uni} = \frac{2.76}{2.61} \approx 0.92,$$

indicating very good performance.

Just to theoretically confirm the results achieved above, we now calculate the minimum variance from the full knowledge of the process and disturbance models by solving Diophantine Eqs. 2.69 using

$$\frac{1}{1 - aq^{-1}} = \sum_{i=0}^{\infty} a^i q^{-i}. \quad (2.78)$$

This gives specifically for the considered case:<sup>5</sup>

$$\begin{aligned} \frac{1}{1 - 0.8q^{-1}} &= 1 + 0.8q^{-1} + 0.64q^{-2} + 0.512q^{-3} = Q_{11} + R_{11}q^{-3}, \\ \frac{q^{-1}}{1 - 0.1q^{-1}} &= q^{-1}(1 + 0.1q^{-1} + 0.01q^{-2}) = Q_{12} + R_{12}q^{-3}, \\ \frac{q^{-1}}{1 - 0.2q^{-1}} &= q^{-1}(1) = Q_{21} + R_{21}q^{-1}, \\ \frac{1}{1 - 0.3q^{-1}} &= 1 + 0.3q^{-1} = Q_{22} + R_{22}q^{-1}, \\ \frac{1}{1 - 0.9q^{-1}} Q_{21} &= S_1 + T_1q^{-1}, \\ \frac{1}{1 - 0.9q^{-1}} Q_{22} &= S_2 + T_2q^{-1}, \end{aligned}$$

---

<sup>5</sup>Remaining error terms are ignored here for simplicity.

where  $Q_{11}$  and  $Q_{12}$  are polynomials in  $q^{-1}$  of order 2,  $Q_{21}$ ,  $Q_{22}$ ,  $S_1$  and  $S_2$  are constants, and  $Q_{ij}$  and  $T_i$  ( $i, j = 1, 2$ ) are proper transfer functions. The solution of these identities yields

$$\begin{aligned} Q_{11} &= 1 + 0.8q^{-1} + 0.64q^{-2}, & R_{11} &= 0.512, \\ Q_{12} &= q^{-1} + 0.1q^{-2}, & R_{12} &= 0.01, \\ Q_{21} &= 0, & R_{21} &= 1, \\ Q_{22} &= 1, & R_{22} &= 0.3, \\ S_1 &= 0, & T_1 &= 0, \\ S_2 &= 1, & T_2 &= 0.9. \end{aligned}$$

Thus, from Eq. 2.70, the primary output  $C_1(k)$  under minimum-variance cascade control is given by

$$C_1(k) = [1 + 0.8q^{-1} + 0.64q^{-2}]\varepsilon_1(k) + [q^{-1} + 1.1q^{-2}]\varepsilon_2(k). \quad (2.79)$$

The minimum variance follows as (Eq. 2.72)

$$\begin{aligned} \sigma_{C_1, MV}^2 &= \text{trace} \left\{ \left( \begin{bmatrix} 1 \\ 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \end{bmatrix} + \begin{bmatrix} 0.8 \\ 1 \end{bmatrix} \begin{bmatrix} 0.8 & 1 \end{bmatrix} \right. \right. \\ &\quad \left. \left. + \begin{bmatrix} 0.64 \\ 1.1 \end{bmatrix} \begin{bmatrix} 0.64 & 1.1 \end{bmatrix} \right) \begin{bmatrix} 1 & 0.1 \\ 0.1 & 1 \end{bmatrix} \right\} \approx 4.56, \end{aligned} \quad (2.80)$$

which is close to the value in Eq. 2.75, estimated only from the simulated closed-loop data and the knowledge of process time delays.

### 2.6.2 Assessment of Different Tuning Strategies

The conventional strategy to tune a cascade control loop is to first tune the secondary controller with the primary controller in the manual mode. Then the primary controller is transferred to automatic, and it is tuned. If the secondary controller is re-tuned for some reason, usually the primary controller must also be re-tuned. It has been devised by Seborg et al. (2004) to tune the slave loop tighter than the master loop to get improved stability characteristics and thus allow larger values of  $K_{c1}$  to be used in the primary control loop. Note that the presence of an integrator in the secondary loop is not strictly necessary since the null steady-state error can be assured by the primary loop controller. If integral action is employed in both the master and the slave controllers, the integrator windup should be carefully handled. A typical approach is to stop the integration of the primary controller when the output of the secondary controller attains its limits (Visioli 2006).



Principally, any (appropriate) tuning method can be applied to both controllers. However, there are some tuning methods explicitly tailored for cascade control systems, such as relay feedback simultaneous tuning, IMC-based simultaneous tuning and SPC-based simultaneous tuning. These will not be described here, but the reader is referred to Visioli (2006: Chap. 9) and the references included therein. As the usual sequential tuning is time-consuming, simultaneous tuning methods should be preferred.

*Example 2.8* Consider the cascade system with the transfer functions

$$G_1(s) = \frac{1}{(s+1)^3}, \quad G_2(s) = \frac{1}{s+1}. \quad (2.81)$$

This system was used by Åström and Hägglund (2006) to show the improved performance for deterministic load disturbance rejection of cascade control, compared with conventional PI control. The parameters of the latter controller were  $K_c = 0.37$  and  $T_I = 2.2$ . For cascade control, a P controller with  $K_c = 5$  was placed in the secondary loop, and a PI controller with  $K_c = 0.55$  and  $T_I = 1.9$  in the primary loop. A high-gain controller is possible in the secondary loop, as its response to the control signal is quite fast.

We now evaluate the performance of both control systems in terms of stochastic disturbance rejection using the MV index. The hypothetical disturbances dynamics and variances are assumed to be the same as in Example 2.7. Under these circumstances, the computed performance index values were  $\eta_{1,\text{uni}} = 0.62$  for the conventional controller and  $\eta_{1,\text{uni}} = 0.85$ ,  $\eta_{1,\text{multi}} = 0.77$  for the cascade control. These results reveal the increase of stochastic performance achieved by the cascade control.

*Example 2.9* An evaluation and comparison of the cascade-control tuning methods mentioned above in the context of CPM is provided using the following example:

$$G_1(s) = \frac{e^{-4s}}{(5s+1)^2}, \quad G_2(s) = \frac{e^{-0.2s}}{s+1}. \quad (2.82)$$

The different PID controller settings derived by Visioli (2006) are given in Table 2.11. This also contains achieved values of the performance indices, the MV index  $\eta_{2,\text{uni}}$  for the inner loop and the MV index  $\eta_{1,\text{multi}}$  for the outer loop based on the multivariable performance analysis. The results confirm again the need for multivariable analysis; otherwise the performance is overestimated.

All tuning methods yields similar and satisfactory, but not excellent, control performance in the primary loop, despite the excellent (stochastic) performance in the inner loop. Therefore, there is still improvement potential for the primary loop from stochastic performance view point. If, for instance, the primary controller is significantly detuned, i.e.,  $\lambda$  is increased, the variance is increased, however, only at the expense of increased rise time.

**Table 2.11** Used controller parameters and assessment results

	Primary controller			Secondary controller			Minimum-variance assessment		
	$K_{c1}$	$T_{I1}$	$T_{D1}$	$K_{c2}$	$T_{I2}$	$T_{D2}$	$\eta_{2,uni}$	$\eta_{1,uni}$	$\eta_{1,multi}$
Initial tuning	1.0	12.0	0	0.5	4.0	0	<b>0.82</b>	<b>0.72</b>	<b>0.64</b>
Relay feedback tuning	1.18	19.0	4.75	0.56	2.14	0	<b>0.95</b>	<b>0.77</b>	<b>0.70</b>
IMC-based tuning 1	0.94	10.0	1.89	3.31	1.06	0.07	<b>0.81</b>	<b>0.71</b>	<b>0.63</b>
IMC-based tuning 2	0.22	8.30	0.56	2.48	1.04	0.03	<b>0.84</b>	<b>0.89</b>	<b>0.83</b>
SPC-based tuning	1.5	8.22	0.91	3.17	1.05	0	<b>0.91</b>	<b>0.78</b>	<b>0.58</b>

From this example we also learn that tuning cascade control should always be driven towards maximising the Harris index (calculated using multivariable analysis) of the primary loop, when the variance is the main point. Thereby, it is not necessary to maximise the Harris index for the secondary loop. This conclusion seems to be not in agreement with the conventional approach of tuning cascade controllers. A similar conclusion was also pointed out by Teo et al. (2005) from their experience on another example.

## 2.7 Summary and Conclusions

Performance assessment based on minimum variance control, as the standard method for evaluating controllers, has been presented in detail. Besides batch calculation, the performance index can also be computed recursively, enabling the use of control charts for online monitoring of changes in controller performance. The following advantages of MV benchmarking contributed to its popularity and usage in the majority of CPM applications:

- Metrics based on MVC are the main criteria used in stochastic performance assessment, providing a direct relationship between the variance of key variables and product quality or energy/material consumption, which are correlated with financial benefits.
- MV benchmarking is easy to apply and implement and remains valuable as an absolute bound on performance against which real controllers can be compared. Performance monitoring should always include at least a look at the Harris index, as a first pass-assessment layer to bring obvious problems to immediate attention.
- Considering the MVC lower bound in setting performance targets will ensure that overly optimistic and conservative performance targets are avoided. MVC information can also be beneficial in incentive studies.

However, one should be aware about some serious drawbacks:

- A well-functioning loop in the process industry has frequently variance well above the minimum variance. Also industrial controllers (usually of the PID-type) do not have always a chance to match the MVC performance.

- Even though, MV control action can lead to highly undesirable, aggressive control and poor robustness.

Principally, MVC-based assessment is useful irrespective of the type of controller installed at the plant. However, tailored versions of MV assessment, such as those for feedback-plus-feedforward control and cascade control, can also be applied when the controller structure is known. Both control strategies are of widespread use in the process industry. The analysis of variance for feedback-plus-feedforward control helps quantify how much the performance of the control loop can be improved by re-tuning the feedforward component or introducing such a component if not yet implemented. For cascade control, it was shown that multivariate performance analysis should be generally applied, since univariate analysis may yield over-estimated loop performance, thus giving misleading conclusions. Also, tuning cascade control should always be driven towards maximising the Harris index (calculated using multivariable analysis) of the primary loop, when the variance is the main point.

Control Performance Management in Industrial  
Automation

Assessment, Diagnosis and Improvement of Control  
Loop Performance

Jelali, M.

2013, XXVII, 480 p., Hardcover

ISBN: 978-1-4471-4545-5