BACHELOR

Independence of the Self-Starting Statistic and Recursive Residuals

Pennings, Gijs P.S.

Award date:
2022

Independence of the Self-Starting Statistic and Recursive Residuals

Bachelor thesis

G. P. S. Pennings

Supervised by Alessandro Di Bucchianico

August 2022

Abstract

Control charts are used to detect changes in a process. They require a benchmark period (i.e., Phase I), but if this is infeasible, self-starting charts can be used instead. For effective use of these charts, the charted statistics must be independent. However, we found that the existing proofs of the independence of the self-starting statistic either show a weaker property (e.g., serial independence) or depend on other complicated results. Therefore, in this thesis, we provide a new, detailed proof which shows that the self-starting statistics are mutually independent. Furthermore, we extend this proof to a regression setting and show that the studentized recursive residuals are independent. Also, we give a practical example that shows how normalized recursive residuals (i.e., studentized recursive residuals transformed to a standard normal distribution) combine control charts with linear regression.

Contents

Chapter 1

Introduction

In this thesis, we will study the independence properties of a statistical tool called control charts and, in the context of linear regression, a special type of residual called recursive residuals. In this chapter, we will first give some background on these topics and demonstrate why they are important (Section 1.1). Then, we will define the goal of this thesis and establish the central research questions (Section 1.2). Finally, we will outline the structure of the remainder of the thesis (Section 1.3).

1.1 Context

Especially in the age of big data, monitoring data streams is important. For example, consider manufacturing processes, where products should conform to the specification, or wind turbines, where the temperature may not exceed an upper limit. For such applications, control charts are a simple but effective tool. In their simplest form, measurements of some process are plotted on the chart, and if the points on the chart exceed predetermined limits, there is likely a problem. An example is shown in Figure 4.1. Control charts are most effective when these measurements are independent, since this minimizes the likelihood of false alarms.

Before control charts can be used for monitoring, their limits must first be calibrated using benchmark data. However, sometimes this is infeasible. For this reason, another type of control chart was developed, called self-starting control charts. Essentially, these charts compute the limits on-the-fly by transforming each measurement using all previous measurements. However, due to this transformation, it is not apparent anymore whether the computed statistics are independent, even if the underlying measurements are. Therefore, since independence is an important theoretical property, it needs to be investigated more carefully.

Sometimes, the measurements are affected by (external) influences. For example, the generator temperature of a wind turbine is affected by the outside temperature.

To detect changes among expected fluctuations, it is important to account for these influences. This can be achieved by plotting the residuals of a linear regression model. Recursive residuals, a special type of residual, are self-starting (i.e., they do not require benchmark data) and independent, which makes them suitable for self-starting control charts.

When combining self-starting control charts with linear regression, we will refer to it as "the regression case". On the other hand, ordinary self-starting control charts are referred to as "the univariate case".

1.2 Problem statement

The goal of this thesis is to formally prove that the statistics used in self-starting control charts are independent, in both the univariate and regression case. To achieve this, we will answer the following research questions.

- Which proofs exist in the literature on the independence of the statistics used in self-starting control charts, in both the univariate and regression case, or of other related concepts?
- Are the existing proofs valid? And if not, how can we correct or extend them?

1.3 Outline of thesis

First, in Chapter 2, we perform a literature review on control charts in general and on proofs of independence in particular. Next, Chapter 3 introduces preliminaries that are used in the later chapters. In Chapter 4, we lay a mathematical foundation for control charts. Then, Chapter 5 presents new proofs of mutual independence of the self-starting statistics using elementary methods. In Chapter 6, a practical example of self-starting control charts is presented. Finally, the work is summarized and possible extensions are suggested in Chapter 7.

The first research question is mainly addressed in Chapter 2, and the second one in Chapter 5. The remaining chapters provide the necessary background and mathematical preliminaries. Furthermore, some notation is presented in Appendix A, Appendix B contains miscellaneous mathematical results, and the code used for this thesis can be found in Appendix C.

Chapter 2

Literature review

In this chapter, we will place this thesis and its results in a broader context. First, we will give an outline of relevant literature on control charts, with a focus on self-starting charts (Section 2.1). Next, we will explore two related concepts, namely slippage tests (Section 2.2) and recursive residuals (Section 2.3), and see how they connect to self-starting control charts. Lastly and most importantly, we will give an overview of proofs of independence for the self-starting statistics and recursive residuals as found in the literature (Section 2.4).

2.1 Control charts

The field of statistical process monitoring[1] (SPM) is said to be founded by Shewhart (1931) when he introduced control charts. It employs a number of statistical tools that are used to monitor processes. These tools are also known under names such as The Magnificent Seven and are part of methodologies like Six Sigma (Montgomery, 2019). An introduction to SPM and control charts is given in Chapter 4. An overview of both mathematical theory and practical applications can be found in Qiu (2013).

Although basic Shewhart charts are suitable to detect large, sudden shifts in the mean, they are less suitable to detect other types of changes. For this purpose, well-known alternatives such as cumulative sum charts (CUSUM) (Page, 1954) and exponentially weighted moving average charts (EWMA) (Roberts, 1959) can be used. Instead of only the current sample, these charts take the entire history of the process into account. In general, these charts are better at detecting small shifts or gradual drifts than Shewhart charts. Nevertheless, due to their simplicity, basic Shewhart charts are still widely used today.

[1] This is the modern name for statistical process control (SPC), as proposed by Woodall (2017). SPC is a misnomer, since "control" suggests that control actions are part of the toolset, which is almost never the case.

Another type of chart that detects changes in the mean, known as the regression chart, combines Shewhart charts with linear regression to remove external influences. These were first proposed by Mandel (1969). A recent review is included in Centofanti et al. (2021). Regression charts are closely related to (linear) profile monitoring, as noted by Kim et al. (2003).

In low-volume manufacturing, basic Shewhart charts are not suitable, since they require a relatively long benchmark period. Alternatively, the self-starting CUSUM charts introduced by Hawkins (1987) can be used, which do not require that distributional parameters are known at the start of production. A few years later, Quesenberry (1991) proposed a similar technique for Shewhart charts, called Q-charts. Their performance is analyzed in Del Castillo and Montgomery (1994) and some improvements are proposed.

2.2 Slippage tests

A field closely related to control charts is that of outlier tests. In particular, a statistic similar to the self-starting statistic[2] is used in an article by Quesenberry and David (1961) on slippage tests. These tests were first introduced by Mosteller (1948) and can be considered a generalization of the one-sample outlier problem. They are used to test whether one of multiple samples (each typically containing more than one measurement) have "slipped", e.g., whether the sample has a different mean. These samples can be compared to rational subgroups for control charts (see Section 4.2).

The results by Quesenberry and David (1961) are essentially a simplification of earlier work by Doornbos et al. (1956). Both approaches are also discussed in a survey by Doornbos (1976). For a complete overview of slippage tests, we refer to Chapter 5 of Barnett and Lewis (1994).

2.3 Recursive residuals

Recursive residuals are a powerful tool in regression diagnostics. They were presented in the seminal paper by Brown et al. (1975), though, according to Farebrother (1978), they date back to the 19th century. They are a special type of residual (see Section 3.3) which are independent and homoscedastic[3]. These properties make them attractive for statistical tests and monitoring. They can be used to test for structural change[4], heteroscedasticity, influential observations, and more. A number of these applications are discussed by Galpin and Hawkins (1984), and a review is given by Kianifard and

[2] Strictly speaking, "the self-starting statistic" (singular) refers to case IV as discussed in Section 2.4.

[3] A collection of random variables is homoscedastic if they have the same (finite) variance. Otherwise, they are heteroscedastic.

[4] An (unexpected) change in the regression coefficients.

Swallow (1996). Recursive residuals are also known in the field of econometrics. For example, see Theil (1971). Furthermore, they are a special case of Kalman filters. This is discussed by Pollock (2003) in more detail.

A statistic essentially equivalent to recursive residuals was defined by Hedayat and Robson (1970), called stepwise residuals. Furthermore, a studentized variant of recursive residuals was introduced by Marr and Quesenberry (1991), called normalized uniform (NU) residuals. These are also discussed in Section 6.5.7 of Quesenberry (1986a). The same statistic is used in the context of linear profile monitoring by Zou et al. (2007), where they are referred to as standardized recursive residuals. Lastly, up to a multiplicative constant, the same statistic is also given by Hawkins (1991).

Recursive residuals are similar to the (sequential) uniform residuals for self-starting control charts defined by Quesenberry (1986b), both in name and approach. This can be explained by the fact that both follow from the conditional probability integral transformation (CPIT) by O'Reilly and Quesenberry (1973), which is discussed in more detail in Section 2.4. This common link to the CPIT suggests that the self-starting statistic and recursive residuals are "structured" in a similar way. In fact, the latter can be seen as a generalization of the former, as will be shown by Corollary 5.14.

2.4 Proofs of independence

The goal of this thesis is to prove that the self-starting statistics are independent, and that independence also holds for their extension to the regression case, i.e., the recursive residuals. Therefore, we will now look more closely at existing proofs of this in the literature.

In the original paper by Quesenberry (1991) on self-starting charts, four cases of the self-starting statistic are introduced in the section on individual measurements. In that paper, they are referred to as Q-statistics. Readers of that paper are referred to the appendix for a proof that the Q-statistics are i.i.d. standard normal random variables. For cases I to III it is stated that independence follows "immediately" from previous work on the CPIT by O'Reilly and Quesenberry (1973), even though the author acknowledges this paper is "rather abstract". For case IV, a succinct proof is given. However, the proof is flawed, since only serial independence (see Definition 3.4) is shown, not mutual independence. The proof has other issues as well, which are discussed in detail throughout Section 5.1.

When the Q-statistic for each of the four cases is defined, a previous paper by Quesenberry (1986b) on outlier detection in control charts is cited. There, the same random variables are used as test statistics. They are transformed to a standard uniform distribution instead of a standard normal distribution, but this does not affect their independence. The k-th statistic is regarded as the (studentized) residual of the k-th measurement based on the sample mean and variance of the first $k-1$ measurements,

so the test statistics are referred to as sequential (uniform) residuals. In that paper, Property 1 states that these are i.i.d., but no further details are given, and O'Reilly and Quesenberry (1973) is cited for the proof.

The CPIT, which is applied in both papers by Quesenberry discussed above, builds on earlier work by Rosenblatt (1952). The transformation is also described in Section 6.2.2 of Quesenberry (1986a), and without measure-theoretic details by Gaudoin (1999). In Examples 4.1 and 4.3 of the original paper by O'Reilly and Quesenberry (1973), the mutual independence of random variables similar to the self-starting statistic and recursive residuals is discussed. However, these concepts are not mentioned explicitly, the random variables do not match the ordinary definitions directly, and some steps are omitted. For these reasons, the proofs are difficult to verify. Therefore, this calls for a more thorough verification.

In an article by Hawkins (1987), the case IV self-starting statistic is used to introduce self-starting CUSUM charts. Again, these are referred to as (studentized) residuals. It is stated that successive statistics are independent, referencing earlier work of Hawkins (1969). This, in turn, cites Quesenberry and David (1961) on slippage tests. Here, Basu's theorem[5] is invoked in order to argue that the test statistics are serially independent. However, details are missing, i.e., it is not clear how Basu's theorem is applied. Quesenberry and David (1961) note that this proof is a simpler alternative to a proof by Doornbos et al. (1956), who prove serial independence by directly computing the density.

In the preceding discussion, we searched for proofs of independence of the self-starting statistic. Alternatively, we can search for a proof that shows the recursive residuals are independent, which then implies independence of the self-starting statistic as a special case. Brown et al. (1975) prove mutual independence, but the variance is assumed to be known (i.e., the distribution depends on σ), making it only comparable to the case II self-starting statistic. Zou et al. (2007) assume that the variance is unknown (i.e., they consider the studentized recursive residuals), but they justify independence by only mentioning Basu's theorem, without presenting further details. This is unsatisfactory. Other sources that discuss studentized recursive residuals, namely Marr and Quesenberry (1991) and Hawkins (1991), both refer to the CPIT without providing any details. Hence, we return to the same work by O'Reilly and Quesenberry (1973).

In conclusion, to the author's knowledge, focusing on the case with unknown variance, there are four proofs of independence in the literature.

- In the context of slippage tests, Quesenberry and David (1961) prove serial independence using a theorem by Basu (1955). Doornbos et al. (1956) also prove serial independence, but achieve this by computing the density.

[5]This well-known theorem states that a complete sufficient statistic is independent of any ancillary statistic. It was presented by Basu (1955).

- Quesenberry (1991) proves serial independence. This approach is further discussed in Section 5.1.2, and extended in Section 5.1.3.
- An application of the CPIT by O'Reilly and Quesenberry (1973) proves mutual independence.

Of these proofs, only the last proves the strongest property of mutual independence. However, as discussed above, the precise application of the CPIT is unclear. Therefore, in Chapter 5, we will give a proof using more elementary methods.

Chapter summary

We introduced three concepts (i.e., self-starting control charts, slippage tests, and recursive residuals) that all involve a type of outlier test, and are closely related. With the goal of proving the mutual independence of the self-starting statistic and of the recursive residuals, we reviewed the existing literature on these topics. We found that all proofs either show serial independence (a weaker condition) or are highly technical and simultaneously terse. Therefore, this topic requires further study.

Chapter 3

Preliminaries

In this chapter, we will present some preliminaries that are used in Chapter 4 and, most importantly, for the proofs in Chapter 5. Since the focus of this thesis is on independence, we will formalize this notion and give a number of results related to it (Section 3.1). Specifically, we will discuss the joint normal distribution and how it can be used to prove independence (Section 3.2). Next, we will introduce linear regression (Section 3.3), which provides the necessary background to define the recursive residuals (Section 3.3.4). Lastly, we will state and prove several updating formulae that are critical to the proofs in Chapter 5 (Section 3.4).

3.1 Independence

The concept of *independence* is central to this thesis. Therefore, in this section, some important definitions and theorems will be presented, with a focus on continuous random variables.

In general terms, two random variables are independent if knowing the value of one does not change the distribution of the other. Since this thesis only deals with probability distributions for which the density exists (e.g., the normal distribution and t-distribution), we will use the following formal characterization.

Definition 3.1 — Let X and Y be continuous random variables with marginal densities f_X and f_Y, respectively, and joint density $f_{X,Y}$. Then X and Y are independent if and only if $f_{X,Y}(x, y) = f_X(x)f_Y(y)$ for all $x, y \in \mathbb{R}$.

The symbol "$\perp\!\!\!\perp$" is used to denote independence, i.e., $A \perp\!\!\!\perp B$ means that the random variables A and B are independent.

When considering more than two random variables, we distinguish between two types of independence. Namely, a collection of random variables is called *pairwise* independent if any pair of them is independent, but it is *mutually* independent if,

informally speaking, each random variable is independent of any combination of other random variables. In other words, the random variables $X_1, \dots, X_n$ are pairwise independent if Definition 3.1 holds for any X_i and X_j ($i \neq j$), but they are only mutually independent under a more general condition, such as the following.

Theorem 3.2 — *Let $X_1, \dots, X_n$ be continuous random variables with a joint density f. Then $X_1, \dots, X_n$ are mutually independent if and only if $f(x_1, \dots, x_n)$ can be written as a product of functions of $x_1, \dots, x_n$ alone, i.e., $f(x_1, \dots, x_n) = g_1(x_1) \cdots g_n(x_n)$.*

Proof. See Appendix B. □

If we take g_i as f_{X_i} (i.e., the marginal density of X_i), this result corresponds directly to a stronger variant of Definition 3.1. However, it is sufficient if g_i equals f_{X_i} up to a multiplicative constant (which is essentially only a normalization issue). This makes the practical application of Theorem 3.2 easier. We will sometimes summarize this result by writing "the density factorizes".

Mutual independence implies pairwise independence, but the converse is *not* true in general. Below, we provide a counterexample to prove this. There exist many counterexamples using discrete random variables, but we give one using normal random variables, taken from Section 2.12 of Romano and Siegel (1986).

Counterexample 3.3 — Let X, Y, Z_0 be mutually independent standard normal random variables, and define $Z = |Z_0| \operatorname{sgn}(XY)$, where

$$\operatorname{sgn}(t) = \begin{cases} -1 & \text{if } t < 0, \\ 0 & \text{if } t = 0, \\ 1 & \text{if } t > 0. \end{cases}$$

Then, Z also has a standard normal distribution. For a proof, see Appendix B. We assumed that $X \perp\!\!\!\perp Y$, and it can also be shown that $X \perp\!\!\!\perp Z$ and $Y \perp\!\!\!\perp Z$, so X, Y, Z are pairwise independent. However, they are not mutually independent, since

$$\mathrm{P}(X < 0, Y < 0, Z < 0) = 0 \neq \frac{1}{8} = \mathrm{P}(X < 0)\,\mathrm{P}(Y < 0)\,\mathrm{P}(Z < 0). \qquad \blacksquare$$

When the type of independence is not explicitly stated, it (usually) refers to mutual independence. This logic also applies to the notion of "i.i.d.", which is therefore an abbreviation of "mutually independent and identically distributed". To eliminate ambiguity, we will always specify whether independence is mutual or pairwise. Furthermore, we will introduce a third type: *serial* independence.

Definition 3.4 — The sequence $X_1, \dots, X_n$ of random variables is serially independent if $X_i \perp\!\!\!\perp X_{i+1}$ for all $i = 1, \dots, n-1$.

This is essentially a weaker version of pairwise independence (and therefore of mutual independence as well). It will be useful when we investigate the independence of random variables which have some inherent order.

A last important result is that "transforming" random variables retains independence. This is formalized by the following theorem.

Theorem 3.5 — *Let $X_1, \dots, X_n$ be mutually independent continuous random variables, and let $g_1, \dots, g_n$ be regular*[1] *functions. Then $g_1(X_1), \dots, g_n(X_n)$ are mutually independent random variables.*

Proof. See Theorem 5.6.13 of Meester (2008). □

3.2 Joint normal distribution

In this thesis, we almost always work with more than one (normally distributed) random variable at once. Therefore, we want to generalize the normal distribution to higher dimensions. Since this distribution and its properties play a critical role in Chapter 5, we will state some important results here. Our presentation follows the exposition of Bingham and Fry (2010).

If a random vector $\mathbf{X} = (X_1, \dots, X_n)^\top$ has a *joint normal* (or *multivariate normal*) distribution, we say that $\mathbf{X}$ is a *normal random vector* and we write

$$\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma).$$

Here, $\mathrm{E}[\mathbf{X}] = \boldsymbol{\mu} \in \mathbb{R}^n$ is the *mean vector*, and $\mathrm{Cov}(\mathbf{X}) = \Sigma \in \mathbb{R}^{n \times n}$ the *covariance matrix*. Note that $\Sigma_{ij} = \mathrm{Cov}(X_i, X_j)$, so Σ is symmetric.

Not every collection of normal random variables is jointly normal, but the following theorems provide sufficient conditions for this. For proofs, see Definition 4.8 and Proposition 4.9 of Bingham and Fry (2010).

Theorem 3.6 — *Let $\mathbf{X} = (X_1, \dots, X_n)^\top$ be a random vector. If $\mathbf{a}^\top \mathbf{X}$ has a (univariate) normal distribution for all $\mathbf{a} \in \mathbb{R}^n$, then $\mathbf{X}$ has a joint normal distribution.*

Theorem 3.7 — *Let $\mathbf{X} = (X_1, \dots, X_n)^\top$ be a normal random vector. Then $A\mathbf{X} + \mathbf{b}$ has a joint normal distribution for all $A \in \mathbb{R}^{m \times n}$ and $\mathbf{b} \in \mathbb{R}^m$.*

Equivalently, Theorem 3.6 states that a collection of random variables is jointly normal if any linear combination of them is normally distributed, and Theorem 3.7 states that any linear transformation of a normal random vector is again jointly normal. An important corollary of the first theorem is that mutually independent normal random variables are always jointly normal.

For our purposes, the following theorem is the most important property of the joint normal distribution.

Theorem 3.8 — *Let $(X_1, \dots, X_n)^\top \sim \mathcal{N}(\boldsymbol{\mu}, \Sigma)$. Then $X_1, \dots, X_n$ are mutually independent if and only if Σ is diagonal.*

[1] See Definition 5.6.11 of Meester (2008). All functions considered in this thesis are regular.

Proof. See Corollary 4.11 of Bingham and Fry (2010). □

In other words, in case of joint normality, pairwise uncorrelatedness implies mutual independence. This is a powerful tool, especially if combined with Theorem 3.5. Since all results in this section are well-known, these theorems will be applied throughout the thesis without reference.

3.3 Linear regression

The goal of this section is to introduce recursive residuals, which are closely related to the self-starting statistic. These are formally defined in Section 3.3.4. First, however, we will introduce the necessary preliminaries, namely (multiple) linear regression and a number of distributional properties. This is largely based on Chapter 3 of Bingham and Fry (2010), although with slightly different notation.

3.3.1 Model

In short, linear regression is used to model the relationship between a *response* variable and one or more *explanatory* variables. We will describe this relation using the following model equation:

$$\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \tag{3.9}$$

where

- $\mathbf{y} \in \mathbb{R}^n$ is the vector of the observed values of the response variable,
- $X \in \mathbb{R}^{n\times p}$ is the *design matrix*, whose rows $\mathbf{x}_1, \dots, \mathbf{x}_n$ contain the observed values of the explanatory variables,
- $\boldsymbol{\beta} \in \mathbb{R}^p$ is the vector of *coefficients*, and
- $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ is the vector of *errors*[2], where we assume $\varepsilon_1, \dots, \varepsilon_n \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$.

Note that n is equal to the number of observations (where a row of X is regarded as one unit), and p the number of coefficients. Often, the first coefficient is chosen as the *intercept*[3], and therefore the first column of X consists of all ones.

[2]There is a subtle but important difference between errors and residuals. The error ε_i is the difference between the observed value y_i and the true value $\mathbf{x}_i\boldsymbol{\beta}$, while the residual $\hat{\varepsilon}_i$ is the difference between the observed value y_i and the fitted value $\mathbf{x}_i\hat{\boldsymbol{\beta}}$. Note that the true value is unobservable, so the error is as well. The residual, which is observable, is an estimate of the error.

[3]For a simple linear regression model, i.e., $y_i = \alpha + \beta x_i + \varepsilon_i$, the intercept is α.

3.3.2 Ordinary least squares

We wish to find an estimate of the coefficients $\boldsymbol{\beta}$, denoted by $\hat{\boldsymbol{\beta}}$, such that the *residuals*, defined as

$$\hat{\varepsilon}_i = y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}, \quad i = 1, \dots, n$$

are as "small" as possible. In particular, we minimize the *sum of squared residuals*[4], i.e.,

$$SSR = \hat{\boldsymbol{\varepsilon}}^\top\hat{\boldsymbol{\varepsilon}} = \sum_{i=1}^{n}\left(y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}}\right)^2 = \left(\mathbf{y} - X\hat{\boldsymbol{\beta}}\right)^\top\left(\mathbf{y} - X\hat{\boldsymbol{\beta}}\right).$$

According to the Gauss-Markov theorem,

$$\hat{\boldsymbol{\beta}} = (X^\top X)^{-1}X^\top\mathbf{y} \tag{3.10}$$

is the "best" estimator of $\boldsymbol{\beta}$, i.e., it has the lowest variance among all linear unbiased estimators. This estimator is also known as the *ordinary least squares* (OLS) estimator. Formula 3.10 requires that $X^\top X$ is invertible. Therefore, we assume that X has full rank p. For a proof of the Gauss-Markov theorem, see Theorems 3.5 and 3.13 of Bingham and Fry (2010).

An important property is that

$$\hat{\boldsymbol{\beta}} \sim \mathcal{N}\left(\boldsymbol{\beta}, \sigma^2(X^\top X)^{-1}\right). \tag{3.11}$$

More details are given in Section 3.3 of Bingham and Fry (2010). Using $\hat{\boldsymbol{\beta}}$, we can also estimate σ^2. In particular, an unbiased estimator is given by

$$\hat{\sigma}^2 = \frac{1}{n-p}SSR, \tag{3.12}$$

where

$$\frac{1}{\sigma^2}SSR \sim \chi^2(n-p). \tag{3.13}$$

This follows from Corollary 3.23 and Theorem 3.26 of Bingham and Fry (2010). Lastly, as shown in Theorem 3.31 of Bingham and Fry (2010), it holds that

$$\hat{\boldsymbol{\beta}} \perp\!\!\!\perp SSR.$$

3.3.3 Notation for partial data

In the next sections, we often perform linear regression on a subset of the data, i.e., the first k observations. For this we need to introduce some notation. We write X_k to

[4]Not to be confused with the *sum of squares for regression*, as defined by Bingham and Fry (2010), which has the same abbreviation. Furthermore, adding to the confusion, the sum of squared residuals is also known as the *sum of squares for (estimate of) error*.

denote the matrix consisting of only the first $k \leq n$ rows of X,[5] i.e.,

$$X_k = \begin{pmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_k \end{pmatrix} \in \mathbb{R}^{k \times p}.$$

Similarly, $\mathbf{y}_{[1:k]}$ denotes the vector of the first k elements of $\mathbf{y}$.[6] Furthermore, we define $\hat{\boldsymbol{\beta}}^{(k)}$ and SSR_k to be the estimate of $\boldsymbol{\beta}$ and SSR based on the first $k \geq p$ observations, respectively. That is,

$$\hat{\boldsymbol{\beta}}^{(k)} = \left(X_k^\mathsf{T} X_k\right)^{-1} X_k^\mathsf{T} \mathbf{y}_{[1:k]}$$

and

$$SSR_k = \left(\mathbf{y}_{[1:k]} - X_k \hat{\boldsymbol{\beta}}^{(k)}\right)^\mathsf{T} \left(\mathbf{y}_{[1:k]} - X_k \hat{\boldsymbol{\beta}}^{(k)}\right). \tag{3.14}$$

Note that for $\hat{\boldsymbol{\beta}}^{(k)}$ (and SSR_k) to exist, we assume that X_k has full rank. It is sufficient to assume that X_p has full rank, since the rank of submatrices is less than or equal to the rank of the full matrix. Hence, $p = \operatorname{rank}(X_p) \leq \operatorname{rank}(X_k)$. Furthermore, due to its dimensions, $\operatorname{rank}(X_k) \leq \min(k, p) = p$. Therefore, $\operatorname{rank}(X_k) = p$ as well.

3.3.4 Recursive residuals

To assess the validity of a regression model, practitioners often perform *residual analysis*, where assumptions such as homoscedasticity are verified. However, (ordinary) residuals are correlated, which means that departures in one part of the model can "spread" to all other residuals. This makes them ineffective. A number of specific issues are discussed by Kianifard and Swallow (1996).

An alternative is to use the *recursive residuals* introduced by Brown et al. (1975), which are a special type of residual that are mutually independent. They are the (scaled) difference between the current observed value and the predicted value based only on the previous values. In particular, values are not used to predict themselves. Essentially, these are one-step-ahead predictive residuals. Other types of independent residuals exist,[7] but recursive residuals have the advantage that they correspond one-to-one to the observations from which they were computed. This is an important property for many statistical (outlier) tests, such as those to detect influential observations. This also makes them suitable for control charts, which is discussed in Section 4.4.

The values of the recursive residuals depend on the order in which they are calculated. In the case of control charts, there is often a natural ordering we can exploit, e.g., time. Since recursive residuals only depend on past measurements, they can be used in a

[5] A notation such as $X_{[1:k],\bullet}$ would be more explicit (and flexible), but also more verbose, which harms readability. Therefore, this notation was chosen. It should not be confused with $\mathbf{x}_k$, which denotes the k-th row of X.

[6] In line with the notation for matrices, $\mathbf{y}_k$ would be more consistent. However, this is too similar to y_k, which denotes the k-th element of $\mathbf{y}$.

[7] For example, the BLUS residuals discussed by Theil (1971).

dynamic environment, where new measurements are continuously received. This is in contrast to ordinary residuals, which can only be used in a static environment, where all measurements are available from the start. It is important to note that, unlike ordinary residuals, there exist no recursive residuals for the first p observations. This is because the initial estimation of the coefficients requires at least p observations.

We will now define them formally. Using the notation from the previous section, the recursive residuals are defined as

$$w_k = \frac{y_k - \mathbf{x}_k \hat{\boldsymbol{\beta}}^{(k-1)}}{\sqrt{1 + \mathbf{x}_k \left(X_{k-1}^\top X_{k-1}\right)^{-1} \mathbf{x}_k^\top}}, \quad k = p+1, \dots, n. \tag{3.15}$$

As will be shown below, they have a normal distribution with mean 0 and variance σ^2. To eliminate the dependency on σ^2 (which is assumed to be unknown), we studentize it using (3.12). That is, we define the *studentized recursive residuals* as

$$w'_k = \frac{w_k}{\sqrt{\frac{1}{k-1-p} SSR_{k-1}}}, \quad k = p+2, \dots, n. \tag{3.16}$$

It is well-known that the (ordinary) recursive residuals are independent (i.e., when σ^2 is known). For example, see Brown et al. (1975). Furthermore, the studentized recursive residuals are independent as well. Proofs of both of these facts are presented in Section 5.2.

In the remainder of this section, we will derive the distributions of these statistics. First, note that y_k and $\mathbf{x}_k \hat{\boldsymbol{\beta}}^{(k-1)}$ are normally distributed and independent, since $\hat{\boldsymbol{\beta}}^{(k-1)}$ is a function of $(y_1, \dots, y_{k-1})$, which is independent of y_k. Therefore, the numerator $y_k - \mathbf{x}_k \hat{\boldsymbol{\beta}}^{(k-1)}$ is normally distributed with mean

$$\mathrm{E}\left[y_k - \mathbf{x}_k \hat{\boldsymbol{\beta}}^{(k-1)}\right] = \mathrm{E}[y_k] - \mathbf{x}_k \, \mathrm{E}\left[\hat{\boldsymbol{\beta}}^{(k-1)}\right] = \mathbf{x}_k \boldsymbol{\beta} + 0 - \mathbf{x}_k \boldsymbol{\beta} = 0,$$

and variance

$$\begin{aligned}
\mathrm{Var}\left(y_k - \mathbf{x}_k \hat{\boldsymbol{\beta}}^{(k-1)}\right) &= \mathrm{Var}(y_k) + \mathrm{Var}\left(\mathbf{x}_k \hat{\boldsymbol{\beta}}^{(k-1)}\right) && (\text{since } y_k \perp\!\!\!\perp \hat{\boldsymbol{\beta}}^{(k-1)}) \\
&= \sigma^2 + \mathbf{x}_k \sigma^2 \left(X_{k-1}^\top X_{k-1}\right)^{-1} \mathbf{x}_k^\top && (\text{by (3.11)}) \\
&= \sigma^2 \left(1 + \mathbf{x}_k \left(X_{k-1}^\top X_{k-1}\right)^{-1} \mathbf{x}_k^\top\right).
\end{aligned}$$

Therefore, if we divide the numerator by the constant $\left(1 + \mathbf{x}_k \left(X_{k-1}^\top X_{k-1}\right)^{-1} \mathbf{x}_k^\top\right)^{1/2}$, we find that

$$w_k \sim \mathcal{N}(0, \sigma^2).$$

Next, we consider the distribution of w'_k. We know that $\hat{\boldsymbol{\beta}}^{(k-1)} \perp\!\!\!\perp SSR_{k-1}$ from the previous section. Furthermore, note that both are functions of $(y_1, \dots, y_{k-1})$, so $y_k \perp\!\!\!\perp$

$\left(\hat{\boldsymbol{\beta}}^{(k-1)}, SSR_{k-1}\right)$. Together, this implies that $y_k - \mathbf{x}_k\hat{\boldsymbol{\beta}}^{(k-1)} \perp\!\!\!\perp SSR_{k-1}$.[8] Therefore, by Definition B.5,

$$w'_k = \frac{\frac{1}{\sigma}w_k}{\sqrt{\frac{1}{k-1-p}\frac{1}{\sigma^2}SSR_{k-1}}} \sim t(k-1-p),$$

since $\frac{1}{\sigma}w_k \sim \mathcal{N}(0,1)$, and $\frac{1}{\sigma^2}SSR_{k-1} \sim \chi^2(k-1-p)$ by (3.13).

3.4 Updating formulae

It is often useful to perform numerical computations in a single pass, i.e., calculate a quantity such as the variance by querying each value X_i only once. This is because computer memory may be too small or too slow. To achieve this, we essentially need a recurrence relation between, for example, the variance of the first n and of the first $n+1$ values. In other words, this recurrence relation allows us to sequentially "update" a quantity if new data arrives. Therefore, these are also known as *updating formulae*. Normally, they are used for efficient numerical algorithms, but they also play a key role in the proofs of independence presented in Chapter 5. Therefore, in this section, we will derive updating formulae for both the sample variance (Section 3.4.1) and the sum of squared residuals (Section 3.4.2).

3.4.1 Sample variance

Let $X_1, X_2, \ldots$ denote observations. Then, define

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad n \geq 1, \tag{3.17}$$

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_n\right)^2, \qquad n \geq 2, \tag{3.18}$$

as the sample mean and (unbiased) sample variance of the first n observations. We will make use of the following updating formulae:

$$\overline{X}_n = \frac{n-1}{n}\overline{X}_{n-1} + \frac{1}{n}X_n, \qquad n \geq 2, \tag{3.19}$$

$$S_n^2 = \frac{n-2}{n-1}S_{n-1}^2 + \frac{1}{n}\left(X_n - \overline{X}_{n-1}\right)^2, \qquad n \geq 3. \tag{3.20}$$

The first follows trivially from the definition; the second is derived from a recurrence relation for the sum of squares published in a note by Welford (1962). For completeness, we will give the proof here as well.

[8] This is similar to the derivation of the distribution of case IV in Section 4.3.1, where more details are provided.

Proof of (3.20). First, note that

$$(n-1)S_n^2 = \sum_{i=1}^{n-1} \underbrace{\left(X_i - \overline{X}_n\right)^2}_{:=A} + \underbrace{\left(X_n - \overline{X}_n\right)^2}_{:=B}.$$

Using (3.19), we find that

$$\begin{aligned} A &= \left(X_i - \tfrac{n-1}{n}\overline{X}_{n-1} - \tfrac{1}{n}X_n\right)^2 \\ &= \left(\left[X_i - \overline{X}_{n-1}\right] - \tfrac{1}{n}\left[X_n - \overline{X}_{n-1}\right]\right)^2 \\ &= \left(X_i - \overline{X}_{n-1}\right)^2 - \tfrac{2}{n}\left(X_i - \overline{X}_{n-1}\right)\left(X_n - \overline{X}_{n-1}\right) + \tfrac{1}{n^2}\left(X_n - \overline{X}_{n-1}\right)^2 \end{aligned}$$

and

$$\begin{aligned} B &= \left(X_n - \tfrac{n-1}{n}\overline{X}_{n-1} - \tfrac{1}{n}X_n\right)^2 \\ &= \left(\tfrac{n-1}{n}\right)^2\left(X_n - \overline{X}_{n-1}\right)^2. \end{aligned}$$

Therefore,

$$\begin{aligned} (n-1)S_n^2 &= \sum_{i=1}^{n-1}\left[\left(X_i - \overline{X}_{n-1}\right)^2 - \tfrac{2}{n}\left(X_i - \overline{X}_{n-1}\right)\left(X_n - \overline{X}_{n-1}\right) + \tfrac{1}{n^2}\left(X_n - \overline{X}_{n-1}\right)^2\right] \\ &\qquad + \left(\tfrac{n-1}{n}\right)^2\left(X_n - \overline{X}_{n-1}\right)^2 \\ &= \sum_{i=1}^{n-1}\left(X_i - \overline{X}_{n-1}\right)^2 - \tfrac{2}{n}\left(X_n - \overline{X}_{n-1}\right)\underbrace{\sum_{i=1}^{n-1}\left(X_i - \overline{X}_{n-1}\right)}_{=0} + \tfrac{n-1}{n^2}\left(X_n - \overline{X}_{n-1}\right)^2 \\ &\qquad + \left(\tfrac{n-1}{n}\right)^2\left(X_n - \overline{X}_{n-1}\right)^2 \\ &= \sum_{i=1}^{n-1}\left(X_i - \overline{X}_{n-1}\right)^2 + \tfrac{n-1+(n-1)^2}{n^2}\left(X_n - \overline{X}_{n-1}\right)^2 \\ &= (n-2)S_{n-1}^2 + \tfrac{n-1}{n}\left(X_n - \overline{X}_{n-1}\right)^2. \end{aligned}$$

Finally, dividing both sides by $n-1$ yields the formula. □

Repeatedly applying (3.20) yields the following alternative formula for the variance.

Proposition 3.21 — *Let $X_1, X_2, \ldots$ denote observations. Then the (unbiased) sample variance of the first $n \geq 2$ observations is given by*

$$S_n^2 = \tfrac{1}{n-1}\sum_{i=2}^{n}\tfrac{i-1}{i}\left(X_i - \overline{X}_{i-1}\right)^2.$$

Proof. We will prove the claim by induction. For the base case $n = 2$, the formula simplifies to $\frac{1}{2}(X_2 - X_1)^2$, which is indeed equal to Definition 3.18, since

$$\begin{aligned} S_2^2 &= \left(X_1 - \tfrac{1}{2}(X_1 + X_2)\right)^2 + \left(X_2 - \tfrac{1}{2}(X_1 + X_2)\right)^2 \\ &= \tfrac{1}{4}(X_1 - X_2)^2 + \tfrac{1}{4}(X_2 - X_1)^2 \\ &= \tfrac{1}{2}(X_2 - X_1)^2. \end{aligned}$$

Now, we assume that the formula holds for $n = k$ (induction hypothesis) and show that it also holds for $n = k + 1$. We verify that

$$\begin{aligned} S_{k+1}^2 &= \frac{k-1}{k} S_k^2 + \frac{1}{k+1}\left(X_{k+1} - \overline{X}_k\right)^2 && \text{(by (3.20))} \\ &= \frac{1}{k}\sum_{i=2}^{k} \frac{i-1}{i}\left(X_i - \overline{X}_{i-1}\right)^2 + \frac{1}{k}\frac{k}{k+1}\left(X_{k+1} - \overline{X}_k\right)^2 && \text{(by IH)} \\ &= \frac{1}{k}\sum_{i=2}^{k+1} \frac{i-1}{i}\left(X_i - \overline{X}_{i-1}\right)^2. \end{aligned}$$

□

Incidentally, this formula provides another way to verify the well-known fact that

$$(n-1)\frac{S_n^2}{\sigma^2} \sim \chi^2(n-1). \tag{3.22}$$

This is because

$$(n-1)\frac{S_n^2}{\sigma^2} = \sum_{i=2}^{n}\left(\frac{1}{\sigma}\sqrt{\frac{i-1}{i}}\left(X_i - \overline{X}_{i-1}\right)\right)^2$$

is a sum of $n-1$ squared independent standard normal random variables, by Lemma 4.1 and Theorem 5.2.

3.4.2 Sum of squared residuals

Similarly to the variance, we will now give an updating formula for the sum of squared residuals. Consider the linear regression model as defined in Section 3.3, but with an arbitrary number of observations. Then

$$SSR_n = SSR_{n-1} + w_n^2, \qquad n \geq p+1, \tag{3.23}$$

where w_n is defined by (3.15).

This formula allows us to efficiently calculate *SSR* when a new observation is made, without refitting the whole model.[9] A proof is given by Brown et al. (1975), referencing a formula from Bartlett (1951). However, the notation is different than our own, and many details are omitted, so we provide our own proof below. First, we will derive updating formulae for $\hat{\boldsymbol{\beta}}^{(n)}$ and $(X_n^\mathsf{T} X_n)^{-1}$, given by (3.25) and (3.26), respectively. Then, these are used to prove (3.23). Similar formulae are proven by Plackett (1950) that make it possible to update these quantities in case of more than one new observation.

Proof of (3.23). It follows from (3.10) that

$$X_k^\mathsf{T} X_k \hat{\boldsymbol{\beta}}^{(k)} = X_k^\mathsf{T} \mathbf{y}_{[1:k]}, \qquad k \geq p. \tag{3.24}$$

[9] That is, if you also update $\hat{\boldsymbol{\beta}}^{(n)}$ using the formula below.

Hence,

$$\begin{aligned} X_n^\top X_n \hat{\boldsymbol{\beta}}^{(n)} &= X_{n-1}^\top \mathbf{y}_{[1:n-1]} + \mathbf{x}_n^\top y_n \\ &= X_{n-1}^\top X_{n-1} \hat{\boldsymbol{\beta}}^{(n-1)} + \mathbf{x}_n^\top y_n \\ &= \big(X_n^\top X_n - \mathbf{x}_n^\top \mathbf{x}_n\big)\hat{\boldsymbol{\beta}}^{(n-1)} + \mathbf{x}_n^\top y_n \qquad \text{(by Lemma B.8)} \\ &= X_n^\top X_n \hat{\boldsymbol{\beta}}^{(n-1)} + \mathbf{x}_n^\top \big(y_n - \mathbf{x}_n \hat{\boldsymbol{\beta}}^{(n-1)}\big), \end{aligned}$$

After multiplying both sides by $\big(X_n^\top X_n\big)^{-1}$ on the left and rearranging the terms, we obtain

$$\hat{\boldsymbol{\beta}}^{(n)} - \hat{\boldsymbol{\beta}}^{(n-1)} = \big(X_n^\top X_n\big)^{-1} \mathbf{x}_n^\top \big(y_n - \mathbf{x}_n \hat{\boldsymbol{\beta}}^{(n-1)}\big). \tag{3.25}$$

Furthermore, by Lemmas B.8 and B.9, we have that

$$\begin{aligned} \big(X_n^\top X_n\big)^{-1} &= \big(X_{n-1}^\top X_{n-1} + \mathbf{x}_n^\top \mathbf{x}_n\big)^{-1} \\ &= \big(X_{n-1}^\top X_{n-1}\big)^{-1} - \frac{\big(X_{n-1}^\top X_{n-1}\big)^{-1} \mathbf{x}_n^\top \mathbf{x}_n \big(X_{n-1}^\top X_{n-1}\big)^{-1}}{1 + \mathbf{x}_n \big(X_{n-1}^\top X_{n-1}\big)^{-1} \mathbf{x}_n^\top}. \end{aligned} \tag{3.26}$$

Next, define

$$\begin{aligned} A &= \mathbf{y}_{[1:n]} - X_n \hat{\boldsymbol{\beta}}^{(n-1)}, \\ B &= X_n \big(\hat{\boldsymbol{\beta}}^{(n)} - \hat{\boldsymbol{\beta}}^{(n-1)}\big) \end{aligned}$$

such that

$$SSR_n = \big(\mathbf{y}_{[1:n]} - X_n \hat{\boldsymbol{\beta}}^{(n)}\big)^\top \big(\mathbf{y}_{[1:n]} - X_n \hat{\boldsymbol{\beta}}^{(n)}\big) = (A - B)^\top (A - B).$$

Note that

$$\begin{aligned} B^\top A &= \big(\hat{\boldsymbol{\beta}}^{(n)} - \hat{\boldsymbol{\beta}}^{(n-1)}\big)^\top X_n^\top \big(\mathbf{y}_{[1:n]} - X_n \hat{\boldsymbol{\beta}}^{(n-1)}\big) \\ &= \big(\hat{\boldsymbol{\beta}}^{(n)} - \hat{\boldsymbol{\beta}}^{(n-1)}\big)^\top \big(X_n^\top \mathbf{y}_{[1:n]} - X_n^\top X_n \hat{\boldsymbol{\beta}}^{(n-1)}\big) \\ &= \big(\hat{\boldsymbol{\beta}}^{(n)} - \hat{\boldsymbol{\beta}}^{(n-1)}\big)^\top \big(X_n^\top X_n \hat{\boldsymbol{\beta}}^{(n)} - X_n^\top X_n \hat{\boldsymbol{\beta}}^{(n-1)}\big) \qquad \text{(by (3.24))} \\ &= \big(\hat{\boldsymbol{\beta}}^{(n)} - \hat{\boldsymbol{\beta}}^{(n-1)}\big)^\top X_n^\top X_n \big(\hat{\boldsymbol{\beta}}^{(n)} - \hat{\boldsymbol{\beta}}^{(n-1)}\big) \\ &= B^\top B. \end{aligned}$$

Furthermore, $B^\top A = (B^\top A)^\top = A^\top B$, since it is symmetric. Hence,

$$SSR_n = A^\top A - A^\top B - B^\top A + B^\top B = A^\top A - B^\top B,$$

where

$$\begin{aligned} A^\top A &= \sum_{i=1}^{n} \big(y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}}^{(n-1)}\big)^2 \\ &= \sum_{i=1}^{n-1} \big(y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}}^{(n-1)}\big)^2 + \big(y_n - \mathbf{x}_n \hat{\boldsymbol{\beta}}^{(n-1)}\big)^2 \\ &= SSR_{n-1} + \big(y_n - \mathbf{x}_n \hat{\boldsymbol{\beta}}^{(n-1)}\big)^2, \end{aligned}$$

and where

$$\begin{aligned} B^\top B &= \left(y_n - \mathbf{x}_n\hat{\boldsymbol{\beta}}^{(n-1)}\right)\mathbf{x}_n\left(X_n^\top X_n\right)^{-1}X_n^\top X_n\left(X_n^\top X_n\right)^{-1}\mathbf{x}_n^\top\left(y_n - \mathbf{x}_n\hat{\boldsymbol{\beta}}^{(n-1)}\right) \\ &= \left(y_n - \mathbf{x}_n\hat{\boldsymbol{\beta}}^{(n-1)}\right)^2\mathbf{x}_n\left(X_n^\top X_n\right)^{-1}\mathbf{x}_n^\top \end{aligned}$$

by (3.25). Hence,

$$SSR_n = SSR_{n-1} + \left(y_n - \mathbf{x}_n\hat{\boldsymbol{\beta}}^{(n-1)}\right)^2\left(1 - \mathbf{x}_n\left(X_n^\top X_n\right)^{-1}\mathbf{x}_n^\top\right).$$

Define $C = \mathbf{x}_n\left(X_{n-1}^\top X_{n-1}\right)^{-1}\mathbf{x}_n^\top$. Then, by (3.26),

$$\begin{aligned} &1 - \mathbf{x}_n\left(X_n^\top X_n\right)^{-1}\mathbf{x}_n^\top \\ &\quad = 1 - \mathbf{x}_n\left(\left(X_{n-1}^\top X_{n-1}\right)^{-1} - \frac{\left(X_{n-1}^\top X_{n-1}\right)^{-1}\mathbf{x}_n^\top\mathbf{x}_n\left(X_{n-1}^\top X_{n-1}\right)^{-1}}{1+C}\right)\mathbf{x}_n^\top \\ &\quad = 1 - C + \frac{C^2}{1+C} \\ &\quad = \frac{1}{1+C}. \end{aligned}$$

Therefore,

$$SSR_n = SSR_{n-1} + \frac{\left(y_n - \mathbf{x}_n\hat{\boldsymbol{\beta}}^{(n-1)}\right)^2}{1+C} = SSR_{n-1} + w_n^2. \qquad \square$$

The following result is similar to Proposition 3.21. An alternative proof of this formula is given by Kianifard and Swallow (1996), who state that "the sum of squares of the $n - p$ LUS residuals[10] is equal to the sum of squares of the n OLS residuals."

Proposition 3.27 — *Consider the linear regression model given by* $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ *as defined in Section 3.3, with the assumptions that* X_p *has full rank and that* $\varepsilon_1, \ldots, \varepsilon_n \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$. *Then, the sum of squared residuals based on the first* $n \geq p$ *observations is given by*

$$SSR_n = \sum_{i=p+1}^{n} w_i^2.$$

Proof. We will prove the claim by induction. For the base case $n = p$, the formula simplifies to 0, which is indeed equal to Definition 3.14, since

$$\begin{aligned} \mathbf{y}_{[1:p]} - X_p\hat{\boldsymbol{\beta}}^{(p)} &= \mathbf{y}_{[1:p]} - X_p\left(X_p^\top X_p\right)^{-1}X_p^\top\mathbf{y}_{[1:p]} \\ &= \mathbf{y}_{[1:p]} - X_pX_p^{-1}(X_p^\top)^{-1}X_p^\top\mathbf{y}_{[1:p]} \\ &= \mathbf{y}_{[1:p]} - \mathbf{y}_{[1:p]} \\ &= \mathbf{0}. \end{aligned}$$

[10] A class of linear unbiased estimators with a scalar covariance matrix, which includes the recursive residuals.

This holds because X_p is square and has full rank, and is therefore invertible. Now, we assume that the formula holds for $n = k$ (induction hypothesis) and show that it also holds for $n = k + 1$. We verify that

$$\begin{aligned} SSR_{k+1} &= SSR_k + w_{k+1}^2 && \text{(by (3.23))} \\ &= \sum_{i=p+1}^{k} w_i^2 + w_{k+1}^2 && \text{(by IH)} \\ &= \sum_{i=p+1}^{k+1} w_i^2. \end{aligned}$$

□

Chapter summary

We established the concept of independence between random variables and presented several results (i.e., Theorems 3.2, 3.5, and 3.8) that will be helpful in proving independence in later chapters. Furthermore, using Counterexample 3.3, we showed the difference between mutual and pairwise independence, the former being a stronger property than the latter.

Next, we defined the recursive residuals and derived their distribution using results from linear regression. We found that the recursive residuals are i.i.d. $\mathcal{N}(0, \sigma^2)$ random variables, and that they can be studentized to eliminate their dependency on σ^2. The k-th studentized recursive residual has a t-distribution with $k - 1 - p$ degrees of freedom.

Lastly, we showed that the sample variance and sum of squared residuals can be efficiently updated. Propositions 3.21 and 3.27 essentially provide alternative formulas of the variance and sum of squared residuals, respectively. We will take advantage of their structure to prove independence in Chapter 5.

Chapter 4

Control charts

In this chapter, we will introduce control charts in more detail. First, we will introduce some terminology and discuss the goal of control charts (Section 4.1). Next, we will outline how control charts are used and which assumptions they rely on (Section 4.2). We will also shortly explore how control charts are linked to hypothesis testing and how this allows us to analyze their performance. Then, we will give some practical limitations of control charts and introduce self-starting charts as a solution to them (Section 4.3). The formulas defined here are the focus of the remainder of this thesis. Lastly, we will present regression charts, a type of control chart that monitors the residuals of a linear regression model (Section 4.4).

4.1 Context

In short, a control chart is a graphical tool that is used to monitor the variability of a process and detect changes. When he first introduced them, Shewhart (1931) differentiated between two *causes* of variability:

- *chance* causes, referring to essentially unavoidable sources of variability inherent to the process (i.e., "background noise");
- *assignable* causes, referring to controllable sources of variability due to malfunctioning parts of the process, such as operator errors or misconfigured machines.

These are also known as *common* and *special* causes, respectively. If a process is subject to assignable causes, it is said to be *out-of-control* (or *unstable*). On the other hand, if only chance causes are present, it is *in-control* (or *stable*). When using control charts (and SPM in general), the ultimate goal is to reduce variability due to assignable causes and ensure that the process is in-control.

Typically, the usage of control charts consists of two distinct phases.

- In Phase I (*retrospective* monitoring), an initial set of process data is collected and analyzed. If unusual "patterns" are found in the data, the process is fine-tuned to eliminate them. This is repeated until a "clean" set of data (i.e., under in-control conditions) is collected, from which the parameters for future monitoring are computed.
- In Phase II (*online* monitoring), the parameters determined in Phase I are used to sequentially monitor new samples during regular production. If a significant deviation is detected, the process is typically stopped and the (assignable) cause is investigated.

This approach works well in high-volume manufacturing, but not for short-run processes where there is too little data available to compute accurate control limits. We are therefore interested in *self-starting charts*, which do not require the use of Phase I. These are discussed in Section 4.3. Since self-starting charts are a variant of the classic Shewhart chart, these will be discussed first, in Section 4.2.

4.2 Shewhart charts

In a control chart, a *quality characteristic* of a process is plotted against the time (or simply the sample number), together with a center line (C), lower control limit (L), and upper control limit (U). A typical chart is shown in Figure 4.1. Commonly, the quality characteristic is a summary statistic computed from batches of products which are grouped in a natural way. These batches are known as *rational subgroups*. The most well-known chart of this type is the *Shewhart $\overline{X}$-chart*, which plots the mean of the subgroups. However, we will focus on charts for rational subgroups of size one. These are known as *individuals charts*.

A frequent choice is to set the control limits at $\mu \pm 3\sigma$, where μ and σ are the mean and standard deviation of the charting statistic, respectively, which are assumed known after Phase I. These *three-sigma limits* were set by Shewhart based on practical experience, and are intended as a heuristic. In particular, when a point is plotted outside the limits, this is taken as evidence that the process is out-of-control. However, the inverse is not necessarily true; even if all points are inside the limits, the process may not be in-control. For example, if many consecutive points are on one side of the center line, this might indicate correlated behavior, and therefore an assignable cause.[1]

It is typically assumed that the measurements are i.i.d. In this thesis, we will make the common assumption that they are normally distributed, which is often justified by the central limit theorem. Although control charts still work well if the normality assumption is broken to a moderate degree (Woodall, 2000; Montgomery, 2019), the

[1] For practical use, many rules exist to systematically detect such patterns. These are called *runs rules*. Perhaps most famous are those by Western Electric (1958).

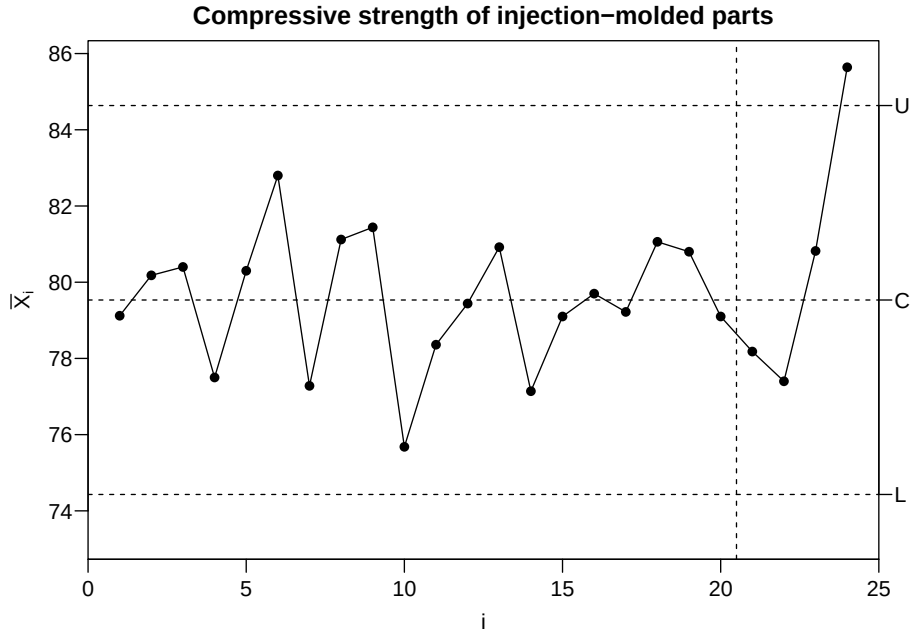

Figure 4.1: An example of an $\overline{X}$-chart based on data from Qiu (2013). It monitors an injection molding process, where the quality characteristic is the compressive strength. Each point is the mean of a sample of 5 measurements. The center line and control limits are determined from the first 20 samples (Phase I) and are visualized as dashed horizontal lines. The boundary between Phase I and II is indicated by a dashed vertical line. Note that it is common practice to connect the points with lines.

independence assumption is much more critical. In particular, Alwan (1992) showed that even mild levels of autocorrelation will significantly increase the number of false alarms.[2] This, in turn, makes it much more difficult to identify assignable causes among chance causes.

Control charts, especially as used in Phase II, are closely linked to hypothesis testing. To illustrate this, we will consider the situation in which an assignable cause induces an abrupt *shift* in the mean of the charting statistic.[3] For this example, we will not use additional decision rules, and simply say that the process is in-control if and only if all points are inside the control limits. Let $X \sim \mathcal{N}(\mu, \sigma^2)$ represent a measurement of the quality characteristic, with true mean μ and variance σ^2. Let μ_0 indicate the in-control mean, determined in Phase I. We will test

$$H_0 : \mu = \mu_0 \qquad \text{vs.} \qquad H_1 : \mu \neq \mu_0.$$

We fail to reject the null hypothesis if the point is plotted inside the control limits, i.e, $X \in [\mu - 3\sigma, \mu + 3\sigma]$, and reject the null hypothesis if it is plotted outside. That is, we

[2] False alarms become even more frequent if runs rules are used. For example, a common rule is to stop the process if a large number of consecutive points are on one side of the center line, which becomes more common if the measurements are autocorrelated.

[3] Other effects are possible as well, such as a (gradual) *drift*. However, they fit less well into a (non-sequential) hypothesis testing framework and will not be considered here.

reject H_0 if

$$\left|\frac{X-\mu}{\sigma}\right| > 3.$$

This test is repeated for each measurement. Under the assumption of normality (which implies the test statistic has a standard normal distribution), the probability of a type I error is equal to

$$\alpha = \Phi(-3) + \big(1 - \Phi(3)\big) \approx 0.0027.$$

Note that a type I error (i.e., a "false alarm") corresponds to a scenario where we conclude the process is out-of-control although it is actually in-control, and that a type II error corresponds to one where we conclude the process is in-control although it is actually out-of-control.

The link to hypothesis testing is especially useful for analyzing the performance of control charts. This is often done by studying the *average run length* (ARL), where the run length is defined as the number of samples collected until the first type I error. Ideally, it is as large as possible. Under the assumptions above (i.e., independence, normality, three-sigma limits, and no runs rules) the run length has a geometric distribution with mean

$$ARL = \frac{1}{\alpha} \approx 370.$$

It is important to note that this simple calculation only holds for basic Shewhart charts under the stated assumptions. For many other types of charts, the computations are much more complex.

Although this link to hypothesis testing is an interesting topic, it is not the focus of this thesis, and therefore we will not go into further detail. However, formal analysis of control charts in the context of hypothesis testing is not without problems, even though it proves to be helpful. This is because it relies on strong assumptions which most often do not hold in practice. Woodall (2000) discusses this topic in more detail.

4.3 Self-starting charts

Control charts as described in the previous section are often used in practice, but Quesenberry (1991) notes the following two issues.

- In low-volume production (e.g., in job shops, where highly-tailored products are manufactured in low numbers) the total number of products may not even reach the number required for accurate parameter estimation during Phase I.[4]
- In general, for all processes, it is beneficial to start monitoring as soon as possible, even if the parameters are not yet known. This allows us to identify assignable causes and bring the process into control at an earlier time.

[4] Jensen et al. (2006) recommend that at a minimum 100 individual measurements are needed in Phase I.

For these reasons, Quesenberry (1991) introduced a type of self-starting chart called Q-charts, which do not require a separate Phase I, but essentially estimate the parameters in real time. This means they can already be used for the very first units of production.

Next, we will formally define the charting statistic. Let $X_1, X_2, \ldots \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$ represent measurements in order of time. We will consider four cases, in which both, either, or neither of the two parameters μ and σ^2 are known at the start of monitoring. In the following definitions, notation from Section 3.4 is used. Furthermore, Φ denotes the CDF of the standard normal distribution, and G_k denotes the CDF of the Student's t-distribution with k degrees of freedom.

(i) If both μ and σ are known, let

$$Q_n = \frac{X_n - \mu}{\sigma}, \quad n \geq 1.$$

(ii) If μ is unknown but σ is known, let

$$Q_n = \frac{1}{\sigma}\sqrt{\frac{n-1}{n}}\, T_n \quad \text{where} \quad T_n = X_n - \overline{X}_{n-1}, \quad n \geq 2.$$

(iii) If μ is known but σ is unknown, let

$$Q_n = \Phi^{-1}\big(G_{n-1}(T_n)\big) \quad \text{where} \quad T_n = \frac{X_n - \mu}{S'_{n-1}}, \quad n \geq 2,$$

where S'_{n-1} is defined in terms of μ, i.e., $(S'_{n-1})^2 = \frac{1}{n-1}\sum_{i=1}^{n-1}(X_i - \mu)^2$.

(iv) If both μ and σ are unknown, let

$$Q_n = \Phi^{-1}\big(G_{n-2}(T_n)\big) \quad \text{where} \quad T_n = \sqrt{\frac{n-1}{n}}\,\frac{X_n - \overline{X}_{n-1}}{S_{n-1}}, \quad n \geq 3.$$

Case I is equivalent to the standard individuals chart. Of the remaining cases, IV is the most interesting, since the assumption for II and III that one of the parameters is known (but not the other) is rather unrealistic.

Note that the mean and variance are estimated using only the previous measurements, i.e., not all available data is used, even though this might yield a better estimate. This is because we assume all previous measurements are in-control, while this is unknown for the current measurement. If we included the current measurement in the estimation, the underlying statistical test would bias toward failure to reject the null hypothesis. In other words, the Q-statistic would be more likely to be inside the control limits.

In all cases, an important result is that $Q_1, Q_2, \ldots \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$. Therefore, we typically set the limits at ± 3. For case I, this fact is trivial.[5] For the other cases, the distribution

[5]The transformation is simply the well-known standardization, and independence follows directly from Theorem 3.5.

will be derived in Section 4.3.1 and independence will be proven in Chapter 5. First, however, we will perform a small-scale simulation study in Section 4.3.2 to confirm that independence is indeed likely the case.

As a side note, a practical benefit of the charting statistics being i.i.d. standard normal is that multiple variables (of the same process) can be plotted on the same chart. On the other hand, the original scale of the data is lost, and with it some context.

4.3.1 Distribution

In this section we will derive the distribution of cases II, III, and IV of the Q-statistic defined in the previous section. First, we will prove the following lemma that will be applied multiple times.

Lemma 4.1 — *Let* $X_1, X_2, \ldots \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$. *Then* $T \coloneqq X_k - \overline{X}_n$ *is normally distributed with* $\mathrm{E}[T] = 0$ *and*

$$\mathrm{Var}(T) = \sigma^2 \cdot \begin{cases} \frac{n-1}{n} & \text{if } k \leq n, \\ \frac{n+1}{n} & \text{if } k > n. \end{cases}$$

Proof. Since T is a linear combination of independent normal random variables, it is normally distributed. Furthermore,

$$\mathrm{E}[T] = \mathrm{E}[X_k] - \frac{1}{n}\sum_{i=1}^{n} \mathrm{E}[X_i] = \mu - \frac{n}{n}\mu = 0.$$

Next, if $k \leq n$,

$$\begin{aligned} \mathrm{Var}(T) &= \mathrm{Var}\left(\frac{n-1}{n}X_k - \frac{1}{n}\sum_{\substack{i=1 \\ i\neq k}}^{n} X_i\right) \\ &= \frac{(n-1)^2}{n^2}\mathrm{Var}(X_k) + \frac{1}{n^2}\sum_{\substack{i=1 \\ i\neq k}}^{n} \mathrm{Var}(X_i) \\ &= \left(\frac{(n-1)^2}{n^2} + \frac{n-1}{n^2}\right)\sigma^2 \\ &= \frac{n-1}{n}\sigma^2. \end{aligned}$$

Otherwise, similarly,

$$\mathrm{Var}(T) = \mathrm{Var}\left(X_k - \frac{1}{n}\sum_{i=1}^{n} X_i\right) = \left(1 + \frac{n}{n^2}\right)\sigma^2 = \frac{n+1}{n}\sigma^2. \qquad \square$$

Now, we examine each case individually. Recall that we assume $X_1, X_2, \ldots \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$. The general idea is to consider the numerator and divide it by its standard deviation to standardize it. Next, for the cases where σ is unknown, we replace it with an estimator. Then, if the resulting statistic does not have a normal distribution, we transform it using Theorem B.6.

(ii) By Lemma 4.1, note that $T_n \coloneqq X_n - \overline{X}_{n-1} \sim \mathcal{N}\Big(0, \sigma^2 \frac{n}{n-1}\Big)$ for $n \geq 2$. Therefore,

$$Q_n = \frac{1}{\sigma}\sqrt{\frac{n-1}{n}}\, T_n \sim \mathcal{N}(0,1).$$

(iii) Note that $X_n - \mu \sim \mathcal{N}(0, \sigma^2)$. Since σ is unknown, we instead divide by an estimator based on the first $n-1$ measurements, i.e.,

$$(S'_{n-1})^2 \coloneqq \frac{1}{n-1}\sum_{i=1}^{n-1}(X_i - \mu)^2.$$

This requires that $n \geq 2$. It holds that[6]

$$(n-1)\frac{(S'_{n-1})^2}{\sigma^2} \sim \chi^2(n-1).$$

Furthermore, note that S'_{n-1} is a function of $(X_1, \dots, X_{n-1})$, which is independent of X_n, so $X_n - \mu \perp\!\!\!\perp S'_{n-1}$. Therefore, by Definition B.5,

$$T_n = \frac{X_n - \mu}{S'_{n-1}} = \frac{\frac{1}{\sigma}(X_n - \mu)}{\sqrt{\frac{1}{\sigma^2}(S'_{n-1})^2}} \sim t(n-1).$$

Then, by the probability integral transformation, $G_{n-1}(T_n) \sim U(0,1)$ and

$$Q_n = \Phi^{-1}\big(G_{n-1}(T_n)\big) \sim \mathcal{N}(0,1).$$

(iv) For this case, the reasoning largely follows the previous cases. By Lemma 4.1, note that $T_n \coloneqq X_n - \overline{X}_{n-1} \sim \mathcal{N}\big(0, \sigma^2 \frac{n}{n-1}\big)$. Since σ is unknown, we instead divide by S_{n-1} as given by (3.18). This requires that $n \geq 3$. By (3.22), it holds that

$$(n-2)\frac{S^2_{n-1}}{\sigma^2} \sim \chi^2(n-2).$$

By Theorem B.10, we have that $\overline{X}_{n-1} \perp\!\!\!\perp S^2_{n-1}$. Furthermore, note that both are functions of $(X_1, \dots, X_{n-1})$, so $X_n \perp\!\!\!\perp (\overline{X}_{n-1}, S^2_{n-1})$. Together, this implies that X_n, $\overline{X}_{n-1}$, and S^2_{n-1} are mutually independent[7] and, in particular, that $X_n - \overline{X}_{n-1} \perp\!\!\!\perp S^2_{n-1}$. Therefore, by Definition B.5,

$$T_n = \sqrt{\frac{n-1}{n}}\,\frac{X_n - \overline{X}_{n-1}}{S_{n-1}} = \frac{\frac{1}{\sigma}\sqrt{\frac{n-1}{n}}\big(X_n - \overline{X}_{n-1}\big)}{\sqrt{\frac{1}{\sigma^2}S^2_{n-1}}} \sim t(n-2).$$

Then, like above, we find that

$$Q_n = \Phi^{-1}\big(G_{n-2}(T_n)\big) \sim \mathcal{N}(0,1).$$

[6] To see this, note that it is a sum of n terms $\Big(\frac{1}{\sigma}(X_i - \mu)\Big)^2$, which are squared standard normal random variables that are independent by Theorem 3.5. Then, the result follows from Definition B.4.

[7] This is because the joint density $f_{X_n, \overline{X}_{n-1}, S^2_{n-1}}$ can first be factorized into $f_{X_n} f_{\overline{X}_{n-1}, S^2_{n-1}}$ and then into $f_{X_n} f_{\overline{X}_{n-1}} f_{S^2_{n-1}}$. This is a convoluted proof of a fact that follows quite intuitively from Theorem B.10 directly.

Figure 4.2: A scatter plot of the sample correlation between self-starting statistics. These were computed from a simulation, which is described in Section 4.3.2.

4.3.2 Simulation of correlation

In this section, we will show using a simple simulation that the self-starting statistics are likely independent. We will focus on case IV, since it is the most complex. In short, we will randomly generate many independent measurements, calculate the self-starting statistics, and compute the sample correlation (i.e., Pearson's correlation coefficient) between them. If the correlation is high, the self-starting statistics are dependent, and if it is low, they may be independent (but this is not guaranteed). Nevertheless, it will provide a good heuristic argument.

The simulation was implemented using Mathematica. It has four parameters: the mean μ and variance σ^2 of the normal distribution, the size n of each sample, and the number of samples m. First, we generate nm random values from a $\mathcal{N}(\mu, \sigma^2)$ distribution, in samples of size n. Then, for each sample, we calculate $T_3, \dots, T_n$ as defined in case IV in Section 4.3. For each $T_3, \dots, T_n$ this yields m random values in total. Next, we compute the sample correlation $\operatorname{corr}(T_i, T_j)$ for each pair, $i \neq j$. These can then be analyzed further. The code for this can be found in Section C.1.

In Figure 4.2 the correlation coefficients for a simulation with $\mu = 5$, $\sigma^2 = 4$, $n = 50$, and $m = 1000$ are plotted. The points are ordered as follows:

$$\operatorname{corr}(T_1, T_2),\ \operatorname{corr}(T_1, T_3),\ \operatorname{corr}(T_2, T_3),\ \operatorname{corr}(T_1, T_4),\ \dots,\ \operatorname{corr}(T_{n-1}, T_n).$$

Note that most coefficients are less than 0.05 in absolute value, which means the self-starting statistics are extremely weakly correlated. In fact, when we increase m, we can lower the correlation to arbitrarily low values. Furthermore, the points

resemble (normally distributed) noise, i.e., no pattern is visible even though the points are ordered. Hence, uncorrelatedness seems to hold for all pairs.

The simulation was only performed for fixed n, but there is no apparent reason why the conclusion would not generalize to any n. This suggests that the self-starting statistics are pairwise uncorrelated. This does not imply independence, but since we assume i.i.d. normality (for which Theorem 3.8 holds), it is reasonable to conjecture that the self-starting statistics are mutually independent.

4.4 Regression charts

For conventional control charts, we make the (implicit) assumption that the true mean is constant. However, the quality characteristic may be dependent on other influences. These are called *covariates* by Centofanti et al. (2021). For example, the outside temperature is a covariate of the generator temperature of a wind turbine. As a result, the mean of the quality characteristic varies based on the value of the covariates. If the covariates are not accounted for, this can introduce a number of issues. For example, if a covariate takes extreme values, we may wrongly conclude that the process is out-of-control. Furthermore, the variance could be overestimated if the effect of the covariates is ignored.

One method to account for covariates is to combine control charts with linear regression (see Section 3.3). The case with simple linear regression (i.e., with one covariate) was discussed by Mandel (1969). In short, we measure the covariates and take them as explanatory variables in a linear regression model[8], where the response variable is the quality characteristic we want to monitor. Then, we can consider two types of control charts.

- Instead of the quality characteristic, we chart the residuals of the linear regression model. We call this the *regression control chart*.
- In the case of simple linear regression, we can chart the response variable against the explanatory variable, together with the regression line. Then, the control limits are parallel to the regression line. This has the advantage that the application of linear regression is evident from the chart, but the disadvantage that the time axis is lost. Furthermore, the chart cannot be generalized to multiple linear regression.

One straightforward implementation of regression charts would be to estimate the coefficients in Phase I and plot the predictive residuals (i.e., the difference between the observed value and the predicted value) in Phase II. This is the approach used by Mandel (1969). However, one issue is that the predictive residuals are dependent, as was shown by, for example, Van Dalen (2018).

[8] In the context of profile monitoring (e.g., see Zou et al. (2007)), this model is called a *profile*.

An alternative is to use the (studentized) recursive residuals, which are independent as will be shown in Section 5.2. Then, the chart effectively becomes a self-starting regression chart. Note that the studentized recursive residuals are not identically distributed (i.e., they differ in the degrees of freedom). Therefore, like the case IV Q-statistic, we will transform them to standard normal random variables. For $k = p+2, \dots, n$ we define

$$\begin{aligned} r_k &= \Phi^{-1}\big(G_{k-1-p}[w'_k]\big) \\ &= \Phi^{-1}\left(G_{k-1-p}\left[\frac{y_k - \mathbf{x}_k\hat{\boldsymbol{\beta}}^{(k-1)}}{\sqrt{\frac{1}{k-1-p}\Big(1+\mathbf{x}_k\big(X_{k-1}^\top X_{k-1}\big)^{-1}\mathbf{x}_k^\top\Big)SSR_{k-1}}}\right]\right) \end{aligned}$$

as the *normalized recursive residuals*. Here, Φ denotes the CDF of the standard normal distribution, and G_k denotes the CDF of the Student's t-distribution with k degrees of freedom. By Theorems 5.13 and B.6 they are i.i.d. $\mathcal{N}(0,1)$ random variables. An example of a regression chart using the normalized recursive residuals is shown in Figure 6.1 in Chapter 6.

Chapter summary

Control charts monitor a process in order to detect assignable causes. If the quality characteristic exceeds the control limits, the process is out-of-control. These control limits are computed in Phase I and are typically set at $\mu \pm 3\sigma$. Control charts are most effective when the measurements are (mutually) independent.

In case there is too little data for Phase I, we can instead use self-starting charts. This type of chart rescales each measurement using all previous measurements such that the charting statistic has a standard normal distribution. Using a simulation study, we showed that it is likely that they are independent. This will be formally shown in Chapter 5.

A third type of control chart, the regression chart, uses linear regression to account for the influence of covariates. Normalized recursive residuals are the regression case generalization of the self-starting statistic. In particular, they have the same property of being i.i.d. standard normal.

Chapter 5

Proofs of independence

This chapter contains the main contribution of this thesis. Namely, we will show in detail that, for each case, the self-starting statistics are mutually independent (Section 5.1). Then, we will extend these proofs and show that the recursive residuals are independent, for both the ordinary and studentized case (Section 5.2).

5.1 Self-starting statistics

In the following sections, we will prove mutual independence for each case of the self-starting statistics, which are defined in Section 4.3. Note that by Theorem 3.5, it is sufficient to show the independence of T_n, which allows us to focus on the key ideas. In fact, the mutual independence of case I follows directly from Theorem 3.5,[1] so it will not be considered below.

Among the remaining cases, we will first consider case II, which is the easiest. Then, we will proceed directly to case IV, since the mutual independence of case III follows from practically the same argument.

5.1.1 Case II

First, we will prove the minor lemma below. It essentially shows the pairwise independence of case II. The proof utilizes a common strategy: we show that the two random variables of interest are jointly normal, and then verify that their covariance is zero. Then, by Theorem 3.8, the random variables are independent.

Lemma 5.1 — *Let* $X_1, X_2, \ldots \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$. *Then, for* $k, n < \ell$ *and* $k, n \leq m$, *the random variables*

$$T_1 \coloneqq X_k - \overline{X}_n \quad \textit{and} \quad T_2 \coloneqq X_\ell - \overline{X}_m$$

[1] Simply note that $Q_n = (X_n - \mu)/\sigma$ is a function of X_n, and that $X_1, X_2, \ldots$ are assumed to be mutually independent.

are independent.

Proof. Note that T_1 and T_2 are jointly normal, since $aT_1 + bT_2$ is normally distributed for all $a, b \in \mathbb{R}$. This is because it is a linear combination of $X_1, X_2, \dots$ which are independent and normal. Furthermore, T_1 and T_2 are uncorrelated, since

$$\begin{aligned}
\operatorname{Cov}(T_1, T_2) &= \operatorname{Cov}\Big(X_k, X_\ell - \overline{X}_m\Big) - \operatorname{Cov}\Big(\overline{X}_n, X_\ell - \overline{X}_m\Big) \\
&= \operatorname{Cov}\Big(X_k, -\tfrac{1}{m}X_k\Big) - \sum_{i=1}^{n} \operatorname{Cov}\Big(\tfrac{1}{n}X_i, -\tfrac{1}{m}X_i\Big) \\
&= -\frac{1}{m}\operatorname{Var}(X_k) + \frac{1}{nm}\sum_{i=1}^{n} \operatorname{Var}(X_i) \\
&= -\frac{\sigma^2}{m} + \frac{n\sigma^2}{nm} \\
&= 0.
\end{aligned}$$

All cross terms disappear due to independence, i.e., $\operatorname{Cov}(X_i, X_j) = 0$ for all $i \neq j$. Now, independence follows from joint normality and uncorrelatedness. □

The preceding lemma is used in the proof of Theorem 5.3, but, most importantly, it is used in the following theorem to show mutual independence of case II.

Theorem 5.2 — *Let $n \geq 2$ and $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$. Then, the random variables*

$$T_k := X_k - \overline{X}_{k-1}, \quad k = 2, \dots, n$$

are mutually independent.

Proof. Note that $T_2, \dots, T_n$ are jointly normal, since $a_2T_2 + \cdots + a_nT_n$ is normally distributed for all $a_2, \dots, a_n \in \mathbb{R}$. This is because it is a linear combination of $X_1, \dots, X_n$ which are independent and normal. Furthermore, $T_2, \dots, T_n$ are pairwise independent by Lemma 5.1.[2] Together, joint normality and pairwise independence imply mutual independence. □

5.1.2 Case IV: Serial independence

Case IV is much more complex than case II. Hence, we will first prove serial independence, and then generalize the proof to mutual independence in Section 5.1.3.

Theorem 5.3 — *Let $n \geq 3$ and $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$. Then, the random variables*

$$T_k := \sqrt{\frac{k-1}{k}}\,\frac{X_k - \overline{X}_{k-1}}{S_{k-1}}, \quad k = 3, \dots, n$$

are serially[3] independent.

[2] In further detail: fix any $2 \leq k, \ell \leq n$ such that $k \neq \ell$. Assume $k < \ell$ without loss of generality. Then $X_k - X_{k-1}$ and $X_\ell - \overline{X}_{\ell-1}$ are independent by Lemma 5.1.

[3] See Definition 3.4.

Quesenberry (1991) presents a proof of Theorem 5.3 in the appendix.[4] It uses a lemma similar to Lemma 5.4, the proof of which is unclear. Basically, it claims that the pairwise independence of three random variables (namely Y_1/Y_2, $Y_1 + Y_2$, and Y_3) implies independence between the first (i.e., Y_1/Y_2) and the other two jointly (i.e., $Y_1 + Y_2$ and Y_3). This is not true in general.[5] For completeness and to circumvent this issue, we will give an alternative proof that directly factorizes the joint density. With this approach, Lemma 1 by Quesenberry (1991) is not necessary anymore.

In short, Lemma 5.4 will be proven as follows. To compute the joint density of W_1 and W_2, we transform the joint density of Y_1, Y_2, and Y_3 using Theorem B.7. Because there are fewer "output" than "input" variables, we introduce an *auxiliary* variable W_3. Since we are not interested in it, we compute the joint density f_{W_1,W_2} from f_{W_1,W_2,W_3} by integrating out W_3. In the end, we can conclude that $W_1 \perp\!\!\!\perp W_2$ if the density factorizes.

Lemma 5.4 — *Let Y_1, Y_2, Y_3 be mutually independent χ^2-distributed random variables with ν_1, ν_2, ν_3 degrees of freedom, respectively. Then, the random variables*

$$W_1 \coloneqq \frac{Y_1}{Y_2} \quad \textit{and} \quad W_2 \coloneqq \frac{Y_3}{Y_1 + Y_2}$$

are independent.

Proof. First, we introduce an auxiliary variable $W_3 \coloneqq Y_3$.[6] We write $\mathbf{Y} = (Y_1, Y_2, Y_3)$ and $\mathbf{W} = (W_1, W_2, W_3)$. Note that $\mathbf{W} = g(\mathbf{Y})$, where $g : \mathbb{R}^3 \to \mathbb{R}^3$ is defined by

$$g(y_1, y_2, y_3) = \left(\frac{y_1}{y_2}, \frac{y_3}{y_1 + y_2}, y_3\right).$$

Next, we will derive $h \coloneqq g^{-1}$ by solving the system of equations given by

$$\begin{cases} w_1 = \frac{y_1}{y_2}, \\ w_2 = \frac{y_3}{y_1+y_2}, \\ w_3 = y_3, \end{cases}$$

for y_1, y_2, y_3. The first and third equation yield $y_1 = w_1 y_2$ and $y_3 = w_3$, respectively. Using the second equation we find that

$$y_2 = \frac{y_3}{w_2} - y_1 = \frac{w_3}{w_2} - w_1 y_2 \implies y_2 = \frac{w_3}{(1+w_1)w_2},$$

[4] Specifically, refer to (7) in Theorem 1.

[5] This is closely related to the fact that pairwise independence does not imply mutual independence; refer to Counterexample 3.3. However, in this case, the claim does hold and can be proven using more precise reasoning. Note that $(Y_1/Y_2, Y_1 + Y_2) \perp\!\!\!\perp Y_3$ (since Y_1, Y_2, Y_3 are assumed to be mutually independent) and that $Y_1/Y_2 \perp\!\!\!\perp Y_1 + Y_2$ (by Lemma 1). Hence, the joint density can be fully factorized in two steps, so the three random variables are in fact mutually independent.

[6] Note that W_3 has a chi-squared distribution, so its support is $[0, \infty)$. This property will be used when integrating out W_3.

which implies that $y_1 = \frac{w_1 w_3}{(1+w_1)w_2}$. Therefore,

$$h(w_1, w_2, w_3) = \left(\frac{w_1 w_3}{(1+w_1)w_2}, \frac{w_3}{(1+w_1)w_2}, w_3\right).$$

The Jacobian matrix of h evaluated at $\mathbf{w} = (w_1, w_2, w_3)$ equals

$$\mathbf{J}_h(\mathbf{w}) = \begin{pmatrix} \frac{w_3}{(1+w_1)^2 w_2} & \frac{-w_1 w_3}{(1+w_1)w_2^2} & \frac{w_1}{(1+w_1)w_2} \\ \frac{-w_3}{(1+w_1)^2 w_2} & \frac{-w_3}{(1+w_1)w_2^2} & \frac{1}{(1+w_1)w_2} \\ 0 & 0 & 1 \end{pmatrix},$$

and therefore

$$|\det(\mathbf{J}_h(\mathbf{w}))| = \left|\frac{-w_3^2}{(1+w_1)^2 w_2^3}\right| = \frac{w_3^2}{(1+w_1)^2 w_2^3},$$

since chi-squared random variables are non-negative.

For the density of Y_i it holds that

$$f_{Y_i}(y) \propto y^{\nu_i/2-1} e^{-y/2}$$

if $y > 0$ and 0 otherwise.[7] Since Y_1, Y_2, Y_3 are mutually independent, it holds for the joint density of $\mathbf{Y}$ that

$$f_{\mathbf{Y}}(y_1, y_2, y_3) \propto y_1^{\nu_1/2-1} y_2^{\nu_2/2-1} y_3^{\nu_3/2-1} \exp\left(-\tfrac{1}{2}(y_1 + y_2 + y_3)\right)$$

if $y_1, y_2, y_3 > 0$ and 0 otherwise. Then, by Theorem B.7, we find that

$$\begin{aligned} f_{\mathbf{W}}(w_1, w_2, w_3) &\propto \left(\frac{w_1 w_3}{(1+w_1)w_2}\right)^{\nu_1/2-1} \left(\frac{w_3}{(1+w_1)w_2}\right)^{\nu_2/2-1} w_3^{\nu_3/2-1} \\ &\quad \cdot \exp\left(-\frac{1}{2}\left[\frac{w_1 w_3}{(1+w_1)w_2} + \frac{w_3}{(1+w_1)w_2} + w_3\right]\right) \cdot \frac{w_3^2}{(1+w_1)^2 w_2^3} \\ &= w_1^{\nu_1/2-1} (1+w_1)^{-(\nu_1+\nu_2)/2} \cdot w_2^{-(\nu_1+\nu_2)/2-1} \\ &\quad \cdot w_3^{(\nu_1+\nu_2+\nu_3)/2-1} \exp\left(-\frac{1+w_2}{2w_2} w_3\right) \end{aligned}$$

if $w_1, w_2, w_3 > 0$ and 0 otherwise. To obtain the joint density of (W_1, W_2), we integrate out W_3. First, let

$$c_1 = \tfrac{1}{2}(\nu_1 + \nu_2 + \nu_3) \quad \text{and} \quad c_2 = \tfrac{1+w_2}{2w_2}.$$

Then, by Lemma B.11, we find that

$$\int_0^\infty w_3^{c_1-1} e^{-c_2 w_3} \, \mathrm{d}w_3 \propto \left(\tfrac{1+w_2}{w_2}\right)^{-(\nu_1+\nu_2+\nu_3)/2}.$$

[7] The symbol $\propto$ denotes proportionality. See Appendix A for the exact definition.

Hence,

$$\begin{aligned} f_{W_1,W_2}(w_1,w_2) &= \int_{-\infty}^{\infty} f_{\mathbf{W}}(w_1,w_2,w_3)\,\mathrm{d}w_3 \\ &\propto w_1^{\nu_1/2-1}(1+w_1)^{-(\nu_1+\nu_2)/2}\cdot w_2^{-(\nu_1+\nu_2)/2-1} \\ &\qquad \cdot \int_0^{\infty} w_3^{(\nu_1+\nu_2+\nu_3)/2-1}\exp\!\left(-\frac{1+w_2}{2w_2}w_3\right)\mathrm{d}w_3 \\ &\propto \underbrace{w_1^{\nu_1/2-1}(1+w_1)^{-(\nu_1+\nu_2)/2}}_{\coloneqq p_1(w_1)}\cdot \underbrace{w_2^{\nu_3/2-1}(1+w_2)^{-(\nu_1+\nu_2+\nu_3)/2}}_{\coloneqq p_2(w_2)}. \end{aligned}$$

if $w_1, w_2 > 0$ and 0 otherwise. Therefore, by Theorem 3.2, it holds that $W_1 \perp\!\!\!\perp W_2$. □

Now, we will reproduce Quesenberry's proof of Theorem 5.3, which uses this lemma. The key observation of this proof is that, using (3.20), the sample variance can be updated with the numerator of the self-starting statistic.

Fix k such that $3 \leq k < n$. We will attempt to show that $T_k \perp\!\!\!\perp T_{k+1}$. Let

$$\begin{aligned} Y_1 &= \frac{1}{\sigma^2}\sum_{i=1}^{k-1}\left(X_i - \overline{X}_{k-1}\right)^2 = \frac{k-2}{\sigma^2}S_{k-1}^2, \\ Y_2 &= \frac{1}{\sigma^2}\frac{k-1}{k}\left(X_k - \overline{X}_{k-1}\right)^2, \\ Y_3 &= \frac{1}{\sigma^2}\frac{k}{k+1}\left(X_{k+1} - \overline{X}_k\right)^2. \end{aligned}$$

It follows from (3.22) that $Y_1 \sim \chi^2(k-2)$ and Lemma 4.1 implies that $Y_2, Y_3 \sim \chi^2(1)$. Next, let

$$W_1 = \frac{Y_2}{\frac{1}{k-2}Y_1} = \frac{k-1}{k}\left(\frac{X_k - \overline{X}_{k-1}}{S_{k-1}}\right)^2$$

and, using (3.20),

$$W_2 = \frac{Y_3}{\frac{1}{k-1}(Y_1+Y_2)} = \frac{\frac{k}{k+1}\left(X_{k+1}-\overline{X}_k\right)^2}{\frac{k-2}{k-1}S_{k-1}^2 + \frac{1}{k}\left(X_k - \overline{X}_{k-1}\right)^2} = \frac{k}{k+1}\left(\frac{X_{k+1}-\overline{X}_k}{S_k^2}\right)^2.$$

It can be shown that Y_1, Y_2, and Y_3 are mutually independent,[8] which implies that $W_1 \perp\!\!\!\perp W_2$ by Lemma 5.4. Therefore, by Theorem 3.5, $\sqrt{W_1} = |T_k|$ and $\sqrt{W_2} = |T_{k+1}|$ are independent as well, but we cannot conclude the same for T_k and T_{k+1}. This is because the square function is not bijective. This means that squaring T_k oversimplifies the problem and Lemma 5.4 is too weak to prove Theorem 5.3, although Quesenberry (1991) glosses over this issue.

[8] We will not go into detail here, but this follows from Lemma 5.1 and joint normality.

To resolve this issue, we state a new lemma that proves independence for statistics modeled after T_k instead of T_k^2. Thus, some input variables have to change from a χ^2 to a normal distribution. The resulting statistics are more complex, but it turns out the same proof strategy still works: using an auxiliary variable we compute the joint density and show that it factorizes.

Lemma 5.5 — *Let $Y \sim \chi^2(k)$ and $Z_1, Z_2 \sim \mathcal{N}(0,1)$. Assume that Y, Z_1, Z_2 are mutually independent. Then, the random variables*

$$V_1 \coloneqq \frac{Z_1}{\sqrt{Y}} \quad \text{and} \quad V_2 \coloneqq \frac{Z_2}{\sqrt{Y + Z_1^2}}$$

are independent.

Proof. First, we introduce an auxiliary variable $V_0 \coloneqq Y$. We write $\mathbf{Z} = (Y, Z_1, Z_2)$ and $\mathbf{V} = (V_0, V_1, V_2)$. Note that $\mathbf{V} = g(\mathbf{Z})$, where $g : \mathbb{R}^3 \to \mathbb{R}^3$ is defined by

$$g(y, z_1, z_2) = \left(y, \frac{z_1}{\sqrt{y}}, \frac{z_2}{\sqrt{y + z_1^2}}\right).$$

Its inverse $h \coloneqq g^{-1}$ is defined by

$$h(v_0, v_1, v_2) = \left(v_0, \sqrt{v_0}v_1, \sqrt{v_0(1 + v_1^2)}v_2\right).$$

The Jacobian matrix of h evaluated at $\mathbf{v} = (v_0, v_1, v_2)$ equals[9]

$$\mathbf{J}_h(\mathbf{v}) = \begin{pmatrix} 1 & 0 & 0 \\ * & \sqrt{v_0} & 0 \\ * & * & \sqrt{v_0(1 + v_1^2)} \end{pmatrix}$$

and therefore

$$|\det(\mathbf{J}_h(\mathbf{v}))| = \left|\sqrt{v_0^2(1 + v_1^2)}\right| = |v_0|\sqrt{1 + v_1^2}.$$

Now, for the densities, note that

$$f_Y(x) \propto x^{k/2-1} \exp\left(-\tfrac{1}{2}x\right), \qquad \text{(if } x > 0 \text{ and 0 otherwise)}$$
$$f_{Z_i}(x) \propto \exp\left(-\tfrac{1}{2}x^2\right).$$

Since Y, Z_1, Z_2 are mutually independent, it holds for the joint density of $\mathbf{Z}$ that

$$f_{\mathbf{Z}}(y, z_1, z_2) \propto y^{k/2-1} \exp\left(-\tfrac{1}{2}\left(y + z_1^2 + z_2^2\right)\right)$$

if $y > 0$ and 0 otherwise. Then, by Theorem B.7, we find that

$$\begin{aligned} f_{\mathbf{V}}(v_0, v_1, v_2) &\propto v_0^{k/2-1} \exp\left(-\tfrac{1}{2}\left(v_0 + v_0 v_1^2 + v_0(1 + v_1^2)v_2^2\right)\right) \cdot |v_0|\sqrt{1 + v_1^2} \\ &= v_0^{k/2}\sqrt{1 + v_1^2} \exp\left(-\tfrac{1}{2}v_0(1 + v_1^2)(1 + v_2^2)\right) \\ &= v_0^{z-1}\sqrt{1 + v_1^2} \exp(-cv_0) \end{aligned}$$

[9]The symbol $*$ is used as a placeholder for irrelevant expressions.

if $v_0 > 0$ and 0 otherwise, where $z \coloneqq \frac{1}{2}k + 1$ and $c \coloneqq \frac{1}{2}(1 + v_1^2)(1 + v_2^2)$. To obtain the joint density of (V_1, V_2), we integrate out V_0. Since $c, z > 0$ and neither depends on v_0, it follows from Lemma B.11 that

$$\begin{aligned} f_{V_1,V_2}(v_1, v_2) &= \int_{-\infty}^{\infty} f_{\mathbf{V}}(v_0, v_1, v_2)\,\mathrm{d}v_0 \\ &\propto \sqrt{1 + v_1^2} \int_0^{\infty} v_0^{z-1} \exp(-cv_0)\,\mathrm{d}v_0 \\ &= \sqrt{1 + v_1^2}\, c^{-z}\, \Gamma(z) \\ &\propto \left(1 + v_1^2\right)^{-(k+1)/2} \left(1 + v_2^2\right)^{-k/2-1}. \end{aligned}$$

Therefore, by Theorem 3.2, we conclude that $V_1 \perp\!\!\!\perp V_2$. □

This new lemma is sufficient to prove Theorem 5.3, in a similar way that we attempted before.

Proof of Theorem 5.3. Fix k such that $3 \leq k < n$. We will show that $T_k \perp\!\!\!\perp T_{k+1}$. Let

$$\begin{aligned} Y &= \frac{1}{\sigma^2} \sum_{i=1}^{k-1} \left(X_i - \overline{X}_{k-1}\right)^2 = \frac{k-2}{\sigma^2} S_{k-1}^2, \\ Z_1 &= \frac{1}{\sigma} \sqrt{\frac{k-1}{k}} \left(X_k - \overline{X}_{k-1}\right), \\ Z_2 &= \frac{1}{\sigma} \sqrt{\frac{k}{k+1}} \left(X_{k+1} - \overline{X}_k\right). \end{aligned}$$

It follows from (3.22) that $Y \sim \chi^2(k-2)$ and Lemma 4.1 implies that $Z_1, Z_2 \sim \mathcal{N}(0, 1)$. Next, let

$$V_1 \coloneqq \frac{Z_1}{\sqrt{Y}} \quad \text{and} \quad V_2 \coloneqq \frac{Z_2}{\sqrt{Y + Z_1^2}}$$

such that

$$\sqrt{k-2}\, V_1 = \frac{Z_1}{\sqrt{\frac{1}{k-2} Y}} = \sqrt{\frac{k-1}{k}} \frac{X_k - \overline{X}_{k-1}}{S_{k-1}} = T_k$$

and, using (3.20),

$$\sqrt{k-1}\, V_2 = \frac{\sqrt{\frac{k}{k+1}} \left(X_{k+1} - \overline{X}_k\right)}{\sqrt{\frac{k-2}{k-1} S_{k-1}^2 + \frac{1}{k} \left(X_k - \overline{X}_{k-1}\right)^2}} = \sqrt{\frac{k}{k+1}} \frac{X_{k+1} - \overline{X}_k}{S_k} = T_{k+1}.$$

If $V_1 \perp\!\!\!\perp V_2$, then $T_k \perp\!\!\!\perp T_{k+1}$ by Theorem 3.5, which concludes the proof. The former follows from Lemma 5.5 if Y, Z_1, Z_2 are mutually independent. To prove this, we define

the random vector

$$\mathbf{U} = \begin{pmatrix} X_1 & -\overline{X}_{k-1} \\ & \vdots \\ X_{k-1} & -\overline{X}_{k-1} \\ X_k & -\overline{X}_{k-1} \\ X_{k+1} & -\overline{X}_k \end{pmatrix}.$$

Note that Z_1 is a function of U_k, Z_2 of U_{k+1}, and Y of $(U_1, \dots, U_{k-1})$. Using Lemma 5.1, we find that $U_k \perp\!\!\!\perp U_{k+1}$ and that

$$\forall i = 1, \dots, k-1 \ : \ U_i \perp\!\!\!\perp U_k \text{ and } U_i \perp\!\!\!\perp U_{k+1}.$$

Furthermore, due to the joint normality of $\mathbf{U}$,[10] pairwise independence implies mutual independence. Therefore $(U_1, \dots, U_{k-1}), U_k, U_{k+1}$ are mutually independent, which implies Y, Z_1, Z_2 are also mutually independent. □

5.1.3 Case IV: Mutual independence

It is possible to generalize Theorem 5.3 to prove pairwise independence, i.e., show that $T_k \perp\!\!\!\perp T_\ell$ for some $k < \ell$. Using the notation from Lemma 5.5, we would need to introduce a fourth random variable, i.e., Y', that "updates" the variance S_k^2 to $S_{\ell-1}^2$. This would require modifying Lemma 5.5 to prove that

$$\frac{Z_1}{\sqrt{Y}} \perp\!\!\!\perp \frac{Z_2}{\sqrt{Y + Z_1^2 + Y'}}$$

Note that Y' would have a chi-squared distribution with $\ell - 1 - k$ degrees of freedom. One disadvantage of this approach, however, is that it requires a second auxiliary variable.

An alternative is to directly generalize the proof to mutual independence, which implies pairwise independence as a special case. For this proof, the most important preliminary is Proposition 3.21. This formula reveals a relationship between the numerator and denominator of the self-starting statistic. In particular, the variance of the first k measurements is essentially the sum of the squared numerators of the first $k-2$ self-starting statistics (and an initial term). This is shown in full detail in the proof of Theorem 5.10.

In order to prove mutual independence, we need to generalize Lemma 5.5. Since we will express the denominator of the self-starting statistic in terms of the numerators, we do not need a χ^2-distributed variable Y to "kick start" the variance. Instead, all input variables can be standard normal, which simplifies the derivation of the density. Furthermore, we will transform n "input" variables to $n-1$ "output" variables, so still

[10] Note that any linear combination of the components of $\mathbf{U}$ is itself a linear combination of $X_1, \dots, X_n$ which are independent and normal.

only one auxiliary variable is necessary. Therefore, in some sense, the proof for mutual independence is more elegant than the proof for pairwise independence.

Now, we will prove the generalized lemma.

Lemma 5.6 — *Let* $Z_1, \dots, Z_n \overset{\text{iid}}{\sim} \mathcal{N}(0,1)$*. Then, the random variables*

$$V_2 \coloneqq \frac{Z_2}{\sqrt{Z_1^2}}, \quad V_3 \coloneqq \frac{Z_3}{\sqrt{Z_1^2 + Z_2^2}}, \quad \dots, \quad V_n \coloneqq \frac{Z_n}{\sqrt{Z_1^2 + \cdots + Z_{n-1}^2}}$$

are mutually independent.

Proof. First, we introduce an auxiliary variable $V_1 \coloneqq Z_1$.[11] We write $\mathbf{Z} = (Z_1, \dots, Z_n)$ and $\mathbf{V} = (V_1, \dots, V_n)$. Note that $\mathbf{V} = g(\mathbf{Z})$, where the components of $g : \mathbb{R}^n \to \mathbb{R}^n$ are defined by

$$\begin{aligned} g_1(\mathbf{z}) &= z_1, \\ g_k(\mathbf{z}) &= z_k \,/\, \sqrt{z_1^2 + \cdots + z_{k-1}^2}, \quad k = 2, \dots, n. \end{aligned}$$

Next, we will derive $h \coloneqq g^{-1}$ by solving the system of equations given by

$$\begin{cases} v_1 = z_1, \\ v_2 = z_2 \,/\, \sqrt{z_1^2}, \\ \quad \vdots \\ v_n = z_n \,/\, \sqrt{z_1^2 + \cdots + z_{n-1}^2}, \end{cases}$$

for $z_1, \dots, z_n$. Define $s_k = \sum_{i=1}^{k} z_i^2$. The k-th equation in the system above implies that

$$z_k = \sqrt{s_{k-1}}\, v_k, \quad k = 2, \dots, n. \tag{5.7}$$

Now, by induction, we will show that

$$s_k = v_1^2 \prod_{i=2}^{k} (1 + v_i^2), \quad k = 1, \dots, n. \tag{5.8}$$

The base case $s_1 = z_1^2 = v_1^2$ holds trivially. Assuming that (5.8) holds for $1 \le k < n$ (induction hypothesis), we will show that it also holds for $k+1$. Indeed,

$$\begin{aligned} s_{k+1} &= s_k + z_{k+1}^2 && \text{(by definition)} \\ &= s_k + s_k v_{k+1}^2 && \text{(by 5.7)} \\ &= s_k(1 + v_{k+1}^2) \\ &= v_1^2 \prod_{i=2}^{k+1} (1 + v_i^2). && \text{(by IH)} \end{aligned}$$

[11] Other definitions for the auxiliary variable are possible as well, such as $V_1 \coloneqq Z_n$. However, this particular choice greatly simplifies the Jacobian matrix later in the proof.

Therefore, (5.7) and (5.8) imply that

$$\begin{aligned} z_1 &= v_1, \\ z_k &= v_k \sqrt{v_1^2 \prod_{i=2}^{k-1} (1 + v_i^2)}, \quad k = 2, \dots, n \end{aligned}$$

and, consequently,

$$\begin{aligned} h_1(\mathbf{v}) &= v_1, \\ h_2(\mathbf{v}) &= \sqrt{v_1^2}\, v_2, \\ h_3(\mathbf{v}) &= \sqrt{v_1^2(1 + v_2^2)}\, v_3, \\ &\vdots \\ h_n(\mathbf{v}) &= \sqrt{v_1^2(1 + v_2^2) \cdots (1 + v_{n-1}^2)}\, v_n. \end{aligned}$$

Next, we want to compute the determinant of $\mathbf{J}_h(\mathbf{v})$, i.e., the Jacobian matrix of h evaluated at $(v_1, \dots, v_n)$. For $i < j$, note that $h_i(\mathbf{v})$ does not involve v_j, so

$$\big(\mathbf{J}_h(\mathbf{v})\big)_{ij} = \frac{\partial h_i}{\partial v_j}(\mathbf{v}) = 0.$$

In other words, $\mathbf{J}_h(\mathbf{v})$ is lower triangular. This implies that its determinant is simply the product of its diagonal. Therefore,

$$\big|\det\big(\mathbf{J}_h(\mathbf{v})\big)\big| = \prod_{i=1}^{n} \frac{\partial h_i}{\partial v_i}(\mathbf{v}) = |v_1|^{n-1} \underbrace{\prod_{i=2}^{n} (1 + v_i^2)^{\frac{n-i}{2}}}_{\coloneqq\, p_1(\mathbf{v})}.$$

Since Z_i are i.i.d. $\mathcal{N}(0,1)$, it holds that

$$f_{\mathbf{Z}}(\mathbf{z}) \propto \prod_{i=1}^{n} e^{-z_i^2/2} = e^{-s_n/2}, \quad \mathbf{z} \in \mathbb{R}^n.$$

Then, by Theorem B.7, we find that

$$f_{\mathbf{V}}(\mathbf{v}) = f_{\mathbf{Z}}\big(h(\mathbf{v})\big)\, \big|\det\big(\mathbf{J}_h(\mathbf{v})\big)\big|. \tag{5.9}$$

To simplify our derivation, we define

$$p_2(\mathbf{v}) = \frac{1}{2} \prod_{i=2}^{n} (1 + v_i^2)$$

such that $e^{-s_n/2} = e^{-p_2(\mathbf{v})\, v_1^2}$ by (5.8). Combining the above, (5.9) resolves to

$$f_{\mathbf{V}}(\mathbf{v}) \propto p_1(\mathbf{v})\, |v_1|^{n-1} e^{-p_2(\mathbf{v})\, v_1^2}.$$

Next, we integrate out V_1 to obtain the joint density of $V_2, \dots, V_n$. Note that $p_1(\mathbf{v})$ and $p_2(\mathbf{v})$ do not depend on v_1, so they are effectively constants. Using Lemma B.12, we

find that

$$\begin{aligned} f_{V_2,\dots,V_n}(v_2,\dots,v_n) &= \int_{-\infty}^{\infty} f_{\mathbf{V}}(\mathbf{v})\,\mathrm{d}v_1 \\ &\propto p_1(\mathbf{v}) \int_{-\infty}^{\infty} |v_1|^{n-1} e^{-p_2(\mathbf{v})\,v_1^2}\,\mathrm{d}v_1 \\ &\propto p_1(\mathbf{w})\big(p_2(\mathbf{w})\big)^{-n/2}. \end{aligned}$$

This can be simplified further, but it is sufficient to note that $f_{V_2,\dots,V_n}$ factorizes into functions of $v_2,\dots,v_n$. This is because $p_1(\mathbf{v})$ and $p_2(\mathbf{v})$ are factorizable and do not depend on v_1. Therefore, by Theorem 3.2, we conclude that $V_2,\dots,V_n$ are mutually independent. □

Using the preceding lemma, we can now prove the mutual independence of the case IV self-starting statistic.

Theorem 5.10 — *Let $n \geq 3$ and $X_1,\dots,X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu,\sigma^2)$. Then, the random variables*

$$T_k := \sqrt{\frac{k-1}{k}}\,\frac{X_k - \overline{X}_{k-1}}{S_{k-1}}, \quad k = 3,\dots,n$$

are mutually independent.

Proof. Let

$$Z_k = \tfrac{1}{\sigma}\sqrt{\tfrac{k-1}{k}}\big(X_k - \overline{X}_{k-1}\big), \quad k = 2,\dots,n.$$

These are the case II self-starting statistics, which are mutually independent by Theorem 5.2 and have a $\mathcal{N}(0,1)$ distribution. Note that, by Proposition 3.21,

$$S_k^2 = \frac{\sigma^2}{k-1}\sum_{i=2}^{k} Z_i^2, \quad k = 2,\dots,n.$$

Hence,

$$T_k = \frac{\sigma Z_k}{\sqrt{\frac{\sigma^2}{k-2}\sum_{i=2}^{k-1} Z_i^2}} = \sqrt{k-2}\,V_k \quad \text{where} \quad V_k := \frac{Z_k}{\sqrt{Z_2^2 + \cdots + Z_{k-1}^2}}$$

for $k = 3,\dots,n$. Since the V_k are independent by Lemma 5.6,[12] the T_k are independent by Theorem 3.5. □

In the end, our proof is substantially different from the one given by Quesenberry (1991). This is because, in summary, it is faulty regarding the following aspects.

- The main text claims that the self-starting statistics are mutually independent, referring to the appendix, but in fact only serial independence is shown.

[12] Compared to Lemma 5.6, note that the indices in this proof are offset by one. That is, Z_k in this proof corresponds to Z_{k-1} in the lemma, and similarly for V_k.

- The proof of Lemma 2 incorrectly assumes that if a random variable is pairwise independent of two other random variables, then it is also independent of any function (i.e., jointly independent) of the other two. Symbolically, for random variables A, B_1, and B_2,

$$A \perp\!\!\!\perp B_1 \wedge A \perp\!\!\!\perp B_2 \implies A \perp\!\!\!\perp (B_1, B_2).$$

- In the proof of Theorem 1, it is claimed that two random variables are independent because their squares are independent. This is not true, however, since the square root is not a bijective function.

5.1.4 Case III

To show mutual independence for case III, we follow almost the same strategy.

Theorem 5.11 — *Let $n \geq 2$ and $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$. Then, the random variables*

$$T_k := \frac{X_k - \mu}{S'_{k-1}}, \quad k = 2, \dots, n$$

where

$$\left(S'_{k-1}\right)^2 = \frac{1}{k-1} \sum_{i=1}^{k-1} (X_i - \mu)^2$$

are mutually independent.

Proof (sketch). Let

$$Z_k = \frac{X_k - \mu}{\sigma}, \quad k = 1, \dots, n$$

such that $Z_1, \dots, Z_n \overset{\text{iid}}{\sim} \mathcal{N}(0, 1)$. Then,

$$T_k = \frac{\sigma Z_k}{\sqrt{\frac{\sigma^2}{k-1} \sum_{i=1}^{k-1} Z_i^2}} = \frac{(k-1) Z_k}{\sqrt{Z_1^2 + \cdots + Z_{k-1}^2}}.$$

Therefore, the T_k are independent by Lemma 5.6. □

5.2 Recursive residuals

The correspondence between Propositions 3.21 and 3.27 reveals that the studentized recursive residuals and the self-starting statistics are similar in structure (this connection will be formalized in Corollary 5.14). Therefore, in Section 5.2.2, we will show that the studentized recursive residuals are mutually independent using a proof similar to Theorem 5.10. First, however, we will show the mutual independence of the (ordinary) recursive residuals in Section 5.2.1.

Note that in order to standardize the (ordinary) recursive residuals, we need to divide them by σ, in a fashion analogous to the case II self-starting statistic. Therefore, we refer to them as the case with "known variance". Likewise, the studentized recursive residuals correspond to case IV, so we refer to them as the case with "unknown variance".

5.2.1 Known variance

The proof of the following theorem is a simplification of the proof by Van Dalen (2018), which is based on the proof by Brown et al. (1975). Although it requires more calculation than, for example, the proof of Theorem 5.2, the underlying idea is again that joint normality and pairwise uncorrelatedness implies mutual independence.

Theorem 5.12 — *Consider the linear regression model given by* $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ *as defined in Section 3.3, with the assumptions that* X_p *has full rank and that* $\varepsilon_1, \dots, \varepsilon_n \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$. *Then, the recursive residuals*

$$w_k = \frac{y_k - \mathbf{x}_k \hat{\boldsymbol{\beta}}^{(k-1)}}{\sqrt{1 + \mathbf{x}_k \left(X_{k-1}^\top X_{k-1}\right)^{-1} \mathbf{x}_k^\top}}, \quad k = p+1, \dots, n$$

are mutually independent.

Proof. By Theorem 3.5, it is sufficient to show that

$$T_k := y_k - \mathbf{x}_k \hat{\boldsymbol{\beta}}^{(k-1)}, \quad k = p+1, \dots, n$$

are mutually independent. Note that

$$\begin{aligned} T_k &= y_k - \mathbf{x}_k \left(X_{k-1}^\top X_{k-1}\right)^{-1} X_{k-1}^\top \mathbf{y}_{[1:k-1]} \\ &= \mathbf{x}_k \boldsymbol{\beta} + \varepsilon_k - \mathbf{x}_k \left(X_{k-1}^\top X_{k-1}\right)^{-1} X_{k-1}^\top \left(X_{k-1}\boldsymbol{\beta} + \boldsymbol{\varepsilon}_{[1:k-1]}\right) \\ &= \mathbf{x}_k \boldsymbol{\beta} + \varepsilon_k - \mathbf{x}_k \boldsymbol{\beta} - \mathbf{x}_k \left(X_{k-1}^\top X_{k-1}\right)^{-1} X_{k-1}^\top \boldsymbol{\varepsilon}_{[1:k-1]} \\ &= \varepsilon_k - \mathbf{x}_k \left(X_{k-1}^\top X_{k-1}\right)^{-1} X_{k-1}^\top \boldsymbol{\varepsilon}_{[1:k-1]}. \end{aligned}$$

Hence, T_k is a linear combination of $\varepsilon_1, \dots, \varepsilon_k$. Furthermore, note that $T_{p+1}, \dots, T_n$ are jointly normal, since $a_{p+1}T_{p+1} + \cdots + a_n T_n$ is normally distributed for all $a_{p+1}, \dots, a_n \in \mathbb{R}$. This is because it is a linear combination of $\varepsilon_1, \dots, \varepsilon_n$ which are independent and normal. Therefore, due to joint normality, mutual independence is implied by pairwise uncorrelatedness. It remains to compute the covariance between T_k and T_ℓ, where

$k < \ell$ (without loss of generality). We find that

$$\begin{aligned}
&\operatorname{Cov}(T_k, T_\ell) \\
&\quad = \operatorname{Cov}\Big(\varepsilon_k - \mathbf{x}_k\big(X_{k-1}^\top X_{k-1}\big)^{-1} X_{k-1}^\top \boldsymbol{\varepsilon}_{[1:k-1]}, \varepsilon_\ell - \mathbf{x}_\ell\big(X_{\ell-1}^\top X_{\ell-1}\big)^{-1} X_{\ell-1}^\top \boldsymbol{\varepsilon}_{[1:\ell-1]}\Big) \\
&\quad = \operatorname{Cov}\Big(\varepsilon_k, \varepsilon_\ell - \mathbf{x}_\ell\big(X_{\ell-1}^\top X_{\ell-1}\big)^{-1} X_{\ell-1}^\top \boldsymbol{\varepsilon}_{[1:\ell-1]}\Big) \\
&\qquad + \operatorname{Cov}\Big(-\mathbf{x}_k\big(X_{k-1}^\top X_{k-1}\big)^{-1} X_{k-1}^\top \boldsymbol{\varepsilon}_{[1:k-1]}, \varepsilon_\ell - \mathbf{x}_\ell\big(X_{\ell-1}^\top X_{\ell-1}\big)^{-1} X_{\ell-1}^\top \boldsymbol{\varepsilon}_{[1:\ell-1]}\Big) \\
&\quad = \operatorname{Cov}\Big(\varepsilon_k, -\mathbf{x}_\ell\big(X_{\ell-1}^\top X_{\ell-1}\big)^{-1} \mathbf{x}_k^\top \varepsilon_k\Big) \\
&\qquad + \operatorname{Cov}\Big(-\mathbf{x}_k\big(X_{k-1}^\top X_{k-1}\big)^{-1} X_{k-1}^\top \boldsymbol{\varepsilon}_{[1:k-1]}, -\mathbf{x}_\ell\big(X_{\ell-1}^\top X_{\ell-1}\big)^{-1} X_{k-1}^\top \boldsymbol{\varepsilon}_{[1:k-1]}\Big) \\
&\quad = \Big(-\mathbf{x}_\ell\big(X_{\ell-1}^\top X_{\ell-1}\big)^{-1} \mathbf{x}_k^\top\Big)\sigma^2 \\
&\qquad + \Big(-\mathbf{x}_k\big(X_{k-1}^\top X_{k-1}\big)^{-1} X_{k-1}^\top\Big)\sigma^2 I_{k-1}\Big(-\mathbf{x}_\ell\big(X_{\ell-1}^\top X_{\ell-1}\big)^{-1} X_{k-1}^\top\Big)^\top \\
&\quad = \Big(-\mathbf{x}_\ell\big(X_{\ell-1}^\top X_{\ell-1}\big)^{-1} \mathbf{x}_k^\top\Big)\sigma^2 \\
&\qquad + \Big(\mathbf{x}_k\big(X_{k-1}^\top X_{k-1}\big)^{-1} X_{k-1}^\top X_{k-1}\big(X_{\ell-1}^\top X_{\ell-1}\big)^{-1} \mathbf{x}_\ell^\top\Big)\sigma^2 \\
&\quad = \Big(-\mathbf{x}_\ell\big(X_{\ell-1}^\top X_{\ell-1}\big)^{-1} \mathbf{x}_k^\top + \mathbf{x}_k\big(X_{\ell-1}^\top X_{\ell-1}\big)^{-1} \mathbf{x}_\ell^\top\Big)\sigma^2 \\
&\quad = 0.
\end{aligned}$$

In the third step above, it is used that $\operatorname{Cov}(\varepsilon_i, \varepsilon_j) = 0$ for all $i \neq j$ due to independence. Furthermore, the fourth step uses that $\operatorname{Cov}(A\boldsymbol{\varepsilon}, B\boldsymbol{\varepsilon}) = A \operatorname{Cov}(\boldsymbol{\varepsilon}) B^\top$. □

5.2.2 Unknown variance

Now, lastly, we will show that the studentized recursive residuals are mutually independent. The proof of this is almost identical to the proof of Theorem 5.10, except that a different updating formula is used. Interestingly, the proof of the "easier" case (i.e., Theorem 5.12) is used in the same way that Theorem 5.2 is applied in the proof of Theorem 5.10.

Theorem 5.13 — *Consider the linear regression model given by* $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ *as defined in Section 3.3, with the assumptions that* X_p *has full rank and that* $\varepsilon_1, \dots, \varepsilon_n \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$. *Then, the studentized recursive residuals*

$$w_k' = \frac{w_k}{\sqrt{\frac{1}{k-1-p} SSR_{k-1}}}, \quad k = p+2, \dots, n$$

are mutually independent. Here, w_k *is defined by* (3.15).

Proof. Let

$$Z_k = \frac{1}{\sigma} w_k, \quad k = p+1, \dots, n.$$

These are the standardized recursive residuals, which are mutually independent by Theorem 5.12 and have a $\mathcal{N}(0,1)$ distribution. Note that, by Proposition 3.27,

$$SSR_k = \sigma^2 \sum_{i=p+1}^{k} Z_i^2, \quad k = p+1, \dots, n.$$

Hence,

$$w'_k = \frac{\sigma Z_k}{\sqrt{\frac{\sigma^2}{k-1-p}\sum_{i=p+1}^{k-1} Z_i^2}} = \sqrt{k-1-p}\, V_k \quad \text{where} \quad V_k \coloneqq \frac{Z_k}{\sqrt{Z_{p+1}^2 + \cdots + Z_{k-1}^2}}$$

for $k = p+2, \dots, n$. Since the V_k are independent by Lemma 5.6, the w'_k are independent by Theorem 3.5. □

Note that Theorem 5.13 is stronger than Theorem 5.10, i.e., the former implies the latter as a special case. This is shown by the following result. The basic idea is to use a linear regression model that only includes an intercept. Then, the first (and only) coefficient is the mean.

Corollary 5.14 — *Let $n \geq 3$ and $A_1, \dots, A_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$.*[13] *Then, the random variables*

$$T_k \coloneqq \sqrt{\frac{k-1}{k}}\, \frac{A_k - \overline{A}_{k-1}}{S_{k-1}}, \quad k = 3, \dots, n$$

are mutually independent.

Proof. Consider the linear regression model given by $\mathbf{y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ as defined in Section 3.3, with the assumption that $\varepsilon_1, \dots, \varepsilon_n \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$. Choose $\mathbf{y} = (A_1, \dots, A_n)^\top$ and $X = \mathbf{1}_n$ such that $\boldsymbol{\beta} = \big(\mu\big) = \mu$.[14] Note that $\operatorname{rank}(X_p) = 1 = p$, so X_p has full rank. Then, $\mathbf{x}_k = \big(1\big) = 1$ and

$$\hat{\boldsymbol{\beta}}^{(k)} = \big(\mathbf{1}_k^\top \mathbf{1}_k\big)^{-1} \mathbf{1}_k^\top \mathbf{y}_{[1:k]} = k^{-1} \sum_{i=1}^{k} A_i = \overline{A}_k$$

for $k = 1, \dots, n$. Hence,

$$SSR_k = \sum_{i=1}^{k} \big(y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}}^{(k)}\big)^2 = \sum_{i=1}^{k} \big(A_i - 1 \cdot \overline{A}_k\big)^2 = (k-1) S_k^2$$

for $k = 2, \dots, n$. Therefore, recursive residuals simplify to

$$w_k = \frac{A_k - 1 \cdot \overline{A}_{k-1}}{\sqrt{1 + 1 \cdot \big(\mathbf{1}_{k-1}^\top \mathbf{1}_{k-1}\big)^{-1} \cdot 1}} = \frac{A_k - \overline{A}_{k-1}}{\sqrt{1 + \frac{1}{k-1}}} = \sqrt{\frac{k-1}{k}} \big(A_k - \overline{A}_{k-1}\big)$$

[13] Since the design matrix is already denoted by X, we instead use A_i to denote measurements. This also means that S^2 is now defined in terms of A_i.

[14] Note that $\boldsymbol{\beta}$ can be considered a normal random variable with zero variance.

for $k = 2, \dots, n$. Furthermore, the studentized recursive residuals simplify to

$$w'_k = \frac{\sqrt{\frac{k-1}{k}}\left(A_k - \bar{A}_{k-1}\right)}{\sqrt{\frac{1}{k-2}(k-2)S^2_{k-1}}} = \sqrt{\frac{k-1}{k}}\,\frac{A_k - \bar{A}_{k-1}}{S_{k-1}}$$

for $k = 3, \dots, n$. Then, independence of $T_k = w'_k$ follows from Theorem 5.13. □

Chapter summary

We proved mutual independence for each case of the self-starting statistic and for both the ordinary and studentized recursive residuals. For case IV and the studentized recursive residuals, mutual independence effectively follows from the updating formulae and Lemma 5.6. The other results were proven by showing joint normality and pairwise uncorrelatedness. An interesting observation is that the independence of the case IV self-starting statistic is a special case of the independence of the studentized recursive residuals.

Chapter 6

Example application

In this chapter, as an example, we will apply the techniques discussed in this thesis to a real-world problem, namely the monitoring of wind turbines. This is based on the work by Kenbeek (2016) in the context of the Dutch national project DAISY (Dynamic Asset Information System). Kenbeek (2016) created a model that can be used to predict failures ahead of time and efficiently schedule maintenance. Following Meeuwis (2017) and Van Dalen (2018), our goal is to find a representative benchmark period (i.e., Phase I) which can be used to create a regression model. For this, we will use self-starting control charts and recursive residuals.

Our analysis uses the same dataset as Kenbeek (2016) of a single wind turbine located in the Netherlands. It consists of 110 variables that were measured every 4 minutes from 2013–06–19, 18:32 to 2015–03–18, 23:56. Before we analyze the data, we first filter it based on findings by Kenbeek (2016).

- There was extensive maintenance on 2013–11–16, so we only consider data before that for the benchmark period.
- To decrease autocorrelation, we subsample the data at a frequency of 4 hours.
- We only consider measurements when the main generator is in use (i.e., when the rotor speed exceeds 25.8 RPM).

Then, only 200 measurements remain (each containing values for the 110 variables).

Using all-subset and stepwise selection methods, Kenbeek (2016) found good regression models for the temperature of various components of the wind turbine. We will focus on the first and simplest model, where the *nacelle*[1] *temperature* (response variable) is predicted from the *environment temperature* (explanatory variable). We will reproduce (the first part of) Method 1 of Van Dalen (2018), where the recursive residuals from this model are plotted. When these exceed the lower or upper limit, this indicates an

[1] The housing that contains all other components.

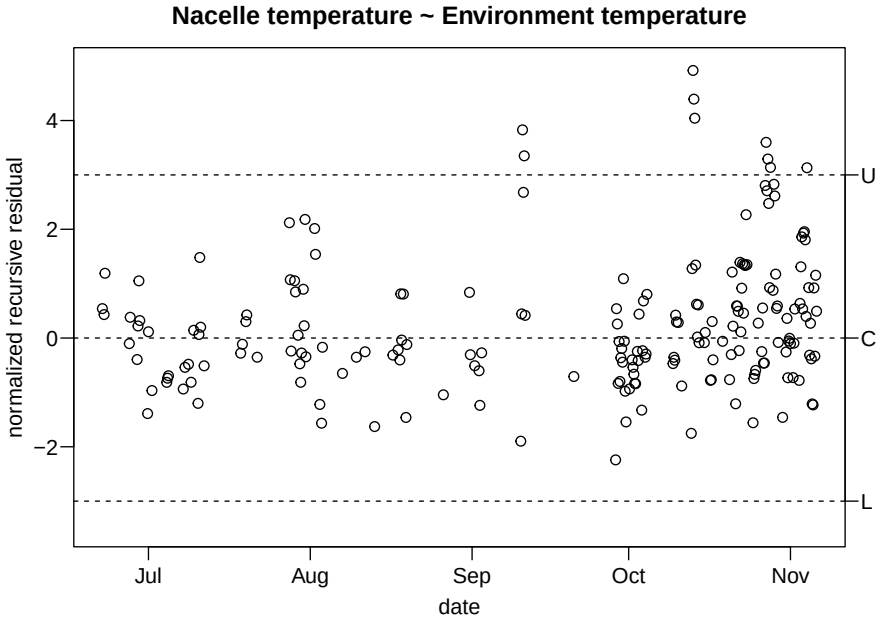

Figure 6.1: The normalized recursive residuals from the model which predicts the nacelle temperature from the environment temperature. The lower control limit, center line, and upper control limit are indicated by dashed lines labeled L, C, and U respectively. Compare this figure to Figure 6.2b of Van Dalen (2018).[3]

out-of-control situation.[2] We can then conclude that the benchmark period should not include the corresponding measurements. However, instead of (ordinary) recursive residuals, we will compute the normalized recursive residuals as defined in Section 4.4.

Recall that control charts are most effective when the measurements are independent. It is important to note that, in this case, there are two potential sources of dependence.

- The underlying data may be dependent. This is certainly true for our application, since the current temperature is strongly correlated with the temperature from 4 minutes before. We decrease this correlation by subsampling the data, although this will not eliminate dependence completely.
- We do not chart the measurements directly, but rather the residuals from a regression model. We use a type of recursive residual which was shown to be mutually independent in Section 5.2, but the same is not true for ordinary or predictive residuals, which were used by Kenbeek (2016).

[2]Van Dalen (2018) set the control limits at ± 3, but this seems to be an oversight. This is because the (ordinary) recursive residuals do not (necessarily) have unit variance. Specifically, in this case, the variance is slightly greater, so the limits should be farther out. Fortunately, the variance is close to 1, so the final conclusion is the same.

[3]As a side note, if the data is not subsampled, Van Dalen (2018) notes that the residuals form clear "layers". This is also the case in Figure 6.1, although it is much less noticeable. Van Dalen (2018) states that these are "inexplicable", but they are likely caused by the fact that the temperatures are integer-valued. As a result, most residuals are close to integers as well.

Now, we will calculate the normalized recursive residuals for the model for the nacelle temperature. These were computed using R with the `strucchange` package. The code, which was adapted from Meeuwis (2017), can be found in Section C.2. The regression chart is shown in Figure 6.1. Since the normalized recursive residuals have a standard normal distribution, we set the control limits at the default ± 3. Because we have 2 coefficients (i.e., the intercept and one for the environment temperature), only $200 - (2 + 1) = 197$ points are plotted.

The limits are first exceeded on 2013–09–10 at 16:00. This might indicate structural change, so this measurement (and all subsequent measurements) will not be included in the benchmark period. Therefore, we take the benchmark period from June 19 to September 9. Coincidentally, this matches the conclusions by Van Dalen (2018), although they are based on much more thorough research. It is more strict than Kenbeek (2016), who took 2013–10–12 as the end of the benchmark period. One flaw of both choices is that the benchmark period takes place almost exclusively in the summer. Therefore, it might not be representative during colder months. More nuanced conclusions are given by Van Dalen (2018).

Chapter summary

We used normalized recursive residuals on real-world data to find an in-control period that can be used for Phase I. For this particular application, it was important to remove the effect of a covariate (i.e., the outside temperature) from the variable we want to monitor (i.e., the nacelle temperature). The normalized recursive residuals are i.i.d. standard normal, which allows them to be plotted on a control chart with limits at ± 3. However, it is important to note that the underlying data may still be correlated. We resolved this by subsampling the data.

Chapter 7

Conclusion

In this final chapter, we will summarize the results of this thesis (Section 7.1) and list some options for future research (Section 7.2).

7.1 Summary of results

The first research question (i.e., "Which proofs exist in the literature on the independence of the statistics used in self-starting control charts, in both the univariate and regression case, or of other related concepts?") was answered in Chapter 2. We found that the statistics used in self-starting control charts, slippage tests, and recursive residuals are all closely related. Consequently, we searched extensively for proofs of independence in the literature on each of these concepts. However, all proofs either show a weaker condition (i.e., serial independence, while our goal is mutual independence) or are, arguably, unclear. Therefore, we set the goal of writing a new proof.

First, however, we established some preliminaries in Chapter 3 and introduced control charts in more detail in Chapter 4. Specifically, we defined the notion of independent random variables and emphasized the difference between mutual and pairwise independence. We also derived two updating formulae that are fundamental to the proofs in Chapter 5. Most importantly, we defined the (ordinary and studentized) recursive residuals and the four cases of the self-starting statistic, which were the central focus of the remainder of the thesis.

In Chapter 5, we answered (the second part of) the other research question (i.e., "Are the existing proofs valid? And if not, how can we correct or extend them?"). First, we amended a proof by Quesenberry (1991) of serial independence of the case IV self-starting statistic. Next, we extended it to mutual independence. Then, we used the same method to prove that the studentized recursive residuals are mutually independent (Theorem 5.13). This is the main contribution of this thesis.

Finally, in Chapter 6, we showed how control charts and recursive residuals can be used in practice. This is based on Kenbeek (2016) and Van Dalen (2018). In short, we replicated a case study by Van Dalen (2018), but instead of the (ordinary) recursive residuals we used the normalized recursive residuals.

7.2 Future research

Finally, we give a number of suggestions for further research.

- Compared to Shewhart charts, (self-starting) regression charts are rather unknown, even though they are a practical choice for many real-world problems. In particular, the studentized recursive residuals are not well-known. This topic is closely related to profile monitoring, on which Woodall et al. (2004) encourage future research.
- In this thesis, we did not closely study the conditional probability integral transform (CPIT), but it is a powerful technique. It would be interesting to verify, in detail, the proof of independence of the self-starting statistics with this transform, or consider new applications of it. A good starting point is Quesenberry (1986a).

Bibliography

L. C. Alwan. Effects of Autocorrelation on Control Chart Performance. *Communications in Statistics – Theory and Methods*, 21(4):1025–1049, Feb. 1992.

V. Barnett and T. Lewis. *Outliers in Statistical Data*. Wiley, 3rd edition, 1994.

M. S. Bartlett. An Inverse Matrix Adjustment Arising in Discriminant Analysis. *Annals of Mathematical Statistics*, 22(1):107–111, Mar. 1951.

D. Basu. On Statistics Independent of a Complete Sufficient Statistic. *Sankhyā*, 15(4): 377–380, Oct. 1955.

N. H. Bingham and J. M. Fry. *Regression: Linear Models in Statistics*. Springer, 2010.

R. L. Brown, J. Durbin, and J. M. Evans. Techniques for Testing the Constancy of Regression Relationships over Time. *Journal of the Royal Statistical Society, Series B*, 37(2):149–192, 1975.

F. Centofanti, A. Lepore, A. Menafoglio, B. Palumbo, and S. Vantini. Functional Regression Control Chart. *Technometrics*, 63(3):281–294, 2021.

E. Del Castillo and D. C. Montgomery. Short-Run Statistical Process Control: *Q*-Chart Enhancements and Alternative Methods. *Quality and Reliability Engineering International*, 10(2):87–97, 1994.

R. Doornbos. *Slippage Tests*. Number 15 in Mathematical Centre Tracts. Mathematisch Centrum, 2nd edition, 1976.

R. Doornbos, H. Kesten, and H. J. Prins. A class of slippage tests. Technical Report S206 (VP8), Mathematisch Centrum, Sept. 1956.

R. W. Farebrother. An Historical Note on Recursive Residuals. *Journal of the Royal Statistical Society, Series B*, 40(3):373–375, 1978.

J. S. Galpin and D. M. Hawkins. The Use of Recursive Residuals in Checking Model Fit in Linear Regression. *The American Statistician*, 38(2):94–105, May 1984.

O. Gaudoin. CPIT Goodness-Of-Fit Tests for Reliability Growth Models. In D. Ionescu and N. Limnios, editors, *Statistical and Probabilistic Models in Reliability*. Birkhäuser, 1999.

D. M. Hawkins. On the Distribution and Power of a Test for a Single Outlier. *South African Statistical Journal*, 3(1):9–15, Jan. 1969.

D. M. Hawkins. Self-starting cusum charts for location and scale. *The Statistician*, 36 (4):299–316, Oct. 1987.

D. M. Hawkins. Diagnostics for Use With Regression Recursive Residuals. *Technometrics*, 33(2):221–234, May 1991.

A. Hedayat and D. S. Robson. Independent Stepwise Residuals for Testing Homoscedasticity. *Journal of the American Statistical Association*, 65(332):1573–1581, Dec. 1970.

W. A. Jensen, L. A. Jones-Farmer, C. W. Champ, and W. H. Woodall. Effects of Parameter Estimation on Control Chart Properties: A Literature Review. *Journal of Quality Technology*, 38(4):349–364, Oct. 2006.

T. A. Kenbeek. Condition Based Prognostics and Diagnostics for Wind Turbines. Bachelor's thesis, Eindhoven University of Technology, 2016.

F. Kianifard and W. H. Swallow. A Review of the Development and Application of Recursive Residuals in Linear Models. *Journal of the American Statistical Association*, 91(433):391–400, Mar. 1996.

K. Kim, M. A. Mahmoud, and W. H. Woodall. On The Monitoring of Linear Profiles. *Journal of Quality Technology*, 35(3):317–328, July 2003.

B. J. Mandel. The Regression Control Chart. *Journal of Quality Technology*, 1(1):1–9, Jan. 1969.

R. L. Marr and C. P. Quesenberry. A NU Test for Serial Correlation of Residuals From One or More Regression Regimes. *Technometrics*, 33(4):441–457, Nov. 1991.

R. Meester. *A Natural Introduction to Probability Theory*. Birkhäuser, 2nd edition, 2008.

L. Meeuwis. Monitoring Wind Turbines with Regression Analysis. Bachelor's thesis, Eindhoven University of Technology, 2017.

D. C. Montgomery. *Introduction to Statistical Quality Control*. Wiley, 8th edition, 2019.

F. Mosteller. A k-Sample Slippage Test for an Extreme Population. *Annals of Mathematical Statistics*, 19(1):58–65, Mar. 1948.

F. J. O'Reilly and C. P. Quesenberry. The Conditional Probability Integral Transformation and Applications to Obtain Composite Chi-Square Goodness-of-Fit Tests. *Annals of Statistics*, 1(1):74–83, Jan. 1973.

E. S. Page. Continuous Inspection Schemes. *Biometrika*, 41(1/2):100–115, June 1954.

R. L. Plackett. Some Theorems in Least Squares. *Annals of Mathematical Statistics*, 37 (1/2):149–157, June 1950.

D. S. G. Pollock. Recursive estimation in econometrics. *Computational Statistics & Data Analysis*, 44(1/2):37–75, 2003.

P. Qiu. *Introduction to Statistical Process Control*. Chapman & Hall, 2013.

C. P. Quesenberry. Some Transformation Methods in Goodness-of-Fit. In R. B. D'Agostino and M. A. Stephens, editors, *Goodness-of-Fit Techniques*. Marcel Dekker, 1986a.

C. P. Quesenberry. Screening Outliers in Normal Process Control Data with Uniform Residuals. *Journal of Quality Technology*, 18(4):226–233, Oct. 1986b.

C. P. Quesenberry. SPC *Q* Charts for Start-Up Processes and Short or Long Runs. *Journal of Quality Technology*, 23(3):213–224, July 1991.

C. P. Quesenberry and H. A. David. Some tests for outliers. *Biometrika*, 48(3/4):379–390, Dec. 1961.

S. W. Roberts. Control Chart Tests Based on Geometric Moving Averages. *Technometrics*, 1(3):239–250, Aug. 1959.

J. P. Romano and A. F. Siegel. *Counterexamples in Probability and Statistics*. Chapman & Hall, 1986.

M. Rosenblatt. Remarks on a Multivariate Transformation. *Annals of Mathematical Statistics*, 23(3):470–472, Sept. 1952.

W. A. Shewhart. *Economic Control of Quality of Manufactured Product*. Van Nostrand, 1931.

H. Theil. *Principles of Econometrics*. Wiley, 1971.

O. Van Dalen. Statistical Monitoring of Wind Turbines. Bachelor's thesis, Eindhoven University of Technology, 2018.

B. P. Welford. Note on a Method for Calculating Corrected Sums of Squares and Products. *Technometrics*, 4(3):419–420, Aug. 1962.

Western Electric. *Statistical Quality Control Handbook*. Mack Printing Company, 2nd edition, 1958.

W. H. Woodall. Controversies and Contradictions in Statistical Process Control. *Journal of Quality Technology*, 32(4):341–350, Oct. 2000.

W. H. Woodall. Bridging the gap between theory and practice in basic statistical process monitoring. *Quality Engineering*, 29(1):2–15, 2017.

W. H. Woodall, D. J. Spitzner, D. C. Montgomery, and S. Gupta. Using Control Charts to Monitor Process and Product Quality Profiles. *Journal of Quality Technology*, 36 (3):309–320, July 2004.

C. Zou, C. Zhou, Z. Wang, and F. Tsung. A Self-Starting Control Chart for Linear Profiles. *Journal of Quality Technology*, 39(4):364–375, Oct. 2007.

Appendix A

Notation

$\mathbf{1}_n$	column vector of n ones
I_n	n-by-n identity matrix
$\mathcal{N}(\mu, \sigma^2)$	normal distribution with mean μ and variance σ^2
$\mathcal{N}(0, 1)$	standard normal distribution
$\chi^2(k)$	chi-squared distribution with k degrees of freedom
$t(k)$	Student's t-distribution with k degrees of freedom
$U(a, b)$	continuous uniform distribution on the interval (a, b)
ϕ	density function of the standard normal distribution
Φ	distribution function of the standard normal distribution
$x \propto y$	x is proportional to y, i.e., $\exists c \in \mathbb{R} \setminus \{0\} : x = cy$
$X \perp\!\!\!\perp Y$	X is independent of Y
$X \sim \mathcal{N}(0, 1)$	X has a standard normal distribution
i.i.d.	mutually independent and identically distributed

Unless stated otherwise, $(a_1, \dots, a_n)$ denotes a row vector and $\mathbf{b} \in \mathbb{R}^n$ a column vector.

As a shorthand, we write $X_1, X_2 \sim \mathcal{N}(0, 1)$ to mean that both X_1 and X_2 have a standard normal distribution.

We define the empty sum as being equal to 0 (i.e., the additive identity), and the empty product to 1 (i.e., the multiplicative identity). For example, $\sum_{n=2}^{1} n = 0$.

Appendix B

Additional preliminaries

Details on Section 3.1

The following is a proof of Theorem 3.2, which is a generalization of, for example, Lemma 5.5.9 of Meester (2008).

Proof of Theorem 3.2. Let f_{X_i} denote the marginal density of X_i. If $X_1, \dots, X_n$ are independent, we simply take $g_i = f_{X_i}$ for each $i = 1, \dots, n$. To prove the converse, we need to show that

$$f(x_1, \dots, x_n) = f_{X_1}(x_1) \cdots f_{X_n}(x_n)$$

given that

$$f(x_1, \dots, x_n) = g_1(x_1) \cdots g_n(x_n) \tag{B.1}$$

for all $x_1, \dots, x_n \in \mathbb{R}$. First, we compute the marginal density of X_1 by integrating over all other X_i, i.e.,

$$\begin{aligned} f_{X_1}(x_1) &= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(x_1, \dots, x_n) \, \mathrm{d}x_2 \cdots \mathrm{d}x_n \\ &= g_1(x_1) \int_{\mathbb{R}} g_2(x_2) \, \mathrm{d}x_2 \cdots \int_{\mathbb{R}} g_n(x_n) \, \mathrm{d}x_n, \end{aligned} \tag{B.2}$$

and similarly for $X_2, \dots, X_n$. Furthermore, by integrating both sides over x_1, we find that

$$\int_{\mathbb{R}} g_1(x_1) \, \mathrm{d}x_1 \cdots \int_{\mathbb{R}} g_n(x_n) \, \mathrm{d}x_n = 1. \tag{B.3}$$

Now, note that

$$\begin{aligned}
&f_{X_1}(x_1)\cdots f_{X_n}(x_n)\\
&\quad= g_1(x_1)\cdots g_n(x_n)\prod_{i=1}^{n}\prod_{\substack{j=1\\ j\neq i}}^{n}\int_{\mathbb{R}} g_j(x_j)\,\mathrm{d}x_j && \text{(by (B.2))}\\
&\quad= g_1(x_1)\cdots g_n(x_n)\left(\prod_{i=1}^{n}\int_{\mathbb{R}} g_i(x_i)\,\mathrm{d}x_i\right)^{n-1}\\
&\quad= g_1(x_1)\cdots g_n(x_n) && \text{(by (B.3))}\\
&\quad= f(x_1,\dots,x_n). && \text{(by (B.1))}
\end{aligned}$$

□

Next, we will fill in the details of Counterexample 3.3. First, we will verify that Z has a standard normal distribution. Using the law of total probability, the distribution function of Z is given by

$$\begin{aligned}
F_Z(z) &= \mathrm{P}(|Z_0|\operatorname{sgn}(XY)\le z)\\
&= \tfrac{1}{2}\cdot\mathrm{P}(|Z_0|\operatorname{sgn}(XY)\le z \mid \operatorname{sgn}(XY)=-1)\\
&\quad + 0\cdot\mathrm{P}(|Z_0|\operatorname{sgn}(XY)\le z \mid \operatorname{sgn}(XY)=0)\\
&\quad + \tfrac{1}{2}\cdot\mathrm{P}(|Z_0|\operatorname{sgn}(XY)\le z \mid \operatorname{sgn}(XY)=1)\\
&= \tfrac{1}{2}\,\mathrm{P}(-|Z_0|\le z) + \tfrac{1}{2}\,\mathrm{P}(|Z_0|\le z)\\
&= \tfrac{1}{2}\Big(\tfrac{1}{2}\,\mathrm{P}(-|Z_0|\le z \mid Z_0\le 0) + \tfrac{1}{2}\,\mathrm{P}(-|Z_0|\le z \mid Z_0>0)\Big)\\
&\quad + \tfrac{1}{2}\Big(\tfrac{1}{2}\,\mathrm{P}(|Z_0|\le z \mid Z_0\le 0) + \tfrac{1}{2}\,\mathrm{P}(|Z_0|\le z \mid Z_0>0)\Big)\\
&= \tfrac{1}{2}\Big(\tfrac{1}{2}\,\mathrm{P}(--Z_0\le z) + \tfrac{1}{2}\,\mathrm{P}(-Z_0\le z)\Big)\\
&\quad + \tfrac{1}{2}\Big(\tfrac{1}{2}\,\mathrm{P}(-Z_0\le z) + \tfrac{1}{2}\,\mathrm{P}(Z_0\le z)\Big)\\
&= \tfrac{1}{2}\,\mathrm{P}(Z_0\le z) + \tfrac{1}{2}\,\mathrm{P}(-Z_0\le z)\\
&= \Phi(z),
\end{aligned}$$

where Φ denotes the CDF of a standard normal distribution. The last step holds by the symmetry of the normal distribution. Now, we will compute the joint distribution

function of X and Z. It holds that

$$\begin{aligned}
F_{X,Z}(x,z) &= \mathrm{P}(X \le x, |Z_0| \operatorname{sgn}(XY) \le z) \\
&= \tfrac{1}{2} \cdot \mathrm{P}(X \le x, |Z_0| \operatorname{sgn}(XY) \le z \mid X < 0) \\
&\quad + 0 \cdot \mathrm{P}(X \le x, |Z_0| \operatorname{sgn}(XY) \le z \mid X = 0) \\
&\quad + \tfrac{1}{2} \cdot \mathrm{P}(X \le x, |Z_0| \operatorname{sgn}(XY) \le z \mid X > 0) \\
&= \tfrac{1}{2} \mathrm{P}(X \le x \mid X < 0) \, \mathrm{P}(|Z_0| \operatorname{sgn}(-Y) \le z) \\
&\quad + \tfrac{1}{2} \mathrm{P}(X \le x \mid X > 0) \, \mathrm{P}(|Z_0| \operatorname{sgn}(Y) \le z) \\
&= \Big(\tfrac{1}{2} \mathrm{P}(X \le x \mid X < 0) + \tfrac{1}{2} \mathrm{P}(X \le x \mid X > 0) \Big) \Phi(z) \\
&= \mathrm{P}(X \le x) \, \Phi(z),
\end{aligned}$$

since $|Z_0| \operatorname{sgn}(Y)$ and $|Z_0| \operatorname{sgn}(-Y)$ both have a standard normal distribution as well. This follows from a derivation very similar to the one above. Note that the joint distribution function factorizes into the marginal distribution functions of X and Z, so they are independent. In an analogous fashion, it can be shown that Y and Z are independent.

Probability distributions

The following definitions are well-known and can be found in any textbook. For example, see pages 32 and 33 of Bingham and Fry (2010).

Definition B.4 — Let $Z_1, \ldots, Z_k$ be i.i.d. standard normal random variables. Then

$$\sum_{i=1}^{k} Z_i^2$$

has a chi-squared distribution with k degrees of freedom.

Definition B.5 — Let Z be a standard normal random variable, and Q a chi-squared random variable with k degrees of freedom. Suppose that $Z \perp\!\!\!\perp Q$. Then

$$\frac{Z}{\sqrt{Q/k}}$$

has a t-distribution with k degrees of freedom.

Transformations of random variables

The following theorem is well-known. For example, it is stated in the second section of Quesenberry (1991).

Theorem B.6 (Probability integral transformation) **—** *Let X be a continuous random variable with distribution function F. Then, the random variable $F(X)$ has a standard uniform distribution.*

Similarly, if Y has a standard uniform distribution and X has a distribution function F, then the random variable $F^{-1}(Y)$ has the same distribution as X.

Theorem B.7 — *Let $\mathbf{X}$ be an n-dimensional continuous random vector with density $f_{\mathbf{X}}$. Then the density of $\mathbf{Y} = g(\mathbf{X})$, where $g : \mathbb{R}^n \to \mathbb{R}^n$ is an invertible, differentiable function, is given by*

$$f_{\mathbf{Y}}(\mathbf{y}) = f_{\mathbf{X}}\big(g^{-1}(\mathbf{y})\big)\,|\det(J)|,$$

where $J \coloneqq \mathbf{J}_{g^{-1}}(\mathbf{y})$ is the Jacobian matrix of g^{-1} evaluated at $\mathbf{y}$.

Proof. See Section 2.2 of Bingham and Fry (2010). □

Linear algebra

Lemma B.8 — *Let $A \in \mathbb{R}^{n\times p}$. For concision, let $\mathbf{a}_1, \dots, \mathbf{a}_n$ denote the rows of A and let*

$$A_\ell = \begin{pmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_\ell \end{pmatrix},$$

i.e., the matrix consisting of the first ℓ rows of A. Then, for $1 < k \leq n$,

$$A_k^\top A_k = A_{k-1}^\top A_{k-1} + \mathbf{a}_k^\top \mathbf{a}_k.$$

Proof. Both sides of the equation are $p \times p$ matrices. We will show they are equal by proving that each entry is equal. First, note that

$$\big(A_k^\top A_k\big)_{ij} = A_{[1:k]i}^\top A_{[1:k]j} = \sum_{\ell=0}^{k} a_{\ell i} a_{\ell j},$$

and similarly for $A_{k-1}^\top A_{k-1}$. Therefore,

$$\big(A_k^\top A_k\big)_{ij} = \sum_{\ell=0}^{k-1} a_{\ell i} a_{\ell j} + a_{ki} a_{kj} = \big(A_{k-1}^\top A_{k-1} + \mathbf{a}_k^\top \mathbf{a}_k\big)_{ij}.$$ □

Although the following formula is named after Sherman and Morrison, in this form it was first published by Bartlett (1951).

Lemma B.9 (Sherman–Morrison formula) — *Let $A \in \mathbb{R}^{p\times p}$ be an invertible matrix and let $u, v \in \mathbb{R}^p$ be row vectors.*[1] *Assume that $1 + vA^{-1}u^\mathsf{T} \neq 0$. Then*

$$\left(A + u^\mathsf{T}v\right)^{-1} = A^{-1} - \frac{A^{-1}u^\mathsf{T}vA^{-1}}{1 + vA^{-1}u^\mathsf{T}}.$$

Proof. It is sufficient to show that the product of $A + u^\mathsf{T}v$ and the above inverse is equal to the identity matrix. Indeed,

$$\begin{aligned}\left(A + u^\mathsf{T}v\right)\left(A^{-1} - \frac{A^{-1}u^\mathsf{T}vA^{-1}}{1 + vA^{-1}u^\mathsf{T}}\right) &= I + u^\mathsf{T}vA^{-1} - \frac{u^\mathsf{T}vA^{-1} + u^\mathsf{T}vA^{-1}u^\mathsf{T}vA^{-1}}{1 + vA^{-1}u^\mathsf{T}} \\ &= I + u^\mathsf{T}vA^{-1} - \frac{u^\mathsf{T}\left(1 + vA^{-1}u^\mathsf{T}\right)vA^{-1}}{1 + vA^{-1}u^\mathsf{T}} \\ &= I + u^\mathsf{T}vA^{-1} - u^\mathsf{T}vA^{-1} \\ &= I.\end{aligned}$$

□

Miscellaneous

The following theorem shows that, in case of normality, the sample mean and variance are independent.

Theorem B.10 — *Let $X_1, \dots, X_n \overset{\text{iid}}{\sim} \mathcal{N}(\mu, \sigma^2)$. Then $S^2 \perp\!\!\!\perp \overline{X}$.*

Proof. Let $A = \left(X_1 - \overline{X}, \dots, X_n - \overline{X}, \overline{X}\right)$. It has a joint normal distribution, since any linear combination of its components (each of which is a linear combination of independent normal random variables) is univariately normal. Since S^2 is a function of $(A_1, \dots, A_n)$, it is sufficient to show that $(A_1, \dots, A_n) \perp\!\!\!\perp A_{n+1}$. Due to joint normality, uncorrelatedness implies independence. Therefore, the claim follows from the fact that

$$\begin{aligned}\operatorname{Cov}\left(\overline{X}, X_i - \overline{X}\right) &= \operatorname{Cov}\left(\overline{X}, X_i\right) - \operatorname{Cov}\left(\overline{X}, \overline{X}\right) \\ &= \operatorname{Cov}\left(\tfrac{1}{n}X_i, X_i\right) - \sum_{i=1}^{n} \operatorname{Cov}\left(\tfrac{1}{n}X_i, \tfrac{1}{n}X_i\right) \\ &= \frac{1}{n}\sigma^2 - \frac{n}{n^2}\sigma^2 \\ &= 0\end{aligned}$$

for all $i = 1, \dots, n$. All cross terms disappear due to independence, i.e., $\operatorname{Cov}\left(X_i, X_j\right) = 0$ for all $i \neq j$. □

[1] Typically, column vectors are used, but our application naturally requires row vectors.

In the following lemmas, Γ denotes the well-known gamma function.

Lemma B.11 — *Let $c, z > 0$. Then,*

$$\int_0^\infty x^{z-1} e^{-cx} \, \mathrm{d}x = c^{-z} \, \Gamma(z).$$

Proof. Substituting $u \coloneqq cx$,

$$\begin{aligned} \int_0^\infty x^{z-1} e^{-cx} \, \mathrm{d}x &= \int_0^\infty \left(\frac{u}{c}\right)^{z-1} e^{-u} \cdot \frac{1}{c} \, \mathrm{d}u \\ &= c^{-z} \int_0^\infty u^{z-1} e^{-u} \, \mathrm{d}u \\ &= c^{-z} \, \Gamma(z). \end{aligned}$$

□

Lemma B.12 — *Let $c, n > 0$ and $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = |x|^{n-1} e^{-cx^2}$. Then,*

$$\int_{-\infty}^\infty f(x) \, \mathrm{d}x = c^{-n/2} \, \Gamma(n/2).$$

Proof. We compute that

$$\begin{aligned} \int_{-\infty}^\infty f(x) \, \mathrm{d}x &= 2 \int_0^\infty x^{n-1} e^{-cx^2} \, \mathrm{d}x && (f \text{ is even}) \\ &= \int_0^\infty (x^2)^{n/2-1} e^{-cx^2} \cdot 2x \, \mathrm{d}x \\ &= \int_0^\infty \left(\frac{u}{c}\right)^{n/2-1} e^{-u} \cdot \frac{1}{c} \, \mathrm{d}u && (\text{substitute } u \coloneqq cx^2) \\ &= c^{-n/2} \int_0^\infty u^{n/2-1} e^{-u} \, \mathrm{d}u \\ &= c^{-n/2} \, \Gamma(n/2). \end{aligned}$$

□

Appendix C

Code

C.1 Simulation of correlation

```
simsize = 1000;
mean = 5;
sd = 2;
n = 50;
X = RandomVariate[NormalDistribution[mean, sd], {simsize, n}];

transform[x_, k_] := (x[[k]] - Mean[Take[x, k-1]]) / StandardDeviation[Take[x, k-1]]
T = Table[transform[X[[i]], k], {k, 3, n}, {i, 1, simsize}];
cor = Table[Correlation[T[[i]], T[[j]]], {i, 1, Length[T]}, {j, 1, i-1}] // Flatten;

ListPlot[cor]
```

C.2 Example application

The following is based on code by Meeuwis (2017). A more extensive implementation can be found in Appendix D of Van Dalen (2018).

```
library(strucchange)

# read and process data
df_raw <- read.csv2('wind-turbine.txt', stringsAsFactors=FALSE)
df_raw[,-1] <- apply(df_raw[,-1], 2, as.numeric)
df_raw$Power[df_raw$Power == 0.1] <- NA  # fix wrong entries
df_raw$Time <- as.POSIXct(strptime(df_raw$Taken, '%d/%m/%Y %T', tz='EST'))

# filter: turbine is running (as opposed to 'emergency' or 'pause')
df <- df_raw[which(df_raw$OpState == 3),]

# filter: main generator is in use
df <- df[which(df$RotorSpeed > 25.8),]

# filter: before major maintenance
df <- df[which(df$Time < '2013-11-16 EST'),]

# filter: subsample every 4 hours to decrease autocorrelation
seq_4h <- seq(min(df$Time), max(df$Time), by='4 hours')
```

```
df <- df[which(is.element(df$Time, seq_4h)),]
rm(seq_4h)

# check that data contains no NA
any(is.na(df$EnvTemp))
any(is.na(df$NacelleTemp))


# For testing:
#   df <- data.frame(EnvTemp=rnorm(200, 4, 3))
#   df$NacelleTemp <- df$EnvTemp + rnorm(nrow(df), 0, 2)
#   df$Time <- 1:nrow(df)

p <- 2

res <- recresid(NacelleTemp ~ EnvTemp, df)
res.var <- cumsum(res^2)
res.var <- res.var / 1:length(res)
res.var <- sqrt(res.var)
res <- res[-1] / res.var[-length(res)]
res <- pt(res, 1:length(res))
res <- qnorm(res)
rm(res.var)

par(
    las=1,
    mgp=c(1.8, .6, 0),
    mar=c(3.25, 3.25, 2.5, 2)
)
plot(
    df$Time[(p+2):nrow(df)], res,
    main='Nacelle temperature ~ Environment temperature',
    xlab='date', ylab='normalized recursive residual',
    ylim=c(-3.5, 5)
)
abline(h=-3, lty='dashed')
abline(h= 0, lty='dashed')
abline(h= 3, lty='dashed')
axis(side=4, at=c(-3, 0, 3), labels=c('L', 'C', 'U'))

# find out-of-control dates
df[which(abs(res) > 3) + p+1,]$Time
res[abs(res) > 3]
```