Maintenance Is Unjustifiable An Improved Inference



MAINTENANCE IS UNJUSTIFIABLE; AN IMPROVED INFERENCE

C. Rijsdijk

Graduation Committee:

| Chairman and secretary: | |
|-------------------------------|--------------------------------|
| Prof.dr. G.P.M.R. Dewulf | University of Twente |
| Supervisor: | |
| Prof.dr.ir. T. Tinga | University of Twente |
| Co-supervisor: | |
| Dr. J.G.M. Heerkens | University of Twente |
| Members: | |
| Prof.dr. H.A. Akkermans | Tilburg University |
| Prof.dr.ir. A. de Boer | University of Twente |
| Prof.dr. R. Curran | Delft University of Technology |
| Prof.dr.ir. L.A.M. van Dongen | University of Twente |
| Prof.dr. U. Kumar | Luleå University of Technology |
| | |

Maintenance is unjustifiable; an improved inference Rijsdijk, Christiaan PhD thesis, University of Twente, Enschede, The Netherlands November 2016

ISBN: 978-90-365-4190-9 DOI: 10.3990/1.9789036541909 http://dx.doi.org/10.3990/1.9789036541909

Printed by Gildeprint, Enschede, The Netherlands Cover design: P.J. de Vries, Bureau Multimedia, NLDA

Copyright $\ensuremath{\mathbb{C}}$ 2016 C. Rijsdijk, Middelburg, The Netherlands. All rights reserved.

MAINTENANCE IS UNJUSTIFIABLE; AN IMPROVED INFERENCE

DISSERTATION

to obtain the degree of doctor at the University of Twente, on the authority of the rector magnificus, prof.dr. H. Brinksma, on account of the decision of the graduation committee, to be publicly defended on Wednesday the 16th of November 2016 at 12:45

by

Christiaan Rijsdijk born on 17 December 1970 in Woensdrecht This dissertation is approved by:

Prof.dr.ir. T. Tinga (supervisor) Dr. J.G.M. Heerkens (co-supervisor)

Abstract

Research motives

In general, decisions appear to be encumbered by subjectivity which is problematic for their validation. In this work, however, we do aim for a validation of decisions. A maintenance policy may seem to be a suitable means of validation because it triggers decisions at a high rate and because the abundant policy violations are typically also recorded. These policy violations may therefore give a glimpse into the counterfactual reality that maintenance policy compliance intends to avoid in the first place. This work demonstrates the feasibility of this unconventional approach to maintenance policy validations. It would be naïve to expect a decisive maintenance policy validation, but at least we purport to improve the justifiability of maintenance.

Approach

We take the viewpoint that inference precision follows from the choice of an argument, an operationalisation and a sampling procedure. We develop a number of candidate arguments and samples. Our iterative journey along these candidates leads to an improved inference.

Our contribution

- We have implemented a maintenance policy validation by a causal argument and a sample from a realistic case study at an improved inference precision;
- We have implemented a maintenance policy validation that relies on evidence about policy violations that from a normative decision theoretical perspective appears to be new;
- We have implemented alternatives for conventional maintenance performance indicators that enable more precise causal inferences in the case study.

Conclusion

To validate a maintenance policy by the proposed approach is very difficult if the only evidence available is from an organisation's recording routines. Therefore, an explicit justification of maintenance cannot easily be obtained. However, the proposed approach showed how to improve the associated inference precision in the specific case study.

Practical implications

This work reveals that conventional maintenance performance indicators are typically insufficient for capturing the variations which will allow us to learn about the system behaviour. We propose and implement some construction rules for maintenance performance indicators that enable us to reveal prima facie causalities from recording routines. Although these construction rules appear to be straightforwardly implementable, they are often violated in the practice of maintenance performance measurement. We therefore argue that organisations could potentially enhance support of their maintenance policy assessments through recording routines; possibly by validating some formal argument, as we do in this work, or else by simply asking: "Where does this peak come from?".

Samenvatting

Onderzoeksmotieven

Besluitvorming is subjectief waardoor een validatie lastig is. Dit werk poogt desondanks besluiten te valideren. Een onderhoudsbeleid kan een geschikte casus voor een dergelijke validatie blijken omdat het besluiten met een hoge frequentie genereert terwijl de veelvuldig voorkomende beleidsovertredingen eveneens worden geregistreerd. Deze beleidsovertredingen kunnen ons toegang geven tot een realiteit die men had willen vermijden met het onderhoudsbeleid. Dit onderzoek kan de praktische toepasbaarheid aantonen van deze onconventionele manier om een onderhoudsbeleid te valideren. Een onweerlegbare validatie van een onderhoudsbeleid is niet te verwachten, maar we kunnen op zijn minst proberen onderhoud preciezer te valideren.

Aanpak

We stellen dat de precisie van een wetenschappelijke redenering volgt uit de keuze voor een argument, een kwantificering en een steekproeftrekking. We ontwikkelen een aantal opties voor het argument en de steekproef. Onze iteratieve zoektocht leidt tot een verbeterde precisie.

Onze bijdrage

- We hebben op basis van de registratieroutines in een bepaalde casus een onderhoudsbeleid nauwkeuriger gevalideerd met een causaal argument;
- We hebben op basis van beleidsovertredingen in een bepaalde casus een onderhoudsbeleid gevalideerd op een wijze die vanuit het perspectief van de besluitvormingstheorie vernieuwend lijkt;
- We hebben alternatieven voor conventionele onderhoudsprestatie-indicatoren toegepast waarmee causale verbanden beter te herkennen zijn.

Conclusie

Het valideren van een onderhoudsbeleid met behulp van registratieroutines is erg lastig volgens de voorgestelde aanpak. Het belang van onderhoud is daarom niet eenvoudig te expliciteren in de praktijk. Echter, in een bepaalde casus hebben we met de voorgestelde aanpak laten zien hoe de precisie van de validatie aan te scherpen is.

Praktische toepasbaarheid

Dit onderzoek toont aan dat de procesvariaties om het systeemgedrag te leren kennen vaak worden uitgemiddeld in conventionele onderhoudsprestatie-indicatoren. Onze constructieregels voor onderhoudsprestatie-indicatoren stellen ons echter in staat prima facie causaliteit te herkennen in registratieroutines. Ondanks het feit dat deze constructieregels gemakkelijk te implementeren zijn, worden ze in de onderhoudspraktijk zelden toegepast. We stellen daarom dat organisaties hun onderhoudsbeleid potentieel beter kunnen ondersteunen met registratieroutines; hetzij door de validatie van een formeel argument, zoals wij doen in dit onderzoek, of door gewoon te vragen: "Waar komt die piek vandaan?".

Acknowledgement

The birth of this work came as a personal surprise. I could not have guessed that a vague interest in philosophy would start taking up most of my weekends and holidays for many a year. I initially just tried projecting a few philosophical ideas onto the maintenance context I was familiar with. It was only gradually that I started feeling increasingly uneasy about the maintainer's dogma that the importance of maintenance is insufficiently recognised. Why would outsiders refuse to recognise the truth?

Over the years, my perception of this problem has wildly altered. For sure, this work is only an intermediate result that is subject to critique. Many people were sceptical about the applicability of this work. Indeed, I was just looking for the observable effects of maintenance and not for a means to improve it, whereas the latter better appeals to the articulated needs of practitioners. However, correspondence with reality is no trivial matter for empirical scientists and practitioners alike.

For sure, people who prospectively intend to improve the world through a maintenance policy are definitely serving societal needs in a better way. The practical importance of revealing the truth in retrospect is questionable. In the past, some people even greatly suffered in their pursuit of truth rather than surrendering to conventional beliefs. I can only say that I am fortunate not having been hampered in my attempts. On the contrary, many discussions with practitioners and fellow researchers greatly helped me in understanding the modest practical applicability of my endeavour. Compelling conclusions do not follow from questionable evidence. Recording routines are typically poorly used to support decisions. Recently, I have been given several opportunities to apply some of my findings to practical cases of decision support.

Many of my part-time students have greatly supported me by reflecting on my naïve ideas. They also provided me with the indispensable data that enabled me to try various arguments at a very early stage. This allowed me to receive fast feedback on the feasibility of my choices. I would never have been able to conceive this research whilst sitting at my desk without this practical guidance. I particularly owe much to André Cornelissen and Omar Lo who provided me with the in-depth information from which I constructed the case study.

Although I entered into the spirit of the research by working as a research hobbyist, it nevertheless did affect my professional life. I had to isolate myself in order to understand and concentrate on the findings of a few brilliant authors. However, urgent matters do not go away by simply ignoring them. As a result, many of my colleagues probably suffered because of my absent-mindedness. I cannot therefore begin to know how much I owe to my close colleagues Erik van der Lichte and Corné Dirne.

As time went on, my perception of the problem radically changed. In retrospect, my quixotism must have brought a smile to many. Ricky Curran allowed me to freely explore the universe while gradually injecting my belief with doubts that were

essential to proceed. My less subtle clashes with Jezdimir Knezevic helped me to elicit the practical applicability of my attempts to observe the effects of maintenance. Finally, Enrico Zio's words "to work on the theoretical aspect of decision" were a tremendous help in arriving at a more precise articulation of the problems with a maintenance policy validation.

Although the idea of a joint PhD never materialised, Flip Wubben introduced me to the Netherlands Defence Academy where I could work on some practical cases related to my research. The justifiability of maintenance is of course not a primary concern of a defence organisation but I am very grateful for the opportunity to test some principles of data driven decision support that I postulated in this work. Participants in the Tools4LCM project like Nick Heerink, Pieter Jansen, Rob Konings, Peet Roovers, Mark Schraven and Sirp-Jan Werkman were a great support in this reality check.

Tiedo Tinga is far more familiar with diagnostics and prognostics than I am. Our discussions revealed that making a choice for an inference follows from the available evidence. The evidence for a maintenance policy validation typically differs from the evidence for diagnostics or prognostics. Still, Tiedo was open to participate by carefully redirecting my texts and by raising critical questions.

At some stage, any research should yield some deliverable although it is not entirely completed. It appeared hard to merge the open mindedness required to explore unresolved issues with the operational pressure to deliver in time. Therefore, I could have benefitted more from comments from Hans Heerkens, the committee members and the various anonymous reviewers that were only involved at a very late stage.

All in all, this endeavour turned out to be less lonely than anticipated.

Middelburg, 2016. Chris Rijsdijk.

Table of contents

| Abstract | i |
|--|-------------------|
| Acknowledgement | iii |
| List of symbols | viii |
| List of figures | X |
| List of tables | xi |
| 1. Introduction | 1 |
| 1.1 Problem statement | 1 |
| 1.2 Research question | 3 |
| 1.3 Aim | 3 |
| 1.4 Approach | 5 |
| 1.4.1 Introduction to the inference objectives | 6 |
| 1.4.2 Choice of an argument | 7 |
| 1.4.3 Choice of an operationalisation | 7 |
| 1.4.4 Choice of a sampling procedure | 9 |
| 1.5 Outline | 9 |
| 2 Preliminary to inference | 11 |
| 2.1 Introduction to scientific inference | 11 |
| 2.2 Assessment of inference precision | 13 |
| 2.3 Introduction to causal inference | 17 |
| 2.3.1 From association to causality | 17 |
| 2.3.2 Non-causality assumptions | 20 |
| 2.3.3 Prima facie causality | 20 |
| 2.3.4 Controversy about causality | 24 |
| 3 Literature review | 27 |
| 3.1 Review of normative decision theory | 28 |
| 3.2 Review of maintenance performance measurement | 34 |
| 3.2.1 Avoid redundancy in performance indicators | |
| 3.2.2 Sample at a rate that allows reconstruction of the original signal | |
| 3.2.3 Balance completeness with efficiency. | |
| 3.3 Review of maintenance policy assessments | |
| 3 3 1 Ontimise maintenance | 40 |
| 3.3.2 Satisfice maintenance | 41 |
| 3 3 3 Knowledge about non-causality assumptions | 42 |
| 3.3.4 Reliability centred maintenance process | 44 |
| 3.4 Review of diagnostics | 47 |
| 3.4.1 Quantitative model based diagnostics | 48 |
| 3.4.2 Qualitative model based diagnostics | 48 |
| 3.4.3 History based diagnostics | <u>49</u> |
| 3.5 Review of prognostics | رب ۱۹ |
| 3.6 Findings regarding the approach | 51 |
| 3.6.1 Findings regarding the choice of an argument | 51 |
| 3.6.2 Findings regarding the choice of an operationalisation | 52 |
| 3.6.3 Findings regarding the choice of a sampling procedure | 52 |
| A Choice of an argument | |
| 4.1 Maintenance ontimisation argument | <i>55</i> |
| 4.11 Claim of the argument | |
| \neg .1.1 Claim Of the argument | |

| | 4.1.2 | Claim about prima facie causality | 59 |
|---|-----------------|--|-------------|
| | 4.1.3 | Sampling issues | 59 |
| | 4.2 M | Iaintenance prognostic argument | 59 |
| | 4.2.1 | Claim of the argument | 60 |
| | 4.2.2 | Claim about prima facie causality | 61 |
| | 4.2.3 | Sampling issues | 62 |
| | 4.3 R | eliability engineering argument | 63 |
| | 4.3.1 | Claim of the argument | 63 |
| | 4.3.2 | Claim about prima facie causality | 65 |
| | 4.3.3 | Sampling issues | 66 |
| | 4.4 N | onparametric argument | 67 |
| | 4.4.1 | Claim of the argument | 67 |
| | 4.4.2 | Claim about prima facie causality | 68 |
| | 4.4.3 | Sampling issues | 70 |
| | 4.5 K | Eview of the arguments on inference precision | /1 |
| | 4.5.1 | Findings regarding the "valid argument" inference objective | 12 |
| | 4.5.2 | Findings regarding the "functional relation" inference objective | 12 |
| | 4.5.5 | Findings regarding the "decideble argument" inference objective | /3 |
| | 4.5.4 | Findings regarding the choice of a compling proceedure | /3 |
| | 4.3.3 | Findings regarding the cloire of a sampling procedure | /4 |
| 5 | 4.3.0 Implor | regarding the craim about prima facte causanty | כו רד |
| 5 | 5 1 C | hoice of an operationalisation | ייייי דד |
| | 5.1 C | Common sense about functionality | // |
| | 512 | Common sense about maintenance policy compliance | 80 |
| | 513 | Findings regarding the "common sense evidence" objective | 00 |
| | 52 C | hoice of a sampling procedure | |
| | 5.2.1 | Alternative for functionality at an increased sampling rate | 85 |
| | 5.2.2 | Alternative for policy compliance at an increased sampling rate | 87 |
| | 5.2.3 | Dichotomous alternative for functionality | 89 |
| | 5.2.4 | Dichotomous alternative for maintenance policy compliance | 91 |
| | 5.2.5 | Findings regarding the choice of a sampling procedure | 92 |
| | 5.3 V | alidation of the arguments | 93 |
| | 5.3.1 | Validation of the maintenance optimisation argument | 93 |
| | 5.3.2 | Validation of the maintenance prognostic argument | 95 |
| | 5.3.3 | Validation of the reliability engineering argument | .107 |
| | 5.3.4 | Validation of the nonparametric argument | .111 |
| | 5.3.5 | Findings regarding the validation of the arguments | .116 |
| | 5.4 A | n improved inference | .119 |
| | 5.4.1 | Weakened maintenance prognostic argument | .119 |
| | 5.4.2 | Extended nonparametric argument | .122 |
| | 5.4.3 | Reduced nonparametric argument | .126 |
| | 5.5 F | indings regarding the "universal argument" inference objective | .131 |
| | 5.5.1 | Background variables influencing functionality | .131 |
| | 5.5.2 | Background variables influencing maintenance policy compliance | .132 |
| _ | 5.5.3 | Split sample validation | .132 |
| 6 | Discus | sion | .137 |
| | 6.1 C | hoice of an argument | .137 |
| | 6.1.1 | Main findings | .137 |

| 6.1.2 | Our contribution regarding the argument selection | |
|--------------|---|-----|
| 6.2 C | hoice of an operationalisation | 140 |
| 6.2.1 | Main findings | 140 |
| 6.2.2 | Our contribution regarding the operationalisation | 141 |
| 6.3 C | hoice of a sampling procedure | 142 |
| 6.3.1 | Main findings | |
| 6.3.2 | Our contribution regarding the sampling procedure | |
| 7 Conclu | usion | |
| 7.1 N | Iaintenance is unjustifiable; an improved inference | |
| 7.2 P | ractical implications | |
| 7.3 F | urther research | |
| 7.3.1 | Predictive maintenance performance | 147 |
| 7.3.2 | Data driven decision support | |
| References | * * | |
| List of publ | ications | |
| | | |

List of symbols

| $ \begin{array}{l} & \leftrightarrow \\ & \leftrightarrow \\ & \rightarrow \\ \mathcal{L} \end{array} $ | Independence Non (prima facie) causality (prima facie) causality Likelihood |
|---|---|
| α | Tentative probability of a prospective outcome (of a decision) |
| 3 | Prediction error |
| θ | Parameters in a body of knowledge that comprises approximating models |
| θ_0 | True best approximating parameters θ |
| θ_{mle} | Maximum likelihood estimation of the parameters θ for some given sample |
| λ | Arrival rate in Little's Law and in the maintenance optimisation argument |
| Ω_{T} | All information available in the universe up to a time T |
| AIC | Akaike's information criterion |
| AIC | Akaike's corrected information criterion |
| AT | Point availability: the probability of finding an item in upstate at a time T |
| - | under given conditions. |
| В | Background variable that is beyond an information set V |
| С | Maintenance resource costs |
| D/d | Queue of delayed maintenance; to be seen as an instance of L/ Value of D |
| do(.) | Do operator (Pearl, 2000) to distinguish a 'within subject dependence' from an 'across subject association' |
| E[.] | Expected value of [.] |
| h(t) | Hazard rate |
| I(g,f) | Kullback-Leibler information; The amount of information lost when |
| | approximating the true probability g(.) with some probability function f(.) |
| K/k | General representation of functionality/ Value of K |
| | Number of estimable parameters in AIC/AIC _c |
| | Sum of K_U+K_U in a hypergeometric distribution |
| Ku | Observed frequency of "wins" among the set of N _U cases |
| MTBF | Mean Time Between Failure |
| MTTR | Mean Time To Repair |
| MTTS | Mean Time To Support |
| L/1 | General representation of maintenance policy compliance/ Value of L |
| L ^x | Leading maintenance performance indicator; to be seen as a component of L |
| LR | Log likelihood ratio |
| N/n | Sample size/ Value of n |
| | Cumulative number of arrivals in Little's Law |
| | Sum of N _U +N _{U'} in a hypergeometric distribution |
| Nu | Observed frequency of identical U in a sample |
| P _{MLE} /p _{mle} | Maximum likelihood estimator P_{U} value of P_{MLE} |
| Pr(.) | Probability of an event (.) |

| pr _Y (0) | Probability function that expresses the probability that a variable Y takes a value 0. |
|---------------------|--|
| P_U/p_u | Bernoulli parameter, given a body of knowledge U/ value of P _U |
| Q/q | Output; to be seen as an instance of K/ Value of Q |
| R _[1,T] | Reliability; the probability of retaining an upstate over a time interval [1,T] under given conditions |
| S/s | Dichotomous maintenance policy compliance variable; to be seen as an |
| | instance of L/ Value of S |
| T/t | Time/ Value of T |
| U | Body of knowledge; to be conceived as some subset of V |
| $U_L(C,K)$ | Utility function of a maintenance policy L that is built on resource costs C and |
| | functionality K |
| V | Information set; |
| W/w | Waiting time or lead time in Little's Law/ value of W |
| X/x | Dichotomous variable that identifies queue membership/ Value of X |
| Y/y | Dichotomous functionality variable; to be seen as an instance of K/ Value of Y |

List of figures

| Figure 1 Discrete choice representation of a decision to carry out maintenance | 4 |
|---|-----|
| Figure 2 Example of a deductive argument with a functional model | 12 |
| Figure 3 Example of a deductive argument with a relational model | 14 |
| Figure 4 Example of an argument that deduces a probability | 15 |
| Figure 5 Inference of a spurious causality | 18 |
| Figure 6 Path graph of non-causality assumptions | 20 |
| Figure 7 Event tree of all possible values in an information set $V = \{l_t, k_t, k_{t+1}\}$ | 23 |
| Figure 8 Path graph of non-causality assumptions in an extended information set | 24 |
| Figure 9 Path graphs of non-causality assumptions in some tentative universes | 43 |
| Figure 10 Technical feasibility rule of time and condition based maintenance | 45 |
| Figure 11 Path graphs of the candidate arguments | 56 |
| Figure 12 Maintenance optimisation argument | 58 |
| Figure 13 Maintenance prognostic argument | 60 |
| Figure 14 Reliability engineering argument | 64 |
| Figure 15 Nonparametric argument | 68 |
| Figure 16 Functionality captured at a daily and a monthly sampling rate | 78 |
| Figure 17 Cumulative distribution of daily output | 79 |
| Figure 18 Queue of delayed maintenance at a daily and a monthly sampling rate | 81 |
| Figure 19 Initial queue and reduced queues of delays | 82 |
| Figure 20 Time series of the case organisation's performance indicators (d,y)[1,m] | 85 |
| Figure 21 Time series of the increased sampling rate alternative (d,q) _[1,t] | 85 |
| Figure 22 Time series of the dichotomous alternative (s,y) _[1,t] | 85 |
| Figure 23 Scatter plots of the one-step-ahead dependence in Q_T and Y_M | 86 |
| Figure 24 Observed frequency of functionalities Q _T and Y _M | 86 |
| Figure 25 Scatter plot of the one-step-ahead dependence in D_T and D_M | 87 |
| Figure 26 Cumulative distribution of the overdue times with a functionality risk | 88 |
| Figure 27 Observed frequency of maintenance policy compliance D_T and D_M | 88 |
| Figure 28 Continuous Q _T and categorical Y _T representations of functionality | 90 |
| Figure 29 Discrete D _T and categorical S _T representations of policy compliance | 91 |
| Figure 30 Path graph of the weakened maintenance prognostic argument | 120 |
| Figure 31 Path graph of the extended nonparametric argument | 123 |
| Figure 32 Prospective uptime, given {s _t ,y _t } | 124 |
| Figure 33 Prospective downtime, given $\{s_t, y_t\}$ | 125 |
| Figure 34 Path graph of the reduced nonparametric argument | 126 |
| Figure 35 K_U/N_U , K/N and the p-value of presumed independence for $(d,y)_{[1,m]}$ | 128 |
| Figure 36 K _U /N _U , K/N and the p-value of presumed independence for $(d,q)_{[1,t]}$ | 129 |
| Figure 37 Prospective uptime (left) and downtime (right) given $\{s_t\}$ | 130 |
| Figure 38 Split sample validation for remaining uptime | 133 |
| Figure 39 Split sample validation for remaining downtime | 133 |
| Figure 40 Split sample validation for stationarity | 134 |

List of tables

| Table 1 Simplified representation of a maintenance scorecard | 4 |
|---|-----|
| Table 2 Survey of choices and inference objectives | 5 |
| Table 3 Inference precision of three tentative arguments | 16 |
| Table 4 Instance of a maintenance scorecard | 35 |
| Table 5 Preliminary inference precision of the candidate arguments | 71 |
| Table 6 Number of possible replications in V | 92 |
| Table 7 Observed frequencies of replications in the case study | 93 |
| Table 8 Relative information loss of some models M3 for (d,y)[1,70] | 100 |
| Table 9 Relative information loss of some models M3 for (d,q)[1,1977] | 101 |
| Table 10 Relative information loss of some models $Pr(Q_T D_T,\theta)$ for $(d,q)_{[1,1977]}$ | 102 |
| Table 11 All possible Bernoulli models M3 and their prima facie causal claim | 103 |
| Table 12 Relative information loss of all Bernoulli models M3 for (s,y)[1,1976] | 106 |
| Table 13 Relative information loss of some models M3 for (d,y)[1,70] | 108 |
| Table 14 Relative information loss of some models M3 for (d,q)[1,1977] | 109 |
| Table 15 Relative information loss of some models $Pr(Q_T D_T,\theta)$ for $(d,q)_{[1,1977]}$ | 110 |
| Table 16 P-values for the sample (s,y) _[1,1976] , given P5, P7 of the NPA | 116 |
| Table 17 Main findings regarding the candidate arguments and samples | 117 |
| Table 18 Relative information loss of all Bernoulli models M3 Y _T =0 | 120 |
| Table 19 Relative information loss of all Bernoulli models M3 Y _T =1 | 121 |
| Table 20 Contingency table of $V = \{s_t, y_t, y_{[t+1,t+x]}\}$ | 124 |
| Table 21 Definite inference precision of the candidate arguments | 138 |

1. Introduction

1.1 Problem statement

For thousands of years, mankind has been developing technology for all kinds of purposes. Gradually, technology has become an indispensable part of our daily life. We almost seem to forget how much we rely on properly functioning technology. But occasionally, we are confronted with failures that disrupt the course of our life. Bit by bit, we learnt that properly functioning technology requires the continuous effort which we now call maintenance.

Until the Second World War, decisions to carry out maintenance were often made on an ad hoc basis. Evident failures were restored and preventive maintenance was often only justified by some ambiguous notion that "grease is cheaper than steel". Our acceptance of failures diminished during the 20th century and the appropriateness of coincidental maintenance became subject to increasing doubts. We developed procedures like reliability centred maintenance (Nowlan & Heap, 1978), (Moubray, 2004) for maintenance policy assessments that established decision rules for time based maintenance, condition based maintenance, corrective maintenance and modifications. The maintenance policy that includes all these decision rules can exert influence on a future yet to be observed, but it cannot manipulate the past. Maintenance policy assessments therefore typically apply some modus ponens reasoning about the future from antecedents about the past. So, a maintenance policy assessment *prospectively* reasons about an unobserved future, whereas we will try to justify it *retrospectively*. Maintenance policy assessments predominantly rely on expert judgement about the prospective future. However, we take the viewpoint that this prospective future should materialise to retain maintenance policy assessments as meaningful to practitioners and empirical scientists.

In this work, we will not develop another framework for a maintenance policy assessment. We simply depart from decisions to carry out maintenance as they occurred, irrespective of the maintenance policy that triggered them. Decisions are choices for actions and maintenance actions differ from other actions by their *intention* to contribute to an item's state "in which it can perform a required function" as stated in the following definition of maintenance:

The combination of all technical and administrative actions, including supervision actions, intended to retain an item in, or restore it to, a state in which it can perform a required function (CEN, 2001), (IEC, 1990).

In this work, we wonder whether this intention of maintenance is observable in retrospect. In other words, did decisions to carry out maintenance contribute to functionality in an observable way? We seek for an inference that gives a better answer to this question than the subjective answers currently available.

About 4% of The Netherlands Gross Domestic Product is annually spent on professional maintenance (NVDO, 2011). However, it seems common sense that maintenance should not be justified by its resource costs. Rather, it should be justified by its contribution to functionality. Is the maintenance sector capable of effectively justifying its existence by its contribution to functionality? Possibly, we could be more precise here.

If functionality were entirely random, like tossing a fair dice, we would abandon any hope for control. Praying would then be as effective as applying a maintenance policy to achieve functionality. The proposition that functionality could be uncontrollable appears to be counterintuitive, but observing this functionality effect through recording routines has, to the best of our knowledge, been ignored up until now. Maintenance optimisation methods often presume that functionality is controllable by a maintenance policy, but they often fail to validate this dependence. Deducing optima from presumptions is not necessarily bad, but should we continue to ignore empirical validation? We pose that correspondence with reality is essential to retain maintenance policy assessments as a scientific discipline *and* as meaningful to practitioners. We should not resort to a maintenance sector that regards its contribution to functionality as though it were some kind of metaphysical belief. We therefore seek a maintenance policy validation.

The justifiability of maintenance relies on the presumption that uncertainty about functionality is epistemic. Epistemic uncertainty is due to things we could know in principle but not in practice. So, we can effectively reduce epistemic uncertainty by increasing our knowledge. We are unaware of attempts to observably reduce uncertainty about functionality through knowledge about a maintenance policy. Prognostic methods aim to reduce epistemic uncertainty about a remaining life, but they conventionally reason from physical variables rather than from the applied maintenance policy. We suspect that the justifiability of maintenance may not follow directly from some well-explored prognostic method:

- Firstly, because prognostics may already effectively reason from associated symptoms whereas we require a causal functionality effect of a maintenance policy. We therefore anticipate that we need to operationalise a notion of causality.
- Secondly, because common sense about operationalising some physical variables typically exceeds common sense about operationalising a maintenance policy. This common sense is essential for any argument to be compelling for reality. Otherwise the argument remains some abstract formalism.
- Thirdly, because common sense about physical laws typically exceeds common sense about models for man-machine interactions. We therefore anticipate that we lack in-depth knowledge about a "law" that relates a maintenance policy to functionality.

In this work, we will extensively discuss the evidence and the presumptions that yield an eventual justifiability of maintenance.

Normative decision theory is known to suffer validation issues. Still, decisions to carry out maintenance may appear to be a promising special case because they (*i*) comprise many routine decisions (*ii*) whose policy compliance is typically recorded and (*iii*) whose functionality effect may straightforwardly follow from an unambiguous physical

variable. In this work, we explore to what extent the generic validation concerns of normative decision theory are surmountable in maintenance cases.

Unlike most research on maintenance policies, we do not intend to manipulate our prospective future by some improved maintenance policy. This attempt to justify maintenance by observing it as it occurs initially just serves a philosophical objective to more precisely approximate reality. However, acquiring more precise knowledge about the causal effects of maintenance may have practical implications. We will show that recording routines could potentially support decisions in a better way.

1.2 Research question

This work departs from a need to justify maintenance. We therefore raise a very simple question:

Is maintenance justified?

In principle, the answer to this question may be:

- A confirmation at an acceptable inference precision: maintenance is justified;
- A negation at an acceptable inference precision: *maintenance is not justified*;
- Neither a confirmation, nor a negation at an acceptable inference precision: *maintenance remains unjustifiable*.

Similar to Gauch (2002), Lakatos (1976) and Popper (2002), we adopt a viewpoint that certainty about scientific claims is lacking. So, we deem that in the end we will have to conclude that "maintenance remains unjustifiable". Resorting to unjustifiability neither contributes to science nor to better maintenance policies, but at least we can try to move further away from unjustifiability. As inference is the process of deriving logical conclusions from known or presumed propositions, the inference precision indicates a degree of certainty about the justification of maintenance here. We therefore pursue an improved inference precision. So, the title of this work reflects our *viewpoint* on what science can claim about maintenance and our *objective*.

1.3 Aim

The research question in Section 1.2 alluded to the problem, but it was not very specific about the scope of this work. In this section, we will outline how we intend to respond to the research question.

Maintenance is not something that just happens, it typically originates from conscious decisions to pursue a better future. Figure 1 shows that a decision "to maintain or not to maintain" provides access to a future with or without maintenance respectively. A decision maker who is indifferent towards a future with or without maintenance would not bother about it. However, we typically do prefer either of these two futures. So, any decision to carry out maintenance originates from some preference for a future with maintenance. Normative decision theory conventionally represents a preference by a utility and the challenge here is to make this subjective preference observable. Maintenance *intends* to intervene in the "natural" course of functionality by definition

(IEC, 1990), (CEN, 2001). We therefore take the viewpoint that this intention should materialise to retain maintenance policy assessments so they are meaningful to practitioners and scientists.

We posit that decisions to carry out maintenance usually result from collaboration within a group. Performance indicators may then enable decision makers to align their individual preferences with the group's preference. Eventually, these performance indicators reflect common sense about the pursued group's preference to be attained through a maintenance policy.



Figure 1 Discrete choice representation of a decision to carry out maintenance

Maintenance performance indicators reflect to what extent some subjective aspiration level has been met. Maintenance performance indicators are typically classified as leading or lagging. Leading indicators quantify maintenance policy compliance and are considered as causal for the future (Figure 1). Lagging indicators quantify the attributes of this future on which we may ground the group's preference. Although the intuition that leading indicators cause lagging indicators is widespread, we are unaware of a validation of this intuition.

| Lagging maintenance performance indicator (result indicator): | "functionality" K "resource costs" C |
|---|---|
| Leading maintenance performance indicator (enabling indicator): | "maintenance policy compliance" L |

Table 1 Simplified representation of a maintenance scorecard

Table 1 is a simplified representation of a maintenance scorecard. In practice, K,C and L are multidimensional vectors. The utility of a maintenance policy is built on the

lagging indicators in Table 1 and is represented by $U_L(C,K)$. This utility could be seen as some cost-effectiveness measure.

The achieved maintenance performance relies on "doing" as much as on "choosing". Normative decision theory poses that "choosing" and "doing" coincide and correspond. Then, maintenance policy violations would be non-existent. However, leading performance indicators for maintenance policy compliance L often show violations that may provide access to the counterfactual reality that maintenance policy compliance is hoping to avoid.

If "choosing" and "doing" were indistinguishable, maintenance resource costs C would be a definitional effect rather than a causal effect of a maintenance policy. We simply ignore any debate on whether resource costs causally or logically depend on a maintenance policy. Instead, we confine ourselves to a validation of a causality between maintenance policy compliance L and functionality K. So, we aim to be more precise about the truth or falsehood of the following proposition:

Maintenance policy compliance causes functionality

We ignore causal inferences that require experimental research. For operating organisations, well-designed experiments are often unattainable, whereas recording routines can be obtained in an efficient manner. We therefore implement some recent ideas about causal inferences from evidence collected by observational research. These causal inferences depart from known antecedents and conclusions and they are labelled as *maintenance policy validations* here. To the best of our knowledge, causal inferences between leading and lagging maintenance performance indicators are unprecedented. The intuition here is that a maintenance crew *intends* to pursue functionality by maintenance policy compliance in line with the definition of maintenance (CEN, 2001), (IEC, 1990). The proposed maintenance policy validation may reveal this intuition at an increased inference precision. A causality extends to a statistical association by providing the essential explanation to support decisions; i.e. we may control prospective effects by manipulating a cause rather than by manipulating just an associated variable. This work may reveal some practical insights to enhance decision support from recording routines.

1.4 Approach

The approach departs from an operationalisation of inference precision that allows us to compare inferences. We have not found common sense about a univariate representation of inference precision. We just arbitrarily define inference precision by the five inference objectives shown in Table 2.

| Choices | Inference objective |
|---------------------------------|-----------------------|
| Choice of an argument | Valid argument |
| | Functional relation |
| Choice of an operationalisation | Common sense evidence |
| | Universal argument |
| Choice of a sampling procedure | Decidable argument |

 Table 2 Survey of choices and inference objectives

To control these inference objectives, we confine ourselves to three choices:

- A choice of an argument;
- A choice of an operationalisation;
- A choice of a sampling procedure.

Gauch (2002), Lakatos (1976) and Popper (2002) suggest that a pursuit of inference precision comprises some trial and error trajectory. We do not believe that this decision problem can be decomposed by optimising the choices in Table 2 separately. Rather, we will iteratively combine arguments, operationalisations and sampling procedures. In Section 2.2, we will demonstrate the assessment of inference precision in a simplified fictitious example. In Section 3.6, we will better embed our approach in related fields of research.

We expect to be unable to find a combination of an argument, an operationalisation and a sampling procedure that entirely fulfils all inference objectives. We therefore resort to an attainable trade-off that may be acceptable. Nor do we pretend that the maintenance policy validation is optimal because we simply cannot assess all candidate arguments, operationalisations and sampling procedures. We just pursue the best inference precision among an arbitrary set of candidates.

We will introduce and discuss the inference objectives in Section 1.4.1. The remainder of Section 1.4 introduces the choice of an argument, the choice of an operationalisation and the choice of a sampling procedure.

1.4.1 Introduction to the inference objectives

We propose to decompose inference precision into the five inference objectives from Table 2. Each of the inference objectives will now be introduced separately.

Valid argument

This inference objective assesses whether a conclusion deductively follows from the other propositions of the argument. The conclusion of an invalid argument is not a necessary consequence from the other propositions of the argument. Arguments are the vehicles along which we reason, as we will further introduce in Section 2.1.

Functional relation

This inference objective assesses whether the antecedents map to a unique conclusion. Any argument may relate its antecedents to some conclusions, but an argument that maps its antecedents to a single conclusion is more precise. This relation (or model) between antecedents and conclusion is often presumed and controversial.

Common sense evidence

This inference objective assesses common sense about the evidence. Without interpretation, an argument remains a mathematical formalism that cannot claim anything about reality. The operationalisation ties an argument to reality. In principle, we are free to choose our operational definitions but if they lack common sense, others would easily refute the inference. For physical properties, a common sense operationalisation is often straightforward. Moreover, physical properties like mass, volume or time are often quantifiable by a single continuous variable. Notions like

maintenance policy compliance and functionality comprise some subjective requirements and their assessment may only follow from some arbitrary (multidimensional) vector of quantities. Inference precision may then suffer from a lack of common sense about the operationalisation (a construct validity issue).

Universal argument

This inference objective assesses whether the argument holds universally, i.e. holds for the entire population. In practice, we only have a stratified sample which hampers the inference of claims that are beyond the sample. A delimited set of recording routines leaves many background variables unobserved that could potentially explain an association within these recording routines. Therefore, this inference objective is important for the essential causal explanation we need for the maintenance policy validation. We do not expect to observe all relevant background variables. This inference objective seems therefore unattainable. We typically alleviate this concern by randomly assigning treatments and an informed selection of the antecedents in the argument.

Decidable argument

This inference objective assesses whether the truth or falsehood of presumptions is identifiable. Ideally, an argument is deductive and it comprises only one presumption while all other propositions are true. The presumption then automatically follows from the argument, which is then considered to be decidable. In Chapter 4, we will show that arguments often comprise several presumptions. Then, the argument is not decisive about the truth or falsehood of these presumptions. We expect that we will unavoidably face presumptions whose truth or falsehood is not identifiable from some stratified sample.

1.4.2 Choice of an argument

Arguments are essential for weighing up pros and cons of a particular presumption. Many arguments may reason about a particular presumption, but this does not imply that we should be indifferent regarding the choice of the argument.

In Chapter 4, we will present some conventional arguments that differ in their presumptions. In Section 4.5, we will review these arguments on their potential inference precision. In the absence of evidence, this review will only be preliminary. In Section 5.3, we will confront these arguments with a real sample of operationalised evidence and in Table 21, we will assess the inference precision of the maintenance policy validation by the candidate arguments and some given evidence. The challenge of the argument selection will be:

to achieve inference precision by choosing an adequate argument.

1.4.3 Choice of an operationalisation

Without an interpretation, an argument reduces to a mathematical formalism that cannot claim anything about reality. So, the argument requires an interpretation. An

operationalisation specifies the evidence; i.e. an operationalisation determines the argument's applicability. We provide some definitions:

A policy specifies a decision rule to be used at all decision epochs (Puterman, 2005).

A "decision" in the above will be defined as:

A decision is a choice of an action that causes the future

An "action" in the above will be confined to a maintenance action by:

Maintenance is the combination of all technical and administrative actions, including supervision actions, intended to retain an item in, or restore it to, a state in which it can perform a required function (CEN, 2001), (IEC, 1990).

And finally, the item's state in the above will be defined as functionality in this work:

Functionality is an ability of an item to perform a required function.

These definitions suggest that a maintenance policy *intends* to contribute to functionality but this contribution is not observable *by definition*. Otherwise, this work would be superfluous. These definitions just strengthen a belief that maintenance policy compliance should cause functionality and that maintenance should be justifiable in the end.

To attain the aim of this work, we need to establish the truth about:

- Maintenance policy compliance;
- Functionality;
- Causality.

Since truth appears to be unattainable, we resort to operational definitions that make these notions observable by common sense at least. In Section 3.2, we will seek common sense about maintenance policy compliance and functionality by exploring a few maintenance performance measurement practices. In Section 5.1, we will follow an organisation's convention to operationalise maintenance policy compliance and functionality. This convention corresponds with the practices found in the literature about maintenance performance measurement.

In Section 2.3, we will introduce the difficulties in operationalising causality. We therefore resort to a notion of *prima facie causality*, i.e. a less strict "causality at first sight". Prima facie causality has been frequently applied under observational research constructs in economics, ecology or biology. We implement prima facie causality in a maintenance decision making context.

We will specifically explore the inferences that *presume* causality between a policy and an effect in Section 3.1 and Section 3.3. In Section 3.4 and Section 3.5 respectively, we will explore inferences that *explain* or *predict* functionality. All these explorations will raise concerns about a maintenance policy validation. Still, we will try to address these concerns in the case study in Chapter 5. We intend to mitigate controversy about the evidence for a maintenance policy validation. The challenge of the operationalisation will be:

to achieve inference precision by establishing common sense about the evidence.

1.4.4 Choice of a sampling procedure

In Section 2.3, we will explain that experimental research constructs are generally more compelling for causality than observational research constructs. However, we will stick to evidence derived from an organisation's recording routines to serve the efficiency of collecting evidence. To pursue inference precision under an observational research construct, we will still be able to choose the sampling rate and the scale of the operationalised evidence.

In Section 3.2, we will criticise the sampling procedure of conventional maintenance performance indicators while proposing some construction rules. In Section 5.2, we will compose alternative samples that represent the organisation's common sense about maintenance performance. In Section 5.3, we will try to validate a maintenance policy from these samples. So finally, we intend to enhance inference precision by an informed choice of a sampling rate and a scale. The challenge of the selection of a sampling procedure will be:

to achieve inference precision by composing a suitable sample, given the constraint on an observational research.

1.5 Outline

This thesis is organised as follows. In Chapter 2, the fundamentals of (causal) inferences will be introduced. Chapter 3 will then position the proposed maintenance policy validation within related areas of research. More specifically, the perspectives of normative decision theory, maintenance performance measurements, maintenance policy assessments, diagnostics and prognostics will be reviewed. In Section 3.6, we will survey the lessons learned regarding our approach. In Chapter 4, the selection of a suitable argument will be discussed. For that purpose, four candidate arguments will be defined and each of them will be evaluated with respect to the inference objectives from Section 1.4.1. In the absence of empirical evidence, this assessment will only be preliminary. Chapter 5 will be concerned with the implementation, discussing both the operationalisation in Section 5.1 and the sampling procedure in Section 5.2. Section 5.3 and Section 5.4 will proceed with the validation of the four arguments using a real case study and Section 5.5 will discuss the influence of background variables. The work will then be discussed critically in Chapter 6, where the results and the contribution will be summarised. Finally, Chapter 7 will arrive at the conclusion and it will indicate directions for future research together with some practical implications.

Table 2 surveyed three choices that all require a theoretical background, a detailed discussion and a validation using a real case, but which also need to be treated in an

integral and iterative way. They are all addressed in several chapters of this work, rather than discussing them in separate chapters.

2 Preliminary to inference

This chapter will revisit some fundamentals on scientific inference and causal inference, but it will also describe the approach followed in this work to obtain an improved inference precision.

In Section 2.1, we will present a refresher on the logic of reasoning. In the context of this work, we will explain that stratified sampling hampers inference precision of universal scientific propositions. We will illustrate how the choice of an argument may influence inference precision.

In Section 2.2, we will illustrate our approach to inference precision. This illustration will be simplified by its omission of operationalisation issues and sampling issues. However, critique on the selected arguments at some given operationalisation and sampling procedure will similarly apply to the maintenance policy validation.

In Section 2.3, we will introduce causal inference. We will explain that causal inferences are particularly problematic, given the constraint on an observational research. We will therefore resort to a notion of prima facie causality (Granger, 1980) that uses knowledge of time to raise credence in causality.

2.1 Introduction to scientific inference

An argument is essential for any reasoning. An argument is any group of propositions of which one is claimed to follow from the others, which are regarded as providing support or grounds for the truth of that one (Copi & Cohen, 2009). We categorise the propositions of an argument in an antecedent, a model and a conclusion. Figure 2 depicts an example of an argument with an antecedent proposition P1, a model M1 and a conclusion C1.

We provide the following definitions:

- Let an information set V comprise the values of all antecedents and the conclusion. In Figure 2, V={1,k}.
- Let the evidence comprise all known propositions.
- Let modus ponens be a model based inference of a conclusion from evidence about the antecedent and the model; for a modus ponens inference by the argument in Figure 2, the evidence is (P1,M1).
- Let modus tollens be a model based inference of an antecedent from evidence about the conclusion and the model; for a modus tollens inference by the argument in Figure 2, the evidence is (M1,C1).
- Let a history based inference be an inference of a model from evidence about the antecedent and the conclusion; for a history based inference by the argument in Figure 2, the evidence is (P1,C1).

- Let a validation be an inference of a claim regarding the soundness of an argument; for a typical validation of the argument in Figure 2, the evidence is (P1,C1) and the model M1 is presumed.
- Let a replication be a duplication of an experiment. In Figure 2, a replication is any observed information set $V=\{1,k\}$ that has been generated by an identical model M1.

| P1 | L = l ;Let maintenance policy compliance be "l" |
|----|--|
| M1 | $(L = l) \rightarrow (K = k)$;If maintenance policy compliance is "l", then functionality is "k" |
| C1 | $\therefore K = k$ |

;Therefore functionality is "k", follows from P1,M1.

In this fictitious illustration, the evidence is a time series $(l,k)_{[1,t]} = \{\{l,k\},...\}$ that only comprises two information sets: $(l,k)_{[1,2]} = \{\{1,1\},\{1,2\}\}$. Below, we comment on inference precision with respect to this specific spatiotemporally constrained evidence:

| 1 I I I | |
|------------------------|---|
| Valid argument: | Yes, because C1 is a necessary consequence of P1,M1. |
| Functional relation: | Yes, because M1 is a functional relationship. |
| Common sense evidence: | In this fictitious example, we presume common sense about the |
| | interpretation of 1 as "maintenance policy compliance" and of k as |
| | "functionality". |
| Universal argument: | Yes, the sample $(l,k)_{[1,2]} = \{\{1,1\},\{1,2\}\}$ universally refutes the argument. |
| Decidable argument: | Yes, only the presumed model M1 is controversial. |
| | |

Figure 2 Example of a deductive argument with a functional model

A valid argument is an argument whose conclusion C1 is a necessary consequence of its antecedent P1 and model M1. A monotonic deductive argument is a valid argument that infers universal claims that are unsusceptible to new evidence (Bandyopadhyay & Forster, 2011). In science, we aim for sound monotonic deductive arguments but we typically only have a spatiotemporally constrained sample of evidence (Gauch, 2002), (Popper, 2002). Such a stratified sample only allows for an existential claim regarding the soundness of an argument that is susceptible to new evidence.

For a typical validation of the argument in Figure 2, model M1:k=f(l) has been presumed arbitrarily whereas we only know a stratified sample of the evidence about proposition P1 and conclusion C1. The stratified sample $(l,k)_{[1,2]}=\{\{1,1\},\{1,2\}\}$ decisively refute the argument in Figure 2 by a single counterexample, but another sample might have existentially confirmed it. Existentially because new evidence may easily overthrow this confirmation. Common sense about the proposition P1 and the conclusion C1 is essential here. Otherwise we may, for example, adopt an ad-hoc auxiliary hypothesis asserting that "this counterexample is not genuine evidence" upon any refutation that we find. This kind of posterior ad hocery prevents an empirical validation of any argument (Popper, 2002), (Lakatos, 1976).

We intend to enhance the justifiability of maintenance by a quest for its observable effects. In Section 1.3, we confined ourselves to a quest in order to be more precise

about:

Maintenance policy compliance causes functionality.

The following example illustrates that the choice of an argument matters:

- An argument like Figure 2 presumes some model M1:1→k=f(l). If the argument in Figure 2 appears to be sound, we would confirm that maintenance policy compliance L causes functionality K (Section 1.3) *and* we may exactly predict functionality K from maintenance policy compliance L. If the argument in Figure 2 appears to be refuted, we could not claim much about the existence of a causality between maintenance policy compliance L and functionality K in general *and* nor could we predict functionality K from maintenance policy compliance L.
- Another argument may presume independence between maintenance policy compliance L and functionality K. This argument does not enhance our capability to predict functionality K. However, both a confirmation *and* a refutation of this independence argument are decisive about the causality between maintenance policy compliance L and functionality K (Section 1.3).

The example above illustrates that if the argument in Figure 2 appears to be unsound, we cannot be very precise about the causality between maintenance policy compliance L and functionality K. Then, a less precise independence argument that lacks predictive capabilities may still better serve inference precision. In this work, we will put forward candidate arguments that we compare on inference precision regarding their claim about the causality between maintenance policy K.

2.2 Assessment of inference precision

This section will illustrate the assessment of the inference precision that we proposed in Section 1.4. Our approach comprised a choice of an argument, an operationalisation and a sampling procedure. Like in many choice problems, considering all possible options for the argument, the operationalisation and the sampling procedure would become intractable. We will therefore delimit, in this tentative example as well as in this work, the number of options that we will consider.

Choice of a sampling procedure.

Since an effect does not precede its cause in time, knowledge about time may attribute to causality in an observational research. We therefore choose a time series $(l,k)_{[1,t]} = \{\{l_1,k_1\},...,\{l_t,k_t\}\}$ rather than some cross-sectional data (l,k). In this tentative example we stick to the sample $(l,k)_{[1,2]} = \{\{1,1\},\{1,2\}\}$ and we will not develop other options regarding the sampling rate, or the scale of the variables as we will do in this work.

Choice of an operationalisation.

In this tentative illustration, we also omit an operationalisation of maintenance policy compliance L and functionality K. We just presume that the evidence $(l,k)_{[1,2]} = \{\{1,1\},\{1,2\}\}$ genuinely reflects common sense about maintenance policy compliance and functionality. In addition, we defer the operationalisation of causality to

Section 2.3. We therefore ignore the inference objective regarding "common sense evidence" here.

Choice of an argument.

In this tentative illustration, we stick to three fictitious arguments:

- The functional deduction in Figure 2;
- The relational deduction in Figure 3;
- The probability deduction in Figure 4.

These three arguments are all valid, but they may not be universally sound. This tentative illustration only considers the inference precision of these three arguments at a given sampling procedure and at a given operationalisation. The best inference precision obtained may therefore easily be overthrown by other unconsidered arguments, sampling procedures and operationalisations.

| P1 | L = l ;Let maintenance policy compliance be "l" |
|----|--|
| M1 | $(L = l) \rightarrow (k - 1 \le K \le k + 1)$;If maintenance policy compliance is "l", then functionality is in "[k-1,k+1]". |
| C1 | $\therefore (k - 1 \le K \le k + 1)$;This conclusion follows from P1,M1. |

In this fictitious illustration, the evidence is a time series $(l,k)_{[1,t]} = \{\{l,k\},...\}$ that only comprises two information sets: $(l,k)_{[1,2]} = \{\{1,1\},\{1,2\}\}$. Below, we comment on inference precision with respect to this specific spatiotemporally constrained evidence:

| Valid argument: | Yes, because C1 is a necessary consequence of P1,M1 |
|------------------------|---|
| Functional relation: | No, because M1 is not a functional relationship. |
| Common sense evidence: | In this fictitious example, we presume common sense about the |
| | interpretation of l as "maintenance policy compliance" and of k as |
| | "functionality". |
| Universal argument: | Yes, but the sample $(1,k)_{[1,2]} = \{\{1,1\},\{1,2\}\}$ only existentially confirms |
| | the argument. |
| Decidable argument: | Yes, only the presumed model M1 is controversial. |

Figure 3 Example of a deductive argument with a relational model

The relational deduction (Figure 3) resembles the functional deduction (Figure 2); it just allows functionality K to be within some upper and lower limit. If the arguments from Figure 2 and Figure 3 were both sound, the functional deduction (Figure 2) would have been preferred because it is more restrictive. If the model M1 was just presumed, the functional deduction (Figure 2) would be easier to falsify than the relational deduction (Figure 3). The spatiotemporally constrained sample of evidence $(1,k)_{[1,2]} = \{1,1\},\{1,2\}\}$ for example, universally falsifies the functional deduction (Figure 2) because for 1=1, two different values of k (1 and 2) are observed. But the same evidence existentially confirms the relational deduction (Figure 3), as the values of K are still within the upper and lower limit of its model M1.

The probability deduction (Figure 4) represents a conventional alternative for a falsified functional deduction (Figure 2). The probability deduction (Figure 4) decomposes functionality K into a deterministic model M1:f(l) and an independent error P4: ε , so that $k=f(l)+\varepsilon$. Both the model M1 and the error P4 do not follow from the evidence or from

| controversy, the probability deduction (Figure 4) becomes undecidable. | | | | |
|--|--|--|--|--|
| P1 | L = l ;Let maintenance policy compliance be "l" | | | |
| M1 | $(L = l) \rightarrow (\hat{K} = \hat{k})$;Presume that if maintenance policy compliance is "1", then functionality is estimated as "k". | | | |
| C1 | : $\hat{K} = \hat{k}$;This conclusion follows from P1,M1. An estimator however is not in the evidence | | | |
| P2 | K = k;Let functionality be "k" | | | |
| Ρ4 | $((K - \hat{K}) = (k - \hat{k})) \Leftrightarrow (L = l)$;Presume that maintenance policy compliance is independent of prediction errors " ε " in K. This means that model M1 captures all information from L about K in its parameters. The evidence P1,P2 does not suffice for P4. | | | |
| M2 | $(P1, M1, P4) \rightarrow (Pr(P2))$;M2 is a common sense definition of a probability that expresses the probability of P2 given P1,M1,P4. | | | |
| C2 | ∴ <i>Pr</i> (<i>P</i> 2) ;Follows from M2 and its antecedents. | | | |

some common sense definition which makes M1 and P4 controversial. Due to this

In this fictitious illustration, the evidence is a time series $(l,k)_{[1,t]} = \{\{l,k\},...\}$ that only comprises two information sets: $(l,k)_{[1,2]} = \{\{1,1\},\{1,2\}\}$. Below, we comment on inference precision with respect to this specific spatiotemporally constrained evidence:

| Valid argument: | Yes, but C1,C2,P4 are not immediately observable |
|------------------------|---|
| Functional relation: | Yes, but C1,C2,P4 are not immediately observable |
| Common sense evidence: | In this fictitious example, we presume common sense about the |
| | interpretation of l as "maintenance policy compliance" and of k as |
| | "functionality". Moreover we presume common sense about the |
| | definition of a conditional probability. |
| Universal argument: | No, C2 only expresses a likelihood that is susceptible to extensions of |
| | the sample $(1,k)_{[1,2]} = \{\{1,1\},\{1,2\}\}.$ |
| Decidable argument: | No, both the model M1 and the presumption P4 are controversial. We |
| - | deem that the model M2 is a common sense definition of a probability |
| | that follows from P1,M1,P4. |

Figure 4 Example of an argument that deduces a probability

By random assignment of P1 treatments, the sampled evidence $(l,k)_{[1,t]}$ would become compelling for the probability deduction (Figure 4) because P1 is known to be the only variable that could eventually associate with P2 then. Any association between P1 and the errors P4 would then become attributable to an incorrect mapping k=f(l), i.e. an incorrect model M1. So, the probability deduction (Figure 4) may then become *existentially* decidable in terms of the likelihood of some presumed model M1 *and* some presumed error distribution P4. Existentially because this likelihood assessment only holds with respect to a specific sample. Particularly at large sample sizes, random assignment of treatments might have been compelling for the probability deduction (Figure 4). However, a sample of two observations $(l,k)_{[1,2]}$ collected by observational research is insufficient to decide about the likelihood of a presumed model M1 and a presumed error distribution P4.

Table 3 surveys the inference precision of the functional deduction (Figure 2), the relational deduction (Figure 3) and the probability deduction (Figure 4). The most precise assessments are in bold.

The soundness of the functional deduction (Figure 2) appears to be very precisely refuted. This refutation is universal; i.e. we will not change our position upon extension of the evidence to $(l,k)_{[1,t]} = \{\{1,1\},\{1,2\},...\}$. However, the functional deduction (Figure 2) only presumes a very specific functional relation. Its refutation allows many alternative (causal) relations to be true.

The soundness of the relational deduction (Figure 3) appears to be existentially confirmed; i.e. it holds for the sample of evidence $(l,k)_{[1,2]}=\{\{1,1\},\{1,2\}\}$ but a single counterexample beyond this sample may overthrow this existential confirmation by a universal refutation. Moreover, the model M1 is not a functional relation that compels the conclusion C1 to a unique value. Rather, the relational deduction (Figure 3) allows the conclusion C1 to be in the range "K=k±1".

| | Functional deduction (Figure 2) | Relational deduction (Figure 3) | Probability deduction (Figure 4) |
|-----------------------|---------------------------------------|---------------------------------------|--|
| Valid argument | Yes | Yes | Yes |
| Functional relation | Yes | No | Yes |
| Common sense evidence | - | - | - |
| Universal argument | Yes: Refuted | Yes: but existentially confirmed | No |
| Decidable argument | Yes | Yes | No |

 Table 3 Inference precision of three tentative arguments

The soundness of the probability deduction (Figure 4) appears to be undecidable because it comprises two controversial presumptions. This indeterminacy is universal; we will not change our position upon the extension of the evidence to $(l,k)_{[1,t]}=\{\{1,1\},\{1,2\},...\}$. Random assignment of treatments would have attributed to the likelihood of the probability deduction (Figure 4) for some presumed model M1 and some presumed error distribution P4. However, a small sample of evidence collected by observational research was insufficient for a compelling likelihood assessment. Therefore, the probability deduction (Figure 4) remained undecidable under this sampling procedure.

To conclude, Table 3 illustrated that validations could be imprecise in different ways. Possibly, we should not be indifferent about where to allocate inference imprecision. This assessment of inference precision has been simplified by just considering a single option for the operationalisation and the sampling procedure. Resembling a typical model selection process, the choice of the argument, the operationalisation and the sampling procedure simultaneously influence the inference precision. Therefore, this problem of which choice to make does not allow for a decomposition. Moreover, we lack in-depth knowledge about some model that maps these choices to inference precision. As a result, we resort to an iterative quest for the most adequate option among an arbitrary set of candidates, as we will show in Section 5.3.

In Table 3, we deferred the operationalisation of causality to Section 2.3. However, the three arguments did not assert the same about the proposition:

Maintenance policy compliance causes functionality $(L \rightarrow K)$

which is at stake in this work. Although the functional deduction (Figure 2) has, most precisely, been refuted, it still allows for many alternative models M1 that also imply $L \rightarrow K$. In other words, the functional deduction (Figure 2) is more restrictive than strictly needed for the causality $L \rightarrow K$. The less precise alternative arguments may in the end assert more about the causality $L \rightarrow K$. In Chapter 4, we will also compare the candidate arguments on their claim about a more modest notion of prima facie causality $L \rightarrow K$ (Table 5). In Table 21, we will survey the inference precision of the candidate arguments together with their claim about the prima facie causality $L \rightarrow K$.

2.3 Introduction to causal inference

In general, variables could be related by:

- Definition like "1 inch=2,54 centimetres";
- Association like "Pr(prospective weather| mercury column)";
- Causality like "disease \rightarrow symptom".

A *definition* is just a claim that only needs common sense to be accepted. A validation of some presumed *association*, generally expressed as a conditional probability, is well-explored in statistics but it allows for many explanations. And finally a *causality* provides the explanation for an association that is needed to predict the effect of a *specific* decision. In this work, we try to explain the evolution of functionality K by maintenance policy compliance L. The challenge here is to assign a causal explanation to an eventual association between maintenance policy compliance L and functionality K.

2.3.1 From association to causality

A statistical association can be represented by the inequality:

$$pr_{K|L}(k|l) \neq pr_{K|L}(k|l')$$

1

This means that the probability to obtain a value k, given some value l rather than l', is not equal. So, the variables K and L are somehow associated. However, this association is explainable by various causal relations like $L \rightarrow K$ (i.e. L causes K), $K \rightarrow L$ or $B \rightarrow (L,K)$. In the latter case, a confounding background variable B is the cause of both L and K. Therefore, Equation 1 is not conclusive about causality. Possibly, L and K are only related by some mediating or confounding variable B as shown in Figure 5. Then,

L and K are independent but they may still appear to be associated with respect to an information set $V = \{l,k\}$.

To reveal a causality $L \rightarrow K$, we should marginalise the effect of all background variables B in the universe Ω , i.e. (B $\in \Omega$). Equation 2 shows this marginalisation that would have identified the causality $L \rightarrow K$.

$$\left(\sum_{\forall B} pr_{K|L,B}(k|l,b) \times pr_{B}(b) \neq \sum_{\forall B} pr_{K|L,B}(k|l',b) \times pr_{B}(b)\right) \equiv (L \to K)$$

The marginalisation over all B is intractable since we typically cannot observe all candidate causes of K. Still, we could eventually extend the information set V with some background variables B to eliminate their confounding or mediating effects by marginalisation. A statistical association between L and K may then appear to be a spurious cause as illustrated in Figure 5.



Figure 5 Inference of a spurious causality

For risk assessments and fault detection systems, statistical associations seem to be appropriate since many applications of prognostics and diagnostics effectively use noncausal symptoms. Vibrations, for example, are often not the cause of an impending bearing failure, but they may still effectively detect or predict it. A statistical association between maintenance policy compliance L and functionality K may similarly appear to be a warning signal but this statistical association does *not* imply that we can steer functionality K by control over maintenance policy compliance L. The identification of the causes of failures in prognostics and diagnostics usually does not follow directly from their arguments. For risk assessment that is not a problem, but it is for decision making. A decision maker would like to know the *prospective* effect of his *specific* choice of an action. We pursue to attribute an observed effect to the maintenance policy in *retrospect*. A statistical association that holds *across* choices does not provide this essential explanation.

Thoughts about causal formalisms date back to structural equations modelling (Wright, 1921), the potential outcomes framework (Neyman, 1923), (Rubin, 1974) and the probability raising approach (Suppes, 1970). An in-depth discussion of these formalisms is beyond our scope. We depart from structural causal modelling (Pearl, 2000) that unifies previously mentioned formalisms. Pearl (2010) defined a causality by augmenting a conditional probability function with a do(.) operator that refers to an action.

 $\left(pr_{K|L}(k|do(l)) \neq pr_{K|L}(k|do(l'))\right) \equiv (L \to K)$ 3

Equation 3 for example asserts that an action do(l) determines the probability of K. Note that equation 1 only claims functionality K and maintenance policy compliance L to be associated *across* decisions to carry out maintenance, whereas Equation 3 imposes a response to a *specific* decision to carry out maintenance which is needed for causality (Greenland, 2011). When choosing an action do(l), we cannot observe the counterfactual effect of do(l') which do(l) was supposed to prevent. So, we can only observe either of the two terms in Equation 3.

We may alleviate this generic concern by distinguishing "choosing" from "doing" in the process of maintenance decision making. Actions may violate or comply with the chosen maintenance policy from which they originate. So, the chosen maintenance policy do(l) may materialise in some alternative do(l'). Eventually, we then reveal the counterfactual effect of do(l') to be avoided by the chosen maintenance policy do(l). In this work, we aim for a maintenance policy validation along this line.

A random assignment of L treatments would have strengthened a belief that L rather than some confounder $B \rightarrow (L,K)$ explains the statistical association in Equation 1 because random assignment of treatments weakens the assumption of equally distributed confounders in the treatment and the control group. However, random assignment of treatments is not decisive about mediators $L \rightarrow B \rightarrow K$ that could alternatively explain the statistical association in Equation 1. Therefore, random assignment of treatments alone does not resolve all controversy about causality assessments. Still, Greenland (2011), Pearl (2000) and Spirtes (2000) favoured experimental research to at least alleviate concerns about confounders.

For various reasons, we cannot randomly assign maintenance policies in the same way as in a controlled experiment. This is mainly due to the fact that we cannot easily intervene in the course of an organisation's operation. We therefore just attempt to use recording routines. Recordings from a so-called observational research are less coercive for causality than similar recordings from a well-designed experimental research. Causal claims based on observational research are therefore less precise. So, our attitude towards these recordings will depend on the sampling procedure. In the next sections, we will adopt some more recent ideas about causal inference in observational research, but we will not entirely resolve the generic troubles with causality assessments.
2.3.2 Non-causality assumptions

Pearl (2010) recommended to explicitly state the non-causality assumptions that allow us to claim a causality from a statistical association within an information set V by a path graph.



Figure 6 Path graph of non-causality assumptions

Figure 6 depicts the path graph that states the non-causality assumptions required to interpret Equation 1 as causal. The non-causality assumptions in Figure 6 may appear to be highly problematic, but at least we are explicit about them. The arrows are candidate causes. The omitted arrows are non-causality assumptions:

 $K_{T+1} \neq L_T$; K_{T+1} does not cause L_T ; this non-causality assumption seems innocuous from a viewpoint that an effect cannot precede its cause in time.

 $B \nleftrightarrow L_T$; B represents all non-redundant variables that could possibly affect K_{T+1} when L_T is held constant; this non-causality assumption prohibits confounders $B \rightarrow (L_T, K_{T+1})$ and mediating causes $L_T \rightarrow B \rightarrow K_{T+1}$. This noncausality assumption implies that L_T comprises unique information about K_{T+1} that is not available otherwise. Since B could be any variable in the universe, this non-causality assumption is not efficiently assessable.

2.3.3 Prima facie causality

We confine the sampling procedure to an observational research that is spatiotemporally constrained. A uniform enforcement of actions $do(l_t)$ over the population is therefore unattainable, so we resort to a modest notion of prima facie (~at first sight) causality (Granger, 1980). A prima facie causality suits our sampling procedure but it still attributes to causality by knowledge about time. For example, prima facie causality weakens the non-causality assumption $K_{T+1} \neq L_T$ in Figure 6.

Granger (1980) operationalised causality with constraint on an observational research. This notion of causality relies on the following principles:

- 1. The past and present may cause the future, but the future cannot cause the past;
- 2. All causal relationships remain constant in direction throughout time;
- 3. A cause comprises unique information about the effect that is not available otherwise.

The following definition of a cause complies with these causality principles.

$$\left(pr_{K_{T+1}|\Omega_{T+1}}(k_{t+1}|\omega_{t+1}) \neq pr_{K_{T+1}|\Omega_{T+1}-L_{T}}(k_{t+1}|\omega_{t+1}-l_{t}) \quad ; \exists l \forall t \forall \omega \right) \equiv (L_{T} \rightarrow K_{T+1}) \\ ; L_{T} \ causes \ K_{T+1}$$

 Ω_{T+1} represents all non-redundant information in the universe up to a time T+1, L_T represents a candidate cause, K_{T+1} represents the effect and ω ,l,k,t refer to the respective values of these variables. Equation 4 addresses the three causality principles as follows:

- The first causality principle imposes that a statistical association between K_{T+1} and L_T is not explainable by $K_{T+1} \rightarrow L_T$ by common sense.
- The second causality principle imposes that $L_T \rightarrow K_{T+1}$ implies $L_{T+A} \rightarrow K_{T+1+A}$ but the positive strength of this causal relationship may change. Therefore, Equation 4 holds at all times t.
- The third causality principle imposes that L_T comprises unique information about K_{T+1} that is not available otherwise. Therefore, Equation 4 holds irrespective of the universe Ω_{T+1} ; i.e. at all ω .

The second and the third causality principle cannot be operationalised since we cannot access all possible values of time T and the universe Ω . We therefore resort to a modest claim of prima facie causality that only holds with respect to some limited information set V. Prima facie causality finds many applications in econometrics, ecology or climate studies which are often observational studies. In this work, we implement prima facie causality in a maintenance decision making context.

Equation 5 defines L_T as a prima facie cause of K_{T+1} with respect to the information set $V=\{U,k_{t+1}\}$.

$$\left(pr_{K_{T+1}|U}(k_{t+1}|u) \neq pr_{K_{T+1}|U-L_T}(k_{t+1}|u-l_t) \quad ; \exists l \forall t \forall u \ (U \subset \Omega_{T+1}) \right) \equiv (L_T \to K_{T+1})$$

$$; \ L_T \ prima \ facie \ causes \ K_{T+1}$$

$$5$$

This definition of a prima facie cause in Equation 5 resembles the definition of a cause in Equation 4, but a prima facie cause only takes the subset U from Ω_{T+1} as a given condition. Therefore, Equation 5 cannot claim that L_T comprises unique information about K_{T+1} . The inequality between the probabilities in Equation 5 and Equation 1 are exchangeable as U={lt}. We provide proof that also holds for an extended body of knowledge U={lt,b}.

Proof

| Let $U = \{l_t, b\}$ | ; The body of knowledge U in Equation 5 has been defined as $\{l_t, b\}$ | | |
|-----------------------|--|--|--|
| Let $K = \{k_{t+1}\}$ | ; K is the event $\{k_{t+1}\}$ | | |
| Let $L=\{l_t\}$ | ; L is a subset of U that comprises the value $\{l_t\}$ | | |
| Let B={b} | ; B is a subset of U that comprises the value {b} | | |
| Let L'=L- $\{l_t\}$ | ; L' is the complementary set of L, i.e. all values that the variable L_T can take | | |
| | except for l _t . | | |
| Let $\{l_t\} \in L'$ | ;lt' is some element from L' | | |

Then, Equation 6 is an alternative representation of the association in Equation 5 that should hold for at least one element in LUL' at all times t and at all possible b to be considered as a prima facie causality:

 $Pr(K|U) \neq Pr(K|U-L)$

Equation 6 straightforwardly deduces to:

$$\frac{Pr(K \cap L \cap B)}{Pr(L \cap B)} \neq \frac{Pr(K \cap L \cap B) + Pr(K \cap L' \cap B)}{Pr(L \cap B) + Pr(L' \cap B)}$$

$$\frac{Pr(K \cap L \cap B)}{Pr(L \cap B)} \left(Pr(L \cap B) + Pr(L' \cap B) \right) \neq Pr(K \cap L \cap B) + Pr(K \cap L' \cap B)$$

$$\frac{Pr(K \cap L \cap B)}{Pr(L \cap B)} Pr(L' \cap B) \neq Pr(K \cap L' \cap B)$$

$$Pr(K|L, B) \neq Pr(K|L', B)$$

$$7$$

The last line of Equation 7 is equivalent to Equation 1. This inequality should hold for at least one element in LUL' at all times t and at all possible b to be considered as a prima facie causality.

In Equation 5, we added quantifiers to the original definition of Granger (1980) in order to be more specific about the assessment of a prima facie causality. These quantifiers specify at which values in the body of knowledge U, the statistical association in Equation 5 should hold to identify a prima facie causality. For the prima facie causality in Equation 8, this implies that the statistical association should hold for *some* 1 and *all* t,k to identify L_T as a prima facie cause of K_{T+1} with respect to the information set $V=\{l_t,k_t,k_{t+1}\}$:

$$\left(pr_{K_{T+1}|L_T,K_T}(k_{t+1}|l_t,k_t) \neq pr_{K_{T+1}|K_T}(k_{t+1}|k_t) \quad ; \exists l \forall t \forall k\right) \equiv (L_T \to K_{T+1})$$
8

Figure 7 illustrates the effect of the quantifiers in Equation 8 on a claim regarding prima facie causality that holds with respect to the information set $V = \{l_t, k_t, k_{t+1}\}$.

Let the branch H of the event tree in Figure 7 represent all *possible* values in this information set $V = \{l_t, k_t, k_{t+1}\}$ and their conditional probabilities. The branch H conceives two *possible* values of the body of knowledge $U = \{l_t, k_t\}$ that are equally probable.

Then, L_T prima facie causes K_{T+1} with respect to the information set $V=\{l_t,k_t,k_{t+1}\}$ by the definition in Equation 8 because *some* value of L_T associates with K_{T+1} at *all* values of K_T that are *possible* in the branch H. This conclusion follows from Equation 9 that confirms the inequality in Equation 8 at $L_T=1$ at the only possible value of $K_T=0$.

$$\begin{pmatrix} pr_{K_{T+1}|L_TK_T}(0|0,0) = pr_{K_{T+1}|K_T}(0|0) \end{pmatrix} \equiv \begin{pmatrix} \frac{1}{2} = \frac{1}{2} \end{pmatrix} \\ \begin{pmatrix} pr_{K_{T+1}|L_TK_T}(0|1,0) \neq pr_{K_{T+1}|K_T}(0|0) \end{pmatrix} \equiv \begin{pmatrix} 0 \neq \frac{1}{2} \end{pmatrix}$$

However, if the event tree H' in Figure 7 represents all *possible* values in the information set $V = \{l_t, k_t, k_{t+1}\}$, L_T would *not* prima facie cause K_{T+1} with respect to the information set $V = \{l_t, k_t, k_{t+1}\}$ by the definition in Equation 8 because *not a single* value

6

of L_T associates with K_{T+1} at *all* values of K_T , that are possible in the event tree H'. The event tree H' conceives four possible values of the body of knowledge U={ l_t,k_t } that are equally probable. Despite the dependence between L_T and K_{T+1} at $K_T=0$ (Equation 9), L_T and K_{T+1} appear to be independent at $K_T=1$:

$$\begin{pmatrix} pr_{K_{T+1}|L_TK_T}(0|0,1) = pr_{K_{T+1}|K_T}(0|1) \end{pmatrix} \equiv \begin{pmatrix} \frac{1}{2} = \frac{1}{2} \end{pmatrix}$$

$$\begin{pmatrix} pr_{K_{T+1}|L_TK_T}(0|1,1) = pr_{K_{T+1}|K_T}(0|1) \end{pmatrix} \equiv \begin{pmatrix} \frac{1}{2} = \frac{1}{2} \end{pmatrix}$$

$$10$$

We therefore deem that L_T does not comprise unique information about K_{T+1} which is troublesome with regard to the third causality principle. In the case of H', knowledge about K_T is also needed to decide about the existence of a dependence between L_T and K_{T+1} .



Figure 7 Event tree of all possible values in an information set V={lt,kt,kt+1}

This example in Figure 7 showed that the possible values of the body of knowledge $U=\{l_t,k_t\}$ are decisive for the existence of a prima facie causality. An individual who knows that $K_T=0$ in a specific case, i.e. who knows that $K_T=1$ is impossible, concludes that L_T prima facie causes K_{T+1} now. This same individual may also deem that both $K_T=0$ and $K_T=1$ are possible in principle and that L_T does not prima facie cause K_{T+1} generally.

In this work, we will implement the definition of the prima facie causality in Equation 5 and we will show the importance of knowledge about the *possible* values of the body of knowledge U in a maintenance case.

2.3.4 Controversy about causality

A prima facie causality raises credence in causality by knowledge about time. This belief may appear to be false as shown in the following example. Let L_T and K_{T+1} be truly independent as shown in Figure 8.



Figure 8 Path graph of non-causality assumptions in an extended information set

Still, if the information set confines to $V = \{l_t, k_{t+1}\}$, we may well *confirm* the prima facie causality:

$$\left(pr_{K_{T+1}|L_T}(k_{t+1}|l_t) \neq pr_{K_{T+1}}(k_{t+1}) \quad \exists l \forall t\right) \equiv (L_T \to K_T)$$
11

However, Figure 8 shows that L_T and K_{T+1} are in fact only associated by the confounder C_{T-1} . Figure 8 shows that if we had known the superset $V' = \{c_{t-1}, l_t, k_{t+1}\}$, we might well have *refuted* the prima facie causality:

$$\left(pr_{K_{T+1}|L_{T},C_{T-1}}(k_{t+1}|l_{t},c_{t-1}) \neq pr_{K_{T+1}|C_{T-1}}(k_{t+1}|c_{t-1}) \quad \exists l \forall t \forall c\right) \equiv (L_{T} \to K_{T})$$
12

The specification of the time in Equation 11 and Equation 12 is essential to identify the confounder $C_{T-1} \rightarrow (L_T, K_{T+1})$. Otherwise, a refutation of Equation 11 and a confirmation of Equation 12 would have been explainable by several causal structures like $C \rightarrow (L,K)$, $L \rightarrow C \rightarrow K$ or $K \rightarrow C \rightarrow L$. The first causality principle in Section 2.3.3 asserting that the future cannot cause the past, excludes the causal structures

 $L_T \rightarrow C_{T-1} \rightarrow K_{T+1}$ and $K_{T+1} \rightarrow C_{T-1} \rightarrow L_T$. The confounder C_{T-1} in its turn, may a appear to be spurious cause upon further extensions of the information set. Any prima facie cause is only an existential claim that holds with respect to an information set V that is just a subset of the universe Ω .

The path graph specifies under what conditions a prima facie causality would still satisfy the "universal argument" inference objective in Section 1.4.1 that we need for a cause. The path graph in Figure 8 for example requires independence (=bidirectional non-causality assumption here) between B, C_{T-1} and B, L_T to interpret the prima facie cause between C_{T-1} and K_{T+1} as a cause. Since we typically cannot enumerate let alone observe B exhaustively, satisfaction of the "universal argument" inference objective is unattainable. Claiming causality from an inferred prima facie causality remains therefore controversial, which affects the inference precision.

The maintenance policy validation demonstrated in this work only relies on an inferred prima facie causality between *future* maintenance performance and *current* maintenance policy compliance. Knowledge about the causal structure is important for decision making. In the case of Figure 8, a decision maker better steers K_{T+1} by control over C_{T-1} than by control over L_T because control over L_T does not affect K_{T+1} . A decision maker could eventually only control the yet to be observed future. Then, knowledge about the prima facie causality between C_{T-1} and K_{T+1} and about the non prima facie causality between L_T and K_{T+1} could already be appreciable. This, in spite of the fact that the decision maker ultimately needs to know the cause of K_{T+1} .

Aven (2011), Parry (1996) and Zio (2009) confirmed that uncertainty about the future may be irreducible (aleatory) or reducible (epistemic) by additional knowledge. The definition of a causality in Equation 4 could be seen as a case of epistemic uncertainty. Although Aven (2011), Parry (1996) and Zio (2009) did not present their case as a causal inference problem, they similarly confirmed that known and unknown mediating and confounding background variables hamper the assessment of eventual uncertainty reductions under an observational research construct. In this work, we will not entirely resolve the imprecision of causal claims from recording routines.

3 Literature review

This chapter will position the maintenance policy validation to five different related areas of research.

In Section 3.1, we will depart from the idea that the maintenance policy validation may be seen as a special case of a normative decision theory validation. Normative decision theory has many ramifications which differ in their representation of preference. We will therefore survey some representations of preference that drive or at least explain decisions. We will conclude that the assessment of an individual's preference appears to be problematic. However, we will posit that policy assessments made by a group require common sense about the group's preference. Moreover, we may extend to conventional policy validations by using evidence about violations which can be efficiently obtained from maintenance cases.

In Section 3.2, we will deepen the common sense about preference mentioned in Section 3.1 by analysing maintenance scorecards that quantify a group's preference by performance indicators. A conventional maintenance scorecard quantifies both maintenance policy compliance (leading indicator) and functionality (lagging indicator). A maintenance scorecard could therefore potentially provide the common sense evidence for the maintenance policy validation. However, maintenance scorecards seem to be designed to show posterior compliance rather than to accommodate causal inferences. We will put forward three construction rules for maintenance performance indicators to enable a maintenance policy validation. This work will reveal the feasibility of these construction rules.

In Section 3.3, we will review some approaches to maintenance policy assessments that typically reason about a yet to be observed future. We will argue that maintenance policy assessments follow a process of satisficing rather than optimising. Satisficing makes fewer demands on a decision maker's psychological and physiological abilities but it is more modest in its claim. We will explain that an argument for the maintenance policy validation, which relies on a presumed independence, may still be informative for both satisficers and optimisers. We will also acknowledge that the maintenance policy validation potentially provides the essential empiricism to maintenance policy assessments, like the reliability centred maintenance process, that predominantly rely on expert judgement.

In Section 3.4 and Section 3.5, we will survey some diagnostic and prognostic arguments that we will also cover in Chapter 4. In diagnostics and prognostics, a single inference that universally outperforms all others seems non-existent. Similarly, our approach to the maintenance policy validation resorts to a case dependent iterative journey along optional inferences. We will acknowledge that maintenance and reliability applications of prognostics and diagnostics conventionally infer functionality from physical variables. For risk assessment, an association or symptom may be as

effective as a cause. However, a maintenance policy validation requires a causal explanation.

Finally in Section 3.6, we will summarise the lessons learned from this literature review.

3.1 Review of normative decision theory

Maintenance is not a natural phenomenon that just befalls us. Rather, maintenance originates from conscious decisions. Figure 1 represented a decision to maintain or not to maintain as a discrete choice problem that provides access to a future with or without maintenance. We are typically not indifferent regarding these futures; i.e. we usually have some preference. The following definition of preference (Savage, 1954), (Broome, 2004) may suffice for this exposition:

A decision maker prefers x_i to x_j means that this decision maker would choose x_i rather than x_i if he were to have a choice between x_i and x_j .

So, any decision to carry out maintenance is somehow preferred over the decision to do nothing. Normative decision theory describes or prescribes that decision makers should decide by their preference. The challenge here is to make preference observable. This section will therefore survey representations of preference; i.e. utilities.

The idea that preferences are subjective is far from new. Bernoulli (transl. 1954) for example argued that preference does not follow from maximising the mathematical expected gain. Rather, the preference for some expected gain may differ for various decision makers. To illustrate the point, we rephrase one of Bernoulli's (transl. 1954) examples:

Example

Consider a discrete choice problem that allows for two alternatives:

Action (a₁) Participate in a lottery that increases a decision maker's initial wealth x by 100 at a probability of 0,95 or by 0 at a probability of 0,05.

Action (a₂) Pay for an insurance fee of 8 to be certain about a wealth increase of 100. So, the decision maker's initial wealth will certainly increase by 92.

The mathematical expectation of wealth, given a choice of action a_1 is:

$$E[X|a_1] = \sum_{\forall i} x_i \times pr_X(x_i) = ((x+100) \times 0.95) + ((x+0) \times 0.05) = x+95$$
13

The mathematical expectation of wealth, given a choice of action a₂ is:

$$E[X|a_2] = \sum_{\forall i} x_i \times pr_{X_i}(x_i) = (x+92) \times 1 = x+92$$
14

Regardless of the initial wealth x, the difference in mathematical expectation is $E[X|a_1]-E[X|a_2]=3$. So, any decision maker should always choose action a_1 if preference is represented by mathematical expectations. However, Bernoulli (transl. 1954) acknowledged that the initial wealth matters for decision makers. He therefore proposed to measure preference by an expected utility. The utility function is in this case assumed to be the logarithm of wealth, expressing a decreased additional utility of every amount of

additional wealth. This means that the expected utility of action a₁ is:

$$E[U|a_{1}] = \sum_{\substack{j=1\\2}}^{n} u(x_{j}) \times pr_{X}(x_{j})$$

$$E[U|a_{1}] = \sum_{\substack{j=1\\j=1}}^{n} log_{10}(x_{j}) \times pr_{X}(x_{j}) = log_{10}(x+100)^{0.95} + log_{10}(x)^{0.05}$$
15

while the expected utility when choosing for action a₂ is:

$$E[U|a_2] = \sum_{j=1}^{1} log_{10}(x_j) \times pr_X(x_j) = log_{10}(x+92)^1$$
16

For a poor decision maker (x=1), the expected utilities equal:

$$E[U|a_1] = log_{10}(1+100)^{0.95} + log_{10}(1)^{0.05} = 1,904$$

$$E[U|a_2] = log_{10}(1+92)^1 = 1,968$$
17

Therefore, the poor decision maker should be risk averse (choose action a_2). A wealthy decision maker however (x=10.000) should be risk seeking (choose action a_1) since:

$$E[U|a_1] = log_{10}(10.000 + 100)^{0.95} + log_{10}(10.000)^{0.05} = 4,0041$$

$$E[U|a_2] = log_{10}(10.000 + 92)^1 = 4,0040$$
18

A decision maker's preference is now not just represented by a difference in mathematical expectations $E[X|a_1]-E[X|a_2]$, but by a difference in expected utility $E[U|a_1]-E[U|a_2]$ that considers the decision maker's initial wealth. In this way, Bernoulli (transl. 1954) conceived subjective preferences.

Bernoulli (transl. 1954) generalised this example to a hypothesis that any increase in wealth, no matter how insignificant, will always result in a utility increase, whereas the magnitude of this utility increase is inversely proportional to the quantity of goods already possessed. Bernoulli (transl. 1954) only claimed this hypothesis of decreasing marginal utility as highly probable by providing a counterexample of two tentative prisoners who could repurchase their freedom for 4000 ducats. A rich prisoner who already owns 2000 ducats may better appreciate a lottery with an eventual gain of 2000 ducats than his poorer fellow. This counterexample of increasing marginal utility requires a definition of a utility function u(.) on various eventual nonmonetary attributes (e.g. freedom). However, it seems plausible that decision makers cannot define their utility function through introspection. This causes an assessment problem for preferences.

The theory of revealed preference (Samuelson, 1938) presumed that preferences straightforwardly follow from choice behaviour provided that a decision maker's choice behaviour is consistent. So, a decision maker should every time choose for x_i from x_i, x_j available to reveal his preference. The theory of revealed preference attracted attention because of its potential to observe preference through choice behaviour whilst excluding an assessment of the decision maker's state of mind (Little I. , 1949). Although the surmise that preferences are revealed by choice behaviour seems innocuous, its validation is highly problematic (Sen, 1973), (Beshears, Choi, Laibson, & Madrian, 2008), (Gruene, 2006). To illustrate the point, decision makers frequently choose x_i

from x_i, x_j available and consecutively choose x_j from x_i, x_j available. Still, it remains undecidable whether this observation falsifies the theory of revealed preference or whether this observation is just an inadmissible case of inconsistent choice behaviour. Sen (1973) therefore argued that the theory of revealed preference requires an assessment of consistency that is founded on a decision maker's state of mind. Little's (1949) idea that choice behaviour alone could reveal preferences is therefore disputable.

Neumann and Morgenstern (1944) alternatively proposed to deduce preference from probabilities rather than from choice behaviour. We revisit Bernoulli's (transl. 1954) example to illustrate this proposal.

Example

Let both the rich and the poor decision maker rank their preference for some optional wealth increase as follows:

$$(u(100) > u(92) > u(0)) \equiv (u(0) < u(92) < u(100))$$
¹⁹

Then, the continuity axiom (Neumann & Morgenstern, 1944) implies:

$$(u(100) > u(92) > u(0)) \to ((\alpha \times u(100)) + ((1 - \alpha) \times u(0)) > u(92)) \exists \alpha: 0 < \alpha < 1 (u(0) < u(92) < u(100)) \to ((\alpha \times u(100)) + ((1 - \alpha) \times u(0)) < u(92)) \exists \alpha: 0 < \alpha < 1$$
 20

which means that there must be some tentative probability α where a decision maker is indifferent towards a choice of action a_1 to participate in a lottery that yields a wealth increase of either 100 or 0 and a choice of action a_2 that yields a certain wealth increase of 92. In Bernoulli's (transl. 1954) example, the decision maker's current wealth determines the strength of a preference for the outcomes {0,92,100} in Equation 20. This special case of known utilities allows us to calculate α :

$$\begin{aligned} \left(u(100) > u(92) > u(0) \right) \\ \to \left(\left(\alpha \log_{10}(x+100) \right) + (1-\alpha) \log_{10}(x+0) > \log_{10}(x+92) \right) \\ \left(u(0) < u(92) < u(100) \right) \\ \to \left(\left(\alpha \log_{10}(x+100) \right) + (1-\alpha) \log_{10}(x+0) < \log_{10}(x+92) \right) \end{aligned}$$

The poor decision maker then becomes a risk seeker at α >0,98 and a risk avoider at α <0,98. The rich decision maker then becomes a risk seeker at α >0,92 and a risk avoider at α <0,92. Since the true probability to win the lottery is 0,95 in Bernoulli's (transl. 1954) example, it is evident that the poor decision maker is risk averse and that the rich decision maker is risk seeking. In Bernoulli's (transl. 1954) special case, the utility that a decision maker assigns to an outcome {0,92,100} is known to be the logarithm of wealth. Neumann and Morgenstern (1944) conversely proposed to calculate the unknown utility from some assessment of α . As opposed to Bernoulli's (transl. 1954) representation of preference by the logarithm of wealth, Neumann and Morgenstern's (1944) representation of preference (by α) confines to the three outcomes {0,92,100} available and it cannot be infinite since α is in <0,1>.

Neumann and Morgenstern (1944) acknowledged that the tentative probability α should be somehow assessable to operationalise the strength of a preference. De Finetti (2008) alleged that it is rational to let the "true" probability then determine choice behaviour as shown in the example above. This "true" probability may follow from observed frequencies. In practice, decisions are often made in ignorance of probabilities and observed frequencies are unattainable for decisions that are deemed as unique.

Savage (1954) extended on Neumann and Morgenstern (1944) by allowing a decision maker to be ignorant about the probability of the effect X. Savage (1954) posed an

action A to be a function that attaches an effect X to some state of the universe S as shown in:

$$A(S) \to X \tag{22}$$

Savage's (1954) omelette example illustrates the meaning of Equation 22.

Example

A decision maker chooses between action a_1 and action a_2 .

| Action (a ₁) | Add a sixth egg that is either good $(S=s_1)$ or bad $(S=s_1')$ that results in a six egg |
|--------------------------|---|
| | omelette $(X=x_1)$ or a wasted omelette $(X=x_1)$. |
| Action (a ₂) | Do not add a sixth egg which results in a five egg omelette and one wasted egg |
| | $(X=x_2)$ |

The preference ordering of X has been given by:

$$(u(x_1) > u(x_2) > u(x_1')) \equiv (u(x_1') < u(x_2) < u(x_1))$$
23

If the "true" probability and the subjective α of X=x₁ were known, the solution to this decision problem would have followed from Equation 20 in line with Neumann and Morgenstern (1944). Knowing the causal mapping from Equation 22, this decision problem may alternatively be solved by knowledge about the current state of the universe S.

$$E[U|a_{1}] = \sum_{j=1}^{n} u(x_{j}) \times pr_{s}(s_{j})$$

$$E[U|a_{1}] = pr_{s}(s_{1}) \times u(x_{1}) + pr_{s}(s_{1}') \times u(x_{1}')$$

$$E[U|a_{2}] = u(x_{2})$$
24

The probability in Equation 24 (Savage, 1954) has been built on the current state of the universe S whereas the probability in Equation 15 (Bernoulli, transl. 1954) and the tentative probability α have been built on a prospective effect X. In principle, uncertainty about the current state of the universe S is eliminable by observing at the moment of deciding whereas prospective effects X remain a matter of belief at that time. However, this potential advantage of Savage (1954) over Bernoulli (transl. 1954) and over Neumann and Morgenstern (1944) comes at the expense of problems to assess the causality in Equation 22.

Jeffrey (1974) opposed Savage's (1954) idea that actions are mappings from states to effects as shown in Equation 22. In his view, Equation 22 is a function that maps an infinite amount of possible states of the universe to definite effects. A decision maker may eventually learn in retrospect how his action associates to the current state of the universe, but he cannot be expected to know the effects of his action on every possible state of the universe. Jeffrey (1974) asserted that Savage's (1954) omelette example is delimited by a finite number of states, by a finite number of effects and by some indepth knowledge about their mapping.

Jeffrey (1974) alternatively posed that decision makers and their actions are an integrated part of the universe that may be represented by some set of propositions. So, an action (A=a_i) and an effect (X=x_j) are merely labels on propositions whose association does not require an explanation like $A \rightarrow X$.

Jeffrey (1974) commented that most of the propositions about the state of the universe are beyond a decision maker's span of control. For example, a decision maker cannot influence the probability of rain. Still, he may or may not take an umbrella. Both the weather X and his action A are ingredients of the state of the universe to be evaluated by Equation 25.

$$E[U|a_i] = \sum_{j=1}^{n} pr_{X|A}(x_j|a_i) \times u(x_j, a_i)$$
25

Jeffrey's (1974) evidential decision theory and its ramifications omit a notion of causality. Equation 25's conditional probability merely expresses to what extent the values of an effect X and an action A coincide. So, evidential decision theories omit knowledge about causalities. Newcomb's problem (Nozick, 1969) illustrates the difference between causal and evidential decision theories.

Example

A decision maker chooses for either action a_1 or action a_2 .

Action (a_1) Take box 1 that either contains \$0 or \$100.

Action (a₂) Take both box 1 that either contains \$0 or \$100 and box 2 that certainly contains \$1 An oracle that appeared to be correct in predicting effects from choices of many decision makers in many cases, alleges that the content of box 1 depends on the decision maker's choice. It will contain \$100 upon action a_1 and \$0 upon action a_2 . Evidential decision theory takes the oracle's past predictions into account since there appears to be some unexplained association between the oracle's prediction and reality. Then, expected utilities follow from:

$$E[U|a_1] = pr(100|a_1) \times u(100) + pr(0|a_1) \times u(0) \approx u(100)$$

$$E[U|a_2] = pr(101|a_2) \times u(101) + pr(1|a_2) \times u(1) \approx u(1)$$
26

Causal decision theory additionally involves an explanation for the association between the oracle's prediction and the decision maker's choice. Eventually, the decision maker still believes that his choice can never cause the content of box 1, despite the observed frequency of correct oracle predictions in the past. Then, the content of box 1 is predetermined regardless of the oracle's prediction and expected utilities follow from Equation 27.

$$E[U|a_1] = pr(100) \times u(100) + pr(0) \times u(0) = \alpha \times u(100)$$

$$E[U|a_2] = pr(101) \times u(101) + pr(1) \times u(1) = \alpha \times u(101) + (1 - \alpha) \times u(1)$$
27

The action a_1 yields a lower expected utility in Equation 27 (E[U| a_1]<E[U| a_2]) irrespective of the unknown probability α whereas the action a_1 yielded a higher expected utility in Equation 26 (E[U| a_1]>E[U| a_2]) due to the oracle's claim regarding the probabilities. This means that a decision maker aiming exclusively for wealth favours action a_2 under causal decision theory and action a_1 under evidential decision theory. Therefore, Newcomb's problem evokes decision indeterminacy.

For prospective decision making, causal decision theories (Gibbard & Harper, 1981), (Lewis, 1981) often have more intuitive appeal than evidential decision theories. Evidential decision theories only require a coinciding of an action a_i and an effect x_j across cases while omitting an explanation. So, evidential decision theory just considers statistical associations irrespective of whether they are explained by oracles or any background variable. Causal decision theories are more compelling because they presume an effect that is uniquely attributable to a specific action. A maintenance policy validation that confirms the causality $L \rightarrow K$ is more coercive, but the association $Pr(K|L) \neq Pr(K)$ would have sufficed already under evidential decision theory.

A cause could be sufficient or necessary to produce an effect. A sufficient cause satisfies: "If I had done a, then x would have occurred" whereas a necessary cause satisfies: "If I hadn't done a, then x would not have occurred". If an action is only a

sufficient cause for an effect, it may be vulnerable to pre-emption as shown in the following example (Lewis, 2000):

Example

Billy and Suzy throw rocks at a bottle. Suzy throws first, or maybe she throws harder. Her rock arrives first. The bottle shatters. When Billy's rock gets to where the bottle used to be, there is nothing there but flying shards of glass. Without Suzy's throw, the impact of Billy's rock on the intact bottle would have been one of the final steps in the causal chain from Billy's throw to the shattering of the bottle. But, thanks to Suzy's pre-empting throw, that impact never happens.

Hitchcock (2013) argues that this example may evoke problems to assess the cause of the shattering of the bottle in retrospect. In the case of pre-emption, the maintenance policy validation would similarly fail to identify an effect of maintenance policy compliance. However, causal decision theory advises Suzy correctly for her *prospective* decision about whether to throw the rock or not. Equation 28 explains that Suzy should be indifferent towards throwing or not throwing if she fully trusts Billy's destructive capabilities (β =1). However, if she does not entirely trust Billy (β <1), she may increase the prospect of a shattered bottle by her own probability α to shatter the bottle.

$$E[U|Throw] = (\alpha + \beta - \alpha\beta) \times u(Shattered) + (1 - \alpha) \times (1 - \beta) \times u(\sim Shattered)$$

$$E[U|\sim Throw] = \beta \times u(Shattered) + (1 - \beta) \times u(\sim Shattered)$$
28

In conclusion, we introduced various representations of preference that could drive or at least explain a decision to carry out maintenance. Bernoulli (transl. 1954) still sought for some model that deduces preference from a prospective outcome but he already realised that these models are disputable. Revealing preference from choice behaviour (Samuelson, 1938) seemed intuitively appealing, but its validation suffered from difficulties to establish *consistent* choice behaviour. We reconcile with Sen (1973) that choice behaviour alone is insufficient to assess preference. Neumann and Morgenstern (1944) alternatively tried to deduce preferences from probabilities. Unfortunately, these probabilities of future effects often appear to be problematic as far as assessability is concerned. Savage (1954) tried to build preferences on the potentially more assessable current state of the universe, but he needed an equally problematic notion of causality to succeed. We also referred to evidential decision theory that may reduce the aim of this work (Section 1.3) to an improved inference of:

Maintenance policy compliance and functionality are associated.

The operationalisation of a causality as introduced in Section 2.3 would then become superfluous. Still, causal decision theories are intuitively more appealing for a decision maker who typically wants to know the prospects of his specific choice. Moreover, Newcomb's problem showed that a decision maker who believes in a non-causality, typically involves this belief in his preference assessment. Even if this belief has been troubled by the observed frequency of correct oracle predictions. We also presented a case of pre-emption to illustrate that a causal decision theory may still advise a decision maker correctly about his prospects despite its validation issues in retrospect. So, decision makers do not exclusively rely on validatable knowledge to assess their preference. On the other hand, the preferred effect of decisions to carry out maintenance should be somehow observable for a meaningful justification of maintenance. We

therefore believe that a maintenance policy validation can be appreciated by scientists as well as by maintenance decision makers.

Normative decision theory did not appear to be prescriptive about what to prefer. For example, a preference for the destruction of the whole world to the scratching of my finger (Hume, 1739) is permissible. However, concerns about the operationalisation of preference are alleviated in the case of collaboration. Collaboration requires that decision makers align their individual preferences somehow with the group's preference. Some common sense about this group's preference must then exist. Since a maintenance policy is typically assessed by a collaborating group, scepticism about the operationalisation of preference may appear to be surmountable.

Normative decision theory typically presumes that "choosing" and "doing" coincide and correspond. However, in maintenance, neither is the case; i.e. the choice to admit an action to the maintenance workload (gatekeeping) may be far ahead of its execution and this execution in its turn may not correspond with the original choice. This lack of maintenance policy compliance may reveal a counterfactual effect that was intended to be avoided. Establishing this counterfactual effect may appear to be a suitable maintenance policy validation that, to the best of our knowledge, has not yet been available in the existing literature on normative decision theory.

3.2 Review of maintenance performance measurement

Maintenance decision making typically involves groups of decision makers. If every group member could independently pursue his disputable preference, the group's preference may not be obtained. The existence of any organisation stems from the choice to collaborate. As illustrated by the prisoner's dilemma, the group's preference would have become easier to attain if the individuals had been able to align their goals. To enable this alignment, an organisation often clarifies its goals in terms of performance indicators (Kaplan & Norton, 1996), (Drucker, 1954). This section will therefore seek common sense about maintenance performance indicators from a review of maintenance scorecards. Furthermore, we evaluate whether some maintenance performance measurement conventions would enable causal inferences.

The definition of maintenance in Section 1.4.3 is informative about the intention of a maintenance policy. This definition does not claim that a maintenance policy has an observable functionality effect *by definition* because:

- The intentions may not be revealed;
- A *required* function is merely a subjective aspiration level instead of a well-defined measurable quantity.

On the other hand, if functionality were truly independent of the applied maintenance policy, functionality would not be a concern in maintenance policy assessments *and* maintenance would become indistinguishable from non-maintenance. In addition, if functionality were merely some subjective construct that lacks any common sense, it would also lack any explanatory power about reality. In this work, we therefore argue for some common sense about an observable functionality effect of a maintenance policy. In this section, we will analyse whether conventional maintenance performance indicators could provide this observable evidence.

Table 4 (Muchiri, Pintelon, Gelders, & Martin, 2011) depicts a typical maintenance scorecard (Haarman & Delahay, 2004), (Weber & Thomas, 2005), (CEN, 2007), (EFNMS; SMRP, 2011), (Blanchard, 2004), (Jones, 2007). The maintenance scorecard in Table 4 reconciles with the simplified maintenance scorecard in Table 1 since it comprises leading indicators for maintenance policy compliance L and lagging indicators for functionality K and resource costs C. Only the univariate representation of L,K,C in Table 1 has become highly dimensional in Table 4. Leading indicators are generally seen as causal for lagging indicators (Kaplan & Norton, 1996). So, a conventional maintenance scorecard potentially comprises the observable evidence for the maintenance policy validation.

| naintenance performance (result indicators) | Equipment performance | Number of failures Failure/breakdown frequency MTBF Availability Overall Equipment Effectiveness Direct maintenance cost Breakdown severity |
|---|-----------------------|---|
| | | Maintenance intensity Maintenance cost component over manufacturing cost Equipment replacement value |
| lg r ors | | Maintenance stock turnover |
| gir | | Cost of personnel |
| Lag | | Cost of subcontractors |
| — ·= | T1 / C / | Cost of supplies |
| Leading maintenance performance indicators (enabling indicators) | Identification | Proactive work |
| | | Reactive work |
| | | Work request response rate |
| | Planning | Planning intensity/rate |
| | 1 mining | Quality of planning |
| | | Planning responsiveness |
| | Scheduling | Scheduling intensity |
| | ~ | Ouality of scheduling |
| | | Schedule realisation rate |
| | Execution | Schedule compliance |
| | | MTTR |
| | | Manpower utilisation rate |
| | | Manpower efficiency |
| | | Work order turnover |
| | | Backlog size |
| | | Quality of execution/rework |
| | Execution | Quality of scheduling Schedule realisation rate Schedule compliance MTTR Manpower utilisation rate Manpower efficiency Work order turnover Backlog size Quality of execution/rework |

Table 4 Instance of a maintenance scorecard

We follow Simon's (1997) observation that an organisation's performance relies on "doing" as much as on "choosing". So, it relies on directing as well as on executing. Section 3.1 already mentioned that normative decision theorists often pose "choosing" and "doing" as indistinguishable. Then, the resource costs C of "doing" would just become a logical rather than a causal effect of a choice. We exclude a definitional debate about whether resource costs should be seen as the price of a choice or as the causal effect of a choice. We confine this work to the validation of a causality between maintenance policy compliance L and functionality K. Since maintenance scorecards typically comprise both a leading indicator L and a lagging indicator K, they may

provide the empirical evidence for a maintenance policy validation. We will proceed on this evidence.

The maintenance scorecards that we analysed are merely used to show posterior compliance to some requirement. However, decisions can only influence the future. So, predicting performance indicators, given a particular action, is what is really needed to accommodate *data driven* decision support. Of course, the current state of the universe could also support decisions as we illustrated by Savage's (1954) omelette example, provided that we also somehow know the causal mapping to future effects (Equation 22). The maintenance policy validation pursues observation of this causality in recording routines.

Possibly, maintenance scorecards could support decisions in a better way. Fluctuations rather than steadiness may reveal causalities that allow us to learn about the system behaviour. Many performance indicators on the typical maintenance scorecard in Table 4 are averages that irreversibly levelled out the informative fluctuations that we need. We therefore suspect that we could potentially learn much more from the underlying recording routines of Table 4.

We analysed Keeney and Raiffa's (1976) desired properties for utility attributes that appear to be ignored in Table 4. In the remainder, we will put forward three construction rules for performance indicators that serve the maintenance policy validation and that are compatible with these desired properties of utility attributes (Keeney & Raiffa, 1976). These construction rules are therefore expected to serve prospective decision making as well as the posterior maintenance policy validation. These construction rules are: (*i*) avoid redundancy, (*ii*) sample at a sufficient rate and (*iii*) balance completeness with efficiency. We will assess whether these construction rules indeed contribute to inference precision in a typical realistic case study (Chapter 5).

3.2.1 Avoid redundancy in performance indicators

A redundant variable is a variable that logically depends on another variable. For example, let the availability, which is one of the indicators in Table 4, be defined as:

$$\bar{A} = \frac{MTBF}{MTBF + MTTR}$$
29

Then, availability would reduce to a deductive consequence of MTBF and MTTR by definition, while the latter two quantities are also separate performance indicators in Table 4. This means that redundancy exists in this maintenance scorecard. That is undesirable, since redundant indicators both contribute to the overall utility function in a similar way. Moreover, a dependence between MTTR and availability should in this case not be interpreted as a causal relation between a leading and a lagging performance indicator. Rather, it is just a logical dependence following from the availability definition in Equation 29.

Another form of redundancy is longitudinal redundancy, which occurs when performance indicators are deduced from an interval that exceeds the sampling interval. For example, let the MTBF in Table 4 be calculated from failure data collected over the last year, whereas the sampling interval for the scorecard is just a month. Then, two consecutive year-to-date MTBF's share 11 out of 12 months of evidence. Any autocorrelation in the MTBF signal, which normally might indicate a (causal) relation between the failures in the past and in the present, is probably entirely definitional. Moreover, this autocorrelation is still uninformative for decision making. Only next month's (limited) contribution to this year-to-date MTBF can be controlled to some extent by a suitable maintenance policy, whereas the remaining contribution (of the 11 preceding months) is a known accomplished fact that cannot be controlled anymore, and should thus be omitted from a utility function u(.).

Longitudinal redundancy is often avoidable. Lead times and queues for example are related by Little's (1961) law that asserts that a time T limiting average of a queue D follows from an arrival N limiting average of a lead time W under the condition that a time limiting average arrival rate λ exists (Stidham, 1974).

$$(\overline{D} = \lambda \times \overline{W}) \equiv \left(\lim_{t \to \infty} \frac{1}{T} \sum_{T=1}^{t} D_T = \lim_{t \to \infty} \frac{N_T}{T} \times \lim_{n \to \infty} \frac{1}{N} \sum_{N=1}^{n} W_N\right)$$
30

A conversion from the instantaneously observed queues D_T and arrivals N_T to the averages in Equation 30 is irreversible. To avoid longitudinal redundancy, we favour an instantaneously observable queue D_T over its corresponding lead time W_N . Lead times rather than queues seem conventional in maintenance performance measurements, but the underlying recording routines may still enable a reconstruction of the queue that corresponds with that lead time. In conclusion, conventional maintenance scorecards could be improved by eliminating logical dependencies.

3.2.2 Sample at a rate that allows reconstruction of the original signal

We did not find any guidance in the existing literature on maintenance performance measurement about sampling rates. However, maintenance scorecards are being sampled at some rate in practice, typically weekly, monthly or quarterly. It seems that a sampling rate is merely opportunistically determined by a meeting frequency or by the effort required to assess a maintenance scorecard.

Adjacent observations often appear to be dependent if they are close enough in time. So, the dependence between adjacent observations follows from the sampling rate. This means that the sampling rate determines whether or not it is possible to reconstruct the original signal. Time series analysis uses this dependence to predict some steps ahead in time (Box, Jenkins, & Reinsel, 2008). In its most simple setup, time series analysis may predict functionality just from its past values. Possibly, this functionality prediction benefits from injecting maintenance policy compliance into the time series model. In that case, the time lags of this predictive model attribute to causality as we illustrated in Figure 8.

We therefore suggest adopting the signal processing convention that prescribes to consciously choose a sampling rate that reconstructs the original signal. Fluctuations rather than averages allow us to learn about the system behaviour. So, it is the amplitudes in the candidate causes that eventually allow us to reveal a prima facie causality.

Instantaneously observable variables are, in principle, recordable at any rate without creating longitudinal redundancy. We therefore suggest substituting the rates (e.g. failure rate) and lead times (e.g. MTTR) in Table 4 by occurrences and queues respectively. Occurrences are phenomena that happen at some instant of time (e.g. a failure) and queues are counts at some instant of time (e.g. the number of faults). Time series of occurrences or queues are convertible to rates or lead times, but that process is irreversible. These levelling out operations therefore yield a loss of information, and should therefore only follow from an analyst's deliberations about an appropriate model. Even when a well-motivated data reduction is performed, still the original time series data should be stored, as that enables reanalysing the data with newly developed analysis methods. In conclusion, conventional maintenance scorecards could be improved by a consciously chosen sampling rate.

3.2.3 Balance completeness with efficiency

In an ideal operationalisation, functionality K and maintenance policy compliance L have a comprehensive univariate representation. However, Keeney and Raiffa (1976) confirmed a requirements engineering convention (Gilb, 2006), (Robertson & Robertson, 2006) that a decision maker's objectives are typically poorly accessible and quantifiable. The trawling for maintenance performance indicators therefore typically leads to a highly dimensional but still incomplete maintenance scorecard.

Since decision makers can only process a limited amount of information, they should prioritise the concerns they want to trace when selecting recording routines. Recording routines therefore just comprise a limited amount of variables that have been traced during a limited interval of time. Therefore, efficiency and completeness are far from trivial in this observational research.

Due to this efficiency constraint, we suspect the maintenance policy validation to be incomplete. We just resort to balancing completeness with efficiency. This balance may shift in time as maintenance scorecards are consistently and periodically recorded. Extended time series may then gradually allow for completer, higher dimensional models. Still, we may fail to collect sufficient evidence to validate even the simplest model within an acceptable time. An acceptable time is typically much shorter than the life of the item, the life of the operating organisation or the life of the analyst. For example high-impact-low-probability effects of a maintenance policy will therefore remain unassessible by the proposed maintenance policy validation.

None of our sources on maintenance performance measurement appeared to use their maintenance scorecards for prospective inferences or for retrospective policy validations. We therefore do not expect that the dimensionality of these maintenance scorecards has been based on the completeness-efficiency balance that we seek. This

work therefore provides a first look at the dimensionality of a maintenance performance management model whose validation is tractable.

Keeney and Raiffa (1976) similarly inferred prospectively what decision makers (should) decide. So, Keeney and Raiffa (1976) must also balance completeness with efficiency. Completeness is in their case needed to capture all concerns of the decision maker, but a minimum number of non-redundant, decomposable utility attributes is needed for efficiency. However, the maintenance policy validation does not rely on expert judgement but on recording routines. Therefore, the balance of completeness and efficiency in the two inferences may well differ.

Incompleteness just delimits the maintenance policy validation to some less comprehensive notions of functionality and maintenance policy compliance. We simply accept that some components of functionality or maintenance policy compliance remain background variables. These background variables are just a subset of the information in the universe Ω_{T+1} that we need for causal claims (see Section 2.3). From this perspective, incompleteness is no more than a contributor to the burden of operationalising causality.

3.3 Review of maintenance policy assessments

A maintenance policy assessment precedes any maintenance policy validation. This section will therefore introduce maintenance policy assessments that pursue a preferred but yet to be observed future. So, maintenance policy assessments apply some modus ponens inference of prospects. A modus ponens inference requires a presumed model that infers the effects of a decision. It is possible that a maintenance policy validation can be reduced to a confirmation of this presumed model which maintenance decision makers were willing to accept. Moreover, the maintenance policy validation may appear to appreciably give support to maintenance policy assessments.

In Section 3.3.1, we will explain that decision makers poorly apply maintenance optimisations. This is plausible, because decision makers are simply not able to find the global optimal maintenance policy among all possible alternatives. Maintenance optimisations predominantly rely on expert judgement. We will explain that there is little reason to believe that maintenance policies are optimal in practice and, moreover, that little effort has been spent on even observing the effect of a maintenance policy. The maintenance policy validation may potentially contribute to this essential empiricism.

In Section 3.3.2, we will introduce satisficing that reconciles far better with the actual decision making process while reducing the analysis burden. However, satisficing typically yields a suboptimal maintenance policy, which is problematic for the maintenance policy validation since policy violations do not necessarily deteriorate functionality. We will argue that there is no reason to pursue a maintenance policy validation of a presumed (optimisation) model that a satisficing decision maker refused to construct in the first place.

In Section 3.3.3, we will illustrate that a maintenance policy validation of a more modest independence presumption enables a simplified representation of the universe which is beneficial for both optimisers and satisficers. Moreover, a maintenance policy that merges satisfied and optimised decision rules may still be validated by an argument that is based on this independence presumption. This work will consider arguments that presume independence as well as arguments that presume dependence. The inference precision of the maintenance policy validation potentially benefits from the choice of either of these arguments.

Finally, Section 3.3.4 will introduce the reliability centred maintenance process that is frequently applied to maintenance policy assessments. The reliability centred maintenance process resembles satisficing rather than optimising and it is predominantly expert driven. The maintenance policy validation may appear to be a meaningful empirical extension to the reliability centred maintenance process.

3.3.1 Optimise maintenance

Maintenance optimisations presume that a maintenance policy causes effects. The most rewarding effect then follows from adherence to the optimal maintenance policy. Therefore, a review of maintenance optimisations may appear to be informative for the maintenance policy validation.

It does not seem very difficult to identify the objectives of a maintenance policy from the simplified maintenance scorecard in Table 1: continuous functionality at zero resource costs. In practice, a decision maker cannot find a maintenance policy that achieves these objectives. Then, a decision maker resorts to a policy that does not entirely fulfil his objectives. Optimisation is about balancing counteracting goals into an attainable compromise, rather than about fictitious objectives.

Optimisation presumes that decision makers are able and willing to collect and process all information needed to identify the attainable optimum. Decision makers often seem incapable of collecting all tentative alternatives, of assigning probabilities and utilities to their effects and of calculating some global maximum (if existing) of their expected utility. In a delimited universe of a fixed number of lotteries with known probabilities and prices like in Bernoulli's (transl. 1954) example in Section 3.1, optimisation may be adequate. In a realistic universe of an undetermined number of lotteries with unknown probabilities and prices, optimisation appears to be intractable (Simon, 1978). Unsurprisingly, Dekker (1996), Horenbeek et al. (2011), Doyle (2002) and Scarf (1997) acknowledged that maintenance optimisations are poorly used in practice. This may be because in a realistic universe optimisation imposes too heavy demands on a decision maker.

Horenbeek et al. (2011) claimed that case studies often only demonstrate the applicability of a developed model, rather than finding a solution to a specific problem of interest for a practitioner. More effort should be spent on putting theory into practice. Dekker and Scarf (1998) concluded that maintenance optimisations are often far from complete for practical decision making. Especially in the multiple component setting, we are only at the beginning. Horenbeek et al. (2011) similarly claimed that

maintenance optimisations are limited to very specific problems. Maintenance optimisations should be multiple objective optimisations whereas single objective optimisations seem dominant (Wang H., 2002), (Horenbeek, Pintelon, & Muchiri, 2011).

Maintenance optimisations have extensively been reviewed (Dekker, 1996), (Pham & Wang, 1996), (Dekker, Wildeman, & Duyn Schouten, 1997), (Wang H., 2002), (Nicolai & Dekker, 2008), (Wang W., 2012), (Sharma, Yadava, & Deshmukh, 2011). However, the library on maintenance optimisations has been growing faster than recordings of their implementations. The reasons for the limited implementation of maintenance optimisations may be found in the lack of data, the lack of knowledge about the models or the lack of maturity in maintenance management (Dekker, 1996), (Dekker & Scarf, 1998). Dekker (1996) expected that a better performance to costs ratio of computers, automatic data capturing and an increasing maturity of maintenance management would contribute to a wider implementation.

This survey indicated that maintenance optimisations predict a future from a given maintenance policy that is attainable. Maintenance optimisations therefore presume a causal model that predicts the future effect of a maintenance policy. This causal model typically relies on expert judgement about prospects. This survey revealed that there is little reason to believe that maintenance policies are optimal in practice. Still, the effects of maintenance optimisations should become somehow observable to make them meaningful to practitioners and scientists. The maintenance policy validation potentially contributes to this essential empiricism.

3.3.2 Satisfice maintenance

Simon (1955), (1956) proposed an alternative choice mechanism that is less demanding for a decision maker's psychological and physiological abilities than an optimisation. This choice mechanism may bear closer resemblance to the actual decision making process. A satisficing decision maker may just start by estimating the effects of the first alternative he arrives at by intuition. If the effects satisfy the requirements derived from his objectives, he stops deliberating. If not, he will continue to look for an alternative that meets his requirements. If it appears very easy to find alternatives that meet requirements, aspiration levels may rise. If requirements appear to be unattainable, aspiration levels lower. Satisficing does not prohibit optimisation in specific cases, it just alleviates some presumptions of optimisation:

- The set of actions is fixed; like in any design problem, generating a fixed set of actions that includes the optimal action often appears to be a burden. At the expense of missing the optimum, a satisficing decision maker just generates actions until he finds a satisfactory one.
- The expected utility, given an action, is known; the attributes of a utility function and its parameters are typically hard to assess. A satisficing decision maker may alleviate the burden of a precise introspection of his trade-offs by just assessing fulfilment of requirements.

- The decision maker should seek the optimal action; deducing an optimum from a highly dimensional model may require a huge calculating effort. A satisficing decision maker just assesses fulfilment of requirements.

Many others (Kahneman & Tversky, 1979), (Hammond, Steward, Brehmer, & Steinmann, 1975), (Hogarth, 1980) similarly recognised that maximising expected utility often does not adequately describe the actual process of decision making.

Satisficing appeals far more to our intuition about the actual process of maintenance policy assessments than optimising does. Satisficers only need to know *that* some effect fulfils a requirement depending on a decision whereas optimisers need to know *how* exactly some effect depends on a decision. A satisficing argument that only refutes an independence presumption may already suffice, whereas an optimising argument needs to confirm a specific dependence presumption. So, the satisficing argument is too modest to enable the functionality predictions that are essential for optimal decisions. Still, a satisficing argument that only relies on a modest independence presumption may already suffice for the maintenance policy validation.

Unlike optimising, satisficing does not require a universal model that specifies the strength and the direction of the causality between maintenance policy compliance and functionality. In fact, satisficing is problematic with regard to the second causality principle (Section 2.3.3) that asserts that causal relationships remain constant in direction throughout time. This is because maintenance policy violations may turn out to be opportunities which enable a satisficer to gradually improve the suboptimal maintenance policy. Satisficers just accept that *some* decisions to carry out maintenance may appear to be unjustified and in this strict sense, a satisficed maintenance policy is not expected to cause functionality. Still, a justified maintenance policy validation in this work could potentially reveal the adequacy of a satisficed maintenance policy. The better a maintenance policy approximates the optimum, the better maintenance policy compliance generally causes functionality.

If maintenance decision makers in practice prefer satisficing to optimising, scientists pursuing a maintenance policy validation may exhibit a similar attitude. Why would scientists be willing to validate a presumed universal causal model that maintenance decision makers refuse to construct in the first place? As a result, an argument that just tests for independence between maintenance policy compliance and functionality may adequately balance the analysis burden with the modesty of the maintenance policy validation. All candidate arguments that we will introduce in Chapter 4 are of use for satisficers, but only the candidate arguments in Section 4.1 and Section 4.2 presume a specific dependence between maintenance policy compliance and functionality that optimisers need.

3.3.3 Knowledge about non-causality assumptions

In Section 3.3.1 and Section 3.3.2, we suggested that satisficing rather than optimising seems to be the applied choice mechanism in maintenance policy assessments. Tentatively, the argument of the maintenance policy validation should just test for the

modest presumption:

Maintenance policy compliance and functionality are independent.

A test for this independence presumption serves the aim of this work in Section 1.3 as well as maintenance decision making. In this section, we will show that independencies enable a simplified representation of the universe that is beneficial for both satificers and optimisers.

Example

- Let the variables A,B,X,Y represent the universe Ω as depicted in Figure 9.
- Let A,B represent variables that are under the decision maker's control.

- Let X,Y represent the decision maker's objectives that a decision maker cannot directly control. Then, the decision maker may wish to know what X,Y he obtains when choosing A,B. So, the conditional probability Pr(X,Y|A,B) is important for the decision maker.



Figure 9 Path graphs of non-causality assumptions in some tentative universes

In the case Figure 9A reflects the true universe, A,B,X,Y are independent variables. The conditional probability Pr(X,Y|A,B) then reduces to:

$$pr_{X,Y|A,B}(x, y|a, b) = pr_X(x) \times pr_Y(y)$$

31

Equation 31 implies that A,B appear to be superfluous conditions for the joint distribution of X,Y. Control over A,B does not contribute to the decision maker's objectives whereas X,Y is a matter of destiny. In this universe, the decision maker does not need any model that assists in choosing the optimal A,B. Including A,B in some information set $V \subseteq \Omega$ does not reduce epistemic uncertainty about X,Y.

In the case Figure 9C reflects the true universe, $A \rightarrow XY$ and $B \rightarrow XY$. The conditional probability Pr(X,Y|A,B) then reduces to:

$$pr_{X,Y|A,B}(x,y|a,b) = pr_{X|A,B}(x|a,b) \times pr_{Y|A,B}(y|a,b)$$
32

Equation 32 implies that the decision maker cannot control X without affecting Y and vice versa. Control over the system is complicated since control over X,Y requires a simultaneous manipulation of A,B. In

this universe, the decision maker needs a complicated model that assists in choosing the optimal A,B. Including A,B in some information set $V \subseteq \Omega$ reduces epistemic uncertainty about X,Y.

In the case Figure 9B reflects the true universe, $A \rightarrow X$ and $B \rightarrow Y$. The decision maker has full control over X,Y by choosing A,B independently. In this universe, the decision maker may decompose the model that assists in choosing the optimal A,B into two simpler models: one to steer X by A and one to steer Y by B as shown in:

$$pr_{X,Y|A,B}(x,y|a,b) = pr_{X|A}(x|a) \times pr_{Y|B}(y|b)$$
33

Equation 33 implies that we could decompose the universe Ω into an information set V={a,x} and V'={b,y}.

In the extreme case of Figure 9A, any maintenance policy that controls A,B appears to be superfluous because A,B,X,Y are presumed to be independent. In the ideal case of Figure 9B, the objectives X and Y are independently controlled by A and B respectively. So, the optimisation problem is decomposable into two simpler optimisation problems $A \rightarrow X$ and $B \rightarrow Y$. Although Figure 9C requires less non-causality assumptions, the objectives X,Y are only attainable by simultaneous control over A,B. So, this more complex optimisation problem in Figure 9C is not decomposable. Therefore, this example showed that knowledge about independencies could reduce the optimisation burden.

This example showed that a universal independence presumption reduces the complexity of the model. In practice, the decision maker does not know the causal structure of the universe, but he may try to infer it. The simpler decomposable model from Equation 33 is more efficient to infer than the more complicated model from Equation 32. Spirtes (2000) and Pearl (2010) similarly reduced the burden of estimating a joint probability distribution by non-causality assumptions (Markov condition). Variables known to be independent can be eliminated from an information set V without an increase in epistemic uncertainty. Suh (2001) similarly illustrated the importance of knowledge about independencies to simplify a design.

A test for independence may reveal *that* functionality K depends on maintenance policy compliance L whereas a test for some dependence may reveal *how* functionality K depends on maintenance policy compliance L. Of course, a test for independence is too modest to be compelling for an optimal maintenance policy, but both optimisers and satisficers benefit from an allowance to simplify their representation of the universe. This allowance to simplify the universe also contributes to the sampling efficiency that we need for the maintenance policy validation.

3.3.4 Reliability centred maintenance process

Reliability centred maintenance seems to be a widely used process to assess a maintenance policy. Moubray (2004) defined reliability centred maintenance as:

A process used to determine what must be done to ensure that any physical asset continues to do what its users want it to do in its present operating context.

This definition tends to support a confirmation of the proposition:

Maintenance policy compliance causes functionality.

when labelling "what must be done" as maintenance policy compliance L and "continues to do what its users want it to do" as functionality K. Moubray (2004) departed from a quest for the item's functions whose performance standards capture the organisation's aspiration levels. This resembles the satisficing approach.

To be admissible to a maintenance policy, a candidate decision to carry out maintenance should pass a technical feasibility rule. Moubray (2004) has been very explicit about these technical feasibility rules that are based on a dichotomous functionality variable that defines the item's state by:

 $Y = \begin{cases} 0, \ Downstate \\ 1, \ Upstate \end{cases} (= required function violated) \\ (= required function fulfilled) \end{cases} 34$

A trigger that makes a decision to carry out maintenance technically feasible may be found in:

- The item's age (time based maintenance);
- The item's downstate (corrective maintenance);
- Any other variable (condition based maintenance).

Corrective maintenance is triggered by a downstate. Controversy about the technical feasibility of corrective maintenance only derives from controversy about the (subjective!) assessment of functionality Y in Equation 34.



Figure 10 Technical feasibility rule of time and condition based maintenance

Time and condition based maintenance are triggered by a steeply increasing hazard rate (Figure 10). A hazard rate follows from unobservable probabilities that are built on some prospective functionality Y.

Underpinning

A hazard rate is a limit, if it exists, of the quotient of the probability that the failure of a repairable item occurs within the time interval δt after a time t, when δt tends to zero, given that the failure has not

occurred in the interval between now (t=0) and the time t:

$$h(t) = \lim_{\delta t \to 0} \left(\frac{F(t + \delta t) - F(t)}{\delta t} \times \frac{1}{R(t)} \right) = \frac{f(t)}{R(t)}$$

$$35$$

In Equation 35, the failure density function f(t) and the reliability function R(t) have been built on a prospective dichotomous state variable Y:

$$h_{t} = \lim_{\Delta t \to 0} \left(\frac{pr_{Y_{T+\Delta T}}(0)}{pr_{Y_{[0,T]}}(1)} \right) = \lim_{\Delta t \to 0} \left(\frac{Pr(Y_{T+\Delta T} = 0)}{Pr(\bigcap_{i=0}^{T}(Y_{i} = 1))} \right)$$
36

Therefore, the hazard rate h_t is a ratio of probabilities on the values of $Y_{[0,T+\Delta T]}$. In the case of a maintenance policy assessment, these probabilities quantify the decision maker's degree of certainty about the future functionality.

In the case of time and condition based maintenance, a decision maker believes that he lowers the hazard rate in [t,2t] by a decision to carry out maintenance at a time t as represented by the dotted line in Figure 10. This lowering of the hazard rate is time bound; i.e. decision makers typically do not want to wait too long for a lower hazard rate. We therefore specifically seek short-term effects of maintenance policy compliance. Moreover, short-term effects also attribute to common sense about causality. Although spatiotemporal proximity is not among the causality principles (Section 2.3.3), a relation between flapping butterfly wings here and a remote hurricane weeks later is unlikely to be causal (Lorenz, 1972). This butterfly effect would typically require many mediating variables. Then, the flapping of the butterfly wing by itself does not uniquely explain this hurricane as required by the third causality principle (Section 2.3.3).

The prospective hazard rates in Figure 10 may be based on the item's age (time based maintenance) or on any other variable (condition based maintenance). Controversy about the technical feasibility of time and condition based maintenance may stem from controversy about the (subjective!) assessment of Y in Equation 34, but *also* from an inability to assess the probabilities for the hazard rate in Equation 36. Moubray (2004) claimed that time based maintenance is rarely technically feasible because a strong dependence between the hazard rate and time appears to be exceptional. In addition, we pose that time by itself is unlikely to be the cause of functionality Y. Then, time is just a symptom that coincidentally associates with the true causes of functionality Y that should be controlled by a technically feasible maintenance. Revealing causalities from condition recordings is a potentially interesting alternative application of the causal inference we pursue in this work.

These technical feasibility rules in a maintenance policy appear to be common sense (Birolini, 2007), (Crespo Marquez, 2007), (Kelly, 2006). So, decisions to carry out maintenance either stem from an observed downstate $Y_T=0$ or from some prospective hazard rate reduction. Maintenance policy violations and functionality seem therefore related in a maintenance policy that is technically feasible. In practice, it is typically applied to the suspected cost drivers or the performance killers that predominantly affect the organisation's aspirations. Resembling the satisficing approach, reliability centred maintenance appears to be an iterative journey to achieve continuous improvement.

However, the reliability centred maintenance process typically relies on fallible expert judgement whilst omitting a validation. The maintenance policy validation in this work may provide the essential empirical support to the reliability centred maintenance process.

3.4 Review of diagnostics

Diagnostics is directed to detecting symptoms of emerging item faults. By detecting and correcting at an early stage, we may avoid disastrous functionality effects. It is challenging to quickly identify symptoms of an emerging item fault among thousands of interacting components that are exposed to varying environmental loads. Diagnostic models may alleviate the burden of manual analysis.

Diagnostics essentially attempts to reveal the true item state from a set of candidate item states. Resembling the maintenance policy validation, diagnostics typically reasons from recording routines. Detecting symptoms that are closely associated with some item states already suffices for diagnostics. For example, some vibrations may diagnose a certain amount of fatigue damage without being the cause of fatigue. So, diagnostics typically solves an identification or a classification problem by revealing the most likely item state, but diagnostics often appears to be inexplicit about the decision to restore or preserve functionality. In many cases, this decision seems obvious by expert judgement. Therefore, it often does not matter that diagnostics remains inexplicit about causality. However, decisions that only control associated symptoms are insufficient. Therefore, the maintenance policy validation looks for an empirical confirmation of the causality between maintenance policy compliance and functionality. In this sense, the maintenance policy validation differs from diagnostic conventions.

Diagnostics matured into a vivid area of research with a wide variety of approaches. This concise introduction adopts the classification from Venkatasubramanian et al. (2003) and Chiang et al. (2001):

- Quantitative model based diagnostics (analytical);
- Qualitative model based diagnostics (knowledge based);
- History based diagnostics (data driven).

This classification does not partition diagnostic arguments into jointly exhaustive and mutually exclusive sets. Venkatasubramanian et al. (2003) did not find a diagnostic argument that universally outperforms all others, whereas the interest in hybrid arguments is increasing.

Similar to our approach to inference precision, Venkatasubramanian et al. (2003) suggested that the choice of a diagnostic argument heavily relies on the sampled evidence and on in-depth knowledge about the model. We do not expect to put forward an inference that is best suited to any maintenance policy validation.

Diagnostics supports operational as well as maintenance decisions. From the perspective of the item, it does not matter whether an operator or a maintainer performs the repair (Nakajima, 1988). Recordings of the maintenance workload may therefore appear to provide incomplete evidence that usually only captures the bigger and more specialised decisions to carry out maintenance.

3.4.1 Quantitative model based diagnostics

Quantitative model based diagnostics presumes an algebraic function that defines the interactions between the antecedents and the conclusion. Therefore, electromechanical systems whose working principles rely on physical laws are often suitable for quantitative model based diagnostics. Man-machine interactions appear generally more difficult to capture in algebraic functions. Quantitative model based diagnostics reasons from observer redundancy or from analytical redundancy. Observer redundancy implies that several sensors measure the same variable. So, observations should be equal. Analytical redundancy relates observations by physical laws. Conflicting observations are then symptomatic for emerging faults.

Quantitative model based diagnostics on large systems may require excessive modelling effort and sensor information may be incomplete. Quantitative model based diagnostics may well perform in predicting behaviour that has not yet been observed. We refer to Frank et al. (2000), Isermann (2005) and Venkatasubramanian et al. (2003) for a more in-depth discussion.

3.4.2 Qualitative model based diagnostics

Qualitative model based diagnostics presumes a qualitative function that defines the interactions between the antecedents and the conclusion. Qualitative functions are less precise than algebraic functions since they are of a categorical nature. They take for example values like (more, equal, less) or (normal, abnormal). At the expense of precision, qualitative model based diagnostics may efficiently approach complex systems that comprise human interference. The applicability of qualitative model based diagnostics is confined to the values of the model's categorical variables. So, it cannot reason about subsets within a category.

Qualitative model based diagnostics supports abductive, inductive and deductive arguments. By abductive reasoning, we arrive at faults that could explain the symptoms. Qualitative model based diagnostics then assists in finding a subset of plausible faults. By inductive reasoning, we arrive at faults that did explain the symptoms in the past. In this way, we may infer some rules that direct us to likely faults. By deductive reasoning, we may refute faults that cannot apply, given some symptoms. Then, we eliminate candidate faults, intending to end up with one remaining fault.

Instances of qualitative model based diagnostics are found in applications of directed graphs, fault tree analysis or qualitative physics. We refer to Venkatasubramanian et al. (2003), Kleer and Brown (1984), Vesely (2002) and Gao et al. (2010) for a more extended introduction.

3.4.3 History based diagnostics

History based diagnostics does not require any a priori knowledge about the model parameters. It just infers associations among past observations. An association does not suffice for a cause, as we saw in Section 2.3. This means that history based diagnostics alone is not compelling for causality. It is unable, therefore, to reason about unobserved changes in the structure of the system or its environment. So, history based diagnostics is vulnerable to future (beyond sample) changes in background variables that delimit any inductive argument. History based diagnostics may comprise qualitative solutions like expert systems (Rich & Venkatasubramanian, 1987) and qualitative trend analysis (Venkatasubramanian, Rengaswamy, Kavuri, & Yin, 2003), (Cheung & Stephanopoulos, 1990), (Villez, Rosen, Anctil, & Vanrolleghem, 2013). Quantitative solutions comprise neural networks (Zorriassatine & Tannock, 1998) and statistical inferences (MacGregor & Cinar, 2012).

3.5 Review of prognostics

Prognostics has many applications including biostatistics, econometrics, operational research, reliability, materials science (Si, Wang, Hu, & Zhou, 2011). In a maintenance decision making context, prognostics may infer prospective functionality from a given choice to maintain or not to maintain. Prognostics therefore comprises some modus ponens reasoning about prospective functionality from given antecedents about the past and present and some type of predictive model. Tentatively, the maintenance policy validation can be reduced to a validation of a prognostic model in retrospect. In this section, we will explain that (*i*) prognostics typically represents prospective functionality by a remaining useful life, that (*ii*) prognostics typically takes physical variables rather than maintenance policy compliance as antecedents and that (*iii*) prognostics does not require its antecedents to cause prospective functionality as required by the maintenance policy validation.

All definitions of prognostics in Sikorska et al. (2011) refer to a dichotomous functionality variable as shown in:

 $Y = \begin{cases} 1, & Life, upstate, state of functioning, ... \\ 0, & Death, downstate, state of failure, ... \end{cases}$

37

Si et al. (2011) recognised that Y is subjective and dependent on context and operational characteristics. Si et al. (2011) did not seek for common sense about Y since their main interest was directed to the modelling methods for remaining useful life (RUL) estimations. A maintenance policy validation, however, requires common sense about functionality Y.

Prognostics may in principle predict any future trajectory of Y, but to serve efficiency, we delimit to a remaining useful life $Y_{[1,T]}=1$ which is often of primary interest. Reliability expresses our degree of certainty about surviving a specific time interval (CEN, 2001), (ARMP-7, 2008), (IEC, 1990):

Reliability is the ability of an item to perform a required function under given conditions for a given time interval.

Uncertainty about a remaining useful life seems to be predominantly approached in a probabilistic way. We omit other approaches to uncertainty (Halpern, 2005), (Aven, 2011). We therefore quantify the "ability" in this reliability definition by a probability function built on a remaining useful life as shown in:

$$R_{[1,T]} = Pr\left(\bigcap_{i=1}^{T} Y_i = 1|U\right)$$
38

In a similar way to causality, the assessment of a probability is encumbered with problems. To explain a decision maker's maintenance policy prospectively, a probability may follow from the decision maker's beliefs about the future as we explained in Section 3.1. To validate this probability retrospectively, we typically use observed frequencies.

The body of knowledge U in Equation 38 is assignable to some fictitious individual. This fictitious individual may for example involve the item's current age to assess his functionality prospects:

$$R_{[T,T+N]} = Pr\left(\bigcap_{i=T}^{T+N} Y_i = 1 | U = \left\{\bigcap_{j=1}^{T-1} Y_j = 1, \dots\right\}\right)$$
39

A remaining useful life in Equation 39 reflects the item's survival till T+N, given that it survived up to T. Jardine et al. (2006) identified that a remaining useful life is the most obvious and widely used predictor variable in prognostics. A remaining useful life indicates how much time is left before a failure occurs.

Banjevic (2009) mentioned that a full distribution of a remaining useful life is sometimes just characterised by its expectation. An alternative for remaining useful life modelling is delay time modelling (Christer, 1973) that infers an expected number of failures given some inspection interval. Wang (2012) reported delay time modelling extensions to multiple component items, to imperfect inspections or repairs and to multiple inspection intervals. Applications of delay time modelling were found in building, manufacturing, energy production, transportation and electronics. Baker and Wang (1991) and Christer et al. (1995) presented some applications that depart from observable evidence. Since delay time modelling is confined to inspections which are only part of a maintenance workload, we will not proceed in this direction.

If reliability $R_{[1,T]}$ is a decreasing continuous function r(t), it is exchangeable with a hazard rate h(t):

$$h(t) = \lim_{\delta t \to 0} \frac{-(r(t+\delta t) - r(t))/\delta t}{r(t)} = \frac{-r'(t)}{r(t)}$$
40

The proportional hazard model (Cox, 1972) and its ramifications (Gorjian, Ma, Mittinty, Yarlagadda, & Sun, 2010) are based on the hazard rate h(t). A hazard model does not require knowledge about the distribution of the item's entire life in the same way as a reliability model would do. Rather, a hazard model allows to be instantaneously adjusted by unforeseen events as they occur during the item's life. Hazard models are better at coping with varying operating conditions than reliability models since they avoid longitudinal redundancy. Therefore, a hazard model complies better with the construction rules for maintenance performance indicators (Section 3.2) than a reliability model or a remaining useful life model. This work will reveal the feasibility of a maintenance policy validation which relies on a model that maps maintenance policy compliance to a hazard rate.

Sikorska et al. (2011) surveyed some classification schemes of reviews on prognostics and recognised little consensus. Reviews on prognostics tend to make predictions from physical variables (Foucher, Boullié, Meslet, & Das, 2002), (Sikorska, Hodkiewicz, & Ma, 2011), (Si, Wang, Hu, & Zhou, 2011), (Jardine, Lin, & Banjevic, 2006), (Peng, Dong, & Zuo, 2010). So, the body of knowledge U in Equation 39 typically comprises values of physical variables. These variables may be causal for the item's downstate (direct condition monitoring information) or just associated (indirect condition monitoring information). For risk identification alone, an association may suffice. But for decision making, control over a variable that is only associated is insufficient.

Abernethy (2006) warned that Equation 39 is already hard to validate empirically for a body of knowledge U that only comprises age due to a lack of replications. Extending this body of knowledge U with values of maintenance policy compliance L is expected to even make sampling efficiency worse. We also did not find prognostic applications that admitted maintenance policy compliance values to their body of knowledge U. Possibly, maintenance policy compliance values appear to be an unattractive element of the body of knowledge U.

3.6 Findings regarding the approach

In Section 1.4, we explained that our approach comprises a choice of (i) an argument, (ii) an operationalisation and (iii) a sampling procedure. We now survey the findings from the literature review regarding these choices.

3.6.1 Findings regarding the choice of an argument

We identified that maintenance stems from decisions. If decisions were independent of any future effect, decision makers would not bother about the choices they make. Decision makers who are bothered do at least have an intuitive causal model in mind which helps them to reason about prospective effects that matter. In Section 3.3, we introduced some practices of maintenance policy assessment from which we concluded that decision makers resist maintenance optimisations. Maintenance optimisations often impose demands on the decision maker's introspective and analytical capabilities that are too heavy. In practice, maintenance decision making seems to take place through a type of satisficing approach, which although alleviating analytical encumbrance, only results in an acceptable rather than an optimal policy. A maintenance policy validation that only relies on:

Maintenance policy compliance and functionality are independent

similarly omits the burden of model selection that satisficing decision makers resist taking. Moreover, knowledge about independence enables both *satisficing* and *optimising* decision makers to simplify their representation of the universe. This work will reveal the feasibility of a maintenance policy validation that relies on this presumed independence.

Decisions can only influence a yet to be observed future. So, decision making therefore resembles a modus ponens reasoning about prospects. Since the prospects are in the unobservable future, they cannot be validated a priori. However, a decision maker may learn from the past and the present to enhance assessment of his prospects. In this way, a maintenance policy validation could provide *data driven* support to decisions to carry out maintenance that are up until now have typically been *expert driven*.

Similar to selecting an appropriate diagnostic or prognostic argument, we do not expect that any single argument will universally outperform all others in inference precision. Depending on the case, a suitable argument needs to be sought iteratively. The candidate arguments for the maintenance policy validation will cover model based and history based approaches that take cardinal as well as categorical samples.

In Section 3.4 and 3.5, we showed that prognostics and diagnostics typically reason from physical variables that are often collected at high sampling rates. Common sense about the operationalisation of physical quantities and about the presumed (physical) models may wildly differ for the maintenance policy validation that relies on maintenance performance recordings and man-machine-models. As a result, prognostic or diagnostic arguments that have been thoroughly explored may seem to be inappropriate. Moreover, symptoms already suffice for a risk assessment, but gaining control over nothing else but a symptom does not suffice to achieve the targeted goal. The maintenance policy validation therefore takes a notion of causality in its argument that is not essential in many prognostic and diagnostic applications.

3.6.2 Findings regarding the choice of an operationalisation

Any decision maker may pursue his disputable preferences. In Section 3.1, we explained that normative decision theories typically suffer validation issues. A maintenance policy may seem to be an exception since it triggers decisions at a high rate and the abundant policy violations are typically also recorded. Maintenance policy compliance essentially measures to what extent "choosing" and "doing" correspond, which may enable us to distinguish the pursued effect from the counterfactual effect. Still, the criterion to identify maintenance policy compliance comprises a subjective requirement.

In Section 3.2, we extended the idea that decisions to carry out maintenance typically involve groups of decision makers. Any organisation stems from a choice to align

individual preferences with the group preference. To allow decision makers to align, organisations use performance indicators to define common sense about preference.

A conventional maintenance scorecard comprises leading and lagging performance indicators. The leading performance indicators may reflect common sense about maintenance policy compliance and the lagging performance indicators may reflect common sense about the pursued effect. So, the maintenance policy validation may use this common sense evidence. Although leading performance indicators are thought to cause lagging performance indicators, we suspect that they are merely used to show fulfilment of requirements in retrospect.

In Section 2.3, we already explained that the operationalisation of causality is particularly problematic under an observational research construct. Because it is often not possible to carry out well-constructed experimental research in the context of maintenance decision making, we resort to a modest notion of prima facie causality that uses time to raise credence in causality.

3.6.3 Findings regarding the choice of a sampling procedure

In a much better way than when using averages, variations can teach us about causal interactions. Nevertheless, conventional maintenance performance indicators are, typically, averages that level out all variations. This is why we deemed conventional maintenance performance indicators as inappropriate for causal inferences in Section 3.2. However, since decisions to carry out maintenance can only influence the future, a dithering decision maker will normally need to have more precise knowledge about prospective causal effects. In order to infer these causal effects more precisely from an organisation's recording routines, we put forward several construction rules for maintenance performance indicators.

This work will verify whether these construction rules for maintenance performance indicators do, indeed, infer (prima facie) causalities more precisely. Despite a correct application of these construction rules, we may still fail to collect sufficient evidence to validate any candidate argument. Eventually, we would existentially conclude that maintenance is unjustifiable.

4 Choice of an argument

Our approach to inference precision requires a set of candidate arguments as explained in Table 2. In this chapter, we will put forward these candidate arguments, which differ in structure and potentially, in inference precision as well. Although the inspiration for these candidate arguments came from typical inferences in reliability engineering, maintenance optimisation and maintenance prognostics, we do not intend to formally define these dynamic research areas using the arguments we present here. We have merely labelled these arguments according to their source of inspiration.

We will confine ourselves to arguments that comprise propositions that are observable by common sense. This delimitation is far from trivial, since many inferences in reliability engineering, maintenance optimisation and maintenance prognostics seem to use arguments that do not explicitly involve observable evidence. Rather, these inferences are only about presumed probabilities whose assessment has been ignored. However, because we are aiming here for a maintenance policy validation, what we require is observable evidence.

For brevity, we will present the candidate arguments as one-step-ahead predictions of functionality K_{T+1} but their extensions to enlarged information sets are straightforward. So, the candidate arguments as we present them, may validate the prima facie causality $L_T \rightarrow K_{T+1}$ with respect to an information set $V = \{l_t, k_t, k_{t+1}\}$:

$$\left(pr_{K_{T+1}|L_{T},K_{T}}(k_{t+1}|l_{t},k_{t}) \neq pr_{K_{T+1}|K_{T}}(k_{t+1}|k_{t}) \quad ; \exists l \forall t \forall k\right) \equiv (L_{T} \to K_{T+1})$$

$$41$$

Equation 41 is just an instance of the prima facie causality definition in Equation 5, given a body of knowledge $U=\{l_t,k_t\}$ that comprises two candidate causes for functionality K_{T+1} :

- Functionality K_T because the sampling rate should allow the original signal to be reconstructed. We therefore expect a dependence between K_T and K_{T+1} that could eventually reduce an association between L_T and K_{T+1} to a spurious cause resembling the case in Figure 8.
- Maintenance policy compliance L_T because this work is about a causality between maintenance policy compliance and functionality.

Ultimately, the composition of the information set V depends on the evidence available. The evidence may allow for extended information sets $V=\{l_t,k_t,k_{t+1},...\}$ but we may also fail to reach sufficient inference precision even at the reduced (minimal) information set $V=\{l_t,k_{t+1}\}$ that we will discuss in Section 5.4.3. In the latter case, we would existentially conclude that maintenance is unjustifiable.

An arbitrary choice of an information set V is known to hamper inference precision. For example, a decision maker could guess the functionality K_{T+1} of some item he does not know. In that case, this decision maker has no evidence on which to base favouring any value of functionality K_{T+1} . However, if this decision maker knew the item and its past, i.e. he had access to a more complete information set V, his guess might have differed wildly. In this example, it was the decision maker's knowledge, rather than the item
itself that determined the prediction of functionality K_{T+1} . Unsurprisingly, the subjective reliability estimates from operators and senior reliability engineers often wildly differ, since their knowledge about the item differs considerably.



Figure 11 Path graphs of the candidate arguments

The candidate arguments differ in their non-causality assumptions (missing arrows) as surveyed by the path graphs in Figure 11. These non-causality assumptions are required to interpret the inferred prima facie causality in Equation 41 as causal. For the relation between maintenance policy compliance L_T and functionality K_{T+1} :

- The maintenance optimisation argument (MOA in Figure 11) does not presume a specific relation between L_T and K_{T+1};
- The maintenance prognostic argument (MPA in Figure 11) presumes a universal model that *defines* either a causality or a non-causality between L_T and K_{T+1};
- The reliability engineering argument (REA in Figure 11) does not presume a specific relation between L_T and K_{T+1} , i.e. since L_T has been held constant, any dependence or even independence between L_T and K_{T+1} may apply;
- The nonparametric argument (NPA in Figure 11) presumes universal independence between L_T and K_{T+1}.

For the relation between functionality K_T and K_{T+1} :

- The maintenance optimisation argument (MOA in Figure 11) does not presume a specific relation between K_T and K_{T+1};
- The maintenance prognostic argument (MPA in Figure 11) presumes a universal model that *defines* either a causality or a non-causality between K_T and K_{T+1};
- The reliability engineering argument (REA in Figure 11) presumes a universal model that *defines* either a causality or a non-causality between K_T and K_{T+1};
- The nonparametric argument (NPA in Figure 11) does not presume a specific relation between K_T and K_{T+1}, i.e. since K_T has been held constant, any dependence or even independence between K_T and K_{T+1} may apply.

For the relation between maintenance policy compliance L_T and functionality K_T:

- The maintenance optimisation argument (MOA in Figure 11) presumes a universal equivalence between L_T and K_T *by definition* which reduces L_T and K_T to redundant variables;
- The maintenance prognostic argument (MPA in Figure 11) presumes universal independence between L_T and K_T; i.e. an eventual association between L_T and K_{T+1} is presumed to be unexplainable by a mediating or by a confounding K_T;
- The reliability engineering argument (REA in Figure 11) presumes universal independence between L_T and K_T; i.e. an eventual association between L_T and K_{T+1} is presumed to be unexplainable by a mediating or by a confounding K_T;
- The nonparametric argument (NPA in Figure 11) does not presume a specific relation between L_T and K_T ; i.e. an eventual association between L_T and K_{T+1} is unexplainable by mediation or by confounding of a constant K_T .

Finally, for the relation between the background variable B and the elements in the information set V:

- The maintenance optimisation argument (MOA in Figure 11) does not presume a specific relation between L_T, B (and K_T, B);
- The maintenance prognostic argument (MPA in Figure 11) presumes universal independence between L_T, B and K_T, B, but B may cause an error in the estimation of K_{T+1};
- The reliability engineering argument (REA in Figure 11) presumes universal independence between K_T and B, but B may cause an error in the estimation of K_{T+1} ;
- The nonparametric argument (NPA in Figure 11) presumes universal independence between L_T and B, but B may cause an error in the estimation of K_{T+1} .

The four variables B, L_T, K_T and K_{T+1} of the maintenance prognostic argument, the reliability engineering argument and the nonparametric argument tentatively allow for 2(3+2+1)=12 non-causality assumptions, i.e. missing arrows in Figure 11. The non-causality assumptions K_{T+1}→L_T and K_{T+1}→K_T seem innocuous by the first causality principle (Section 2.3.3) asserting that an effect cannot precede its cause in time and the existence of the causality K_{T+1}→B is just irrelevant in this work. However, a non-causality assumption with an unobserved background variable B cannot be validated empirically which affects inference precision. The reliability engineering argument and the nonparametric argument alleviate the burden of non-causality assumptions that involve B by holding some variable constant to rule out its eventual causal effects. However, a requirement of a constant implies an additional constraint on the composition of the sample (1,k)_[1,t] that may not be satisfied under an observational research construct.

A preference for the candidate arguments is expected to be driven by:

- The amount of in-depth knowledge about the non-causality assumptions in Figure 11;

- The efficiency to collect a sample $(l,k)_{[1,t]}$ that meets the requirements. From the candidate arguments in Figure 11, the maintenance prognostic argument appears to be most restrictive in its non-causality assumptions and most permissive about its sampling requirements. The nonparametric argument on the contrary, appears to be most restrictive in its sampling requirements and most permissive in its noncausality assumptions.

This chapter will introduce the candidate arguments by (i) explicitly stating their propositions, by (ii) discussing their claim about the prima facie causality in Equation 41, and by (iii) discussing their sampling issues. We conclude with a preliminary assessment of the inference precision of a maintenance policy validation based on these candidate arguments in Section 4.5

4.1 Maintenance optimisation argument

This section will introduce the maintenance optimisation argument by (*i*) presenting its propositions in Section 4.1.1, by (*ii*) discussing its claim regarding the prima facie causality in Equation 41 in Section 4.1.2 and by (*iii*) discussing its sampling efficiency in Section 4.1.3. We labelled this argument as the maintenance optimisation argument because we found many maintenance optimisations that presumed queue membership (present/absent) and functionality (upstate/downstate) as being equivalent. In that case, a policy to control the queue L_T equivalently controls functionality K_T .

4.1.1 Claim of the argument

This section will explain what the maintenance optimisation argument claims about the relation between its antecedent P1 and its conclusion C1.

- P1 $(L_T = l_t)$;Observe maintenance policy compliance l_t
- M1 $(L_T = l_t) \leftrightarrow (K_T = k_t)$;Presume that l_t and k_t are equivalent
- C1 $\therefore K_T = k_t$;Follows from P1,M1

| Valid argument: | Yes |
|------------------------|--|
| Functional relation: | Yes, this equivalence relation maps every l_t to a single value k_t |
| Common sense evidence: | - |
| Universal argument: | Yes, some stratified sample $(l,k)_{[1,t]}$ could universally refute the argument. |
| Decidable argument: | Yes, only the presumed model M1 is controversial. |
| | |

Figure 12 Maintenance optimisation argument

Figure 12 depicts the maintenance optimisation argument, where P1 indicates the proposition, M1 indicates the model and C1 the conclusion. The proposition P1 and the conclusion C1 straightforwardly follow from common sense about a sample $(l,k)_{[1,t]}$, but the presumed model M1 is controversial.

The stratified sample $(l,k)_{[1,t]}$ may existentially confirm the model M1 but it may also comprise a single counterexample that universally refutes the model M1. So, particularly in the case of a refutation, the maintenance optimisation argument is *decisive* about its presumed model M1.

The maintenance optimisation argument just presumes an equivalence relation between maintenance policy compliance L_T and functionality K_T but it does not presume anything about the future functionality K_{T+1} . The maintenance optimisation argument therefore lacks the predictive capabilities that are essential for decision support.

4.1.2 Claim about prima facie causality

This section will explain what the maintenance optimisation argument claims about the prima facie causality in Equation 41. An equivalence relation (like the model M1) is more compelling than the causality principles (Section 2.3.3). For example as opposed to an equivalence relation, a causality is not symmetric $(L \rightarrow K) \equiv (K \rightarrow L)$ and neither reflexive $(L \rightarrow L)$. Still, the maintenance optimisation argument is potentially informative about the prima facie causality in Equation 41. If sound, the maintenance optimisation argument would refute the prima facie causality in Equation 41 because maintenance policy compliance L_{T} is just a superfluous redundancy in an information set V={ $l_{t,k_{t,k_{t+1}}}$. Granger (1980) therefore explicitly prohibited redundant variables (like L_T, K_T) to be separate elements in an information set V. If unsound, the maintenance optimisation argument would allow for any dependence or independence relation between L_T and K_{T+1} . So, the maintenance optimisation argument cannot confirm the prima facie causality in Equation 41. However, we suspect that functionality and maintenance policy compliance are not related in the same way as centimetres and inches. So, we expect to find counterexamples that refute the maintenance optimisation argument. The presumed equivalence relation of the maintenance optimisation argument is therefore expected to be unnecessarily restrictive and not decisive about the prima facie causality in Equation 41. The path graph of the maintenance optimisation argument (MOA) in Figure 11 shows that the information set V_{MOA} simply comprises a single element that may relate in any way to the background variable B. So, the maintenance optimisation argument omits a need to arbitrarily operationalise a notion of causality.

4.1.3 Sampling issues

We suspect that a stratified sample $(l,k)_{[1,t]}$ suffices to universally refute the maintenance optimisation argument while leaving the prima facie causality in Equation 41 undetermined. The maintenance optimisation argument admits all information sets $V=\{l_t,k_t\}$ from the sample $(l,k)_{[1,t]}$.

4.2 Maintenance prognostic argument

This section will introduce the maintenance prognostic argument by (*i*) presenting its propositions in Section 4.2.1, by (*ii*) discussing its claim regarding the prima facie causality in Equation 41 in Section 4.2.2 and by (*iii*) discussing its sampling efficiency in Section 4.2.3. We labelled this argument as the maintenance prognostic argument because we found many attempts to reduce epistemic uncertainty about an item's remaining useful life by knowledge about physical variables. It is possible that we

might similarly be able to predict the item's functionality from maintenance policy compliance.

4.2.1 Claim of the argument

This section will explain what the argument claims about the relation between its antecedents and its conclusion.

- **P3** $(K_T = k_t, L_T = l_t)$;Observe functionality k_t and maintenance policy compliance l_t.
- M2 $P3 \rightarrow (\hat{K}_{T+1} = \hat{k}_{t+1})$;Presume a model M2 that estimates k_{t+1} from P3 ; If M2 has nonzero parameters for L_T , the second causality principle has been satisfied, i.e. a cause remains constant in direction throughout time ; If M2 has nonzero parameters for L_T , the third causality principle has been satisfied, i.e. a cause comprises unique information about the effect that is not available otherwise ; Note that the model M2 is more compelling than needed for the causality principles because it universally maps the antecedent P3 to a single conclusion C2.
- C2 $\therefore \hat{K}_{T+1} = \hat{k}_{t+1}$;Follows from P3,M2
- P4 $K_{T+1} = k_{t+1}$;Observe functionality k_{t+1} .
- P5 P4 → P3
 ; Presume this non-causality assumption P5.
 ; P5 satisfies the first causality principle, i.e. P4 cannot cause P3 since an effect does not precede its cause in time
- **P6** $(P4 C2) \nleftrightarrow P3$;Presume that prediction errors are independent of M2's domain, i.e. all information from L_T, K_T about K_{T+1} has been captured in M2's model parameters.
- M3 $(P3, M2, P5, P6) \rightarrow Pr(P4)$;Presume this definition of the probabilities in Equation 41. P5 just strengthens a belief in a causal interpretation of this probability.
- C3 ∴ Pr(P4)
 ;Concludes that any information set V={P3,P4} has been conceived as a replication. So, the maintenance prognostic argument holds for all l,k,t which is more restrictive than strictly needed for the prima facie causality in Equation 41.

| Valid argument: | Yes, but C2,C3,P5,P6 are not immediately observable | | |
|------------------------|--|--|--|
| Functional relation: | Yes, but M2,M3 map to conclusions C2,C3 that are not immediately observable. | | |
| Common sense evidence: | - | | |
| Universal argument: | No, some stratified sample $(l,k)_{[1,t]}$ does not suffice to observe the probability C3, but the maintenance prognostic argument addressed the | | |
| | three causality principles for $L_T \rightarrow K_{T+1}$ in M2,P5. | | |
| Decidable argument: | No, both the model M2 and the presumption P6 are controversial. | | |

Figure 13 Maintenance prognostic argument

The maintenance prognostic argument in Figure 13 comprises the presumptions M2,P5,P6,M3. Presumption P5 and model M3 do not appear to be controversial. Presumption P5 just reflects common sense about the causality principle that an effect P4 does not precede its cause P3 in time and model M3 appears to be a common sense operationalisation of a probability. Only model M2 and presumption P6 seem to be controversial. Controversy about both the model M2 and the presumption P6 regarding the errors makes the maintenance prognostic argument *undecidable*. This controversy could tentatively be settled by means of in-depth knowledge about a universal "manmachine-law" that determines the structure of model M2. In conventional prognostics, controversy about the model M2 is often mitigated by in-depth knowledge about "universal laws of physics". Presumption P6 in the maintenance prognostic argument alternatively denotes as:

$$(P4 = C2 + P6 = M2(P3) + P6) \equiv (k_{t+1} = f(l_t, k_t) + \varepsilon_{t+1}) \qquad ; \forall l \forall t \forall k \qquad 42$$

Equation 42 shows that the distribution of functionality P4: $K_{T+1}=k_{t+1}$ is the sum of the model M2: $f(l_t,k_t)$ and the error distribution P6: ϵ_{t+1} that are *universal*. The probability C3, which follows from the distribution of functionality P4: $K_{T+1}=k_{t+1}$ in Equation 42, will then require in-depth knowledge about the model M2 and the error distribution P6: ϵ_{t+1} . If we were to have this in-depth knowledge, the causality between maintenance policy compliance L_T and functionality K_{T+1} would have followed from a universal equation 42 and maintenance would have been justifiable. So, it is the controversy about the model M2 and the presumption P6 that is problematic for the justifiability of maintenance. However, we suspect to only know some stratified sample $(l,k)_{[1,t]}$. Then, the maintenance prognostic argument may *existentially* claim the likelihood of some *arbitrarily presumed* model M2 and some *arbitrarily presumed* error distribution P6. These existential claims are vulnerable to extensions of the sample $(l,k)_{[1,t]}$ and to more likely presumptions about the model M2 and the error distribution P6.

The maintenance prognostic argument is geared to reducing the epistemic uncertainty about future (beyond sample) functionality, i.e. reducing the error term ε_{t+1} . To reduce the error term ε_{t+1} , a dependence between the antecedent P3 and the proposition P4 must exist.

For prognostics to be effective, it is essential to know the nonzero parameters of the model M2 which map the antecedent P3 to the estimate C2. In-depth knowledge about the parameters of the model M2 enables reasoning about a yet to be observed future by the maintenance prognostic argument which is crucial for prospective decision making. In this work, however, we do not pursue a *prospective* inference but a *retrospective* validation.

4.2.2 Claim about prima facie causality

This section will explain what the maintenance prognostic argument claims about the prima facie causality in Equation 41. A universally sound maintenance prognostic argument could potentially claim a causality $L_T \rightarrow K_{T+1}$ because its model M2 and its presumption P5 cover all three causality principles (Section 2.3.3). The model M2 even specifies *what* maintenance policy compliance L_T uniquely attributes to functionality

 K_{T+1} whereas we only need to know *that* maintenance policy compliance L_T uniquely attributes to functionality K_{T+1} to justify maintenance. The maintenance prognostic argument therefore claims more than is strictly needed for the maintenance policy validation.

If sound, the maintenance prognostic argument would be decisive about a specific causality by the parameters of the model M2 rather than just about the existence of a prima facie causality between L_T and K_{T+1} . If unsound, the maintenance prognostic argument would only refute a very specific causality, but it would still allow *that* L_T and K_{T+1} are causally related.

We suspect there is a lack of the in-depth knowledge required to settle the controversy about a *universal* model M2 that defines a causality or a non-causality between maintenance policy compliance L_T and functionality K_{T+1} by its parameters. Therefore, we expect that the model M2 will remain controversial.

Eventually, the maintenance prognostic argument could *existentially* claim the likelihood of some presumed model M2 *and* some presumed error distribution P6. However, we can only assess this likelihood of a limited subset of options for the model M2 and the error distribution P6 from a potentially infinite set of available options. Therefore, the maintenance prognostic argument is expected to be imprecise about the prima facie causality in Equation 41 because model uncertainty and parameter uncertainty cannot be resolved.

The path graph of the maintenance prognostic argument in Figure 11 specifies the noncausality assumptions that we need to interpret an existential confirmation of the prima facie causality in Equation 41 as being causal. These non-causality assumptions do not just follow from inferred statistical associations and as a consequence the path graphs in Figure 11 are not entirely testable (Pearl, 2010), (Spirtes, Glymour, & Scheines, 2000). The independencies (=bidirectional non-causality assumption in Figure 11) between L_T, B and K_T, B are untestable because the background variable B remains unobserved and it is simply in-depth knowledge about the first causality principle (Section 2.3.3), which asserts that the future cannot cause the past, that made us omit the arrows to the left in $K_T \rightarrow K_{T+1}$ and $L_T \rightarrow K_{T+1}$. However, the independence between L_T and K_T is straightforwardly testable from a sample $(l,k)_{[1,t]}$ which potentially allows us to be more precise about the causal interpretation of an existentially confirmed prima facie causality in Equation 41 by the maintenance prognostic argument.

4.2.3 Sampling issues

In the absence of in-depth knowledge about the model M2 and the error distribution P6, the stratified sample $(l,k)_{[1,t]}$ does not suffice to validate the maintenance prognostic argument. Random assignment of treatments would have mitigated controversy about the treatment being the cause of some statistical association. Then, the following presumptions might have become tenable:

- A significant difference in the observed functionality $P4:K_{T+1}=k_{t+1}$ for the treatment group and the control group uniquely relies on the treatment;
- the error distribution P6 for the treatment and the control group is identical.

These presumptions would have enabled likelihood assessments of some presumed *universal* model M2. However, we have ignored costly experimental research to validate a maintenance policy despite these potential merits. As a result, the maintenance prognostic argument is expected to remain undecidable about the causality $L_T \rightarrow K_{T+1}$.

In Section 2.3.3, we introduced a more modest *prima facie* causality that only holds with respect to an information set V. Since it would be odd to conceive identical information sets V as distinguishable replications, we propose conceiving a replication which would then be a subset of V. In the case of the prima facie causality in Equation 41, the information set V comprises the following elements:

- The element $\{k_{t+1}\}$ in V, claimed by the proposition P4, has been conceived as a trial from the probability function M3 in the maintenance prognostic argument;
- The elements $\{l_t,k_t\}$ in V, claimed by the proposition P3, are independent of the prediction errors P6=P4-C2 in the maintenance prognostic argument.

From the above, it follows that any information set V could be seen as a replication from the probability function M3 in the maintenance prognostic argument. So, the observed frequency of the replications in the sample $(l,k)_{[1,t]}$ is t-1. The sampling efficiency to obtain a sufficient number of replications seems high because the evidence $(l,k)_{[1,t]}$ is not scattered over several replications.

Finally, conventional prognostics relies on condition monitoring data that are sampled by supervisory control and data acquisition (SCADA) systems at a high rate. This allows us to efficiently collect time series that could reconstruct the original signal. However, a maintenance policy validation relies on maintenance performance recordings that are typically only tagged by calendar date. This daily sampling rate may not suffice to reconstruct the signal or to efficiently capture enough evidence. A typical prognostic argument may therefore not suffice for the maintenance policy validation.

4.3 Reliability engineering argument

This section will introduce the reliability engineering argument by (*i*) presenting its propositions in Section 4.3.1, by (*ii*) discussing its claim regarding the prima facie causality in Equation 41 in Section 4.3.2 and by (*iii*) discussing its sampling efficiency in Section 4.3.3. The reliability argument has been deduced from parametric (Abernethy, 2006) as well as nonparametric (Kaplan & Meier, 1958), (Coolen, Coolen-Schrijner, & Yan, 2002) inferences of reliability.

4.3.1 Claim of the argument

This section will explain what the reliability engineering argument claims about the relation between its antecedents and its conclusion.

The reliability engineering argument in Figure 14 only differs from the maintenance prognostic argument by its reduction of model M2's domain from proposition P3 to $K_T=k_t$. Therefore, the conclusion C3 of the reliability engineering argument only holds under a given maintenance policy compliance $L_T=l_t$.

 $\mathbf{P3} \quad (K_T = k_t, L_T = l_t)$

;Observe functionality k_t and maintenance policy compliance $l_t. \label{eq:observe}$

- M2 $(K_T = k_t) \rightarrow (\hat{K}_{T+1} = \hat{k}_{t+1})$;Presume a model M2 that estimates k_{t+1} from P3 ; If M2 has nonzero parameters for K_T, the second causality principle has been satisfied, i.e. a cause remains constant in direction throughout time ; If M2 has nonzero parameters for K_T, the third causality principle has been satisfied, i.e. a cause comprises unique information about the effect that is not available otherwise ; Note that the model M2 is more compelling than needed for the causality principles because it universally maps the antecedent P3 to a single conclusion C2. However, the M2 does not presume a specific relation between L_T and K_{T+1}.
- C2 $\therefore \hat{K}_{T+1} = \hat{k}_{t+1}$;Follows from P3,M2
- P4 $K_{T+1} = k_{t+1}$;Observe functionality k_{t+1} .
- P5 P4 → P3
 ; Presume this non-causality assumption P5.
 ; P5 satisfies the first causality principle, i.e. P4 cannot cause P3 since an effect does not precede its cause in time
- **P6** $(P4 C2) \leftrightarrow (K_T = k_t)$

;Presume that prediction errors are independent of M2's domain, i.e. all information from K_T about K_{T+1} has been captured in M2's model parameters.

- **M3** $((K_T = k_t), M2, P5, P6) \rightarrow Pr(P4|(L_T = l_t))$;Presume this definition of the probabilities in Equation 41. P5 just strengthens a belief in a causal interpretation of this probability.
- C3 $\therefore Pr(P4|(L_T = l_t))$

;Concludes that any information set $V = \{P3, P4\}$ of equal $\{l_t\}$ has been conceived as a replication. To confirm the prima facie causality in Equation 41, C3 differs at various values l_t .

| Valid argument: | Yes, but C2,C3,P5,P6 are not immediately observable | | |
|------------------------|--|--|--|
| Functional relation: | Yes, but M2,M3 maps to conclusions C2,C3 that are not immediately | | |
| | observable. | | |
| Common sense evidence: | - | | |
| Universal argument: | No, the stratified sample (l,k) _[1,t] does not suffice to observe the probability | | |
| | C3 and the reliability engineering argument only addressed the three | | |
| | causality principles for $K_T \rightarrow K_{T+1}$ in M2,P5. | | |
| Decidable argument: | No, both the model M2 and the presumption P6 are controversial. | | |
| | | | |

Figure 14 Reliability engineering argument

Presumption P6 in the reliability engineering argument alternatively denotes as:

$$(P4 = C2 + P6 = M2(K_T = k_t) + P6) \equiv ((k_{t+1} = f(k_t) + \varepsilon_{t+1})) \quad ; L_T = l_t, \forall t, \forall k$$
43

Equation 43 shows that the distribution of functionality $P4:K_{T+1}=k_{t+1}$ is the sum of the model M2: $f(k_t)$ and the error distribution $P6:\varepsilon_{t+1}$ that are *universal* provided that $L_T=l_t$. The probability C3 that follows from the distribution of functionality P4: $K_{T+1}=k_{t+1}$ in Equation 43 then requires in-depth knowledge about the model M2 and the error distribution P6. If we were to have this in-depth knowledge, the causality between maintenance policy compliance L_T and functionality K_{T+1} would have followed from a

universal equation 43 and maintenance would have been justifiable. So, it is the controversy about the model M2 and the presumption P6 that is problematic for the justifiability of maintenance. However, we suspect to only know some stratified sample $(l,k)_{[1,t]}$. Then, the reliability engineering argument may *existentially* claim the likelihood of some *arbitrarily presumed* model M2 and some *arbitrarily presumed* error distribution P6. These existential claims are vulnerable to extensions of the sample $(l,k)_{[1,t]}$ and for more likely presumptions about the model M2 and the error distribution P6.

Equation 43 holds for *a single value* 1 and *all* t,k reconciling with the definition of reliability in Section 3.5 that similarly referred to "given conditions" like a constant maintenance policy compliance $L_T=l_t$. The universal model M2:f(k_t) may or may not depend on maintenance policy compliance $L_T=l_t$.

The *universal* model M2: $f(k_t)$ in the reliability engineering argument is simpler than the equivalent *universal* model M2: $f(l_t,k_t)$ in the maintenance prognostic argument. Eventually, the model M2: $f(k_t)$ is less controversial than the model M2: $f(l_t,k_t)$. Otherwise, the reliability engineering argument remains equally undecidable due to the controversy about the universal model M2: $f(k_t)$ and about the universal error distribution P6: ϵ_{t+1} while delimiting the applicability of its claim to a constant maintenance policy compliance $L_T=l_t$. We suspect that the latter will be true.

4.3.2 Claim about prima facie causality

This section will explain what the reliability engineering argument claims about the prima facie causality in Equation 41. The reliability engineering argument claims a causality $K_T \rightarrow K_{T+1}$ because its model M2 and its presumption P5 cover all three causality principles (Section 2.3.3). So, conventional reliability engineering is geared to predict the item's life from its age under given conditions. The reliability engineering argument from Figure 14 could similarly be seen as some hazard rate model that predicts the item's one-step-ahead functionality K_{T+1} from its current functionality K_T under a given maintenance policy compliance $L_T=l_t$. Therefore, the reliability engineering argument does not claim a specific relation between L_T and K_{T+1} and consequently neither about the prima facie causality in Equation 41.

Tentatively, we have in-depth knowledge about the *universal* model M2 at a given $L_T=l_t$ and $L_T=l_t$ '. If these models M2 differ, we would conclude *that* maintenance policy compliance L_T causes functionality K_{T+1} . Even if we only knew the universal model M2 given a single $L_T=l_t$, we could still assess the likelihood of the assumption that this model M2 also generated the observations at some alternative $L_T=l_t$ '. In this way, the reliability engineering argument may still claim *that* L_T and K_{T+1} are related but not *how* L_T and K_{T+1} are related.

We suspect a lack of in-depth knowledge that could settle the controversy about even a single *universal* model M2. We therefore expect that the reliability engineering argument will not be compelling for the causality between L_T and K_{T+1} .

Eventually, the reliability engineering argument could only *existentially* claim the likelihood of some presumed model M2 *and* some presumed error distribution P6 at a given $L_T=l_t$. However, we can only assess this likelihood of a limited subset of candidate models M2 and error distributions P6 from a potentially infinite set of available options. Therefore, the reliability engineering argument is expected to be imprecise about the prima facie causality in Equation 41.

The path graph of the reliability engineering argument in Figure 11 specifies the noncausality assumptions that we need to interpret an existential confirmation of the prima facie causality in Equation 41 as being causal. These non-causality assumptions do not just follow from inferred statistical associations and the path graphs in Figure 11 are therefore not entirely testable (Pearl, 2010), (Spirtes, Glymour, & Scheines, 2000). The independence (=bidirectional non-causality assumption in Figure 11) between K_T and B is untestable as the background variable B remains unobserved, and it is only in-depth knowledge about the first causality principle (Section 2.3.3), which asserts that the future cannot cause the past, that made us omit the arrows to the left in K_T \rightarrow K_{T+1} and L_T \rightarrow K_{T+1}. However, the independence between L_T and K_T is straightforwardly testable from a sample (l,k)_[1,t] which potentially allows us to be more precise about the causal interpretation of an existentially confirmed prima facie causality in Equation 41 by the reliability engineering argument.

4.3.3 Sampling issues

In the absence of in-depth knowledge about the model M2 and the error distribution P6, the stratified sample $(l,k)_{[1,t]}$ does not suffice to validate the reliability engineering argument. Random assignment of treatments would have mitigated controversy about the treatment being the cause of some statistical association. In that case, the following presumptions might have become tenable:

- A significant difference in the observed functionality $P4:K_{T+1}=k_{t+1}$ for the treatment group and the control group uniquely relies on the treatment;

- the error distribution P6 for the treatment and the control group is identical. These presumptions would have enabled likelihood assessments of some presumed *universal* model M2. However, we have ignored costly experimental research to validate a maintenance policy despite these potential merits. As a result, the reliability engineering argument is expected to remain undecidable about the causality $L_T \rightarrow K_{T+1}$.

In Section 2.3.3, we introduced a more modest *prima facie* causality that only holds with respect to an information set V. Since it would be odd to conceive identical information sets V as distinguishable replications, we propose conceiving a replication which would then be a subset of V. In the case of the prima facie causality in Equation 41, the information set V comprises the following elements:

- The element $\{k_{t+1}\}$ in V, claimed by the proposition P4, has been conceived as a trial from the probability function M3 in the reliability engineering argument;
- The element {k_t} in V, claimed by the proposition P3, is independent of the prediction errors P6 in the reliability engineering argument;
- The element $\{l_t\}$ in V, claimed by the proposition P3, may relate to the proposition P4 in any way.

From the above, it follows that any information set V with identical $\{l_t\}$ could be seen as a replication from the probability function M3 in the reliability engineering argument. The observed frequency of the replications V= $\{l_t,k_t,k_{t+1}\}$ with identical $\{l_t\}$ in the sample $(l,k)_{[1,t]}$ may tentatively not even exceed one. The sampling efficiency to obtain a sufficient number of replications could appear to be low because the evidence $(l,k)_{[1,t]}$ could eventually be scattered over an infinite number of possible replications.

Still, the reliability engineering argument may *existentially* claim the likelihood of some presumed model M2 and some presumed error distribution P6. To existentially *refute* the prima facie causality in Equation 41, the sample $(l,k)_{[1,t]}$ should comprise sufficient replications of every possible value l_t . Since a finite sample $(l,k)_{[1,t]}$ cannot cover an infinite sample space, maintenance policy compliance must have a less precise categorical scale. To existentially *confirm* the prima facie causality in Equation 41, the sample $(l,k)_{[1,t]}$ should comprise sufficient replications of at least two different values l_t . It is possible that the stratified sample $(l,k)_{[1,t]}$ fails to comply with this additional constraint on its composition. In that case, the maintenance policy validation by the reliability engineering argument would fail.

In conclusion, the reliability engineering argument is less efficient in its sampling because it typically discards many infrequently observed replications from the sample $(l,k)_{[1,t]}$. Because of the constraints on an observational research, where control over the composition of the sample $(l,k)_{[1,t]}$ is impossible, this concern is realistic. Reliability engineering is known to often suffer efficiency problems (Abernethy, 2006), which implies that the number of replications may not suffice in a typical sample $(l,k)_{[1,t]}$.

4.4 Nonparametric argument

This section will introduce the maintenance prognostic argument by (*i*) presenting its propositions in Section 4.4.1, by (*ii*) discussing its claim regarding the prima facie causality in Equation 41 in Section 4.4.2 and by (*iii*) discussing its sampling efficiency in Section 4.4.3. The nonparametric argument instantiates an exact conditional approach that compares two independent binomial proportions (Lin & Yang, 2009).

4.4.1 Claim of the argument

This section will explain what the nonparametric argument claims about the relation between its antecedents and its conclusion.

The nonparametric argument in Figure 15 comprises presumptions P5,P7,M4. The presumption P5 and the model M4 do not appear to be controversial. The presumption P5 just reflects common sense about the causality principle that an effect P4 does not precede its cause P3 in time and the model M4 appears to be a common sense operationalisation of a probability. Therefore, the nonparametric argument appears to be *decidable* because only the presumption P7 appears to be controversial. However, the conclusion C4 is a not immediately observable probability of P4, given the presumptions P5 and P7 and a constant k_t .

- **P3** $(K_T = k_t, L_T = l_t)$;Observe functionality k_t and maintenance policy compliance l_t.
- P4 $K_{T+1} = k_{t+1}$;Observe functionality k_{t+1} .
- P5 P4 → P3
 ; Presume this non-causality assumption P5.
 ; P5 satisfies the first causality principle, i.e. P4 cannot cause P3 since an effect does not precede its cause in time
- **P7** $(L_T = l_t) \not\rightarrow P4$

; Presume this non-causality assumption P7.

; P7,P5 imply universal independence between $(L_T=l_t)$ and P4

;If P7 holds at some time, second causality principle has been refuted, i.e. a cause does not remain constant in direction throughout time.

;If P7 holds for some k_t , third causality principle has been refuted, i.e. a cause does not comprise unique information about the effect that is not available otherwise.

- M4 $(P5, P7) \rightarrow Pr(P4|(K_T = k_t))$;Presume this definition of the probability of P4, given P5,P7. This probability does *not* depend on maintenance policy compliance L_T.
- $\mathbf{C4} \quad \therefore \Pr(P4|(K_T = k_t))$

; Concludes that any information set $V=\{P3,P4\}$ of equal k_t has been conceived as a replication. To confirm the prima facie causality in Equation 41, P7 is false at all possible values k_t .

| Valid argument: | Yes, but C4,P5,P7 are not immediately observable | | |
|------------------------|---|--|--|
| Functional relation: | Yes, but M4 maps to a conclusion C4 that is not immediately observable | | |
| Common sense evidence: | - | | |
| Universal argument: | No, but the nonparametric argument does assess the probability that the | | |
| | sample $(1,k)_{[1,t]}$ comes from a data generating process that satisfies P5,P7. | | |
| Decidable argument: | Yes, only the presumption P7 is controversial. | | |
| - | | | |

Figure 15 Nonparametric argument

The presumptions P5 and P7 claim a *universal* independence between maintenance policy compliance L_T and functionality K_{T+1} at a given K_T . This independence assumption omits the model uncertainty and the parameter uncertainty of the maintenance prognostic argument and the reliability engineering argument. If this independence assumption were universally true, maintenance policy compliance L_T would not have caused functionality K_{T+1} and maintenance would have been unjustified. So, it is the controversy about this independence presumption that makes maintenance unjustifiable. The nonparametric argument does not conclusively *decide* about the truth or falsehood of the controversial presumption P7. The nonparametric argument just deems presumption P7 more or less *likely*.

4.4.2 Claim about prima facie causality

This section will explain what the nonparametric argument claims about the prima facie causality in Equation 41.

In the tentative case that we would have had in-depth knowledge about presumption P7, we would straightforwardly decide about the prima facie causality in Equation 41 without the need for any argument. Since we suspect that this in-depth knowledge is lacking, the nonparametric argument may infer the presumption P7 from propositions that we deem true. The nonparametric argument only *existentially* claims the likelihood of presumption P7 at a given K_T :

$$\left(\mathcal{L}\left(\left(pr_{K_{T+1}|L_{T},K_{T}}(k_{t+1}|l_{t},k_{t}) = pr_{K_{T+1}|K_{T}}(k_{t+1}|k_{t}) \right) | (l,k)_{[1,t]} \right) < \rho ; \exists l \forall t \forall k \right)$$

$$\cong (L_{T} \to K_{T+1})$$

$$44$$

Equation 44 shows that the prima facie causality in Equation 41 may also follow from a low likelihood, i.e. $\mathcal{L} < \rho$, of a presumed independence at some l and all t,k.

Let independence be likely $\mathcal{L} \ge \rho$ at all or only some values of functionality K_T in Equation 44. Then, maintenance policy compliance L_T would not *uniquely* attribute to functionality K_{T+1} irrespective of K_T which would cause a problem for the third causality principle (Section 2.3.3), which requires that a cause comprises *unique* information about the effect that is not available otherwise. Consequently, we would deem the prima facie causality in Equation 41 as unlikely.

Alternatively, let independence be unlikely $\mathcal{L} < \rho$ at all values of functionality K_T in Equation 44. Then, we would *existentially* claim that presumption P7 is unlikely irrespective of K_T. Consequently, we would deem the prima facie causality in Equation 41 as likely. A straightforward assessment of the observed proportions or means of functionality K_{T+1} could be informative about the strength or the direction of the prima facie causality in Equation 41. So, the nonparametric argument lacks the predictive capabilities of some model M2 that are needed to support prospective decision making, but it may serve the maintenance policy validation in retrospect.

The path graph of the nonparametric argument in Figure 11 specifies the non-causality assumptions that we need to interpret an existential confirmation of the prima facie causality in Equation 41 as being causal. These non-causality assumptions do not just follow from inferred statistical associations and the path graphs in Figure 11 are therefore not entirely testable (Pearl, 2010), (Spirtes, Glymour, & Scheines, 2000). The independence (=bidirectional non-causality assumption in Figure 11) between L_T and B is untestable as the background variable B remains unobserved and it is simply in-depth knowledge about the first causality principle (Section 2.3.3), which asserts that the future cannot cause the past that made us omit the arrows to the left in $K_T \rightarrow K_{T+1}$ and $L_T \rightarrow K_{T+1}$. Then, the independence between L_T and K_{T+1} directly follows from the proposition P7 of the nonparametric argument. Tentatively, an inferred independence within an information set V may appear to be spurious similar to Figure 8. Then, "true" dependencies like $L_T \rightarrow K_{T+1}$ and $L_T \rightarrow B \rightarrow K_{T+1}$ should rule each other out which is quite a strong presumption (faithfulness condition (Spirtes, Glymour, & Scheines, 2000)). In this respect, the nonparametric argument is more compelling for a causal interpretation of an existentially inferred claim regarding the prima facie causality in Equation 41 than any of the other candidate arguments.

4.4.3 Sampling issues

In the absence of in-depth knowledge about the presumption P7, the stratified sample $(l,k)_{[1,t]}$ does not suffice to validate the nonparametric argument. Random assignment of treatments would have mitigated controversy about the treatment being the cause of some statistical association. This would have enabled likelihood assessments of some *universal* presumption P7. However, we have ignored costly experimental research to validate a maintenance policy despite these potential merits. As a result, the nonparametric argument is expected to remain undecidable about the causality $L_T \rightarrow K_{T+1}$.

In Section 2.3.3, we introduced that a more modest *prima facie* causality only holds with respect to an information set V. Since it would be odd to conceive identical information sets V as distinguishable replications, we propose conceiving a replication which would then be a subset of V. In the case of the prima facie causality in Equation 41, the information set V comprises the following elements:

- The element $\{k_{t+1}\}$ in V, claimed by the proposition P4, has been conceived as a trial from the probability function M4 in the nonparametric argument;
- The element $\{l_t\}$ in V, claimed by the proposition P3, is independent of the proposition P4 in the nonparametric argument;
- The element $\{k_t\}$ in V, claimed by the proposition P3, may relate to the proposition P4 in any way.

From the above, it follows that any information set V with identical $\{k_t\}$ could be seen as a replication from the probability function M4 in the nonparametric argument. The observed frequency of the replications V= $\{l_t, k_t, k_{t+1}\}$ with identical $\{k_t\}$ in the sample $(l,k)_{[1,t]}$ may tentatively not even exceed one. The sampling efficiency to obtain a sufficient number of replications could appear to be low because the evidence $(l,k)_{[1,t]}$ could eventually be scattered over an infinite number of possible replications.

To existentially refute the prima facie causality in Equation 41, the sample $(l,k)_{[1,t]}$ should comprise sufficient replications of:

- various lt, to assess the likelihood of independence between LT and KT+1 at
- a constant k_t , to avoid a presumption regarding the relation between K_T and K_{T+1} and to prevent K_T from being a mediator or a confounder of an eventual dependence between L_T and K_{T+1} .

We typically do not a priori know which values k_t yield the larger likelihoods. We possibly need to repeat this validation of the nonparametric argument at different values k_t until we find one that existentially claims the independence between L_T and K_{T+1} .

To existentially confirm the prima facie causality in Equation 41, the sample $(l,k)_{[1,t]}$ should comprise sufficient replications of:

- various l_t , to assess the likelihood of independence between L_T and K_{T+1} at
- a constant k_t , to avoid a presumption regarding the relation between K_T and K_{T+1} and to prevent K_T from being a mediator or a confounder of an eventual dependence between L_T and K_{T+1} and
- at every possible k_t to ensure the third causality principle (Section 2.3.3), which asserts that a cause comprises unique information about the effect that is not available otherwise.

Since a finite sample $(l,k)_{[1,t]}$ cannot cover an infinite sample space, functionality K_T must have a less precise categorical scale to existentially confirm the prima facie causality in Equation 41. Eventually, the stratified sample $(l,k)_{[1,t]}$ fails to comply with these additional constraints on its composition. In that case, the maintenance policy validation by the nonparametric argument would fail.

In conclusion, the nonparametric argument is least efficient in its sampling because it typically discards many infrequently observed replications from the sample $(l,k)_{[1,t]}$. Because of the constraints on an observational research, where control over the composition of the sample $(l,k)_{[1,t]}$ is impossible, this concern is realistic.

4.5 Review of the arguments on inference precision

| | Maintenance optimisation argument (Section 4.1) | Maintenance prognostic argument (Section 4.2) | Reliability engineering argument (Section 4.3) | Nonparametric argument (Section 4.4) |
|---|--|--|---|--|
| Valid argument | Yes | Yes | Yes | Yes |
| Functional relation | Yes | Yes | Yes | Yes |
| Common sense evidence | - | - | - | - |
| Universal argument | Yes, if refuted | No | No | No |
| Decidable argument | Yes | No | No | Only in likelihood |
| Decidable about $L_T \rightarrow K_{T+1}$ w.r.t $\{l_t, k_t, k_{t+1}\}$ | No, if refuted | No | No | Only in likelihood |

This section will preliminarily assess the inference precision of a maintenance policy validation by the candidate arguments.

 Table 5 Preliminary inference precision of the candidate arguments

We will only discuss the inference objectives from Section 1.4.1 that are affected by the choice of the argument:

- Valid argument;
- Functional relation;
- Universal argument;
- Decidable argument.

Without a realistic sample $(l,k)_{[1,t]}$, the remarks regarding inference precision are only preliminary.

Table 5 reveals that the arguments are all valid and they all comprise functional models. Therefore, these inference objectives are not selective for the candidate arguments. The

maintenance optimisation argument is straightforwardly testable. A preference for the other candidate arguments appears to be driven by:

- The amount of in-depth knowledge about the model M2 and the error distribution P6;
- The efficiency to collect a sample $(l,k)_{[1,t]}$ that meets the requirements.

Possibly, both in-depth knowledge and the sampling efficiency may turn out to be insufficient. We would then decide that maintenance is unjustifiable by any of these arguments.

4.5.1 *Findings regarding the "valid argument" inference objective*

This inference objective assesses whether a conclusion deductively follows from the other propositions of the argument. The candidate arguments comprise *explicit* models that imply the conclusion of these arguments. We ignore arguments that leave the choice of a model to some random generator, like artificial neural networks or genetic algorithms. Nor have we widened the field to include mixed model arguments here.

4.5.2 Findings regarding the "functional relation" inference objective

This inference objective assesses whether the antecedents map to a unique conclusion.

The model M1 of the maintenance optimisation argument is an equivalence relation that inherits all properties of the functional relation $k_t=f(l_t)$. Both the antecedent and the consequent immediately follow from common sense about the observable evidence P1,C1.

The model M2 of the maintenance prognostic argument and the reliability engineering argument is a functional relation, but its conclusion is a not immediately observable estimate of a one-step-ahead functionality.

The model M3 of the maintenance prognostic argument and the reliability engineering argument is a functional relation that appears to be a common sense operationalisation of a probability function. However, probabilities are not immediately observable and they may only be estimated when in-depth knowledge about the model M2 and about the error distribution P6 is available. This in-depth knowledge is beyond the information set V.

The model M4 of the nonparametric argument is a functional relation that appears to be a common sense operationalisation of a probability function. A probability is a not immediately observable quantity.

So, all models in the candidate arguments are functional relations that map their domain to a unique conclusion. However, only model M1 appears to be straightforwardly decidable by some sample $(l,k)_{[1,t]}$ whereas all the other arguments comprise not immediately observable propositions that may turn out to be difficult to estimate.

4.5.3 Findings regarding the "universal argument" inference objective

This inference objective assesses whether the argument holds universally, i.e. holds for the entire population.

The maintenance optimisation argument *defines* an equivalence between maintenance policy compliance L and functionality K. A single counterexample in a stratified sample $(l,k)_{[1,t]}$ would already suffice to universally refute this equivalence.

The other arguments cannot make universal claims from a stratified sample $(l,k)_{[1,t]}$. Given the constraint on the prima facie causality in Equation 41 that delimits knowledge to the information set V={ l_t,k_t,k_{t+1} }, the stratified sample $(l,k)_{[1,t]}$ may still comprise enough replications to test the likelihood of some presumption regarding this prima facie causality. For such an *existential* claim, the maintenance prognostic argument and the reliability engineering argument still require an *arbitrary* set of candidate models M2 and error distributions P6. We may not be able to resolve this model uncertainty and parameter uncertainty that seem inapplicable to the nonparametric argument. The nonparametric argument presumes independence between L_T and K_{T+1} which omits the burden of model selection. So, the nonparametric argument directly infers an *existential* claim regarding this independence that maps to a prima facie causality by Equation 44.

So, we suspect that only the maintenance optimisation argument will be refuted *universally*. The nonparametric argument omits model uncertainty and parameter uncertainty about its *existential* claim regarding independence. For the maintenance prognostic argument and the reliability engineering argument, we suspect that model uncertainty and parameter uncertainty hamper an *existential* claim regarding a specific dependence.

4.5.4 Findings regarding the "decidable argument" inference objective

This inference objective assesses whether the truth or falsehood of presumptions is identifiable.

The maintenance optimisation argument only presumes a controversial model M1, whereas P1 and C1 follow from common sense observable evidence. Some stratified sample $(l,k)_{[1,t]}$ could therefore universally refute or existentially confirm the maintenance optimisation argument. The maintenance optimisation argument is therefore *decidable*.

The maintenance prognostic argument comprises two controversial propositions that make the argument undecidable. In-depth knowledge about some "man-machine-law" may settle the controversy about the model M2. Alternatively, random assignment of treatments might have weakened a presumed equality of the error distributions P6. Neither in-depth knowledge, nor random assignment of treatments seem applicable to this maintenance policy validation. The maintenance prognostic argument will then remain *undecidable*.

The reliability engineering argument resembles the maintenance prognostic argument in this respect since it only excludes maintenance policy compliance L_T from the domain of model M2. Therefore the reliability engineering argument is expected to be *undecidable*.

The nonparametric argument only comprises a single controversial proposition which makes the argument *decidable* in principle. However, the nonparametric argument only concludes with a not immediately observable probability.

So, we expect that only the maintenance optimisation argument will be decidable, whereas the nonparametric argument will only be decidable in terms of likelihood. The maintenance prognostic argument and the reliability engineering argument are expected to remain undecidable.

4.5.5 Findings regarding the choice of a sampling procedure

In this section, we will only survey the additional constraints that a candidate argument imposes on the composition of a sample $(l,k)_{[1,t]}$ that has been collected by observational research. Since we have no control over the composition of the sample, these additional constraints potentially obstruct the maintenance policy validation by a particular argument.

Any sample $(l,k)_{[1,t]}$ may universally refute or existentially confirm the model M1 of the maintenance optimisation argument. Any information set $\{l_t,k_t\}$ from $(l,k)_{[1,t]}$ could be seen as a replication. So, the maintenance optimisation argument does not discard any component in the sample $(l,k)_{[1,t]}$. The sampling for the maintenance optimisation argument is therefore expected to be *efficient*.

The maintenance prognostic argument takes any information set $V = \{l_t, k_t, k_{t+1}\}$ as a replication. So, the maintenance prognostic argument does not discard any information set V from the sample $(l,k)_{[1,t]}$. Therefore, the sampling for the maintenance prognostic argument is expected to be *efficient*.

The reliability engineering argument takes any information set $V=\{l_t,k_t,k_{t+1}\}$ with identical $\{l_t\}$ as a replication. Eventually, the observed frequency of these replications in the sample $(l,k)_{[1,t]}$ does not suffice for a maintenance policy validation by the reliability engineering argument. This means that replications in a sample $(l,k)_{[1,t]}$ may be discarded because their observed frequency in the sample $(l,k)_{[1,t]}$ is too low for a significant claim. The sampling for the reliability engineering argument is therefore *less efficient*.

To existentially confirm the prima facie causality in Equation 41 by the reliability engineering argument, the sample $(l,k)_{[1,t]}$ should comprise sufficient replications of at least two well-chosen values of maintenance policy compliance L_T. To existentially refute the prima facie causality in Equation 41 by the reliability engineering argument, the sample $(l,k)_{[1,t]}$ should comprise sufficient replications of every element in the sample space of maintenance policy compliance L_T. The latter is only possible if this sample space is finite. The sampling for the reliability engineering argument therefore *lacks efficiency* if maintenance policy compliance has a cardinal scale.

The nonparametric argument takes any information set $V=\{l_t,k_t,k_{t+1}\}$ with identical $\{k_t\}$ as a replication. Eventually, the observed frequency of these replications in the sample $(l,k)_{[1,t]}$ does not suffice for a maintenance policy validation by the nonparametric argument. This means that replications in a sample $(l,k)_{[1,t]}$ may be discarded because their observed frequency in the sample $(l,k)_{[1,t]}$ is too low for a significant claim. In addition, replications where $\{l_t\}$ does not vary enough may be discarded due to their inability to test for independence between L_T and K_{T+1} . The sampling for the nonparametric argument is therefore *least efficient*.

To existentially refute the prima facie causality in Equation 41 by the nonparametric argument, the sample $(l,k)_{[1,t]}$ should comprise sufficient replications of varying $\{l_t\}$ at some well-chosen $\{k_t\}$. To existentially confirm the prima facie causality in Equation 41 by the nonparametric argument, the sample $(l,k)_{[1,t]}$ should comprise sufficient replications of varying $\{l_t\}$ at every possible $\{k_t\}$. The latter is only possible if the sample space of functionality K_T is finite. The sampling for the nonparametric argument therefore *lacks efficiency* if functionality has a cardinal scale.

To establish common sense about a replication in the sample $(l,k)_{[1,t]}$, we only used the information set $V = \{l_t, k_t, k_{t+1}\}$ that we already used for the prima facie causality in Equation 41. So, this definition of a replication is not an additional encumbrance to the presumption regarding the prima facie causality in Equation 41. Still, one may argue that the sample $(1,k)_{[1,t]}$ comprises more information than the information set $V = \{l_t, k_t, k_{t+1}\} = \{P3, P4\}$ of the prima facie causality in Equation 41. Eventually, a strong time-dependent evolution of L and K is problematic as regards the presumption that every information set $V = \{l_{t}, k_{t}, k_{t+1}\}$ in the sample $(l, k)_{[1,t]}$ could be seen as a replication. In Section 5.5.3, we will apply a split sample test to get a glimpse into this timedependent behaviour in a realistic case study. Note that independence between L_T and K_{T+1} at *some* time already suffices to refute the prima facie causality in Equation 41 due to the second causality principle (Section 2.3.3), i.e. a cause remains constant in direction throughout time. However, time is just one of the many candidate background variables that may reduce an inferred prima facie causality to a spurious cause as shown in Figure 8. So, a claim about the prima facie causality in Equation 41 does not entirely suffice for the maintenance policy validation.

So, we expect that the maintenance optimisation argument and the maintenance prognostic argument will admit all information sets $V = \{l_t, k_t, k_{t+1}\}$ from the sample $(l,k)_{[1,t]}$. The reliability engineering argument and the nonparametric argument impose constraints on the composition of the sample $(l,k)_{[1,t]}$ that may appear to be problematic in practice.

4.5.6 Findings regarding the claim about prima facie causality

This section will preliminarily revisit what the candidate arguments claim about the prima facie causality in Equation 41.

If the maintenance optimisation argument appears to be sound, it refutes the prima facie causality in Equation 41 because it presumes that L_T and K_T are redundant variables. If unsound, the maintenance optimisation argument cannot claim much about the prima facie causality in Equation 41. We suspect that the latter is true.

A stratified sample $(l,k)_{[1,t]}$ does not decide about the soundness of the other arguments. Therefore, *universal* claims about causality are similarly problematic but we may resort to an *existential* claim about prima facie causality.

The maintenance prognostic argument takes all three causality principles (Section 2.3.3) in its propositions, which potentially enables even a claim regarding a causality $L_T \rightarrow K_{T+1}$ if it were universally sound. Still, we suspect there is a lack of in-depth knowledge about an *arbitrary* set of candidate models M2 and about an *arbitrary* set of error distributions P6 to make the maintenance prognostic argument decidable. Therefore, the maintenance prognostic argument may only existentially claim the likelihood of an arbitrarily presumed model M2 and an arbitrarily presumed error distribution P6 that define the prima facie causality in Equation 41 by their parameters. Still, this existential claim is expected to remain subject to model uncertainty and parameter uncertainty.

The reliability engineering argument only differs from the maintenance prognostic argument by reducing the domain of the model M2 to $K_T=k_t$. Therefore, all concerns regarding the selection of the model M2 and the selection of the error distribution P6 in the maintenance prognostic argument equally apply. Furthermore, the reliability engineering argument imposes an additional constraint on the composition of a sample $(l,k)_{[1,t]}$ which potentially requires us to discard rarely observed replications. We therefore suspect that the reliability engineering argument is even less precise about the prima facie causality in Equation 41 than the maintenance prognostic argument.

The nonparametric argument may *existentially* claim the likelihood of the prima facie causality in Equation 41. However, the nonparametric argument imposes severe constraints on the composition of a sample $(l,k)_{[1,t]}$ which potentially requires us to discard rarely observed replications *and* frequently observed replications where $\{l_t\}$ does not vary enough. It is possible that we will fail to efficiently collect an adequate sample for the nonparametric argument.

In conclusion, in the absence of a realistic sample, the nonparametric argument seems to be most precise about the maintenance policy validation because it is the only candidate argument that decides about the likelihood of the prima facie causality in Equation 41 as shown in Table 5.

5 Implementation of the inference

This chapter will present the maintenance policy validation in a realistic case study.

In Section 5.1, we will assess to what extent the case organisation's common sense about maintenance policy compliance and functionality satisfies the construction rules for maintenance performance indicators in Section 3.2. We will suggest some improvements and we will reveal some concerns regarding a lack of common sense.

In Section 5.2, we will compose three candidate samples that could instantiate the case organisation's common sense about maintenance policy compliance and functionality. We will specifically evaluate whether these samples comply with the additional sampling constraints from the reliability argument and the nonparametric argument that we introduced in Section 4.3 and Section 4.4 respectively.

In Section 5.3, we will try to validate the candidate arguments from Chapter 4 by the candidate samples from Section 5.2. Eventually, the evidence also comprises in-depth knowledge about the argument's presumptions. For this specific case study, we will assess the sufficiency of the evidence to decide about the soundness of the candidate arguments and about the prima facie causality $L_T \rightarrow K_{T+1}$ with respect to an information set $V = \{l_t, k_t, k_{t+1}\}$. This will reveal differences in inference precision.

In Section 5.4, we will expand on the maintenance policy validation by the preferred argument and the preferred sample in this case study. We will extend the information set V to seek for long-term effects of maintenance policy compliance. We will also reduce the information set V to better compare the candidate samples.

In Section 5.5, we will discuss the influence of background variables that might have biased the maintenance policy validation in this specific case study. We may possibly find a certain degree of controversy about the evidence and we will also apply a split sample test that may reveal a lack of stationarity.

5.1 Choice of an operationalisation

This section will concentrate on the operationalisation of maintenance policy compliance L and functionality K in a realistic case study. In the previous chapters, we already introduced some potential obstructions for a common sense operationalisation:

- Assessments of causality are problematic, particularly in an observational research, as the influence of background variables cannot be managed statistically (Section 2.3);
- Decision making is encumbered with subjectivity (Section 3.1);
- Conventional maintenance performance indicators do not really accommodate validations of causal claims (Section 3.2).

The case organisation uses a maintenance scorecard based on Haarman and Delahay (2004). This maintenance scorecard defines common sense about the objectives of a maintenance policy in terms of (i) asset utilisation, (ii) cost control, (iii) resource allocation and (iv) health, safety and environment.

A maintenance policy that achieves these objectives (i.e. zero incidents, no costs, 100% asset utilisation) seems inaccessible. The case organisation therefore resorts to some attainable requirements (goals or aspiration levels). These requirements seem to be assessed iteratively by experience rather than by explicit indifference curves (Pareto, transl. 1971). By reconciling with the case organisation's requirements on its performance indicators, we hope to mitigate the concerns regarding subjectivity as discussed in Section 3.1. In Section 5.1.1 and 5.1.2, we will therefore seek for common sense about the performance indicators for functionality K and maintenance policy compliance L respectively.

5.1.1 Common sense about functionality

This section will attempt to reveal the case organisation's common sense about functionality. We will identify some improvement opportunities.

The case organisation uses a single performance indicator for the "asset utilisation" objective. We label this performance indicator for functionality by "monthly availability" \overline{Y}_M . Although availability is typically a probability that expresses a degree of belief in a yet to be observed upstate, in this case study the "monthly availability" is some retrospectively observed proportion of uptime over total time. Figure 16 depicts the *monthly* availability and the *daily* output (e.g. the produced amount) on which it is built. Evidently, monthly availability does not adequately capture the variation in the daily output signal, as it has a much lower sampling rate.



Figure 16 Functionality captured at a daily and a monthly sampling rate

Functionality does not just follow from some physical variable, it is also encumbered with a subjective requirement. This requirement typically specifies a threshold value for this physical variable that distinguishes an upstate from a downstate.

In this case study, the objective for the "asset utilisation" was to constantly maximise output rather than to satisfy a varying demand. Then, an assumption that inherent output

becomes observable does not appear to be problematic. The case organisation adopted the reliability engineering convention of a dichotomous state variable Y_T as shown in:

$$Y_T = \begin{cases} 1 & if \ daily \ output \ > 16 \ (upstate) \\ 0 & if \ daily \ output \ \le 16 \ (downstate) \end{cases}$$

$$45$$

The value of Y_T then states whether a required functionality has been fulfilled or not. The distribution of daily output in Figure 17 shows that daily output seems to take either a high (above 16) or a low (around 0) value. The dichotomous variable Y_T in Equation 45 therefore seems an acceptable representation of functionality that ignores subtle output fluctuations.



Figure 17 Cumulative distribution of daily output

The case organisation's performance indicator for functionality averages all functionality Y_T to some monthly availability \overline{Y}_M by:

$$\bar{Y}_{M} = \frac{1}{m} \times \sum_{i=1}^{m} Y_{i} \qquad ; m = "number of days in a month"$$
46

Equation 46 levels out many of the potentially informative fluctuations in output since similar values of monthly availability \overline{Y}_M may come from very different output trajectories. The longitudinal redundancy concern that was discussed in Section 3.2.1 is inapplicable to this case study because the monthly availability \overline{Y}_M keeps pace with a monthly sampling rate.

The maintenance scorecard in Table 4, comprised various performance indicators for functionality like availability, MTBF, breakdown frequency or number of failures. These all seem to be deductions from the dichotomous time series $y_{[1,t]}$. However, these deductions are irreversible. Apparently, a conversion from a time series $y_{[1,t]}$ to the performance indicators for functionality in Table 4 implies a loss of information. Moreover, this conversion would introduce redundancies that do not exist in the time series $y_{[1,t]}$.

The case organisation's maintenance scorecard also includes lagging indicators for the objectives of "cost control", "resource allocation" and "health, safety and environment".

We will not assess whether the adopted scorecard (Haarman & Delahay, 2004) is complete in reflecting the genuine organisational objectives, nor whether these lagging indicators are complete and non-redundant representations of these objectives. To serve efficiency, we just consider functionality. So, we direct our first attempt to validate a maintenance policy by a simple bivariate sample $(l,k)_{l,l}$. Like in any model selection problem (Burnham & Anderson, 2010), such a simple argument is an arbitrary choice to balance the completeness of an information set V with the efficiency of collecting sufficient evidence as discussed in Section 3.2.3. As functionality is just one of the maintenance performance indicators that are subject to trade-offs, it may occur that some maintenance policy violations even enhance functionality but at the same time deteriorate the unobserved total utility (i.e. the weighted combination of all indicators). This does not mean that a maintenance policy validation can no longer be put into practical use. In Section 3.3.2, we explained that maintenance policy assessments typically follow some satisficing process that yields a sub-optimal maintenance policy in which violations may turn out to enhance utility. Still, the maintenance policy validation may provide this essential empirical feedback regarding the effect of the currently applied maintenance policy.

The case organisation's convention to operationalise functionality by the monthly availability \overline{Y}_M may improve by (*i*) implementing an alternative to better capture the potentially informative fluctuations in daily output; (*ii*) adopting the times series $y_{[1,t]}$ as the single indicator to remove existing redundancy in the functionality indicators on the maintenance scorecard in Table 4.

5.1.2 *Common sense about maintenance policy compliance*

This section will attempt to reveal the case organisation's common sense about the quantification of maintenance policy compliance on its maintenance scorecard. We will identify some improvement opportunities.

The case organisation uses three leading indicators for the "resource allocation" objective:

- L¹: monthly mean proportion of timely completed maintenance actions;
- L²: queue of delayed maintenance actions;
- L³: backlog of uncompleted maintenance actions (expressed in hours).

We ignored L^3 because the case organisation has not defined a requirement (an acceptable amount of backlog) on this leading indicator. If the required value of L^3 remains unspecified, we lack a criterion for maintenance policy compliance. Leading indicators L^1 and L^2 suffer from a definitional dependence as L^2 is closely coupled to L^1 by definition. This dependence leads to a redundancy concern as explained in Section 3.2.1. As opposed to L^1 , which is a monthly averaged value, L^2 is instantaneously observable. We therefore propose to operationalise maintenance policy compliance by the queue of delays L^2 .

Figure 18 shows the case organisation's convention to recording L^2 at a monthly sampling rate. Figure 18 also depicts L^2 at a daily sampling rate, which immediately reveals that a monthly sampling rate does not capture all potentially informative perturbations that are revealed at a daily sampling rate.

Besides the leading indicators L^1 and L^3 that we ignored, we may conceive many other possible policy violations. We do not, however, intervene in the case organisation's course of operations by asking for evidence that is beyond the scope of the recording routines. As a result, the applicability of the maintenance policy validation in this case study is confined to an output effect resulting from delayed maintenance. Figure 18 shows that the signals of the delays contain some peaks that are promising because they may eventually detect causal responses. By confining ourselves to delayed maintenance and output, we sacrifice completeness for efficiency as explained in Section 3.2.3.



Figure 18 Queue of delayed maintenance at a daily and a monthly sampling rate

Every decision to carry out maintenance is time bound. If we were allowed to defer maintenance actions till infinity, there would be no need for resources. Every decision can therefore be considered as being violated if a certain delay of the associated action occurs. Other maintenance policy violations only apply to a very specific subset of a maintenance policy. Alignment errors, for example, only apply to decisions to (re-)align. The evidence for delays seems to be obtainable in an efficient way, since many organisations already record required and actual completion dates of maintenance actions. Finally, Figure 18 confirms that delays appear in abundance. The counterfactual reality seems therefore attainable.

This case study comprises a set of maintenance actions that were not completed at some time during a particular interval of 1977 days. To assess a queue of delayed maintenance at a particular time from that set, all maintenance actions should comprise:

- A birthdate (T_B) , i.e. the date on which the decision was taken;
- A required completion date (T_R) ;
- The actual completion date (T_C):
- A job description (to enable removing maintenance actions that are deferrable till infinity without affecting functionality).

To build a queue of delays on these maintenance actions, an indicator function $X_{T,m}$ determines whether the mth maintenance action is a queue member (i.e. has been delayed) at a time T:

$$X_{T,m} = \begin{cases} 1, & \text{if } T_{R,m} < T \le T_{C,m} \\ 0, & \text{otherwise} \end{cases}$$

47

Then, the queue length can be expressed as:

$$D_T = \sum_{\forall m} X_{T,m} \tag{48}$$

Section 5.1.1 explained that functionality is just an incomplete operationalisation of the objectives of the maintenance policy, as we did not admit "cost control" and "health, safety and environment" indicators from the case organisation's maintenance scorecard. We therefore propose to remove all maintenance actions that are deferrable till infinity without affecting the "asset utilisation" objective. This removal factor follows from expert judgement of the job description. To avoid a tautology, this removal factor should rely on a believed functionality risk and *not* on posterior knowledge about functionality.

For the case study, the survey below shows that the removal factors reduced the evidence by about one third:

- The sample comprised 6342 maintenance actions that were in a state of not having been completed at some time during an interval of 1977 days.
- Removing timely completed maintenance actions reduced this set to 2740 delayed maintenance actions.
- Removing maintenance that is known to be deferrable till infinity without affecting functionality reduced this set further to 2209 delayed maintenance actions with a functionality risk.

Figure 19 shows the time series of the initial queue and the queues after applying these two removal factors.



Figure 19 Initial queue and reduced queues of delays

In hindsight, we could have skipped the removal of delayed maintenance without functionality risk because Figure 19 shows that the two queues of delays show a similar trend. Still, there is good reason to remove these maintenance actions. Delayed maintenance that is known to be not affecting functionality, can only spuriously correlate with functionality. Provided that common sense exists about the expert judgement regarding the functionality risk, this removal factor should be retained. Figure 19's initial queue of uncompleted maintenance actions evolves differently than the two queues of delays. Although the queue of uncompleted maintenance actions may also relate to functionality, this relation is not compelling for the causal effect of maintenance policy compliance. Removing compliant maintenance provides the essential access to the counterfactual reality. Besides achieving a required completion date, we may similarly attempt other requirements imposed on a maintenance policy by common sense. So, the applicability of the maintenance policy validation is not necessarily confined to the delays that we took from the case organisation's maintenance scorecard.

Common sense about the required completion date of maintenance appears to be problematic. This case organisation quantified a required completion date by a categorical variable. So, maintenance should be completed within either 1,7 or 30 days or deferred until a scheduled shutdown. A field expert claimed that these categories are imprecise in representing "genuine required completion dates". However, we choose to follow the group policy as determined by the case organisation rather than some policy as determined by the field expert's individual preferences. We realise that this controversy hampers inference precision.

The case organisation's convention to operationalise maintenance policy compliance by the performance indicators L^1 , L^2 and L^3 may improve by (*i*) defining a requirement (an acceptable backlog) on performance indicator L^3 , (*ii*) eliminating the definitional dependence between L^1 and L^2 and (*iii*) abandoning monthly averages like L^1 that are vulnerable to longitudinal redundancy as the sampling rate increases.

5.1.3 Findings regarding the "common sense evidence" objective

The case organisation uses a maintenance scorecard that includes leading and lagging performance indicators. The case organisation does not attempt to predict these maintenance performance indicators. Rather, it just qualitatively reviews their year-to-date values.

The case organisation's maintenance scorecard includes fewer redundancies (Section 3.2.1) than the maintenance scorecard from Table 4 but we found one in the leading performance indicators L^1 and L^2 . Longitudinal redundancy appears to be absent, but an increased sampling rate would introduce it when the monthly means are not adjusted accordingly.

The case organisation's convention to measure maintenance performance does not foster causal inferences, specifically because the sampling rate is not conducive to reconstructing the original signals of maintenance policy compliance and functionality. Informative peaks in the candidate cause that should have an effect have been missed or levelled out. The case organisation's sampling rate deserves reconsideration because it does not allow reconstruction of the original signal as required in Section 3.2.2.

The case organisation does not infer models from its maintenance performance indicators. So, these maintenance performance indicators merely reflect posterior satisfaction of goals. The completeness efficiency balance from Section 3.2.3 has

therefore never been a concern. By confining the maintenance policy validation to functionality, we are incomplete in the case organisation's objectives since we exclude objectives regarding "cost control", "resource allocation", "health, safety and environment". However, our attempt to validate some relationship between maintenance policy compliance and functionality alludes to the *intention* of maintenance *by definition* (CEN, 2001), (IEC, 1990). Moreover, completeness is unattainable as explained in Section 3.2.3.

The case organisation implicitly complies with a notion of satisficing by having defined subjective but attainable requirements on its maintenance performance indicators. The subjective requirement on functionality does not appear to be controversial within the case organisation. However, the subjective requirement on maintenance policy compliance is subject to critique. This work does not resolve controversy about the recordings of due dates of maintenance work orders. The maintenance policy validation in this case study therefore does not entirely comply with the inference objective about "common sense evidence" in this respect.

5.2 Choice of a sampling procedure

In this section, we will compose three candidate samples built on the same recording routines:

- The case organisation's conventional performance indicators (Figure 20);
- An alternative at an increased sampling rate (Figure 21);
- A dichotomous alternative (Figure 22).

Section 3.2.2 explained the importance of a sampling rate that enables reconstruction of the original signal. The plot of daily output and monthly availability (Figure 16) and the plot of delayed maintenance (Figure 18) showed that the case organisation's convention of recording maintenance policy compliance and functionality can be improved with respect to the sampling rate. To avoid longitudinal redundancy, an increased sampling rate also requires an adjustment of the case organisation's maintenance performance indicators. In this section, we will construct an alternative approach to measure the case organisation's common sense of maintenance policy compliance and functionality. We suspect that this alternative at an increased sampling rate serves inference precision.

Section 4.5.5 explained that the reliability engineering argument and the nonparametric argument additionally constrain the composition of the sample. In this section, we will assess to what extent the candidate samples satisfy these constraints.

Figure 20, Figure 21 and Figure 22 depict the candidate samples that we will consider for the maintenance policy validation. Figure 20 shows the case organisation's convention to represent functionality by monthly availability \overline{Y}_M and maintenance policy compliance by the monthly queue of delays D_M . Figure 21 shows a cardinal alternative that is expected to suit the maintenance prognostic argument. Functionality is now represented by the daily output Q_T and maintenance policy compliance by the daily queue of delays D_T . Finally, Figure 22 shows a dichotomous alternative that is expected to suit the reliability engineering argument and the nonparametric argument. Now functionality is expressed by the dichotomous variable Y_T from Equation 45 and policy compliance is represented by the dichotomous variable S_T that we will introduce in Equation 49.



Figure 20 Time series of the case organisation's performance indicators (d,y)[1,m]



Figure 21 Time series of the increased sampling rate alternative (d,q)[1,t]



Figure 22 Time series of the dichotomous alternative (s,y)[1,t]

Section 5.2.1 and Section 5.2.2 will discuss functionality Q_T maintenance policy compliance D_T in Figure 21 respectively. Section 5.2.3 and Section 5.2.4 will discuss functionality Y_T and maintenance policy compliance S_T in Figure 22 respectively.

5.2.1 Alternative for functionality at an increased sampling rate

The case organisation's convention to measure functionality has been built on daily output recordings. Figure 16 already showed that monthly availability \overline{Y}_M did not

reconstruct the daily output Q_T on which it was built. We therefore consider daily output Q_T as an alternative representation of the common sense about functionality.



Figure 23 Scatter plots of the one-step-ahead dependence in Q_{T} and Y_{M}

Figure 23 shows the one-step-ahead dependence in both daily output and monthly availability. The monthly availabilities \overline{Y}_M and \overline{Y}_{M+1} seem independent whereas daily output Q_T and Q_{T+1} seem dependent as can be concluded from the many data points on the diagonal $Q_T=Q_{T+1}$. So, Figure 23 reconfirms that knowledge about \overline{Y}_M is uninformative about \overline{Y}_{M+1} whereas Q_T predicts Q_{T+1} rather well.

Figure 23 also confirms that daily output Q_T either takes a high or a low value. By further increasing the sampling rate, more spikes in the output signal may be revealed. Still, daily output outperforms monthly availability in its reconstructive capabilities. At least, daily output Q_T adequately captures the major drops in functionality.



Figure 24 Observed frequency of functionalities $Q_{T} \mbox{ and } Y_{M}$

In Section 4.4.3, we conceived an information set $V = \{l_t, k_t, k_{t+1}\}$ with identical $\{k_t\}$ as a replication for the nonparametric argument (Figure 15). Since any empirical validation requires a sufficient number of replications, the observed frequencies of functionality K_T in the candidate samples is important. Figure 24 shows these observed frequencies of the corresponding daily output Q_T and monthly availability \overline{Y}_M . The observed frequencies in the sample $(d, \overline{y})_{[1,70]}$ range over [0,16] and those in the sample $(d,q)_{[1,1977]}$ range over [0,30]. In addition, the nonparametric argument requires that these [0,16] respective [0,30] replications (=V given $\{k_t\}$) should comprise sufficient instances of different $\{l_t\}$ to test for independence between maintenance policy compliance L_T and functionality K_{T+1} . However, we found many identical $U = \{l_t, k_t\}$ in the samples $(d, \overline{y})_{[1,70]}$

and $(d,q)_{[1,1977]}$. We therefore deem the samples $(d,\overline{y})_{[1,70]}$ and $(d,q)_{[1,1977]}$ as inadmissible evidence for a maintenance policy validation by the nonparametric argument.

Even in the tentative case that the samples $(d, \overline{y})_{[1,70]}$ and $(d,q)_{[1,1977]}$ would have had a sufficient number of different {l_t} at a constant {k_t}, Figure 24 still shows that $\overline{y}_{[1,70]}$ and $q_{[1,1977]}$ are scattered over 19 and 1750 different values respectively, which is just a subset of the infinite sample space of these functionality variables. In Section 4.4.3, we explained that the nonparametric argument only existentially confirms the prima facie causality in Equation 41 if all possible values kt are in the sample. This means that the cardinal samples $(d, \overline{y})_{[1,70]}$ and $(d, q)_{[1,1977]}$ could never existentially confirm the prima facie causality in Equation 41 by the nonparametric argument, but they may refute it or leave it undetermined. In the context of this work, the maintenance policy validation by the nonparametric argument and a cardinal sample could then only infer that maintenance is either unjustified or unjustifiable but *not* that it is justified. However, the maintenance policy validation by the maintenance prognostic argument and a cardinal sample could still existentially confirm or refute the prima facie causality in Equation 41 at some degree of inference precision. Therefore, the cardinal samples $(d, \bar{y})_{[1,70]}$, $(d,q)_{[1,1977]}$ are better suited to the maintenance prognostic argument than to the nonparametric argument.

We deem the cardinal functionality $\bar{y}_{[1,m]}$ and $q_{[1,t]}$ as unnecessarily precise for the nonparametric argument. We doubt whether adjacent functionality values in Figure 24 should really be seen as different replications. The inadmissibility of the cardinal samples $(d,\bar{y})_{[1,70]}$, $(d,q)_{[1,1977]}$ is therefore insufficient reason to disqualify the nonparametric argument.

5.2.2 Alternative for policy compliance at an increased sampling rate

The case organisation's convention expresses maintenance policy compliance by a monthly queue of delayed maintenance. Figure 18 already showed that a daily queue D_T better reconstructs the original signal. We therefore consider a daily queue D_T as an alternative representation of the organisation's common sense about maintenance policy compliance.



Figure 25 Scatter plot of the one-step-ahead dependence in D_T and D_M

Figure 25 confirms that a one-step-ahead dependence in the daily queue D_T appears to be much stronger than the dependence in the monthly queue D_M . The overdue times of the 2209 delays with a functionality risk are shown in Figure 26. The overdue time is the number of days that has passed since the required completion date. As the typical overdue time is several to many days, this confirms that a daily sampling rate suffices to reconstruct the queue signal.



Figure 26 Cumulative distribution of the overdue times with a functionality risk

Figure 26 shows that the queue is poorly refreshed within a day, since a large portion of the maintenance actions is overdue for at least 10 days. So, a maintenance action that resides in the queue today will often still be there tomorrow. Figure 25 therefore shows that D_T is a strongly autoregressive signal. About 400 maintenance actions are even overdue for more than one month. Still, Figure 25 shows a poor autocorrelation in D_M , which is mainly due to some outliers. By suspending these few outliers, we may also recognise an autoregressive property of D_M in Figure 25. However, this suspension would hamper a causal inference because these outliers are the informative peaks in Figure 18 that could potentially most clearly trigger a functionality effect. Since the case organisation only records maintenance actions by calendar date, a daily sampling rate is the highest attainable rate for maintenance policy compliance in this case study. However, the overdue times in Figure 26 also suggest that an increase of the sampling rate is superfluous.



Figure 27 Observed frequency of maintenance policy compliance D_T and D_M

In Section 4.3.3, we conceived an information set $V = \{l_t, k_t, k_{t+1}\}$ with identical $\{l_t\}$ as a replication for the reliability engineering argument (Figure 14). Since any empirical validation requires a sufficient number of replications, the observed frequencies of maintenance policy compliance L_T in the candidate samples is important. Figure 27 shows the observed frequencies of the corresponding daily queue D_T and monthly queue

D_M. The observed frequencies in the sample $(d, \overline{y})_{[1,70]}$ range over [0,4] and those in the sample $(d,q)_{[1,1977]}$ range over [0,114]. A maintenance policy validation by the reliability engineering argument discards many rarely observed replications, but we still deem the samples $(d,\overline{y})_{[1,70]}$, $(d,q)_{[1,1977]}$ as admissible to the reliability engineering argument in principle. Evidently, the observed frequencies of the organisation's convention $d_{[1,m]}$ are less promising than those of the alternative $d_{[1,t]}$.

Figure 27 shows the observed frequencies of the corresponding daily queue $D_{\rm T}$ and monthly queue D_M . In Figure 27, $d_{[1,m]}$ and $d_{[1,t]}$ are scattered over 41 and 123 different values respectively, which is just a subset of the infinite sample space of these maintenance policy compliance variables. In Section 4.3.3, we explained that the reliability engineering argument only existentially refutes the prima facie causality in Equation 41 if all possible values lt are in the sample. This means that the cardinal samples $(d, \overline{y})_{[1,70]}$ and $(d,q)_{[1,1977]}$ could never existentially refute the prima facie causality in Equation 41 by the reliability engineering argument, but they may confirm it or leave it undetermined. In the context of this work, the maintenance policy validation by the reliability engineering argument and a cardinal sample could then only infer that maintenance is either justified or unjustifiable but *not* that it is unjustified. However, the maintenance policy validation by the maintenance prognostic argument and a cardinal sample could still existentially confirm or refute the prima facie causality in Equation 41 at some degree of inference precision. Therefore, the cardinal samples $(\mathbf{d}, \overline{\mathbf{v}})_{[1,70]}$, $(\mathbf{d}, \mathbf{q})_{[1,1977]}$ are better suited to the maintenance prognostic argument than to the reliability engineering argument.

Due to the observational research, we have no control over the observed frequencies in Figure 27. The most frequent replications in $d_{[1,m]}$ and $d_{[1,t]}$ are all rather close to the mean of D_M and D_T respectively, whereas Figure 20 and Figure 21 suggest that particularly the major perturbations in maintenance policy compliance appear to be the most promising for a functionality effect. It is not counterintuitive to presume that the queue of delays with a functionality risk is just one of many causes of functionality. Therefore, we would not be surprised if subtle changes in the queue D_T do not find a functionality response. The most frequent replications in $d_{[1,m]}$ and $d_{[1,t]}$, that are the most promising replications from a statistical viewpoint, might therefore not be the preferred ones for the maintenance policy validation.

We deem the cardinal maintenance policy compliance $d_{[1,m]}$ and $d_{[1,t]}$ as unnecessarily precise for the reliability engineering argument. We doubt whether adjacent values of maintenance policy compliance in Figure 27 should be seen as different experiments. The reliability engineering argument may infer more precisely from a less precise categorical representation of maintenance policy compliance that does not scatter the evidence over so many different replications.

5.2.3 Dichotomous alternative for functionality

The case organisation adopted a dichotomous functionality Y_T as defined by Equation 45 (i.e. daily output larger or smaller than 16). Figure 28 shows the time series of both daily output $q_{[1,1977]}$ and dichotomous functionality $y_{[1,1977]}$. Figure 28 shows that the dichotomous functionality Y_T adequately captures the most important transitions from a

high daily output Q_T (Y_T=1) to a low daily output Q_T (Y_T=0) while ignoring the less important fluctuations in Q_T . We therefore consider Y_T as an appropriate dichotomous alternative for the common sense about functionality in terms of output.



Figure 28 Continuous QT and categorical YT representations of functionality

A dichotomous functionality Y_T appears to be widely explored. For example, any remaining useful life prediction relies on a dichotomous life or death variable. From the many possible trajectories of functionality, a state holding time (duration) of an upstate or a downstate seems the most important for decision making. These state holding times may therefore efficiently capture the behaviour of interest over a time interval. For the cardinal functionality $\bar{y}_{[1,70]}$ and $q_{[1,1977]}$ on the contrary, we typically cannot reduce the functionality trajectories of interest to some finite number. For the case study, Figure 24's observed frequencies leave little hope to even infer a two-day-ahead trajectory of functionality.

In Section 4.4.3, we conceived an information set $V=\{s_t, y_t, y_{t+1}\}$ with identical $\{y_t\}$ as a replication for the nonparametric argument (Figure 15). Since any empirical validation requires a sufficient number of replications, the observed frequencies of functionality Y_T in the candidate samples is important. From Figure 28, it follows that the time series $y_{[1,t]}$ comprises 307 replications of $Y_T=0$ and 1669 replications of $Y_T=1$. In addition, the nonparametric argument requires that these 307 and 1669 replications (=V given $\{y_t\}$) respectively should comprise sufficient instances of $\{s_t\}$ and $\{s_t'\}$ to compare the observed frequencies of functionality Y_{T+1} , given $U=\{s_t, y_t\}$ and $U'=\{s_t', y_t\}$. For the dichotomous sample $(s, y)_{[1,1976]}$, the sample space of U is finite, i.e. $U=\{\{0,0\},\{0,1\},\{1,0\},\{1,1\}\}$ and the observed frequencies of these U's are 277, 1643, 30 and 25 respectively. We therefore deem the dichotomous sample $(s, y)_{[1,1976]}$ as admissible evidence for a maintenance policy validation by the nonparametric argument.

The nonparametric argument requires that the dichotomous sample $(s,y)_{[1,1976]}$ comprises a sufficient number of every element in the sample space of $U=\{s_t,y_t\}$ to existentially confirm or refute the prima facie causality in Equation 41 at some degree of inference precision. Evidently, this similarly holds for the maintenance prognostic argument that takes any information set $V=\{s_t,y_t,y_{t+1}\}$ from the dichotomous sample $(s,y)_{[1,1976]}$ as a replication. Therefore, the dichotomous sample $(s,y)_{[1,1976]}$ equally suits the maintenance prognostic argument and the nonparametric argument.

5.2.4 Dichotomous alternative for maintenance policy compliance

We may consider a dichotomous variable S_T that captures the larger leaps in the queue D_T . For the complex multiple component item in this case study, we may expect that the composition of the queue is as important as its length. A cardinal queue D_T or D_M may scatter the sample over many different values. Some less precise dichotomous S_T that extracts the larger fluctuations may ultimately serve inference precision. We therefore define some S_T by Equation 49, separating large queue extensions (by more than 4) of the daily queue D_T from smaller and possibly less interesting queue extensions.

$$S_T = \begin{cases} 0, & \text{if } D_T - D_{T-1} \le 4\\ 1, & \text{if } D_t - D_{T-1} > 4 \end{cases}$$

$$49$$

Figure 29 shows that maintenance policy compliance S_T captures the observed explosions in the queue very well. The case organisation does not use some S_T . Thus, maintenance policy compliance S_T is just an additional construct that may serve the maintenance policy validation. Still, we consider S_T as the dichotomous alternative for the common sense about maintenance policy compliance.



Figure 29 Discrete D_T and categorical S_T representations of policy compliance

In Section 4.3.3, we conceived an information set $V=\{s_t, y_t, y_{t+1}\}$ with identical $\{s_t\}$ as a replication for the reliability engineering argument (Figure 14). Since any empirical validation requires a sufficient number of replications, the observed frequencies of maintenance policy compliance S_T in the candidate samples is important. From Figure 29, it follows that the time series $s_{[1,t]}$ comprises 1920 replications of $S_T=0$ and 55 replications of $S_T=1$. We therefore deem the dichotomous sample $(s,y)_{[1,1976]}$ as admissible evidence for a maintenance policy validation by the reliability engineering argument.

The reliability engineering argument requires that the dichotomous sample $(s,y)_{[1,1976]}$ fully covers the sample space of maintenance policy compliance S_T to test for the prima facie causality in Equation 41. Since $S_T=0$ and $S_T=1$ are both represented in the dichotomous sample $(s,y)_{[1,1976]}$, the reliability engineering argument could existentially confirm or refute the prima facie causality in Equation 41 at some degree of inference precision. Evidently, this similarly holds for the maintenance prognostic argument that takes any information set $V=\{s_t,y_t,y_{t+1}\}$ from the dichotomous sample $(s,y)_{[1,1976]}$ as a replication. Therefore, the dichotomous sample $(s,y)_{[1,1976]}$ equally suits the maintenance prognostic argument and the reliability engineering argument.
5.2.5 *Findings regarding the choice of a sampling procedure*

This section will present the potential effect of the candidate samples on the inference precision of the maintenance policy validation. We conceived the information set $V=\{l_t,k_t,k_{t+1}\}$ of the prima facie causality in Equation 41 as leading for the definition of a replication. The reliability engineering argument and the nonparametric argument prohibit particular causal explanations by holding some element in the information set V constant. This means that the reliability engineering argument only takes information sets $V=\{l_t,k_t,k_{t+1}\}$ with identical $\{l_t\}$ as a replication and that the nonparametric argument only takes information sets $V=\{l_t,k_t,k_{t+1}\}$ with identical $\{l_t\}$ as a replication and that the nonparametric argument only takes information sets $V=\{l_t,k_t,k_{t+1}\}$ with identical $\{l_t\}$ as a replication for the dichotomous sample $(s,y)_{[1,t]}$, the sample space of both maintenance policy compliance S_T and functionality Y_T is $\{0,1\}$. Therefore, the number of possible replications for the reliability engineering argument and for the nonparametric argument is also two. For the cardinal samples $(d,\bar{y})_{[1,m]}$ and $(d,q)_{[1,t]}$, the sample space of maintenance policy compliance and functionality is infinite. Therefore, the number of possible replications for the reliability engineering argument and for the nonparametric argument is also infinite (Table 6).

As a result, the reliability engineering argument cannot existentially refute the prima facie causality in Equation 41 by the samples $(d,\bar{y})_{[1,m]}$ and $(d,q)_{[1,t]}$ because it cannot test for independence at all d_m and d_t respectively. Similarly, the nonparametric argument cannot existentially confirm the prima facie causality in Equation 41 by the samples $(d,\bar{y})_{[1,m]}$ and $(d,q)_{[1,t]}$ because it cannot assess the likelihood of independence at all \bar{y}_m and q_t respectively.

From Table 6, it follows that only the dichotomous sample $(s,y)_{[1,t]}$ could either existentially confirm or existentially refute the prima facie causality in Equation 41 by the maintenance prognostic argument, the reliability engineering argument and the nonparametric argument.

| | Maintenance optimisation argument | Maintenance prognostic argument | Reliability engineering argument | Nonparametric argument |
|---|---|---------------------------------------|--|--|
| Definition of a replication | All $\{l_{t,k_t}\}$ | All $\{l_t, k_t, k_{t+1}\}$ | All $\{l_t, k_t, k_{t+1}\}$ given $\{l_t\}$ | All $\{l_t, k_t, k_{t+1}\}$ given $\{k_t\}$ |
| $(\mathbf{d}, \overline{\mathbf{y}})_{[1,t]}$ | 1 | 1 | 00 | 00 |
| (d , q) _[1,t] | 1 | 1 | 00 | 00 |
| (s , y) _[1,t] | 1 | 1 | 2 | 2 |

Table 6 Number of possible replications in V

Table 6 only lists the number of possible replications, whereas Table 7 lists the observed frequency of the replications in the specific case study. The observed frequencies for the maintenance optimisation argument and the maintenance prognostic argument directly follow from the length of the time series. The observed frequencies for the reliability engineering argument follow from Figure 27 and Figure 29 and the observed frequencies for the nonparametric argument follow from Figure 24 and Figure 28.

| | Maintenance optimisation argument | Maintenance prognostic argument | Reliability engineering argument | Nonparametric argument |
|--|---|---------------------------------------|---|--|
| Definition of a replication | All $\{l_t\}$ | All $\{l_t, k_t, k_{t+1}\}$ | $ \begin{array}{l} All \; \{l_t, k_t, k_{t+1}\} \\ given \; \{l_t\} \end{array} $ | All $\{l_t, k_t, k_{t+1}\}$ given $\{k_t\}$ |
| $(\mathbf{d},\!\bar{\mathbf{y}})_{[1,70]}$ | 70 | 69 | [0,4] | [0,16] |
| (d , q)[1,1977] | 1977 | 1976 | [0,114] | [0,30] |
| (s,y)[1,1976] | 1976 | 1975 | 55,1920 | 307,1668 |

Table 7 Observed frequencies of replications in the case study

Unsurprisingly, the observed frequencies range from zero where the number of possible replications is infinite. Moreover, the observed frequencies in the samples $(d,q)_{[1,1977]}$ and $(s,y)_{[1,1976]}$ that have been collected at a daily sampling rate seem to be more prolific than the observed frequencies in the sample $(d,\overline{y})_{[1,70]}$ that follow the case organisation's convention. The sample $(s,y)_{[1,1976]}$ in the case study includes all possible replications. So, the sample $(s,y)_{[1,1976]}$ potentially complies with the additional constraints on its composition as required for maintenance policy validation by the reliability engineering argument and the nonparametric argument.

5.3 Validation of the arguments

In this section, we will attempt to validate the candidate arguments from Chapter 4 by the samples $(d,\bar{y})_{[1,70]}$, $(d,q)_{[1,1977]}$ and $(s,y)_{[1,1976]}$ from Section 5.2. In Table 5, we already preliminarily compared the arguments from Chapter 4 on inference precision. We will now expand on this comparison by involving the evidence about the case study.

A spatiotemporally constrained sample $(l,k)_{[1,t]}$ only comprises *some* l,t,k which is insufficient for the prima facie causality in Equation 41 that holds at *some* l and *all* t,k. The observational research construct does not support *universal* claims about the population $(l,k)_{[1,\infty)}$. Still, we could existentially claim the likelihood of the prima facie causality in Equation 41. We will infer this likelihood by using the maintenance prognostic argument, the reliability engineering argument and the nonparametric argument in this section. The *universal* refutation of the maintenance optimisation argument will straightforwardly follow from the sample $(l,k)_{[1,t]}$.

We will conclude this section with a choice of the argument and the sample that most precisely infers a claim regarding the prima facie causality in Equation 41 in this realistic case study.

5.3.1 Validation of the maintenance optimisation argument

The maintenance optimisation argument does not require a notion of causality. Its model M1 just asserts that functionality K_T and maintenance policy compliance L_T are

equivalent.

$$k_t = f(l_t) \tag{50}$$

The samples $(d,\bar{y})_{[1,70]}$, $(d,q)_{[1,1977]}$ and $(s,y)_{[1,1976]}$ refute Equation 50 since they all include counterexamples of maintenance policy values l_t that map to several functionality values k_t and conversely; i.e. Figure 20, Figure 21 and Figure 22 refute Equation 50. This refutation is universal. So, enlarged samples will not change our position. We therefore conclude that the case organisation's common sense *definition* of maintenance policy compliance and functionality does not comply with the model M1 of the maintenance optimisation argument.

Many maintenance optimisations *define* an upstate/downstate as equivalent to the absence/presence of a maintenance queue. This alternative definition would refute all recorded maintenance actions in the case study as inadmissible evidence since none of these 6342 maintenance actions complied with this equivalence. We may resolve this conflict in *definitions* by decomposing the item into components. At least, the widespread applications of fault trees and reliability block diagrams suggests some resilience of an item's functionality against component faults that may in fact coincide with the presence of a maintenance action.

Let us therefore explore an attempt to *unify* the case organisation's definition of maintenance with the alternative definition of these maintenance optimisations. Then, the following equivalence would exist:

$$(K_1, \dots, K_m = k_1, \dots, k_m) \leftrightarrow (X_1, \dots, X_m = x_1, \dots, x_m)$$

$$51$$

where the series X_i represents presence or absence of an uncompleted maintenance action and the series K_i represents the functionality of the item's components. Then, we should *firstly* assess the logic K=f(K_i) that maps the component functionality K_i to the item's functionality K. For cases where combinatorial logic does not suffice, Rauzy (2008) and Remenyte-Prescott and Andrews (2009) extended to incoherency and Vesely (2002), DiStefano and Puliafito (2008) and Xu et al. (2009) extended to dynamic logic. Still, Zio (2009) remarked that components increasingly operate in networks (system of systems), the behaviour of which may seem to be chaotic and only subjectively assessable. In addition, Bucci et al. (2008), Yuge and Yanagi (2008), Schüller (1997) and Bobbio et al. (2003) warned about state space explosions that make an explicit statement about a presumed mapping K=f(K_i) a tedious job. Never mind that we may simply be unable to efficiently collect the evidence for its empirical validation. Otherwise, K=f(K_i) remains arbitrary expert judgement that affects inference precision.

Secondly, Bouissou and Bon (2003), Vaurio (1998), Walter et al. (2008) and Rehage and Carl (2005) recognised that faults may propagate within an item. So, past values of the series K_i influence functionality predictions. For a posterior validation of Equation 51, predictions are not needed but they are needed in maintenance decision making that can only influence the future. Even under a presumed Markov property that only takes the current value of the vector K_T in a body of knowledge $U=\{k_{1,t},...,k_{m,t}\}$ may already appear to be too large to efficiently collect the evidence to enable the empirical validation of the one-step-ahead functionality vector K_{T+1}:

$$pr_{K_{i,T+1}|K_T}(k_{i,t+1}|k_{1,t},\dots,k_{m,t})$$
52

Finally, the case organisation's common sense definition of maintenance includes some counterexamples like online inspections that can only be executed on an item in upstate. Therefore, we will never find an Equation 51 that unifies the case organisation's common sense definition with the alternative definition of many maintenance optimisations. We choose to retain the case organisation's common sense definition of maintenance, which implies a universal refutation of the maintenance optimisation argument as already suspected in Table 5.

5.3.2 Validation of the maintenance prognostic argument

Figure 13 showed that the maintenance prognostic argument comprises all three causality principles (Section 2.3.3). Therefore, the maintenance prognostic argument would have been decidable about causality if it were sound. In Section 4.2.1, we already suspected being deprived of in-depth knowledge about a universal model M2 and about a universal error distribution P6 in a typical case study. We therefore resort to some *existential* claim about the likelihood of some presumed model M2 and some presumed error distribution P6. Likelihood ratios generally favour the highest dimensional model from a set of nested candidate models because it would be entirely coincidental that maximum likelihood estimations yield a zero parameter. This section will alternatively illustrate an information theoretical approach to model selection that is less supportive to high dimensional models. Still, this information theoretical approach will not resolve both the model uncertainty and the parameter uncertainty of the maintenance prognostic argument which affects inference precision.

The maintenance prognostic argument takes any information set V={P3,P4} of the prima facie causality in Equation 41 as a replication (Table 7). For the samples $(d,\bar{y})_{[1,70]}$ and $(d,q)_{[1,1977]}$, we will show that the model uncertainty and the parameter uncertainty are irresolvable. We will just evaluate some linear regression models, which leaves uncertainty about the adequacy of other possible candidate model families. However, for the sample $(s,y)_{[1,1976]}$ we will show that the model uncertainty is potentially eliminable without additional presumptions. Then, we will evaluate all Bernoulli models that are possible within an information set V={ $l_{t,k_t,k_{t+1}}$ } of the prima facie causality in Equation 41.

This section will proceed with a concise and informal introduction to Kullback-Leibler information that underlies the model selection criteria that we use: Akaike's Information Criterion AIC (Akaike, 1973) and Akaike's Corrected Information Criterion AIC_c (Hurvich & Tsai, 1989). We will then demonstrate how AIC and AIC_c could decide about the existence of a prima facie causality for the samples $(d,\bar{y})_{[1,70]}$, $(d,q)_{[1,1977]}$ and $(s,y)_{[1,1976]}$. We will finally review the initial assessment of the inference precision of a maintenance policy validation by the maintenance prognostic argument in Table 5.

Let $g(l_t,k_t)$ represent the true probability of functionality K_{T+1} .

Let $f(l_t,k_t|\theta)$ represent an approximating probability function of functionality K_{T+1} that follows from the antecedents of the model M3 in Figure 13 that comprise the parameters θ .

Then, the Kullback-Leibler information I(g,f) follows from:

$$I(g,f) = \iint \ln\left(\frac{g(l_t,k_t)}{f(l_t,k_t|\theta)}\right)g(l_t,k_t)\delta l_t\delta k_t = E_{g(l_t,k_t)}\left[\ln\left(\frac{g(l_t,k_t)}{f(l_t,k_t|\theta)}\right)\right]$$
53

Equation 53 reads as: "The amount of information lost when the true probability $g(l_t,k_t)$ of functionality K_{T+1} has been approximated by the probability function $f(l_t,k_t|\theta)$ ". If the approximating probability function were true, i.e. $f(l_t,k_t|\theta)=g(l_t,k_t)$, the Kullback-Leibler information would be zero I(g,f)=0. The notation $g(l_t,k_t)$ just conveys that the integration is over l_t,k_t , the full truth does not need human concepts like probability functions and parameters. Equation 53 can alternatively be expressed as the difference between two expectations:

$$I(g, f) = E_{g(l_t, k_t)} [ln(g(l_t, k_t))] - E_{g(l_t, k_t)} [ln(f(l_t, k_t | \theta))]$$
54

For the presumed approximating probability function $f(l_t,k_t|\theta_0)$, the parameters θ_0 that minimise the loss of information follow from the derivative:

$$\frac{\delta(I(g,f))}{\delta\theta} = \frac{\delta(E_{g(l_t,k_t)}[ln(g(l_t,k_t))] - E_{g(l_t,k_t)}[ln(f(l_t,k_t|\theta))])}{\delta\theta} = 0$$
55

Provided that I(g,f) is continuous and convex in θ_0 :

$$\frac{\delta^2 \left(I(g, f) \right)}{\delta \theta_0^2} > 0 \tag{56}$$

The true probability $g(l_t,k_t)$ is independent of human conceptions like a presumed best approximating probability function $f(l_t,k_t|\theta_0)$. Even in the fictitious case of a true probability function, $f(l_t,k_t|\theta_0)=g(l_t,k_t)$, the true parameters θ_0 would have been a constant. Therefore, the first term in Equation 54 rules out as a constant C that could be ignored in its derivative (Equation 55):

$$I(g, f)_{\theta_0} = C - E_{g(l_t, k_t)} \Big[ln \big(f(l_t, k_t | \theta_0) \big) \Big]$$
57

Equation 57 shows the amount of information lost when the true probability $g(l_t,k_t)$ has been approximated by the probability function $f(l_t,k_t|\theta_0)$ where θ_0 represents the true best approximating parameters in $f(l_t,k_t|\theta_0)$. So, the model selection target is to find some best approximating probability function $f(l_t,k_t|\theta_0)$ among a set of candidates that minimises Equation 57. If the true probability $g(l_t,k_t)$ were known, Equation 57 would have been an adequate model selection criterion to compare various approximating probability functions $f(l_t,k_t|\theta)$ and to assess their best parameters θ_0 by Equation 55. Since the true probability $g(l_t,k_t)$ is typically unknown, Equation 57 is not straightforwardly assessable, but we may estimate the best parameters θ_0 from some "good" sample $(l_t,k_t|_{t_1,t_1})$. Akaike (1973) therefore tried to find a rigorous way to estimate an expectation of the last term in Equation 57 from a "good" sample $(l,k)_{[1,t]}$:

$$E_{g((l,k)_{[1,t]})}\left[E_{g(l_{t},k_{t})}\left[ln\left(f((l,k)_{[1,t]}|\theta\right)\right)\right]\right]$$
58

Akaike (1973) realised that the most likely parameters θ_{mle} that are estimated from a "good" sample $(l,k)_{[1,t]}$ are biased estimates for the expected loss of information:

$$E_{g((l,k)_{[1,t]})}[I(g,f)_{\theta_0}] < C - E_{g((l,k)_{[1,t]})}\left[E_{g(l_t,k_t)}\left[ln\left(f((l,k)_{[1,t]}|\theta_{mle})\right)\right]\right]$$
59

However, under specific conditions, an asymptotically unbiased estimation of the expected loss of information in Equation 59 follows from:

$$E_{g((l,k)_{[1,l]})}[I(g,f)_{\theta_0}] = C + \left(K - \ln\left(\mathcal{L}(\theta_{mle}|(l,k)_{[1,l]})\right)\right)$$
60

Where ln(.) represents the log likelihood function and K the number of estimable parameters θ . As opposed to Equation 57, Equation 60 provides a model selection criterion that is straightforwardly estimable from a "good" sample (l,k)_[1,t]. Akaike (1973) proposed a model selection criterion AIC that is based on minimising the expected loss of information in Equation 60 by:

$$AIC = 2 \times \left(K - ln\left(\mathcal{L}(\theta_{mle}|(l,k)_{[1,t]}) \right) \right)$$
61

A "good" sample should comprise random trials from the true probability $g(l_t,k_t)$. A stratified time series $(l,k)_{[1,t]}$ collected by observational research can therefore not be deemed as "good". In Section 4.2.3, we already mentioned that it would be odd to distinguish identical information sets V as different experiments in a test for a prima facie causality. We therefore initially presume the non-causalities that hold for the information set V as shown in Figure 11. These presumed non-causalities will only be validated in Section 5.5.3 by exploiting the information about time in the sample $(l,k)_{[1,t]}$.

A "good" sample should also be sufficiently large because as the number of estimable parameters K increases in comparison with the sample size N, the model selection criterion AIC is known to become negatively biased, i.e. the last term in Equation 59 is underestimated by the last term of Equation 60. This bias can lead to overfitting. We therefore consider a small sample correction AIC_c that is exactly unbiased if the true probability $g(l_t,k_t)$ is a linear regression model (Hurvich & Tsai, 1989). As the sample size N increases, AIC_c reduces to AIC:

$$AIC_{c} = 2 \times \left(\frac{KN}{N - K - 1} - \ln\left(\mathcal{L}(\theta_{mle}|(l,k)_{[1,t]})\right)\right)$$

$$62$$

The relative ΔAIC score with respect to the best approximating probability function $f_{min}(l_t,k_t|\theta_0)$ among an arbitrary set of alternatives $f_i(l_t,k_t|\theta_0)$ is assessable by:

$$\Delta AIC_{f_i} = AIC_{f_i} - AIC_{f_{min}} \tag{63}$$

Equation 63 expresses the relative loss of information if the true probability $g(l_t,k_t)$ were approximated by the probability function $f_i(l_t,k_t|\theta_0)$ rather than by the best approximating probability function $f_{min}(l_t,k_t|\theta_0)$ in that arbitrary set of alternatives. If the relative ΔAIC score is small, i.e. $\Delta AIC < 2$, the probability function $f_i(l_t,k_t|\theta_0)$ would still be a relatively good approximation with respect to the best approximating probability function $f_{min}(l_t,k_t|\theta_0)$. If the relative ΔAIC score is big, i.e. $\Delta AIC > 10$, the candidate probability function $f_i(l_t,k_t|\theta_0)$ should be ignored (Burnham & Anderson, 2010). As we could typically only evaluate a finite number of approximating probability functions $f_i(l_t,k_t|\theta_0)$ from an infinite set of possible candidates, the outcome of Equation 63 may be overthrown by an up-to-now unconsidered better approximating probability function $f'_{min}(l_t,k_t|\theta_0)$.

AIC allows for comparisons of non-nested candidate models (i.e. linear and non-linear models) as long as they all map to the same conclusion. So, mappings to k_{t+1} and $ln(k_{t+1})$ are not immediately comparable. Moreover, an AIC score only holds for a specific sample. So, we cannot say which of the samples $(d,\bar{y})_{[1,70]}$, $(d,q)_{[1,1977]}$ and $(s,y)_{[1,1976]}$ has the lowest expected information loss. In the remainder of this section, we will compare the expected information loss of some candidate models M3: $f_i(l_t,k_t|\theta_0)$ for the samples $(d,\bar{y})_{[1,70]}$, $(d,q)_{[1,1977]}$ and $(s,y)_{[1,1976]}$ separately.

The maintenance prognostic argument in Figure 13 comprises an approximating probability function M3:Pr(P4|P3,M2,P6) that has three antecedents:

- P3: A proposition that straightforwardly follows from common sense about the sample $(l,k)_{[1,t]}$.
- M2: A controversially presumed model that maps the proposition P3 to the conclusion C2, i.e. an estimation of the functionality K_{T+1} .
- P6: A controversially presumed error distribution that is independent of the proposition P3. Equation 42 already showed that P4=M2(P3)+P6.

As explained in Section 4.2.1, the maintenance prognostic argument remains undecidable because of the controversy about the universal model M2 and about the universal error distribution P6. Still, we could existentially claim the likelihood of some arbitrarily presumed model M2 and some arbitrarily presumed error distribution P6.

The samples $(d,\bar{y})_{[1,70]}$ and $(d,q)_{[1,1977]}$ allow for an infinite set of possible presumptions regarding the model M2 and the error distribution P6. In this exposition, we only select some linear regression models that presume (*i*) linearity for the model M2 and (*ii*) a normal error distribution P6. These additional presumptions regarding the parameters θ do *not* rely on in-depth knowledge about the case study. The inferred claim regarding the prima facie causality in Equation 41 therefore only holds with respect to the arbitrary set of linear regression models that we considered.

In Section 4.2.3, we conceived every information set $V=\{P3,P4\}=\{l_t,k_t,k_{t+1}\}$ as a replication to validate a claim regarding the prima facie causality in Equation 41 by the maintenance prognostic argument. For the cardinal samples $(d,\bar{y})_{[1,70]}$ and $(d,q)_{[1,1977]}$, a replication could include the following linear regression model families:

$$\begin{array}{ll} M3_0: (k_{t+1} = p_0 + \sigma N(0,1) &) \to (L_T \not \to K_{T+1}) \\ M3_1: (k_{t+1} = p_0 + p_1 k_t + \sigma N(0,1) &) \to (L_T \not \to K_{T+1}) \\ M3_2: (k_{t+1} = p_0 + p_1 k_t + p_2 l_t + \sigma N(0,1) &) \to (L_T \to K_{T+1}) \end{array}$$

$$\begin{array}{ll} 64 \\ \end{array}$$

The linear regression models in Equation 64 imply an approximating probability function M3: Pr(P4|P3,M2,P6) for functionality K_{T+1} with the parameters $\theta = \{p_0,p_1,p_2,\sigma^2\}$:

$$M3: pr_{K_{T+1}|L_T,K_T}(k_{t+1}|l_t,k_t,\theta) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{\varepsilon_{l+1}^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(k_{t+1}-(p_0+p_1k_t+p_2l_t))^2}{2\sigma^2}}$$
65

The linear regression models M3₀ and M3₁ refute the prima facie causality in Equation 41 by the parameter P₂=0 which implies that maintenance policy compliance L_T cannot have any effect on functionality K_{T+1}. On the other hand, the linear regression model M3₂ confirms the prima facie causality in Equation 41 by the parameter P₂≠0 which implies that maintenance policy compliance L_T does affect functionality K_{T+1}. Since we could not compose an exhaustive set of possible approximating probability functions for the cardinal samples $(d,\bar{y})_{[1,70]}$ and $(d,q)_{[1,1977]}$, any subset of approximating probability functions (like those in Equation 64) is insufficient to conclusively validate a maintenance policy by the maintenance prognostic argument. The set of regression models in Equation 64 does not originate from in-depth knowledge that could alleviate the model uncertainty that hampers inference precision. Still, we could existentially claim the prima facie causality in Equation 41 with respect to this arbitrary set of regression models in Equation 64. We just proceed with an assessment of the AIC scores of the linear regression models in Equation 64:

The likelihood of a linear regression model with parameters θ follows from:

$$\mathcal{L}(\theta|(l,k)_{[1,t]}) = pr((l,k)_{[1,t]}|\theta) = \prod_{i=1}^{t-1} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\varepsilon_{i+1}^2}{2\sigma^2}}$$
66

So, the log likelihood function is:

$$ln\left(\mathcal{L}\left(\theta|(l,k)_{[1,t]}\right)\right) = -\frac{t-1}{2}ln(2\pi) - \frac{t-1}{2}ln(\sigma^2) - \sum_{i=1}^{t-1}\frac{\varepsilon_{i+1}^2}{2\sigma^2}$$
67

The maximum likelihood parameters θ_{mle} follow from the derivative of Equation 67:

$$\frac{\delta\left(\ln\left(\mathcal{L}\left(\theta|(l,k)_{[1,l]}\right)\right)\right)}{\delta\theta} = 0$$
68

Provided that Equation 68 is continuous and concave in θ_{mle} :

$$\frac{\delta^2 \left(ln \left(\mathcal{L} \left(\theta_{mle} | (l, k)_{[1, t]} \right) \right) \right)}{\delta \theta_{mle}^2} < 0$$
⁶⁹

The most likely parameters θ_{mle} enable an assessment of the AIC scores of the linear

regression models in Equation 64:

$$AIC = 2 \times \left(K - \ln \left(\mathcal{L}(\theta_{mle} | (l, k)_{[1,t]}) \right) \right)$$

$$70$$

The model M3₀ has two estimable parameters $\theta = \{p_0, \sigma^2\}$, i.e. K=2 and M3₂ has four estimable parameters $\theta = \{p_0, p_1, p_2, \sigma^2\}$, i.e. K=4. This implies that the higher dimensional model M3₂ could only attain a preferred lower AIC score if it is sufficiently more likely than the model M3₀. If we were to alternatively select a linear regression model from Equation 64 by a log likelihood ratio LR, the prima facie causality in Equation 41 would become nearly irrefutable because it would be entirely coincidental that the maximum likelihood estimation of the parameter P₂ is exactly zero. Still, we will assess the log likelihood ratio LR with respect to the most parsimonious linear regression model M3₀:

$$LR = ln\left(\mathcal{L}(M3_{i}|(l,k)_{[1,t]})\right) - ln\left(\mathcal{L}(M3_{0}|(l,k)_{[1,t]})\right)$$
71

Table 8 now surveys the results for the N=69 replications that we found in the sample $(d,\bar{y})_{[1,70]}$:

- The maximum likelihood estimation of the parameters θ for each linear regression model in Equation 64;
- The number of estimable parameters K for each linear regression model in Equation 64;
- The relative $\triangle AIC$ score and the relative $\triangle AIC_c$ score that follow from Equation 63 for each linear regression model in Equation 64;
- The log likelihood LR of each linear regression model in Equation 64 that follows from Equation 71.

Table 8 shows that all candidate models are nearly equally likely since LR \approx 0. So, the empirical support as implied by the relative Δ AIC scores (Burnham & Anderson, 2010) in Table 8 is nearly entirely driven by the number of parameters K in Equation 61. Then, the relative Δ AIC scores favour the most parsimonious model M3₀ (with the smaller number of parameters for the sample (d, \bar{y})_[1,70]. The model M3₀ considers functionality \bar{Y}_{M+1} as independent trails from a normal distribution. So, the models M3₁ and M3₂ appear to be overfitting.

| | MLE distribution of P4 | =M2(P3)+P6 | K | LR | ΔΑΙC | AAIC _c |
|-----------------|--|-------------------------------------|---|------|------|--------------------------|
| M30 | $\bar{y}_{m+1} = 0.83 + 0.20N(0$ | 2 | 0 | 0 | 0 | |
| M31 | $\bar{y}_{m+1} = 0.84 - 0.02\bar{y}_m$ - | + 0,20 <i>N</i> (0,1) | 3 | 0,01 | 2 | 2 |
| M3 ₂ | $\bar{y}_{m+1} = 0.88 - 0.04\bar{y}_m + 0, \dots d_m + 0.20N(0,1)$ | | | 0,22 | 4 | 4 |
| | Level of empirical | $\Delta AIC, \Delta AIC_{c}$ scores | | | | |
| | support for a model | | | | | |
| | Substantial: | 0-2 | | | | |
| | Considerably less: | 4-7 | | | | |
| | Essentially none: | >10 | | | | |
| | | | | | | |

Table 8 Relative information loss of some models M3 for (d,y)[1,70]

This confirms our suspicion that the case organisation's convention to maintenance performance does not allow for predictions by some linear model, nor do the scatter plots in Figure 23 suggest that some nonlinear model would have allowed for these predictions. Tentatively, the empirical support for the model M3₀ may be overthrown by some unconsidered candidate model but the evolution of D_M and \overline{Y}_M in Figure 20 does not allude to obvious alternative candidate models. Still, we deem this result as indecisive about the existence of the prima facie causality in Equation 41 because the monthly sampling rate of the sample $(d, \overline{y})_{[1,70]}$ evidently did not adequately reconstruct the original signals as required in Section 3.2.2. Possibly, the functionality response to maintenance policy compliance just remained invisible due to the low sampling rate. The maintenance policy validation by the maintenance prognostic argument and the sample $(d, \overline{y})_{[1,70]}$ remains indecisive about the prima facie causality in Equation 41. We therefore conclude:

Maintenance policy compliance L_T may or may not prima facie cause functionality K_{T+1} with respect to the information set $V = \{l_t, k_t, k_{t+1}\}$.

| | MLE distribution of P4 | =M2(P3)+P6 | K | LR | ΔΑΙC | ΔAICc |
|-----------------|---|-------------------------------------|---|------|-------|-------|
| M30 | $q_{t+1} = 18,69 + 7,72N(0$ | $q_{t+1} = 18,69 + 7,72N(0,1)$ | | | 17574 | 17574 |
| M31 | $q_{t+1} = 3,53 + 0,81q_t +$ | 4,52 <i>N</i> (0,1) | 3 | 8781 | 14 | 14 |
| M3 ₂ | $q_{t+1} = 3,80 + 0,81q_t - 0,01d_t + 4,52N(0,1)$ | | | 8789 | 0 | 0 |
| | Level of empirical | $\Delta AIC, \Delta AIC_{c}$ scores | | | | |
| | support for a model | , | | | | |
| | Substantial: | 0-2 | | | | |
| | Considerably less: | 4-7 | | | | |
| | Essentially none: | >10 | | | | |
| | | | | | | |

 Table 9 Relative information loss of some models M3 for (d,q)[1,1977]

We show analogous results for the N=1976 replications that we found in the sample $(d,q)_{[1,1977]}$ in Table 9. As opposed to the sample $(d,\bar{y})_{[1,70]}$, the sample $(d,q)_{[1,1977]}$ shows extremely little empirical support for the most parsimonious model M3₀. Evidently, the sample $(d,q)_{[1,1977]}$ gives a much better reconstruction of the original signal as required in Section 3.2.2.

The results in Table 9 strongly support the model $M3_2$ that existentially confirms the prima facie causality in Equation 41. This existential confirmation may be overthrown by some unconsidered candidate model that outperforms the model $M3_2$ in expected information loss. This concern is realistic for the sample $(d,q)_{[1,1977]}$ because the scatter plot in Figure 23 already showed that the dependence between Q_T and Q_{T+1} is far from linear. Moreover, we know that the output, on which functionality Q_{T+1} has been built, is physically constrained between some upper and lower limit, whereas maintenance policy compliance D_T is in $[0,\infty)$. Figure 21 shows that functionality Q tends to reside at either a high or a low output and that maintenance policy compliance D evolves like some random walk with some outliers. We therefore deem none of the linear regression models in Table 9 as likely. However, under the acceptance of this model uncertainty, the maintenance policy validation by the maintenance prognostic argument and the

sample $(d,q)_{[1,1977]}$ would existentially confirm the prima facie causality in Equation 41. We therefore conclude:

If model uncertainty has been accepted, maintenance policy compliance D_T would prima facie cause functionality Q_{T+1} with respect to the information set $V = \{d_t, q_t, q_{t+1}\}$.

In Figure 11, we specified (by omitting arrows) the non-causality assumptions of the maintenance prognostic argument that allow us to interpret an existential confirmation of the prima facie causality in Equation 41 as being causal. Pearl (2010) and Spirtes (2000) already warned that the path graphs like those in Figure 11 do not conclusively follow from statistical associations. For example the presumed independencies between L_T , B and K_T , B remain untestable because the background variable B remains unobserved, which delimits any causal interpretation of an inferred prima facie cause. However, the path graph of the maintenance prognostic argument in Figure 11 also presumes an independence between L_T and K_T that is straightforwardly testable. This independence between L_T and K_T prevents functionality K_T from being a mediator $L_T \rightarrow K_T \rightarrow K_{T+1}$ or from being a confounder $K_T \rightarrow (L_T, K_{T+1})$. In the latter case, maintenance policy compliance would not cause functionality K_T or K_{T+1} despite an existential confirmation of the prima facie causality in Equation 41.

We now proceed with a test for the independence between D_T and Q_T by the maintenance prognostic argument and the sample $(d,q)_{[1,1977]}$.

| [1 | r_{D_T,Q_T} | $r_{D_T,Q_{T+1}}$ | | ۲ 1,00 | -0,24 | -0,221 | |
|---------------------|-----------------|-------------------|---|--------|-------|--------|----|
| r_{D_T,Q_T} | 1 | $r_{Q_T,Q_{T+1}}$ | = | -0,24 | 1,00 | 0,81 | 72 |
| $r_{D_{T},Q_{T+1}}$ | r_{Q_T,Q_T+1} | 1 | | L-0,22 | 0,81 | 1,00 | |

The correlation matrix in Equation 72 shows a strong linear dependence between functionality Q_T and Q_{T+1} but in fact functionality Q just resides at either a high or a low value as shown in Figure 21. Although the best approximating function M3₂ also confirmed a linear dependence between maintenance policy compliance D_T and functionality Q_{T+1} , Equation 72 shows that this is actually the weakest linear dependence in the information set $V=\{d_t,q_t,q_{t+1}\}$.

| | MLE distribution of Q ₁ | , | | K | LR | ΔΑΙC | AAIC _c |
|-----------------|--|--------------------------------------|--|---|-----|------|--------------------------|
| M3 ₃ | $q_t = 18,69 + 7,72N(0,1)$ | | | 2 | 0 | 1377 | 1377 |
| M34 | $q_t = 20,37 - 0,06d_t + 2$ | $b_t = 20,37 - 0,06d_t + 7,50N(0,1)$ | | | 690 | 0 | 0 |
| | Level of empirical support for a model | $\Delta AIC, \Delta AIC_c$ scores | | | | | |
| | Substantial: | 0-2 | | | | | |
| | Considerably less: | 4-7 | | | | | |
| | Essentially none: | >10 | | | | | |
| | | | | | | | |

Table 10 Relative information loss of some models Pr(Q_T|D_T, θ) for (d,q)_[1,1977]

Table 10 compares an arbitrary set of two candidate linear regression models that existentially confirm $(M3_3)$ or existentially refute $(M3_4)$ the independence between D_T and Q_T . We did not involve functionality Q_{T+1} because Q_{T+1} is known not to influence

the relation between D_T and Q_T due to the first causality principle (Section 2.3.3) asserting that the future cannot cause the past. Table 10 shows no empirical support for the model M3₃ that existentially confirms the independence between D_T and Q_T as required by the path graph of the maintenance prognostic argument in Figure 11. This result allows for the existence of the alternative causal structures $Q_T \rightarrow (D_T, Q_{T+1})$ and $D_T \rightarrow Q_T \rightarrow Q_{T+1}$ that is problematic for making a causal interpretation of the inferred prima facie causality in Equation 41 as implied by the best approximating probability function M3₂ in Table 9. We therefore ultimately deem that the maintenance policy validation by the maintenance prognostic argument and the sample $(d,q)_{[1,1977]}$ is insufficient for the justification of maintenance that we pursue in this work.

For the dichotomous sample $(s,y)_{[1,1976]}$, it would be inappropriate to conceive the distribution of functionality Y_{T+1} as some linear regression model that allows for an infinite sample space of Y_{T+1} . A Bernoulli model is the obvious choice of the distribution of the dichotomous functionality Y_{T+1} . The approximating probability function Pr(P4|P3,M2,P6) in the maintenance prognostic argument in Figure 13 would then reduce to some Bernoulli model with parameters θ that holds for some given body of knowledge U={ s_t,y_t } that has been claimed by proposition P3:

$$M3: pr_{Y_{T+1}|S_T,Y_T}(1|U = \{s_t, y_t\}, \theta) = p_u^{y_{t+1}}(1 - p_u)^{1 - y_{t+1}}$$
73

Since the body of knowledge U={s_t,y_t} has a finite sample space, i.e. {{0,0},{0,1},{1,0},{1,1}}, the maintenance prognostic argument becomes potentially decidable about these four presumed Bernoulli parameters $\theta = \{p_{\{0,0\}}, p_{\{0,1\}}, p_{\{1,0\}}, p_{\{1,1\}}\}$. Table 11 surveys *all* possible equalities and inequalities among these four Bernoulli parameters θ .

| | Candidat | e Bernoulli | model | | Causal implication |
|--------------------------|-----------------|-----------------|-----------------|---------------|---|
| M3 _A : | $p_{\{0,0\}};$ | $p_{\{0,1\}};$ | $p_{\{1,0\}};$ | $p_{\{1,1\}}$ | $\rightarrow (S_T \rightarrow Y_{T+1})$ |
| M3 _B : | $p_{\{0,0\}} =$ | $p_{\{0,1\}};$ | $p_{\{1,0\}};$ | $p_{\{1,1\}}$ | $\rightarrow (S_T \rightarrow Y_{T+1})$ |
| M3 c: | $p_{\{0,0\}} =$ | $p_{\{1,0\}};$ | $p_{\{0,1\}};$ | $p_{\{1,1\}}$ | $\rightarrow (S_T \not\rightarrow Y_{T+1})$ |
| M3 _D : | $p_{\{0,0\}} =$ | $p_{\{1,1\}};$ | $p_{\{0,1\}};$ | $p_{\{1,0\}}$ | $\rightarrow (S_T \rightarrow Y_{T+1})$ |
| M3 _E : | $p_{\{0,0\}};$ | $p_{\{0,1\}} =$ | $p_{\{1,0\}};$ | $p_{\{1,1\}}$ | $\rightarrow (S_T \rightarrow Y_{T+1})$ |
| M3 _F : | $p_{\{0,0\}};$ | $p_{\{0,1\}} =$ | $p_{\{1,1\}};$ | $p_{\{1,0\}}$ | $\rightarrow (S_T \not\rightarrow Y_{T+1})$ |
| M3 _G : | $p_{\{0,0\}};$ | $p_{\{0,1\}};$ | $p_{\{1,0\}} =$ | $p_{\{1,1\}}$ | $\rightarrow (S_T \rightarrow Y_{T+1})$ |
| М3 н: | $p_{\{0,0\}};$ | $p_{\{0,1\}} =$ | $p_{\{1,0\}} =$ | $p_{\{1,1\}}$ | $\rightarrow (S_T \not\rightarrow Y_{T+1})$ |
| M3 _I : | $p_{\{0,1\}};$ | $p_{\{0,0\}} =$ | $p_{\{1,0\}} =$ | $p_{\{1,1\}}$ | $\rightarrow (S_T \not\rightarrow Y_{T+1})$ |
| M3 _J : | $p_{\{1,0\}};$ | $p_{\{0,0\}} =$ | $p_{\{0,1\}} =$ | $p_{\{1,0\}}$ | $\rightarrow (S_T \not\rightarrow Y_{T+1})$ |
| М3к: | $p_{\{1,1\}};$ | $p_{\{0,0\}} =$ | $p_{\{0,1\}} =$ | $p_{\{1,1\}}$ | $\rightarrow (S_T \not\rightarrow Y_{T+1})$ |
| M3 _L : | $p_{\{0,0\}} =$ | $p_{\{0,1\}};$ | $p_{\{1,0\}} =$ | $p_{\{1,1\}}$ | $\rightarrow (S_T \rightarrow Y_{T+1})$ |
| М3м: | $p_{\{0,0\}} =$ | $p_{\{1,0\}};$ | $p_{\{0,1\}} =$ | $p_{\{1,1\}}$ | $\rightarrow (S_T \not\rightarrow Y_{T+1})$ |
| M3 _N : | $p_{\{0,0\}} =$ | $p_{\{1,1\}};$ | $p_{\{0,1\}} =$ | $p_{\{1,0\}}$ | $\rightarrow (S_T \rightarrow Y_{T+1})$ |
| M30: | $p_{\{0,0\}} =$ | $p_{\{0,1\}} =$ | $p_{\{1,0\}} =$ | $p_{\{1,1\}}$ | $\rightarrow (S_T \not\rightarrow Y_{T+1})$ |

Table 11 All possible Bernoulli models M3 and their prima facie causal claim

The Bernoulli model $M3_A$ presumes that every element in θ is unique and the other extreme Bernoulli model $M3_O$ presumes that all four elements in θ are equal. Table 11 shows that the existence of the prima facie causality in Equation 41 follows from the equality or the inequality of these Bernoulli parameters θ .

The prima facie causality in Equation 41 only exists in those cases where the Bernoulli parameter P_U that produced functionality Y_{T+1} depends on maintenance policy compliance S_T at both values of functionality Y_T . For example, the Bernoulli model M_{3C} has different Bernoulli parameters at $Y_T=1$ ($p_{\{0,1\}}\neq p_{\{1,1\}}$) and the same Bernoulli parameter at $Y_T=0$ ($p_{\{0,0\}}=p_{\{1,0\}}$). Therefore, M_{3C} refutes the prima facie causality in Equation 41 because the Bernoulli parameter P_U only depends on maintenance policy compliance S_T at $Y_T=1$ and not at $Y_T=0$. This is problematic for the third causality principle (Section 2.3.3), which asserts that maintenance policy compliance S_T comprises unique information about functionality Y_{T+1} that is not otherwise available.

The set of candidate models in Table 11 is *complete*. Therefore, the inferred relative Δ AIC scores that reveal the best approximating model do not merely hold with respect to some arbitrary set of candidate models. This makes the maintenance prognostic argument *decidable* in terms of AIC scores for the sample (s,y)_[1,1976] provided that we are willing to evaluate all candidate Bernoulli models. In this specific case, we proceed with an assessment of the relative Δ AIC scores of all candidate Bernoulli models in Table 11:

Let N_U be the observed frequency of functionality Y_{T+1} that come from the same Bernoulli process. In the case of M3_O where all information sets V={s_t,y_t,y_{t+1}} come from the same Bernoulli process, N_U is just the observed frequency of these information sets V={s_t,y_t,y_{t+1}}. In the other extreme case of M3_A where all information sets V={s_t,y_t,y_{t+1}} given U={s_t,y_t} come from the same Bernoulli process, N_U is a vector of four elements each representing the observed frequency of information sets V={s_t,y_t,y_{t+1}} given U={s_t,y_t}.

Let K_U be the observed frequency of $Y_{T+1}=1$ among the N_U observations. Let every instance of the N_U cases be generated by a Bernoulli process with parameter P_U ;

Then, the likelihood (=probability of observing the sample $(s,y)_{[1,t]}$, given some candidate Bernoulli model in Table 11), is given by:

$$\mathcal{L}(Bern(\theta)|(s,y)_{[1,1976]}) = \prod_{\forall p_u} \prod_{i=1}^{n_u} p_u^{y_{i+1}} \times (1-p_u)^{(1-y_{i+1})}$$
74

So, the log likelihood function is:

$$ln\left(\mathcal{L}(Bern(\theta)|(s,y)_{[1,1976]})\right) = \sum_{\forall p_u} k_u ln(p_u) + (n_u - k_u)ln(1 - p_u)$$
75

The maximum likelihood estimation of a Bernoulli parameter P_U in θ follows from the derivative of Equation 75:

$$\frac{\delta k_u ln(p_u) + (n_u - k_u)ln(1 - p_u)}{\delta p_u} = 0 \qquad \qquad \rightarrow p_{mle} = \frac{k_u}{n_u} \qquad \qquad 76$$

Provided that Equation 76 is continuous and concave in pmle:

$$\frac{\delta^2 (k_u ln(p_{mle}) + (n_u - k_u) ln(1 - p_{mle}))}{\delta p_{mle}^2} < 0 \qquad \qquad \rightarrow -\frac{n_u^2}{k_u} - \frac{k_u^2 (n_u - k_u)}{(k_u - n_u)^2} < 0 \qquad \qquad 77$$

The most likely parameters θ_{mle} enable an assessment of the AIC scores of the candidate Bernoulli models in Table 11:

$$AIC = 2 \times \left(-\left(\sum_{\forall p_{mle}} k_u ln(p_{mle}) + (n_u - k_u) ln(1 - p_{mle}) \right) + K \right)$$

$$78$$

The Bernoulli model M3₀ has a single estimable parameter $\theta = \{p\}$, i.e. K=1 and the Bernoulli model M3_A has four estimable parameters $\theta = \{p_{\{0,0\}}, p_{\{0,1\}}, p_{\{1,0\}}, p_{\{1,1\}}\}$, i.e. K=4. This implies that the higher dimensional model M3_A could only attain a preferred lower AIC score if it is sufficiently more likely than the model M3₀. If we were to alternatively select a Bernoulli model from Table 11 by a log likelihood ratio LR, the prima facie causality in Equation 41 would become nearly irrefutable because it would be entirely coincidental that the maximum likelihood estimations of the Bernoulli parameters θ in M3_A are exactly equal. Still, we will assess the log likelihood ratio LR with respect to the most parsimonious Bernoulli model M3₀:

$$LR = ln\left(\mathcal{L}(M3_i|(s,y)_{[1,1976]})\right) - ln\left(\mathcal{L}(M3_0|(s,y)_{[1,1976]})\right)$$
79

Table 12 now surveys the results for the N=1975 information sets V={ s_t,y_t,y_{t+1} } that we found in the sample (s,y)_[1,1976]:

- The maximum likelihood estimation of the Bernoulli parameters θ for each Bernoulli model in Table 11;
- The number of estimable parameters K for each Bernoulli model in Table 11;
- The relative \triangle AIC score that follows from Equation 63 and Equation 78 for each Bernoulli model in Table 11;
- The log likelihood LR of each Bernoulli model in Table 11 that follows from Equation 79.

The relative Δ AIC scores in Table 12 show the strongest empirical support for the Bernoulli model M3_F. The Bernoulli model M3_F implies that maintenance policy compliance S_T positively associates with functionality when the item is in downstate (Y_T=0) as p_{0,0}>p_{1,0}. However, the Bernoulli model M3_F also implies that maintenance policy compliance S_T is independent of functionality Y_{T+1} when the item is in upstate (Y_T=1) as p_{0,1}=p_{1,1}. For the prima facie causality in Equation 41, the positive association should be revealed at all values of Y_T to ensure that maintenance policy compliance S_T uniquely determined functionality Y_{T+1}. The Bernoulli model M3_F therefore existentially refutes the prima facie causality in Equation 41 in terms of AIC scores. On the other hand, the Bernoulli model M3_A that existentially confirms the prima facie causality in Equation 41 also receives substantial empirical support.

The candidate Bernoulli models $M3_F$ are just a subset of the candidate Bernoulli models $M3_A$, i.e. $M3_F$ is nested in $M3_A$. Therefore, the most likely Bernoulli model $M3_F$ cannot attain a higher likelihood than the most likely Bernoulli model $M3_A$.

| | MLE distributio | n of P4=M2(P3)+I | 26 | | | K | LR | ΔΑΙC |
|-----------------|-------------------------------------|-----------------------------------|---------------------------------|---------------------|-----------------------------------|--------------|--------------|--------------|
| M3 _A | $p_{\{0,0\}} = \frac{58}{277}$ | $p_{\{0,1\}} = \frac{1585}{1643}$ | $p_{\{1,0\}} = -\frac{1}{3}$ | $\frac{1}{10}$ p | $_{\{1,1\}} = \frac{24}{25}$ | 4 | 451 | 2 |
| M3 _B | $p_{\{0,0\}} =$ | $p_{\{0,1\}} = \frac{1643}{1920}$ | $p_{\{1,0\}} = -\frac{1}{3}$ | $\frac{1}{30}$ p | $_{\{1,1\}} = \frac{24}{25}$ | 3 | 52 | 798 |
| M3c | $p_{\{0,0\}} =$ | $p_{\{1,0\}} = \frac{59}{307}$ | $p_{\{0,1\}} = \frac{158}{164}$ | $\frac{15}{3}$ p | $_{\{1,1\}} = \frac{24}{25}$ | 3 | 448 | 7 |
| M3 _D | $p_{\{0,0\}} =$ | $p_{\{1,1\}} = \frac{82}{302}$ | $p_{\{0,1\}} = \frac{158}{164}$ | $\frac{15}{3}$ p | $\{1,0\} = \frac{1}{30}$ | 3 | 421 | 61 |
| M3 _E | $p_{\{0,1\}} =$ | $p_{\{1,0\}} = \frac{1586}{1673}$ | $p_{\{0,0\}} = \frac{58}{27}$ | $\frac{3}{7}$ p | $_{\{1,1\}} = \frac{24}{25}$ | 3 | 365 | 173 |
| M3 _F | $p_{\{0,1\}} =$ | $p_{\{1,1\}} = \frac{1609}{1668}$ | $p_{\{0,0\}} = \frac{58}{27}$ | $\frac{3}{7}$ p | $\{1,0\} = \frac{1}{30}$ | 3 | 451 | 0 |
| M3 _G | $p_{\{1,0\}} =$ | $p_{\{1,1\}} = -\frac{25}{55}$ | $p_{\{0,0\}} = \frac{58}{27}$ | $\frac{3}{7}$ p | $_{\{0,1\}} = \frac{1585}{1643}$ | 3 | 422 | 59 |
| МЗн | $p_{\{0,0\}} = \frac{58}{277}$ | $p_{\{0,1\}} =$ | $p_{\{1,0\}} =$ | р | $_{\{1,1\}} = \frac{1610}{1698}$ | 2 | 365 | 171 |
| M3ı | $p_{\{0,1\}} = \frac{1585}{1643}$ | $p_{\{0,0\}} =$ | $p_{\{1,0\}} =$ | p | $_{\{1,1\}} = \frac{83}{332}$ | 2 | 416 | 70 |
| M3 _J | $p_{\{1,0\}} = -\frac{1}{30}$ | $p_{\{0,0\}} =$ | $p_{\{0,1\}} =$ | р | $ _{\{1,1\}} = \frac{1667}{1945}$ | 2 | 51 | 799 |
| М3к | $p_{\{1,1\}} = -\frac{24}{25}$ | $p_{\{0,0\}} =$ | $p_{\{0,1\}} =$ | р | $ _{\{1,0\}} = \frac{1664}{1950}$ | 2 | 2 | 898 |
| $M3_{L}$ | $p_{\{0,0\}} =$ | $p_{\{0,1\}} = \frac{1643}{1920}$ | $p_{\{1,0\}} =$ | p | $ _{\{1,1\}} = \frac{25}{55}$ | 2 | 23 | 855 |
| М3м | $p_{\{0,0\}} =$ | $p_{\{1,0\}} = \frac{59}{307}$ | $p_{\{0,1\}} =$ | p | $ _{\{1,1\}} = \frac{1609}{1668}$ | 2 | 448 | 5 |
| $M3_{\rm N}$ | $p_{\{0,0\}} =$ | $p_{\{1,1\}} = \frac{82}{302}$ | $p_{\{0,1\}} =$ | p | $ _{\{1,0\}} = \frac{1586}{1673}$ | 2 | 335 | 231 |
| M30 | $p_{\{0,0\}} =$ | $p_{\{0,1\}} =$ | $p_{\{1,0\}} =$ | p | $ _{\{1,1\}} = \frac{1668}{1975}$ | 1 | 0 | 899 |
| | Level of | ΔΑΙC, ΔΑΙC | | | Contingen | cy table | | |
| | empirical support for a model | scores | | | ST=0 S YT=0 Y | Sт=1 /т=0 | ST=0 YT=1 | ST=1 YT=1 |
| | Substanti | al: 0-2 | | Y _{T+1} =0 | 219 | 58 | 29 | 1 |
| | Considerably les | ss: 4-7 | | Y _{T+1} =1 | 58 | 1585 | 1 | 24 |
| | Essentially nor | ie: >10 | | | | | | |

Table 12 Relative information loss of all Bernoulli models M3 for (s,y)[1,1976]

In Table 12, the most likely Bernoulli models $M3_A$ and $M3_F$ are indeed about equally likely (LR≈451) and the more parsimonious model $M3_F$ is only a better approximation of functionality Y_{T+1} because it has fewer parameters K in Equation 61. So, although the Bernoulli model $M3_A$ and $M3_F$ both carry substantial empirical support to approximate functionality Y_{T+1} , we ultimately conclude:

Maintenance policy compliance S_T *does* **not** *prima facie cause functionality* Y_{T+1} *with respect to the information set* $V = \{s_t, y_b, y_{t+1}\}$.

This finding is *not* easily overthrown since we considered all possible Bernoulli models for functionality Y_{T+1} that are possible within the information set $V = \{s_t, y_t, y_{t+1}\}$

resembling the prima facie causality in Equation 41. In this respect, the inference precision by the maintenance prognostic argument favours the dichotomous sample $(s,y)_{[1,1976]}$ rather than the cardinal samples $(d,\bar{y})_{[1,70]}$ and $(d,q)_{[1,1977]}$ that require knowledge beyond the information set V to choose a candidate model M3:Pr(P4|P3,M2,P6). On the other hand, a dichotomous operationalisation $(s,y)_{[1976]}$ is less precise than the cardinal operationalisations $(d,\bar{y})_{[1,70]}$ and $(d,q)_{[1,1977]}$. In this case study, we deemed that the inference precision suffers less from a dichotomous operationalisation than from the unavoidable model uncertainty of a cardinal operationalisation. In other cases, a "good" cardinal candidate model M3:Pr(P4|P3,M2,P6) may be preferred.

This section showed that the case organisation's convention to operationalise maintenance performance by the sample $(d,\bar{y})_{[1,70]}$ did not enable predictions of functionality \overline{Y}_{M+1} by the maintenance prognostic argument. By adopting either of the alternative samples $(s,y)_{[1,1976]}$ or $(d,q)_{[1,1977]}$ the case organisation would have been able to make these predictions. Although the relative ΔAIC and ΔAIC_c scores deduced for the sample $(d,q)_{[1,1977]}$ existentially confirmed the prima facie causality in Equation 41, we ultimately deemed this finding imprecise due to the model uncertainty and due to the existential refutation of the independence between D_T and Q_T that was required for a causal interpretation of this existentially confirmed prima facie causality in Equation 41.

Table 5 already suspected that the maintenance prognostic argument would be undecidable due to controversy about the model M2 and the error distribution P6. The samples $(d,\bar{y})_{[1,70]}$, $(d,q)_{[1,1977]}$ confirmed our preliminary dithering about the maintenance prognostic argument in Table 5. However, the sample $(s,y)_{[1,1976]}$ appeared to resolve the model uncertainty of the maintenance prognostic argument provided that we are willing to evaluate all candidate Bernoulli models. Within the delimitations of the prima facie causality in Equation 41, the maintenance prognostic argument appeared to be *decidable* by the sample $(s,y)_{[1,1976]}$ in terms of AIC scores. The maintenance policy validation by the maintenance prognostic argument and the sample $(s,y)_{[1,1976]}$ appeared to be the most precise, but it existentially refuted the prima facie causality in Equation 41.

5.3.3 Validation of the reliability engineering argument

Figure 14 showed that the reliability engineering argument includes all three causality principles (Section 2.3.3) only for $K_T \rightarrow K_{T+1}$. It does not presume a specific relation between maintenance policy compliance L_T and functionality K_{T+1} . Still, the reliability engineering argument would have been decidable about the prima facie causality in Equation 41 if it were sound. In Section 4.3.1, we already suspected being deprived of in-depth knowledge about a universal model M2 and the presumption P6 on the errors at some given maintenance policy compliance L_T in a typical case study. We therefore resort to some *existential* claim as we have already demonstrated for the maintenance prognostic argument.

For the case study, a reduction of the domain from $M2:f(l_t,k_t)$ to $M2:f(k_t)$ does not really resolve controversy. So, the reliability engineering argument seems similarly encumbered with the model selection problem that we already discussed for the

maintenance prognostic argument in Section 5.3.2. In this exposition, we confine ourselves again to the candidate linear regression models in Equation 64 for the cardinal samples $(d,\bar{y})_{[1,70]}$, $(d,q)_{[1,1977]}$ and to the candidate Bernoulli models in Table 11 for the dichotomous sample $(s,y)_{[1,1976]}$.

For the cardinal samples $(d,\bar{y})_{[1,70]}$, $(d,q)_{[1,1977]}$, we simply confine ourselves to two (fixed) maintenance policy compliance values that are (*i*) very different and that (*ii*) frequently occur in the samples as shown in Figure 27.

| | MLE distribution of P4 | =M2(P3)+P6 | K | LR | ΔΑΙC | ΔAICc |
|-----------------|---|--|------|----|------|-------|
| M30 | $\bar{y}_{m+1} = 0.83 + 0.15 \sim N($ | (0,1) | 2 | 0 | 0 | 0 |
| M31 | $\bar{y}_{m+1} = 0,71 + 0,14\bar{y}_m -$ | $v_{n+1} = 0.71 + 0.14\bar{y}_m + 0.15N(0.1)$ $v_{n+1}v_{n+2} = -2.97 + 4.00\bar{y}_m + 0.08N(0.1)$ | | | | 5 |
| M3 ₂ | $\overline{y}_{m+1 D_M=2} = -2,97 + \frac{1}{\overline{y}_{m+1 D_M=26}} = 0,94 - 0,5$ | 6 | 2,93 | 2 | 84 | |
| | Level of empirical | $\Delta AIC, \Delta AIC_{c}$ scores | | | | |
| | support for a model | | | | | |
| | Substantial: | 0-2 | | | | |
| | Considerably less: | 4-7 | | | | |
| | Essentially none: | >10 | | | | |
| | | | | | | |

 Table 13 Relative information loss of some models M3 for (d,y)[1,70]

Table 13 depicts the findings for the values $D_M=2$ and $D_M=26$ that both occur four times in the sample $(d,\bar{y})_{[1,70]}$. Clearly, the reliability engineering argument inefficiently uses the sample $(d,\bar{y})_{[1,70]}$ because it discards much of the potentially useful information in $(d,\bar{y})_{[1,70]}$. Table 7 already showed that the observed frequency of the replications in the sample $(d,\bar{y})_{[1,70]}$ ranges over [0,4] for the reliability engineering argument. Figure 27 already showed that 38 out of 41 distinguishable replications occur fewer than three times in the sample $(d,\bar{y})_{[1,70]}$. For those replications, it is impossible to estimate the three parameters in $\bar{y}_{m+1}=p_0+p_1\bar{y}_m+\sigma N(0,1)$. We therefore just involved the N=8 replications at $D_M=2$ and $D_M=26$ that most frequently occurred in the sample $(d,\bar{y})_{[1,70]}$ while at the same time being very different. Note that the maintenance policy validation by the maintenance prognostic argument and the sample $(d,\bar{y})_{[1,70]}$ in Table 8 relied on N=69 replications.

Table 13 reveals that the model M3₀ that considers functionality \overline{Y}_{M+1} as independent trials from a normal distribution achieves the lowest AIC and AIC_c scores. However, the corresponding AIC and AIC_c scores differ and should be distrusted. AIC is only asymptotically precise as the number of replications N tends to infinity whereas N=8 in this case. AIC_c is precise for a finite sample on the presumption that the true model is among the candidate linear regression models, which would seem to be a strong presumption in this case. Even if we were to follow these distrusted AIC and AIC_c scores in Table 13, the models M3₁ and M3₂ again tend to be overfitting. This confirms our suspicion that the case organisation's convention to maintenance performance does not allow for predictions by some linear model.

In Section 4.3.3, we already explained that the reliability engineering argument cannot existentially refute the prima facie causality in Equation 41 if maintenance policy

compliance has a cardinal scale. Therefore, the empirical support for models $M3_0$ and $M3_1$ in Table 13 is not existentially refuting the prima facie causality in Equation 41 and only empirical support for the model $M3_2$ would have existentially confirmed it. So, even if we just follow the distrusted AIC and AIC_c scores in Table 13, the existence of the prima facie causality in Equation 41 has been left undetermined.

In Section 5.3.2, we already deemed the maintenance policy validation by the maintenance prognostic argument and the sample $(d,\bar{y})_{[1,70]}$ as indecisive about the prima facie causality in Equation 41 due to the monthly sampling rate that did not adequately reconstruct the original signals as required in Section 3.2.2. We now showed that the maintenance policy validation by the reliability engineering argument and the sample $(d,\bar{y})_{[1,70]}$ is even less precise. We therefore conclude:

Maintenance policy compliance L_T may or may not prima facie cause functionality K_{T+1} with respect to the information set $V = \{l_t, k_t, k_{t+1}\}$.

Table 14 depicts the findings for the values $D_T=3$ and $D_T=24$ that occur 114 and 61 times respectively in the sample $(d,q)_{[1,1977]}$. Clearly, the reliability engineering argument also very inefficiently uses the sample $(d,q)_{[1,1977]}$. Table 7 already showed that the observed frequency of the replications in the sample $(d,q)_{[1,1977]}$ ranges over [0,114] for the reliability engineering argument. Figure 27 already showed that 53 out of 123 distinguishable replications occur fewer than three times in the sample $(d,q)_{[1,1977]}$. For those replications, it is impossible to estimate the three parameters in $q_{t+1}=p_0+p_1q_t+\sigma N(0,1)$. We therefore just involved the N=175 replications at $D_T=3$ and $D_T=24$ that frequently occur while at the same time being very different. Note that the maintenance policy validation by the maintenance prognostic argument and the sample $(d,q)_{[1,1977]}$ in Table 9 relied on N=1976 replications.

| | MLE distribution of P4 | =M2(P3)+P6 | K | LR | ΔΑΙC | ΔAICc |
|-----------------|--|---|---|-----|------|-------|
| M30 | $q_{t+1} = 19,73 + 6,16N(0$ |),1) | 2 | 0 | 774 | 774 |
| M3 1 | $q_{t+1} = 7,63 + 0,62q_t +$ | $t_{t+1} = 7,63 + 0,62q_t + 4,63N(0,1)$ | | | 80 | 80 |
| M3 ₂ | $q_{t+1 D_T=3} = 4,29 + 0,78q_t + 3,62N(0,1)$ $q_{t+1 D_T=24} = 10,83 + 0,48q_t + 5,87N(0,1)$ | | | 391 | 0 | 0 |
| | Level of empirical | $\Delta AIC, \Delta AIC_c$ scores | | | | |
| | support for a model | | | | | |
| | Substantial: | 0-2 | | | | |
| | Considerably less: | 4-7 | | | | |
| | Essentially none: | >10 | | | | |
| | | | | | | |

 Table 14 Relative information loss of some models M3 for (d,q)[1,1977]

The results in Table 14 are consistent with those in Table 9. Again the model $M3_2$ that existentially confirms the prima facie causality in Equation 41 receives the strongest empirical support. This existential confirmation may be overthrown by some unconsidered candidate model that outperforms the model $M3_2$ in expected information loss. However, under the acceptance of this model uncertainty, the maintenance policy validation by the reliability engineering argument and the sample $(d,q)_{[1,1977]}$ would existentially confirm the prima facie causality in Equation 41.

We therefore conclude:

If model uncertainty has been accepted, maintenance policy compliance D_T would prima facie cause functionality Q_{T+1} with respect to the information set $V = \{d_t, q_t, q_{t+1}\}$.

The non-causality assumptions of the reliability engineering argument do not differ from those of the maintenance prognostic argument. Still, the requirement of a constant maintenance policy compliance D_T prohibits the following causal explanations for the observed dependencies in the information set $V=\{d_t,q_t,q_{t+1}\}$: D_T \rightarrow Q_T, D_T \rightarrow Q_{T+1}, D_T \rightarrow (Q_T,Q_{T+1}). So, the stronger support for the model M3₁ as compared to the model M3₀ in Table 14 cannot be explained by the confounder D_T \rightarrow (Q_T,Q_{T+1}) or the mediator Q_T \rightarrow D_T \rightarrow Q_{T+1} which attributes to a causal interpretation of the inferred prima facie causality Q_T \rightarrow Q_{T+1} with respect to the information set $V=\{d_t,q_t,q_{t+1}\}$. Still, the path graph of the reliability engineering argument in Figure 11 requires independence between D_T and Q_T to allow for a causal interpretation of the inferred prima facie causality in Equation 41 as implied by the strongest empirical support for the model M3₂.

Table 15 compares the two candidate models that existentially confirm (M3₃) and existentially refute (M3₅) the independence between D_T and Q_T. We did not involve functionality Q_{T+1} because Q_{T+1} is known not to influence the relation between D_T and Q_T due to the first causality principle (Section 2.3.3) asserting that the future cannot cause the past. Table 15 shows no empirical support for the model M3₃ that existentially confirms the independence between D_T and Q_T as required by the path graph of the reliability engineering argument in Figure 11. This result allows for the existence of the alternative causal structures Q_T→(D_T,Q_{T+1}) and Q_T→D_T→Q_{T+1} that are problematic for a causal interpretation of the inferred prima facie causality in Equation 41 as implied by the best approximating probability function M3₂ in Table 14. We therefore ultimately deem that the maintenance policy validation by the reliability engineering argument and the sample (d,q)_[1,1977] is insufficient for the justification of maintenance that we pursue in this work.

| | MLE distribution of Q | ſ | | K | LR | ΔΑΙC | AAIC _c |
|-----------------|--|--|--|---|-------|------|--------------------------|
| M3 ₃ | $q_t = 19,46 + 6,52N(0,1)$ | | | 2 | 0 | 25 | 25 |
| M35 | $q_{t D_T=3} = 19,66 + 5,63$ $q_{t D_T=24} = 19,07 + 7,97$ | $q_{t D_T=3} = 19,66 + 5,63N(0,1)$ $q_{t D_T=24} = 19,07 + 7,97N(0,1)$ | | | 14,63 | 0 | 0 |
| | Level of empirical support for a model Substantial: | ΔAIC, ΔAIC _c scores | | | | | |
| | Considerably less: Essentially none: | 4-7 >10 | | | | | |

Table 15 Relative information loss of some models Pr(Q_T|D_T, θ) for (d,q)_[1,1977]

Finally, a maintenance policy validation by the reliability engineering argument and the sample $(s,y)_{[1,1976]}$ equals the maintenance policy validation by the maintenance prognostic argument and the sample $(s,y)_{[1,1976]}$ because the sample space of S_T only

contains two values, i.e. $\{0,1\}$. Therefore, no evidence from the sample $(s,y)_{[1,1976]}$ will be discarded.

This section showed that the reliability engineering argument causes concerns resembling the concerns that were raised for the maintenance prognostic argument. Moreover, the reliability engineering argument very inefficiently used the samples $(d,\bar{y})_{[1,70]}$ and $(d,q)_{[1,1977]}$ because many replications did not occur frequently enough to estimate the parameters of some candidate model M3. However, the maintenance policy validation by the reliability engineering argument and the sample $(s,y)_{[1,1976]}$ is exactly the same as was explained for the maintenance prognostic argument in Section 5.3.2.

Table 5 already suspected that the reliability engineering argument would be undecidable due to controversy about the model M2 and the error distribution P6. The samples $(d,\bar{y})_{[1,70]}$, $(d,q)_{[1,1977]}$ confirmed our preliminary dithering about the reliability engineering argument in Table 5. However, a maintenance policy validation by the reliability engineering argument and the sample $(s,y)_{[1,1976]}$ appeared to resolve the model uncertainty of the reliability engineering argument. Within the delimitations of the prima facie causality in Equation 41, the reliability engineering argument appeared to be *decidable* by the sample $(s,y)_{[1,1976]}$ in terms of AIC scores.

5.3.4 Validation of the nonparametric argument

The nonparametric argument in Figure 15 falsifies the second and the third causality principle if it were sound. Since universal independence does not follow from a stratified sample $(l,k)_{[1,t]}$, the nonparametric argument just infers an *existential* claim regarding the likelihood of the presumptions P5 and P7. In this section, we will only assess this likelihood for the dichotomous sample $(s,y)_{[1,1976]}$ because we already disqualified the cardinal samples $(d,\bar{y})_{[1,70]}$ and $(d,q)_{[1,1977]}$ for the maintenance policy validation by the nonparametric argument in Section 5.2.1.

The nonparametric argument may be instantiated by an exact conditional approach to compare two independent binomial proportions that reasons as follows:

Let N_U be the observed frequency of identical U={ l_t,k_t } in the sample (l,k)_[1,t-1]; Let N_U' be the observed frequency of identical U'={ l_t,k_t } in the sample (l,k)_[1,t-1]; Let us arbitrarily define a "win" as some subset of the sample space of K_{T+1} that is "of interest". The complementary set of a "win" is a "loss". So, we essentially build a dichotomous ("win" or "loss") variable on K_{T+1}. The dichotomous functionality Y_{T+1} as defined in Equation 45, could instantiate a "win" or "loss" definition.

Let K_U be the observed frequency of $V = \{U, "win"\}$ in the sample $(l,k)_{[1,t]}$;

Let K_U ' be the observed frequency of $V = \{U', "win"\}$ in the sample $(l,k)_{[1,t]}$;

Let every "win" among the $N_{\rm U}$ replications be generated by a Bernoulli process with parameter $P_{\rm U};$

Let every "win" among the N_U ' replications be generated by a Bernoulli process with parameter P_U ';

The latter two presumptions seem strong, but for a prima facie causality that is confined to an information set $V = \{l_t, k_t, k_{t+1}\}$, it would be odd to distinguish identical information

sets as different experiments. We therefore argue that these two presumptions do not enlarge the encumbrance of a presumption regarding the prima facie causality in Equation 41. In Section 4.4.3, we explained that the nonparametric argument takes any information set $V=\{l_t,k_t,k_{t+1}\}$ with identical $\{k_t\}$ as a replication for the prima facie causality in Equation 41. So, information sets V that comprise either U or U' that differ in $\{l_t\}$ but not in $\{k_t\}$, should be seen as replications.

Then, the joint distribution of K_U,K_U' is given by:

$$pr_{K_{U},K_{U'}}(k_{u},k_{u'}) = \binom{n_{u}}{k_{u}}\binom{n_{u'}}{k_{u'}}p_{u}^{k_{u}}(1-p_{u})^{(n_{u}-k_{u})}p_{u'}^{k_{u'}}(1-p_{u'})^{(n_{u'}-k_{u'})}$$
80

Because information sets V that comprise either U or U' have been posited as replications, maintenance policy compliance L_T must be independent of functionality K_{T+1} at a given K_T as implied by the presumptions P5 and P7 of the nonparametric argument. So, the parameters P_U and $P_{U'}$ of the Bernoulli processes are presumed to be equal and Equation 80 transforms to:

$$pr_{K_{U},K_{U'}}(k_{u},k_{u'}) = \binom{n_{u}}{k_{u}}\binom{n_{u'}}{k_{u'}}p^{(k_{u}+k_{u'})}(1-p)^{(n_{u}+n_{u'}-k_{u}-k_{u'})}$$
81

Since N_U and N_U' are disjoint counts, $N=N_U+N_U'$ and $K=K_U+K_U'$ holds for the union. Then, the distribution of the joint number of "wins" K is given by:

$$pr_{K}(k) = \binom{n}{k} p^{k} (1-p)^{(n-k)} = \binom{n}{k} p^{(k_{u}+k_{u'})} (1-p)^{(n_{u}+n_{u'}-k_{u}-k_{u'})}$$
82

The observed frequencies K_U, K_U', N_U, N_U' are straightforwardly assessable from the sample $(l,k)_{[1,t]}$. Still, the distributions in Equation 81 and Equation 82 require an estimation of the Bernoulli parameter P. However, the joint distribution of K_U, K_U' conditioned on the distribution of the joint number of "wins" K does *not* require an estimation of the Bernoulli parameter P:

$$\frac{pr_{K_{u},K_{u'}}(k_{u},k_{u'})}{pr_{K}(k)} = \frac{\binom{n_{u}}{k_{u}}\binom{n_{u'}}{k_{u'}}p^{(k_{u}+k_{u'})}(1-p)^{(n_{u}+n_{u'}-k_{u}-k_{u'})}}{\binom{n}{k}p^{(k_{u}+k_{u'})}(1-p)^{(n_{u}+n_{u'}-k_{u}-k_{u'})}} = \frac{\binom{n_{u}}{k_{u}}\binom{n_{u'}}{k_{u'}}}{\binom{n}{k}}$$
83

The hypergeometric distribution in Equation 83 is a direct consequence of the observed frequencies K_U, K_U', N_U, N_U' and the presumed equality of the Bernoulli parameters: $P_U=P_U$. Since the observed frequencies K_U, K_U', N_U, N_U' follow from common sense, Equation 83 expresses the likelihood of the controversial presumption $P_U=P_{U'}$ without an estimation of the Bernoulli parameter $P=P_U=P_U'$.

The statistical significance of this independence follows from the p-value. The p-value generally indicates the probability to observe frequencies that are at least as extreme as the ones actually observed. Loosely phrased, an extreme frequency c on the presumption of $P_U=P_{U'}$ is any observed frequency K_U , $K_{U'}$ among the N_U , $N_{U'}$ cases that enlarges the difference $|(K_U/N_U)-(K/N)|$. More precisely, the p-value follows from:

$$p \text{ value} = \begin{cases} \sum_{\forall c \ge k_{u'}} \frac{\binom{n_{u'}}{c} \binom{n_u}{k-c}}{\binom{n}{k}} & \text{ if } \left(\frac{k_{u'}}{n_{u'}}\right) \ge \left(\frac{k-1}{n}\right) \\ \sum_{\forall c \ge k_u} \frac{\binom{n_u}{c} \binom{n_{u'}}{k-c}}{\binom{n}{k}} & \text{ if } \left(\frac{k_u}{n_u}\right) \ge \left(\frac{k-1}{n}\right) \end{cases} \end{cases}$$

If the p-value in Equation 84 is *above* some significance level α , the observed frequency K_U , $K_{U'}$ among the N_U , $N_{U'}$ cases was apparently not that extreme since more extreme values are likely. The presumption of $P_U=P_{U'}$ that implied Equation 83 has then *existentially* been *confirmed*. Conversely, if the p-value in Equation 84 is *below* some significance level α , the observed frequency K_U , $K_{U'}$ among the N_U , $N_{U'}$ cases was apparently extreme since more extreme values are unlikely. The presumption of $P_U=P_{U'}$ that implied Equation 83 has then *existentially* been *refuted*. Conventionally, some test statistic (like the Wald-statistic) that requires an estimation of the Bernoulli parameter $P=P_U=P_{U'}$ identifies the extreme values c in Equation 84. To show that these extreme values c also directly ensue from the hypergeometric distribution in Equation 83, we present an alternative proof:

84

Proof

This proof essentially asserts under what conditions a number of "wins" ($c=k_u+1$) in the hypergeometric distribution in Equation 83 yields a likelihood that is at least as low as the observed frequency K_U , K_U among the N_U , N_U cases.

$$\frac{\binom{n_u}{k_u+1}\binom{n_{u'}}{k_{u'}-1}}{\binom{n}{k}} \le \frac{\binom{n_u}{k_u}\binom{n_{u'}}{k_{u'}}}{\binom{n}{k}}$$
85

We ignore the constant denominator which reduces Equation 85 to:

$$\begin{pmatrix} \frac{n_{u}!}{(k_{u}+1)! (n_{u}-k_{u}-1)!} \end{pmatrix} \begin{pmatrix} \frac{n_{u'}!}{(k_{u'}-1)! (n_{u'}-k_{u'}+1)!} \end{pmatrix} \\ \leq \begin{pmatrix} \frac{n_{u}!}{k_{u}! (n_{u}-k_{u})!} \end{pmatrix} \begin{pmatrix} \frac{n_{u'}!}{k_{u'}! (n_{u'}-k_{u'})!} \end{pmatrix} \\ \begin{pmatrix} \frac{n_{u}-k_{u}}{k_{u}+1} \end{pmatrix} \leq \begin{pmatrix} \frac{n_{u'}-k_{u'}+1}{k_{u'}} \end{pmatrix}$$

$$\begin{pmatrix} \frac{n_{u}-k_{u}}{k_{u}+1} \end{pmatrix} \leq \begin{pmatrix} \frac{(n-n_{u})-(k-k_{u})+1}{k-k_{u}} \end{pmatrix}$$

$$86$$

Equation 86 is true when the denominator satisfies:

$$\left((k_u+1) \ge (k-k_u)\right) \equiv \left(k_u \ge \left(\frac{k-1}{2}\right)\right)$$
87

And the numerator satisfies:

$$((n_u - k_u) \le (n - n_u) - (k - k_u) + 1)$$

$$2n_u \le n - k + 2k_u + 1$$

$$88$$

To identify the minimum value of N_U in Equation 88, K_U should take the minimal value of zero. However, K_U has been bound by Equation 87.

$$\left(2n_u \le n - k + 2\left(\frac{k-1}{2}\right) + 1\right) \equiv \left(n_u \le \left(\frac{n}{2}\right)\right)$$

$$89$$

Equation 87 and Equation 89 then specify that the proportion K_U/N_U should satisfy:

$$\left(\frac{k_u}{n_u}\right) \ge \left(\frac{k-1}{n}\right) \tag{90}$$

Equation 90 states the criterion for which values of K_U the value (c=k_u+1) will be identified as "more unlikely" as indicated in Equation 84. The extreme values of K_U ' are similarly identifiable.

Only the dichotomous sample $(s,y)_{[1,1976]}$ potentially suffices for the maintenance policy validation by the nonparametric argument according to Section 5.2.3. Still, the nonparametric argument may only existentially confirm or refute the prima facie causality in Equation 41. We now repeat the earlier presented inference by the nonparametric argument for the sample $(s,y)_{[1,1976]}$.

Let N_U be the observed frequency of $U=\{s_t, y_t\}=\{0,1\}$ in the sample $(s,y)_{[1,t-1]}$, i.e. $N_U=1643$;

Let N_U ' be the observed frequency of U'={st', yt}={1,1} in the sample (s,y)_[1,t-1], i.e. $N_{U'}=25$;

Let a "win" for the sample $(s,y)_{[1,1976]}$ follow from the case organisation's common sense in Equation 45:

91

$$Y_{T+1} = \begin{cases} 0; downstate, "loss" \\ 1; upstate, "win" \end{cases}$$

Let K_U be the observed frequency of V={U,"win"} in the sample $(s,y)_{[1,1976]}$, i.e. K_U =1585;

Let K_U ' be the observed frequency of $V = \{U', "win"\}$ in the sample $(s,y)_{[1,1976]}$, i.e. $K_U = 24$;

Let $K=K_U+K_U=1609$ and $N=N_U+N_U=1668$ as shown in Equation 82.

Let presumption P5, P7 be true. Note that P5 is uncontroversial.

Then, the p-value is:

$$p \ value = \sum_{\forall c \ge k_u} \frac{\binom{n_u}{c}\binom{n_{u'}}{k-c}}{\binom{n}{k}} \qquad if \ \binom{k_u}{n_u} \ge \binom{k-1}{n}$$

$$p \ value = \sum_{\forall c \ge 1585} \frac{\binom{1643}{c}\binom{25}{1609-c}}{\binom{1668}{1609}} \qquad if \ \binom{1585}{1643} \ge \binom{1609-1}{1668}$$

$$p \ value = \sum_{c=1585}^{1609} \frac{\binom{1643}{c}\binom{25}{1609-c}}{\binom{1668}{1609}} = 0,5963$$

The p-value expresses the probability of observing proportions that are at least as extreme as the observed frequencies K_U , $K_{U'}$ among the N_U , $N_{U'}$ cases, given the truth of the presumptions P5 and P7. High p-values that are above some arbitrary significance level existentially confirm the presumptions P5 and P7. Conversely, low p-values that are below this arbitrary significance level existentially refute the presumptions P5 and P7. So, Equation 92 existentially confirms the presumptions P5 and P7 at an arbitrary significance level α =0,01. The presumptions P5 and P7 only imply independence between maintenance policy compliance S_T and functionality Y_{T+1} . In Equation 44, we already explained how this independence relates to the prima facie causality in Equation 41. Applied to this case study, Equation 44 transforms to:

$$\left(\mathcal{L}\left(\left(pr_{Y_{T+1}|S_{T},Y_{T}}(1|0,y_{t})=pr_{Y_{T+1}|,Y_{T}}(1|y_{t})\right)|(s,y)_{[1,1976]}\right)<\rho;\exists s\forall t\forall y\right)\cong(S_{T}\to Y_{T+1})$$
93

Equation 93 shows that an existential confirmation of the prima facie causality in Equation 41 follows from a low likelihood at some s and all y,t. Although Equation 92 already provided a counterexample of likely independence at $Y_T=1$, we still list the results for both $Y_T=0$ and $Y_T=1$ in Table 16.

If both p-values in Table 16 were below an arbitrary significance level α =0,01 indicating that the independence between maintenance policy compliance S_T and functionality Y_{T+1} would be unlikely for all y_t, we would have existentially refuted the prima facie causality between maintenance policy compliance S_T and functionality Y_{T+1} with respect to the information set V={s_t,y_t,y_{t+1}}. Since none of the p-values in Table 16 is below α =0,01, we conclude:

Maintenance policy compliance S_T does **not** prima facie cause functionality Y_{T+1} with respect to the information set $V = (s_t, y_t, y_{t+1})$.

The sample $(s,y)_{[1,1976]}$ comprised all elements of the sample space of the information set V. Still, the observed frequencies N_U of identical U={s_t,y_t} in the sample $(s,y)_{[1,1976]}$ ranged from 25 to 1643 (Table 16) which might have contributed to the existential refutation of the prima facie causality in Equation 41 at an arbitrary significance level α =0,01. These significance levels allow for comparisons of different samples like $(d,\overline{y})_{[1,70]}$, $(d,q)_{[1,1977]}$ *in principle* whereas the relative Δ AIC or Δ AIC_c scores do not allow for comparisons between the different samples like $(d,\overline{y})_{[1,70]}$, $(d,q)_{[1,1977]}$ and $(s,y)_{[1,1976]}$.

The path graph of the nonparametric argument in Figure 11 showed that an eventual dependence between S_T and Y_T would not have been problematic for a causal interpretation of an existential confirmation of the prima facie causality in Equation 41. Since Y_T has been held constant, Y_T cannot act as a confounder $Y_T \rightarrow (S_T, Y_{T+1})$ or as a mediator $S_T \rightarrow Y_T \rightarrow Y_{T+1}$ to alternatively explain the dependence between S_T and Y_{T+1} that we did *not* actually observe in this case study (Table 16).

The preliminary assessment of inference precision in Table 5 already indicated that the nonparametric argument would be decidable in terms of likelihood. However, the nonparametric argument also imposed severe constraints on the composition of the sample.

| Number of identical U={st,yt} | Number of identical V={U, upstate} upstate: Y _{T+1} =1 | Observed proportion | p-value |
|--|---|--|---------|
| $\begin{array}{llllllllllllllllllllllllllllllllllll$ | $\begin{array}{llllllllllllllllllllllllllllllllllll$ | K _U /N _U =0,209 K _U '/N _U '=0,033 (K-1)/N =0,189 | 0,0105 |
| $\begin{array}{ll} N_{U}\!\!=\!\!N_{\{0,1\}} &=\!\!1643 \\ N_{U}\!\!=\!\!N_{\{1,1\}} &=\!\!25 \\ N\!\!=\!\!N_{U}\!\!+\!\!N_{U}\!\!=\!\!1668 \end{array}$ | $\begin{array}{lll} K_{U}\!\!=\!\!K_{\{0,1\}} &=\!\!1585\\ K_{U}'\!\!=\!\!K_{\{1,1\}} &=\!\!24\\ K\!\!=\!\!K_{U}\!\!+\!K_{U}'\!\!=\!\!1609 \end{array}$ | $\begin{array}{ll} K_{U}/N_{U} &=\!\!0.965 \\ K_{U}{}^{\prime}/N_{U}{}^{\prime}\!=\!\!0.960 \\ (K\!\!-\!\!1)/N =\!\!0.964 \end{array}$ | 0,5963 |

Table 16 P-values for the sample (s,y)[1,1976], given P5, P7 of the NPA

The case study confirmed that the sample $(s,y)_{[1,1976]}$ complied with these constraints but the samples $(d,\bar{y})_{[1,70]}$ and $(d,q)_{[1,1977]}$ appeared to be inadmissible evidence.

5.3.5 Findings regarding the validation of the arguments

In Section 5.3, we reduced the maintenance policy validation to the assessment of the prima facie causality in Equation 41. We confronted candidate arguments with candidate samples to confirm the preliminary assessment of the inference precision in Table 5 by a realistic case study. We now summarise the main findings in Table 17:

- The "decidable argument" inference objective for the candidate arguments and samples because the candidate arguments mainly differed in the "decidable argument" inference objective in Table 5.
- The sufficiency of the sampling efficiency for the candidate arguments and samples to assess whether the sampling issues from Section 4.5.5 apply to this specific case study.
- The claim of the candidate arguments and samples about the prima facie causality in Equation 41 to confirm the preliminary findings in Table 5 regarding the aim of this work.

We will now discuss the columns in Table 17. Note that Table 17 already includes three improved arguments (final three rows), that will be introduced in Section 5.4.

The "decidable argument" inference objective

The maintenance optimisation argument only comprised one controversially presumed model M1. The maintenance optimisation argument is therefore decidable, but it has been refuted by all samples.

For the samples $(d,\bar{y})_{[1,70]}$, $(d,q)_{[1,1977]}$, we lacked the in-depth knowledge that would have settled the controversy about the model M2 and the error distribution P6 of the maintenance prognostic argument and the reliability engineering argument. We naively resorted to a model selection from some linear regression model families that did not yield any acceptable approximating model. Only for the sample $(s,y)_{[1,1976]}$, we have been able to resolve model uncertainty within the information set V={ s_t, y_t, y_{t+1} } of the prima facie causality in Equation 41. Therefore, only for the sample $(s,y)_{[1,1976]}$, the maintenance prognostic argument and the reliability engineering argument appeared to be decidable.

| | Decidable argument | Sufficient sampling efficiency | Decidable about $L_T \rightarrow K_{T+1}$ w.r.t $\{l_{t,k_t,k_{t+1}}\}$ | |
|--|---|---|---|--|
| maintenance optimisation argument (Section 4.1) | Yes, but M1 is refuted | $(d,\bar{y})_{[1,70]}$: Yes $(d,q)_{[1,1977]}$: Yes $(s,y)_{[1,1976]}$: Yes | No | |
| maintenance prognostic argument (Section 4.2) | Only in AIC for (s,y) _[1,1976] . | $(d,\bar{y})_{[1,70]}$: Yes $(d,q)_{[1,1977]}$: Yes $(s,y)_{[1,1976]}$: Yes | Only in AIC for (s,y) _[1,1976] . | |
| reliability engineering argument (Section 4.3) | Only in AIC for (s,y) _[1,1976] . | $(\mathbf{d}, \overline{\mathbf{y}})_{[1,70]}$: No $(\mathbf{d}, \mathbf{q})_{[1,1977]}$: Yes $(\mathbf{s}, \mathbf{y})_{[1,1976]}$: Yes | Only in AIC for (s,y) _[1,1976] . | |
| nonparametric argument (Section 4.4) | Only in likelihood | $(d, \overline{y})_{[1,70]}$: No $(d,q)_{[1,1977]}$: No $(s,y)_{[1,1976]}$: Yes | Only in likelihood for (s,y) _[1,1976] . | |
| weakened maint. prognostic argument (Section 5.4.1) | Only in AIC for (s,y) _[1,1976] . | $(\mathbf{d}, \overline{\mathbf{y}})_{[1,70]}$: No $(\mathbf{d}, \mathbf{q})_{[1,1977]}$: No $(\mathbf{s}, \mathbf{y})_{[1,1976]}$: Yes | Only in AIC for (s,y) _[1,1976] . | |
| extended nonparametric argument (Section 5.4.2) | Only in likelihood | $(\mathbf{d}, \overline{\mathbf{y}})_{[1,70]}$: No $(\mathbf{d}, \mathbf{q})_{[1,1977]}$: No $(\mathbf{s}, \mathbf{y})_{[1,1976]}$: Yes | Only in likelihood for (s,y) _[1,1976] . | |
| reduced nonparametric argument (Section 5.4.3) | Only in likelihood | $(d,\bar{y})_{[1,70]}$: Yes $(d,q)_{[1,1977]}$: Yes $(s,y)_{[1,1976]}$: Yes | Only in likelihood for (d,q) _[1,1977] ,(s,y) _[1,1976] . | |

The nonparametric argument only comprised one controversial presumption P7. Therefore, the nonparametric argument appeared to be decidable in terms of likelihood.

Table 17 Main findings regarding the candidate arguments and samples

Sufficient sampling efficiency

The maintenance optimisation argument did not constrain the composition of the sample. All candidate samples therefore appeared to be admissible.

The maintenance prognostic argument did not constrain the composition of the sample. All candidate samples therefore appeared to be admissible. However, the sample $(d,\bar{y})_{[1,70]}$ did not reconstruct the signals of maintenance policy compliance and functionality well enough to be compelling for the prima facie causality in Equation 41.

We disqualified the sample $(d, \overline{y})_{[1,70]}$ for the reliability engineering argument due to a lack of observed frequencies of the replications.

We disqualified the samples $(d, \overline{y})_{[1,70]}$ and $(d,q)_{[1,1977]}$ for the nonparametric argument due to a lack of observed frequencies of the replications.

In conclusion, only the sample $(s,y)_{[1,1976]}$ appeared to be admissible to all candidate arguments. The samples $(d,q)_{[1,1977]}$ and $(s,y)_{[1,1976]}$ were taken at a higher sampling rate which provided a richer view on the interactions between the elements in the information set V.

Decidable about $L_T \rightarrow K_{T+1}$ with respect to $\{l_t, k_t, k_{t+1}\}$

Since not all candidate arguments claim the same about prima facie causality, we also survey whether these arguments could decide about the prima facie causality in Equation 41.

The maintenance optimisation argument is not a causal argument that could only existentially refute the prima facie causality in Equation 41 upon its existential confirmation. However, the maintenance optimisation argument has universally been refuted by all candidate samples. Still, we found it important to show that a definitional equivalence between maintenance policy compliance L and functionality K conflicted with the case organisation's common sense about L and K.

For the sample $(d,\bar{y})_{[1,70]}$, the maintenance prognostic argument appeared to be undecidable about the prima facie causality in Equation 41 due to a too low sampling rate that insufficiently reconstructed the evolution of maintenance policy compliance and functionality. For the sample $(d,q)_{[1,1977]}$ and an arbitrary set of candidate models, the maintenance prognostic argument existentially confirmed the prima facie causality in Equation 41 but a causal interpretation of this result has been problematic (Table 10). For the sample $(s,y)_{[1,1976]}$, the maintenance prognostic argument appeared to existentially refute the prima facie causality in Equation 41 in terms of expected information loss (AIC).

The reliability engineering argument appeared to be somewhat superfluous as it reconciled with the concerns of the maintenance prognostic argument at a lower sampling efficiency.

For the sample $(s,y)_{[1,1976]}$, the nonparametric argument existentially refuted the prima facie causality in Equation 41 in terms of likelihood at some arbitrary significance level α =0,01 (Table 16).

In conclusion, we deem a maintenance policy validation by the nonparametric argument and the sample $(s,y)_{[1,1976]}$ as most suitable because of (*i*) the computational efficiency, i.e. not all candidate models should be assessed as in the maintenance prognostic argument, (*ii*) the potential to compare different samples at a reduced information set $V=\{l_t,k_{t+1}\}$, (*iii*) the ability to quantify its ambiguity about the prima facie causality in Equation 41 by a likelihood whereas the AIC and AIC_c scores in Section 5.3.2 only compared models on expected information loss when approximating functionality K_{T+1} and (*iv*) the ability to both existentially confirm and existentially refute the prima facie causality in Equation 41 (*v*) its more compelling causal interpretation of an existential confirmation of the prima facie causality in Equation 41 whereas the maintenance prognostic argument and the reliability engineering argument would require an additional test for independence between L_T and K_T as we did in Table 10 and Table 15.

Table 17 also supports the iterative approach to inference precision that we introduced in Section 1.4. From the candidate arguments, the maintenance optimisation argument claimed the *most precise* relation between maintenance policy compliance and functionality. The case organisation's maintenance scorecard *more completely operationalised* the goals of the maintenance policy than any bivariate sample $(1,k)_{[1,t]}$. From the candidate samples, the cardinal sample $(d,q)_{[1,1977]}$ *most precisely reconstructed* the original signals. Still, the most precise inference did not follow from combining the most precise argument, operationalisation and sampling procedure.

5.4 An improved inference

In this section, we will pursue an improved inference precision of the maintenance policy validation by (*i*) weakening the non-causality assumptions of the maintenance prognostic argument in Section 5.4.1 to abandon the need to test for the independence between L_T and K_T upon an existential confirmation of the prima facie causality in Equation 41, by (*ii*) extending the information set V of the nonparametric argument in Section 5.4.2 to seek for long-term functionality responses and by (*iii*) delimiting the information set to $V=\{l_t, k_{t+1}\}$ of the nonparametric argument in Section 5.4.3 to make all candidate samples admissible to the maintenance prognostic argument.

5.4.1 Weakened maintenance prognostic argument

This section will present an alternative path graph that weakens the non-causality assumptions of the maintenance prognostic argument in Figure 11.

The maintenance prognostic argument in Figure 11 required independence between L_T and K_T to exclude K_T as a mediator $L_T \rightarrow K_T \rightarrow K_{T+1}$ or as a confounder $K_T \rightarrow (L_T, K_{T+1})$ that would reduce the prima facie causality in Equation 41 to a spurious cause. The path graph of the weakened maintenance prognostic argument excludes this spurious cause by taking functionality K_T as a constant while still allowing for any relation between L_T and K_T (Figure 30). So, the path graph in Figure 30 is weakened since it does not require the independence between L_T and K_T as a condition for a causal interpretation of an existentially confirmed prima facie causality in Equation 41. The propositions of the weakened maintenance prognostic argument and the propositions of the reliability engineering argument are almost identical; only K_T and L_T should be exchanged in Figure 14. The sampling issues of the weakened maintenance prognostic argument and the nonparametric argument are similar since both arguments similarly define a replication by the information set $V=\{l_t, k_t, k_{t+1}\}$ with identical $\{k_t\}$. Then, it follows from Table 17 that only the dichotomous sample $(s, y)_{[1,1976]}$ is admissible to the weakened maintenance prognostic argument.



Figure 30 Path graph of the weakened maintenance prognostic argument

We now proceed with a maintenance policy validation by the weakened maintenance prognostic argument and the sample $(s,y)_{[1,1976]}$ as introduced in Section 5.3.2.

Let a replication be defined as any information set $V = \{s_t, y_t, y_{t+1}\}$ with identical $\{y_t\}$ in the sample $(s, y)_{[1,1976]}$.

Let the functionality Y_T be known.

Then, Table 18 and Table 19 present an exhaustive set of Bernoulli models to be considered. In the absence of in-depth knowledge about the true probability, we resort to an assessment of the AIC and AIC_c scores of these Bernoulli models as explained in Section 5.3.2. By presuming the functionality Y_T as a known constant, the number of candidate Bernoulli models M3 drops from 15 (Table 11) to 4 (Table 18 and Table 19 together). So, the presumption of a known functionality Y_T , reduces the analysis burden to deduce the AIC and the AIC_c scores for every possible Bernoulli model M3.

| | AAIC |
|--------------------------------|--|
| 4 | 0 |
| 0 | 5 |
| ncy table Y _T =0 | ST=1 YT=0 |
| 219 | 29 |
| 58 | 1 |
| | |
| r | 4 0 ncy table Yr=0 219 58 |

Table 18 Relative information loss of all Bernoulli models M3|Y_T=0

Table 18 shows considerably less empirical support for the Bernoulli model $M3_Q$ that existentially refutes the prima facie causality in Equation 41, given that the item is currently down $Y_T=0$. We therefore conclude:

If the item is known to be down, maintenance policy compliance S_T would prima facie cause functionality Y_{T+1} with respect to the information set $V = \{s_t, y_t, y_{t+1}\}$.

However, Table 19 shows substantial empirical support for both the Bernoulli model $M3_R$ and $M3_S$ that contradict each other regarding the prima facie causality in Equation 41, given that the item is currently up $Y_T=1$. The candidate Bernoulli model $M3_S$ is just a subset of the candidate Bernoulli model $M3_R$, i.e. $M3_S$ is nested in $M3_R$. Therefore, the most likely Bernoulli model $M3_S$ cannot attain a higher likelihood than the most likely Bernoulli model $M3_R$. In Table 19, the most likely Bernoulli model $M3_R$ is indeed slightly more likely than the Bernoulli model $M3_S$ (LR=2). Still, the more parsimonious model $M3_S$ is a better approximation of functionality Y_{T+1} because it has less parameters K in Equation 61. So, although the Bernoulli model $M3_R$ and $M3_S$ both carry substantial empirical support to approximate functionality Y_{T+1} , we ultimately conclude:

If the item is known to be up, maintenance policy compliance S_T would **not** prima facie cause functionality Y_{T+1} with respect to the information set $V = \{s_t, y_t, y_{t+1}\}$.

| | MLE distribution of | | K | LR | ΔΑΙΟ | |
|-----------------|---|-----------------------------------|---------------------|-----------------------------|-----------------------------------|-------------------------------------|
| M3 _R | $p_{\{0,1\}} = \frac{1585}{1643}$ | $p_{\{1,1\}} = \frac{24}{25}$ | | 2 | 2 | 2 |
| M3s | $p_{\{0,1\}} =$ | $p_{\{1,1\}} = \frac{1609}{1668}$ | | 1 | 0 | 0 |
| | Level of empirical support for a model | ΔAIC, ΔAICc scores | | Conting S _T = | ency table 0 Y _T =1 | S _T =1 Y _T =1 |
| | Substantial: | 0-2 | Y _{T+1} =0 | | 58 | 1 |
| | Considerably less: | 4-7 | Y _{T+1} =1 | | 1585 | 24 |
| | Essentially none: | >10 | | | | |

Table 19 Relative information loss of all Bernoulli models M3|Y_T=1

The most likely parameters of the Bernoulli model $M3_P$ suggest that maintenance policy compliance $S_T=0$ contributes to the item's functionality $Y_{T+1}=1$, given that the item is down $Y_T=0$:

$$p_{mle|U=\{s_t, y_t\}=\{0,0\}} = \frac{58}{277} \approx 0,21$$

$$p_{mle|U=\{s_t, y_t\}=\{1,0\}} = \frac{1}{30} \approx 0,03$$
94

The most likely parameters of the Bernoulli model M3_R suggest independence between

maintenance policy compliance $S_T=0$ and the item's functionality $Y_{T+1}=1$, given that the item is up $Y_T=1$:

$$p_{mle|U=\{s_t, y_t\}=\{0,1\}} = \frac{1585}{1643} \approx 0,96$$

$$p_{mle|U=\{s_t, y_t\}=\{1,1\}} = \frac{24}{25} = 0,96$$
95

We may wonder whether it is realistic to presume that a decision maker who controls maintenance policy compliance S_T knows the current functionality Y_T . In Section 3.3.4 for example, we mentioned that corrective maintenance has been triggered by faults that are currently known. So, it may well be that decision makers know the current functionality Y_T in practice. Resembling the event tree in Figure 7, the number of *possible* values of the functionality Y_T then reduces to either up $Y_T=1$ or down $Y_T=0$. Then, the weakened maintenance prognostic argument would better approximate reality than the maintenance prognostic argument while reducing the analysis burden.

This section showed that the weakened maintenance prognostic argument more precisely validates the maintenance policy than the maintenance prognostic argument in this specific case study. At least the existence of the prima facie causality in Equation 41 has existentially been confirmed when the item is known to be down $Y_T=0$. The causal interpretation of this existential confirmation may, unlike the maintenance policy validation by the maintenance prognostic argument, only be overthrown by some unobserved background variable B.

5.4.2 Extended nonparametric argument

This section will present an alternative path graph that may reveal long-term functionality responses to maintenance policy compliance as the maintenance policy validation by the nonparametric argument and the sample $(s,y)_{[1,1976]}$ in Section 5.3.4 only confined to a one-day-ahead prediction of functionality. Eventually, a causal effect of a maintenance policy is not revealed within a day.

Figure 31 specifies the non-causality assumptions of the extended nonparametric argument that are equivalent to those of the nonparametric argument (NPA) in Figure 11. The propositions of the extended nonparametric argument and the propositions of the nonparametric argument are almost identical; only K_{T+1} should be replaced by $K_{[T+1,T+X]}$ in Figure 15. The sampling issues are similar because the extended nonparametric argument similarly defines a replication by the information set $V=\{l_t,k_t,k_{[t+1,t+x]}\}$ with identical $\{k_t\}$. Then, it follows from Table 17 that only the dichotomous sample $(s,y)_{[1,1976]}$ is admissible to the extended nonparametric argument. We now proceed with a maintenance policy validation by the extended nonparametric argument and the sample $(s,y)_{[1,1976]}$ as introduced in Section 5.3.4.

Even for the dichotomous sample $(s,y)_{[1,1976]}$, the number of possible trajectories of $Y_{[t+1,t+x]}$ rapidly explodes as x increases.



Figure 31 Path graph of the extended nonparametric argument

Reconciling with the prognostic convention to operationalise prospective functionality by a remaining useful life (Section 3.5), we therefore similarly confine ourselves to the following two trajectories of $Y_{[t+1,t+x]}$ that appear to be most important to a decision maker:

$$Y_{[T+1,T+x]} = \begin{cases} 0, & \bigcap_{i=1}^{x} Y_{T+i} = 0 & ; prospective downtime \\ 1, & \bigcap_{i=1}^{x} Y_{T+i} = 1 & ; prospective uptime \end{cases}$$
96

Equation 96 reflects the prospective uptime and the prospective downtime over an interval [t+1,t+x] which is quantifiable by the dichotomous variable $Y_{[T+1,T+x]}$. Then, maintenance policy compliance S_T prima facie causes functionality $Y_{[T+1,T+x]}$ with respect to the information set $V=\{s_t, y_t, y_{[t+1,t+x]}\}$ by:

$$\left(\mathcal{L}\left(pr_{Y_{[.]}|S_{T},Y_{T}}\left(y_{[.]}|0,y_{t}\right) = pr_{Y_{[.]}|S_{T},Y_{T}}\left(y_{[.]}|1,y_{t}\right)|(s,y)_{[1,t]}\right) < \rho; \exists s \forall t \forall y\right) \cong \left(S_{T} \to Y_{[.]}\right)$$

$$97$$

Equation 97 is very similar to Equation 44, but a prospective uptime or downtime $Y_{[.]}$ may be more informative for the maintenance policy validation and for practical decision making.

Table 20 surveys the observed frequencies of all possible values of the information set $V=\{s_t, y_t, y_{[t+1,t+x]}\}$ in the sample $(s, y)_{[1,1976]}$. The observed frequencies in the first row (x=1) implied the p-values in Table 16. The p-values for the extended information set $V=\{s_t, y_t, y_{[t+1,t+x]}\}$ are similarly assessable as discussed in Section 5.3.4.

| $V = \{s_t, y_t, y_{[t+1,t+x]}\}:$ | (0,0,0) | (0,0,1) | (0,1,0) | (0,1,1) | (1,0,0) | (1,0,1) | (1,1,0) | (1,1,1) |
|------------------------------------|---------|---------|---------|---------|---------|---------|---------|---------|
| | | | | | | | | |
| Observed frequency (x=1): | 219 | 58 | 58 | 1585 | 29 | 1 | 1 | 24 |
| Observed frequency (x=2): | 184 | 56 | 34 | 1529 | 29 | 1 | 1 | 22 |
| Observed frequency (x=3): | 160 | 53 | 25 | 1474 | 27 | 1 | 1 | 22 |
| Observed frequency $(x=4)$: | 138 | 51 | 23 | 1422 | 25 | 1 | 1 | 21 |
| Observed frequency $(x=5)$: | 117 | 49 | 21 | 1372 | 24 | 1 | 1 | 20 |
| Observed frequency (x=6): | 100 | 46 | 20 | 1325 | 20 | 1 | 1 | 19 |
| Observed frequency (x=7): | 84 | 44 | 17 | 1279 | 18 | 1 | 1 | 19 |
| Observed frequency (x=8): | 72 | 44 | 15 | 1233 | 14 | 1 | 1 | 19 |
| Observed frequency (x=9): | 60 | 44 | 12 | 1191 | 13 | 1 | 1 | 15 |
| Observed frequency (x=10): | 55 | 42 | 6 | 1151 | 11 | 1 | 1 | 11 |
| Observed frequency (x=11): | 51 | 40 | 6 | 1111 | 8 | 1 | 1 | 9 |
| Observed frequency (x=12): | 47 | 38 | 4 | 1071 | 7 | 1 | 1 | 9 |
| Observed frequency (x=13): | 42 | 37 | 4 | 1032 | 7 | 1 | 1 | 9 |
| Observed frequency (x=14): | 38 | 37 | 4 | 995 | 7 | 1 | 0 | 7 |
| Observed frequency (x=15): | 34 | 36 | 4 | 957 | 7 | 1 | 0 | 7 |
| Observed frequency (x=16): | 31 | 36 | 3 | 919 | 7 | 1 | 0 | 7 |
| Observed frequency (x=17): | 28 | 35 | 3 | 882 | 7 | 0 | 0 | 7 |
| Observed frequency (x=18): | 25 | 32 | 3 | 848 | 7 | 0 | 0 | 7 |

Table 20 Contingency table of V={st,yt,y[t+1,t+x]}

Let prospective uptime in Equation 96 be a "win". Then, Figure 32 depicts the observed proportions of this "win" for every possible element of the sample space $\{s_t, y_t\}$ claimed by proposition P3 in the extended nonparametric argument. Figure 32 also shows the p-values to identify the significance of the difference in these proportions as explained in Section 5.3.4. The right-hand graph in Figure 32 shows that maintenance policy compliance S_T may well be independent from prospective uptime when Y_T=0, i.e. the item is down, because significance levels are above α =0,01. The sample $(s,y)_{[1,1976]}$ therefore existentially refutes prima facie causality by Equation 97.





Alternatively, let prospective downtime in Equation 96 be a "win". Then, Figure 33 depicts the observed proportions of this "win" for every possible element of the sample space of the body of knowledge U={ s_t, y_t }. Figure 33 also shows the p-values to identify the significance of the difference in these proportions as explained in Section 5.3.4. The right-hand graph in Figure 33 shows that maintenance policy compliance S_T may well be independent from prospective downtime when Y_T=1, i.e. the item is up, because

significance levels are above α =0,01. The sample (s,y)_[1,1976] therefore existentially refutes prima facie causality by Equation 97.



Figure 33 Prospective downtime, given {st,yt}

In conclusion, a maintenance policy validation by the nonparametric argument and the sample $(s,y)_{[1,1976]}$ existentially supports the claims that:

Maintenance policy compliance S_T *does* **not** *prima facie cause prospective* uptime with respect to the information set $V = \{s_t, y_t, y_{[t+1,t+x]}\}$ when x is in [1,18].

Maintenance policy compliance S_T does **not** prima facie cause prospective downtime with respect to the information set $V = \{s_t, y_t, y_{[t+1,t+x]}\}$ when x is in [1,18].

Still, we may wonder whether a decision maker who controls maintenance policy compliance S_T is unaware about the item's current functionality Y_T . In Section 3.3.4, we mentioned for example that corrective maintenance has been triggered by faults that are currently known. So, it may well be that decision makers do know the current Y_T in practice. Resembling the event tree in Figure 7, the number of possible values of Y_T then reduces to either up $Y_T=1$ or down $Y_T=0$ and our position regarding prima facie causality may change. If the item is known to be up $Y_T=1$ at the moment that S_T is being controlled, the right-hand graph in Figure 32 can be neglected.

Then, we may assert at a significance level α =0,01:

If the item is known to be up, maintenance policy compliance S_T would prima facie cause remaining uptime with respect to the information set $V = \{s_t, y_t, y_{[t+1,t+x]}\}$ when x is in [1,18].

If the item is known to be down $Y_T=0$ at the moment that S_T is being controlled, the right-hand graph in Figure 33 can be neglected.

Then, we may assert at a significance level α =0,01:

If the item is known to be down, maintenance policy compliance S_T would prima facie cause remaining downtime with respect to the information set $V = \{s_t, y_b, y_{[t+1,t+x]}\}$ when x is in [1,18].

In addition, the observed proportions in the left-hand graphs of Figure 32 and Figure 33 indicate the direction of the causality. Maintenance policy compliance $S_T=0$ extends remaining uptime and it shortens remaining downtime.

This section showed that knowledge about the item's current functionality Y_T at the moment of controlling S_T determines our attitude towards the prima facie causality by Equation 97. Given a known current functionality Y_T , a maintenance policy validation by the extended nonparametric argument and the sample $(s,y)_{[1,1976]}$ would existentially confirm the prima facie causality between maintenance policy compliance S_T and prospective functionality $Y_{[.]}$. The extended nonparametric argument improved the inference precision of the nonparametric argument by revealing long-term functionality responses to maintenance policy compliance S_T .

5.4.3 Reduced nonparametric argument

This section will present an alternative path graph that admits all candidate samples to the nonparametric argument.



Figure 34 Path graph of the reduced nonparametric argument

This allows us to compare the candidate samples on their inference precision. In Section 5.3.2, we already validated the maintenance policy by the maintenance prognostic argument and all candidate samples, but the Δ AIC scores only held for a specific candidate sample. This obstructed a comparison across samples. In this section, we will

just pursue this mutual comparison of the candidate samples (and not a further improvement of the inference precision as such).

Figure 34 specifies the non-causality assumptions of the reduced nonparametric argument that, as opposed to those of the nonparametric argument (NPA) in Figure 11, conceive functionality K_T as an unknown and independent background variable B. The propositions of the reduced nonparametric argument and the propositions of the nonparametric argument are almost identical; only K_T should be omitted in Figure 15. In Section 4.4.3, we already explained that the nonparametric argument required a less precise categorical quantification of functionality K_T to existentially confirm the prima facie causality in Equation 41. However, the information set of the reduced nonparametric argument delimits to $V=\{l_t,k_{t+1}\}$ as shown in Figure 34. Then, a replication could also be conceived as any information set $V=\{l_t,k_{t+1}\}$ and following a similar reasoning as expounded in Section 4.4.3, the requirement of a constant $\{k_t\}$ in a replication $V=\{l_t,k_{t+1}\}$ becomes superfluous. As a result, all candidate samples become potentially admissible to the reduced nonparametric argument.

The non-causality assumptions in Figure 34 indicate that the reduced nonparametric argument presumes independence between maintenance policy compliance L_T and functionality K_{T+1} . To test for this independence, the sample $(l,k)_{[1,t]}$ should still comprise sufficient different values of maintenance policy compliance L_T to existentially confirm a null response in functionality K_{T+1} . The observed frequencies of the various values of maintenance policy compliance in Figure 27 already confirmed that the cardinal samples $(d,\bar{y})_{[1,70]}$ and $(d,q)_{[1,1977]}$ comprise different values of maintenance policy compliance in Figure 27 already confirmed that the cardinal samples $(d,\bar{y})_{[1,70]}$ and $(d,q)_{[1,1977]}$ appear to be indeed admissible to the nonparametric argument in this specific case study. Of course, the dichotomous sample $(s,y)_{[1,1976]}$ was already admissible as indicated in Table 17.

In the remainder of this section, we will present the maintenance policy validation by the reduced nonparametric argument by the candidate samples $(d,\overline{y})_{[1,70]}$, $(d,q)_{[1,1977]}$ and $(s,y)_{[1,1976]}$ and we compare these validations on their inference precision. This maintenance policy validation will assess the following prima facie causality:

$$\left(\mathcal{L}\left(\left(pr_{K_{T+1}|L_{T}}(k_{t+1}|l_{t}) = pr_{K_{T+1}}(k_{t+1})\right)|(l,k)_{[1,t]}\right) < \rho; \exists l \forall t\right) \cong (L_{T} \to K_{T+1})$$
98

For the sample $(d,\bar{y})_{[1,70]}$, an inference of a claim regarding the prima facie causality in Equation 98 by the reduced nonparametric argument may proceed as follows:

Let N_U be the observed frequency of U={d_m}={2} in the sample $(d,\bar{y})_{[1,m-1]}$, i.e. N_U=4; Let N_U' be the observed frequency of the complement of U, i.e. N_U=65; Let a "win" for the sample $(d,\bar{y})_{[1,70]}$ be defined by "all values of \overline{Y}_{M+1} equal or below a particular threshold value \bar{y}_{m+1} " as shown in:

$$\bar{Y}_{M+1} = \begin{cases} \leq \bar{y}_{M+1}; "win" \\ > \bar{y}_{M+1}; "loss" \end{cases} \quad for \ (d, \bar{y})_{[1,70]} \end{cases}$$
99
Let K_U be the observed frequency of V={U,"win"} in the sample $(d,\bar{y})_{[1,m]}$, i.e. K_U ranges from 0 to 4;

Let K_U ' be the observed frequency of $V = \{U', "win"\}$ in the sample $(d, \overline{y})_{[1,m]}$, i.e. $K_{U'}$ ranges from 0 to 65;

Then, $K=K_U+K_{U'}$ ranges from 0 to 69 and $N=N_U+N_{U'}=69$.

Let presumption P5, P7 be true. Note that P5 is uncontroversial.



Figure 35 Ku/Nu, K/N and the p-value of presumed independence for (d,y)[1,m]

Then, Figure 35 existentially confirms independence between functionality and maintenance policy compliance at $D_M=2$ as the difference between the observed proportions is not significant at an arbitrary level $\alpha=0,01$, i.e. the p-value always exceeds $\alpha=0,01$. As shown in Figure 27, the observed frequency of the replications N_U does not exceed four which is fairly small for any significant result. We therefore deem the maintenance policy validation by the reduced nonparametric argument and the sample $(d,\bar{y})_{[1,70]}$ as indecisive about the prima facie causality in Equation 98. We therefore conclude:

Maintenance policy compliance D_M may or may not prima facie cause functionality \overline{Y}_{M+1} with respect to the information set $V = \{d_m, \overline{y}_{m+1}\}$.

For the sample $(d,q)_{[1,t]}$, an inference of a claim regarding the prima facie causality in Equation 98 by the reduced nonparametric argument may proceed as follows:

Let N_U be the observed frequency of U={d_t}={3} in the sample $(d,q)_{[1,t-1]}$, i.e. N_U=114; Let N_U' be the observed frequency of the complement of U, i.e. N_U=1860; Let a "win" for the sample $(d,q)_{[1,1977]}$ be defined by "all values of Q_{T+1} that are equal or below a particular threshold value q_{t+1}" as shown in:

$$Q_{T+1} = \begin{cases} \leq q_{T+1}; "win" \\ > q_{T+1}; "loss" \end{cases} \quad for \ (d,q)_{[1,1977]}$$
 100

Let K_U be the observed frequency of V={U,"win"} in the sample $(d,q)_{[1,t]}$, i.e. K_U ranges from 0 to 114;

Let K_U ' be the observed frequency of $V = \{U', "win"\}$ in the sample $(d,q)_{[1,t]}$, i.e. $K_{U'}$ ranges from 0 to 1860;

Then, $K=K_U+K_{U'}$ ranges from 0 to 1975 and $N=N_U+N_{U'}=1975$. Let presumption P5, P7 be true. Note that P5 is uncontroversial.



Figure 36 K_U/N_U, K/N and the p-value of presumed independence for (d,q)_[1,t]

Then, Figure 36 existentially refutes independence between functionality and maintenance policy compliance at $D_T=3$ as the difference between the observed proportions is significant at an arbitrary level α =0,01. However, the observed proportions in Figure 36 indicate that the distributions of functionality Q_{T+1} differ in shape rather than in expectation. So, a short queue of delayed maintenance $D_T=3$ appears to reduce the variation in functionality Q_{T+1} rather than to change the mean of functionality Q_{T+1} . Still, the maintenance policy validation by the reduced nonparametric argument and the sample (d,q)_[1,1977] existentially confirms the prima facie causality in Equation 98. We therefore conclude:

Maintenance policy compliance D_T prima facie causes functionality Q_{T+1} with respect to the information set $V = \{d_t, q_{t+1}\}$

For the sample $(s,y)_{[1,1976]}$, we just duplicate the inference of a prospective uptime and a prospective downtime in Section 5.4.2 to also reveal eventual long-term effects. This maintenance policy validation then transforms the prima facie causality in Equation 98 to:

$$\left(\mathcal{L} \left(pr_{Y_{[T+1,T+X]}|S_T} (y_{[t+1,t+x]}|0) = pr_{Y_{[T+1,T+X]}|S_T} (y_{[t+1,t+x]}|1) \right) < \rho; \exists s \forall t \right)$$

$$\cong \left(S_T \to Y_{[T+1,T+X]} \right)$$
 101

Equation 101 just reduces to Equation 98 as $Y_{T+1,T+X}$ reduces to Y_{T+1} , i.e. X=1.

The left-hand side in Figure 37 existentially refutes independence between maintenance policy compliance at S_T and prospective uptime as the difference between the observed proportions is significant at an arbitrary level α =0,01. The observed proportions indicate that maintenance policy compliance (S_T =0) tends to extend prospective uptime.

The right-hand side in Figure 37 existentially refutes independence between maintenance policy compliance at S_T and prospective downtime as the difference between the observed proportions is again significant at an arbitrary level α =0,01. The

observed proportions indicate that maintenance policy compliance ($S_T=0$) tends to shorten prospective downtime.



Figure 37 Prospective uptime (left) and downtime (right) given $\{s_t\}$

Therefore, the maintenance policy validation by the reduced nonparametric argument and the sample $(s,y)_{[1,1976]}$ existentially confirms the prima facie causality in Equation 101. We therefore conclude:

Maintenance policy compliance S_T *prima facie causes prospective uptime with respect to an information set* $V = \{s_t, y_{[t+1,t+x]}\}$ *for x in* [1,18].

Maintenance policy compliance S_T *prima facie causes prospective downtime with respect to an information set* $V = \{s_t, y_{[t+1,t+x]}\}$ *for x in [1,18].*

Although the p-values in Figure 37 are substantially lower than the corresponding p-values in Figure 32 and Figure 33, this maintenance policy validation by the reduced nonparametric argument and the sample $(s,y)_{[1,1976]}$ should not be seen as more precise. The proportions in the left and right-hand graph in Figure 32 and Figure 33 suggested a strong dependence between Y_T and $Y_{[T+1,T+x]}$. It is hard to accept an information set V that excludes a known variable (like Y_T) that is also known to be strongly dependent.

This section compared the maintenance policy validations by the reduced nonparametric argument and the samples $(d,\bar{y})_{[1,70]}$, $(d,q)_{[1,1977]}$ and $(s,y)_{[1,1976]}$. These samples have all been deduced from the same recording routines. Again, the case organisation's convention to maintenance performance $(d,\bar{y})_{[1,70]}$ led to an indecisive result that was attributable to the low observed frequencies of the replications. The samples $(d,q)_{[1,1977]}$ and $(s,y)_{[1,1976]}$ existentially confirmed a prima facie causality between maintenance policy compliance and functionality at a significance level α =0,01. Still, the maintenance policy validation by the reduced nonparametric argument and the sample $(s,y)_{[1,1976]}$ appeared to be most compelling for the prima facie causality in Equation 98 as it attained the lowest p-values. Possibly, the sample $(d,q)_{[1,1977]}$ would have attained lower p-values at some other, but less frequently observed value of maintenance policy

compliance D_T (Figure 27). Clearly, this additional analysis effort to eventually test for the prima facie causality in Equation 98 at other observed values of maintenance policy compliance D_T is omissible in the case of the sample $(s,y)_{[1,1976]}$. Moreover, the sample $(s,y)_{[1,1976]}$ also provided an insight into long-term functionality responses to maintenance policy compliance that seemed less obviously attainable in the case of the sample $(d,q)_{[1,1977]}$. In conclusion, the maintenance policy validation by the reduced nonparametric argument and the sample $(s,y)_{[1,1976]}$ was the most precise whilst requiring less of an analysis effort and giving better insight into the long-term effects.

5.5 Findings regarding the "universal argument" inference objective

Ideally, scientific arguments hold universally, but in practice an operating organisation only has a stratified sample of recording routines available. If we could presume that an argument holds irrespective of any background variable B like in the maintenance optimisation argument (MOA in Figure 11), we might still universally refute the argument by a stratified sample as we did in Section 5.3.1. All other arguments did specify some independence between L_T , B or K_T , B (Figure 11, Figure 30 and Figure 31) that remained unassessible. These presumptions of independence were essential to interpret an existentially inferred claim regarding the prima facie causality in Equation 41 as being causal. However, confounding or mediating background variables B almost certainly exist but we are simply unable to even enumerate them all, nor can we presume equality of their distributions due to random assignment of treatments. The maintenance policy validation therefore remains vulnerable to these background variables B.

In this section, we will capture some of the suspicions from a field expert about sample bias and we will verify the recording routines underlying the candidate samples with some in-depth knowledge about the case study. The magnitude of these concerns is unassessible from the evidence available. A split sample validation may reveal time-dependent behaviour and it may reveal an efficiency improvement; i.e. the applicability of the maintenance policy validation benefits from an ability to infer similar results from a shorter time series sample $(l,k)_{[1,t]}$.

5.5.1 Background variables influencing functionality

The functionality in the case study has been built on a physical variable from an instrument that has not been subject to known calibration issues. We are therefore unaware of known and avoidable background variables that biased the evolution of functionality K. However, a field expert suspected that maintenance policy compliance L is not regarded as a dominant cause of functionality K. This suspicion implies that maintenance policy compliance L may not be an attractive element of the body of knowledge U to predict functionality K. The possible presence of unobserved background variables B potentially heavily biased the maintenance policy validation.

5.5.2 Background variables influencing maintenance policy compliance

It would be quite naïve to presume that all members in the queue of delayed maintenance equally attribute to functionality. The composition of the queue is therefore another suspected cause of functionality. The maintenance recordings usually include additional information which will allow refined removal factors. For example, in this case study we could have specified the queue by discipline (electrical, mechanical,...), by the item's involved components, by job type (preventive, corrective,...) or by urgency. Eventually, it reveals that some modified queue much more strongly prima facie causes functionality. For a decision maker, it may also be more appropriate to only manipulate some subset of the maintenance policy. So, it could be worthwhile to implement additional removal factors and to infer the associated (prima facie) causality. This extension potentially identifies the most influential maintenance policy violations. As long as the refreshment of the queue suffices, these extensions are possible. In this work, however, we seek for a generic justifiability of maintenance. Then, these removal factors would appear as additional antecedents in the argument that require model parameters or sufficient replications. We therefore initially ignored the evidence that enables further partitioning of the queue of delayed maintenance.

The case study seems unaffected by software conversions, changed operational intents or changes in the operating and maintenance crew during the time interval spanned by the sample $(l,k)_{[1,t]}$, but this does not imply integrity of the maintenance recordings. In the case study, a field expert suspected, for example, that there was a herding effect to register completions batch-wise at a "convenient time". Indeed, the number of days with at least one arrival (1181) considerably exceeds the number of days with at least one completion (921). This finding supports the expert's suspicion.

The case study comprised 74 transitions from a "bad" functionality ($Y_T=0$) to a "good" functionality ($Y_T=1$). We presume that these transitions require human intervention that should be seen as maintenance. However, 16 out of these 74 transitions are not associated with a recording of the completion of some maintenance action. It could well be the case that our stratified sample is incomplete or lacks integrity.

5.5.3 Split sample validation

In Section 4.5.5, we already mentioned that a sample $(l,k)_{[1,t]}$ comprises more information than the information set V={P3,P4} of the prima facie causalities that we explored. The candidate samples in Figure 20, Figure 21 and Figure 22 show, for example, some symptoms of non-stationary drifts that have not been described by the information set V={P3,P4}. The presumed definitions of a replication in Table 6 that we used in all maintenance policy validations so far are therefore problematic. So, we did not verify the second causality principle (Section 2.3.3), i.e. a causality remains constant in direction throughout time. In this section, we will implement a split sample validation to (*i*) test whether we could increase the efficiency of the maintenance policy validation by showing similar results from a shorter time series $(l,k)_{[1,t]}$ and to (*ii*) test whether our results are time dependent. We will only duplicate the maintenance policy validation by the nonparametric argument and the sample $(s,y)_{[1,1976]}$ while presuming knowledge about the current functionality Y_T as discussed in Section 5.4.2. A partitioning of the sample $(s,y)_{[1,1976]}$ then yields Figure 38 and Figure 39, which correspond to the left-hand graphs in Figure 32 and Figure 33 respectively.



Figure 38 Split sample validation for remaining uptime

The observed proportions in the remaining uptime for both halves in Figure 38 are problematic for the second causality principle (Section 2.3.3), which requires that causalities should remain constant in direction throughout time because during the first 8 days, maintenance policy compliance $S_T=0$ seems to extend remaining uptime in the first half and to shorten remaining uptime in the second half. However, the p-values indicate that this difference is insignificant. Still, maintenance policy compliance $S_T=0$ tends to extend remaining uptime after 10 days in both halves. However, the p-values have grown to barely acceptable values. An existential confirmation of the dependence between S_T and $Y_{[.]}$ by Figure 38 at an arbitrary significance level $\alpha=0,01$ has become barely tenable. We therefore posit that we cannot improve the efficiency of the maintenance policy validation by reducing the time series by half.



Figure 39 Split sample validation for remaining downtime

The trends in the remaining downtime for both halves in Figure 39 are also similar. Again maintenance policy compliance $S_T=0$ tends to shorten remaining downtime. However, the p-values have grown to barely acceptable values. An existential confirmation of the dependence between S_T and $Y_{[.]}$ by Figure 39 at an arbitrary significance level α =0,01 has become barely tenable. We therefore posit that we cannot improve the efficiency of the maintenance policy validation by reducing the time series by half.

Still, the difference between the observed proportions in Figure 38 shows that the remaining uptime tends to shorten upon $S_T=1$ after about 10 days in both halves. Similarly, the difference between the observed proportions in Figure 39 shows that remaining downtime tends to extend upon $S_T=1$ during the first 10 days in both halves. So, although the differences in the observed proportions in Figure 38 and Figure 39 are barely significant, they are not problematic for the second causality principle which requires that a causality remains constant in direction throughout time.



Figure 40 Split sample validation for stationarity

This does not mean that the split sample validation would support stationarity of the sample $(s,y)_{[1,1976]}$. We therefore alternatively tested for independence of uptimes and downtimes across halves. Figure 40 just depicts the p-values since the corresponding proportions have already been depicted in Figure 38 and Figure 39. Figure 40 shows that the presumption of stationarity has been refuted at an arbitrary significance level α =0,01.It can therefore be concluded that:

- Given U={st,yt}={0,0}, the proportions in Figure 39 indicate that remaining downtime in the second half is significantly shorter (α=0,01);
- Given U={s_t,y_t}={0,1}, the proportions in Figure 38 indicate that remaining uptime in the second half is significantly shorter (α =0,01).

Since we do not believe that membership of a particular half on its own caused this nonstationarity, Figure 40 triggers a quest for background variables.

In this section, we surveyed some expert suspicions regarding background variables that could bias the maintenance policy validation applied to the case study. We also showed that we only selectively used the information from the recording routines of the case organisation. This information allows for extensions to more refined compositions of the maintenance queue. The outcome of the split sample validation was not problematic for the second causality principle, but it showed some symptoms of non-stationarity in the sample $(s,y)_{[1,1976]}$. Since time in itself is unlikely to explain this non-stationarity, it can be concluded that the inference precision suffers from background variables.

6 Discussion

This section will survey and critically discuss our achievements with respect to an improved inference. We will organise them along the choices in Table 2, i.e. the choice of an argument, of an operationalisation and of a sampling procedure. The structured approach along these choices is original and allowed us to compare a number of arbitrarily selected options. Moreover, it allowed us to propose a way to quantify the precision of a causal inference. Although this quantification could be considered to be based on common sense, it has not been described precisely in scientific literature before. In the next subsections, for each of these choice problems, the contribution of the present work will be stated.

6.1 Choice of an argument

An argument is an essential vehicle for our reasoning. We considered several arguments that differ in structure and potentially also in inference precision. This section will survey our achievements regarding the choice of an argument.

6.1.1 Main findings

We have constructed a set of candidate arguments that differ in their presumptions. In Table 5, we preliminarily compared these candidate arguments on their inference precision and in Section 5.3, we confronted these candidate arguments with the samples $(d,\bar{y})_{[1,70]}$, $(d,q)_{[1,1977]}$ and $(s,y)_{[1,1976]}$ from a realistic case study. Table 21 now surveys our definite findings regarding the inference objectives for this specific case study.

Table 21 generally confirms the preliminary assessment of inference precision in Table 5. To serve the aim of this work, we now survey what the candidate arguments, given their inference precision in the case study, assert about the causality between maintenance policy compliance and functionality.

The maintenance optimisation argument has universally been refuted. This refutation applies only to a very specific (definitional) equivalence of functionality and maintenance policy compliance that is more coercive than a (prima facie) causality. Only a confirmation of the maintenance optimisation argument would have been decisive for the prima facie causality between maintenance policy compliance and functionality.

If the maintenance prognostic argument were sound, it would have been very compelling for the prima facie causality between maintenance policy compliance and functionality because its model M2 eventually maps maintenance policy compliance to an estimate of prospective functionality. For prospective decision making about a yet to be observed future, a sound maintenance prognostic argument would have been

appreciable for its predictive capabilities to make beyond sample claims. Unfortunately, any sample $(l,k)_{[1,t]}$ collected by observational research appeared insufficient to establish the soundness of the maintenance prognostic argument. We therefore resorted to an *existential* claim about the expected information loss of some arbitrarily presumed model. Only for the sample $(s,y)_{[1,1976]}$ could we compose a finite set of candidate model families that were uncontroversial if a replication would reduce to some subset of the information set V like in a prima facie causality. So, only for the sample $(s,y)_{[1,1976]}$ did the maintenance prognostic argument become *decidable* in terms of AIC scores. Although the AIC scores only compared candidate models on their expected information loss when approximating functionality Y_{T+1} , we ultimately concluded that the maintenance policy validation by the maintenance prognostic argument and the sample $(s,y)_{[1,1976]}$ was also *decidable* about a prima facie causality.

| | Maintenance optimisation argument (Section 4.1) | Maintenance prognostic argument (Section 4.2) | Reliability engineering argument (Section 4.3) | Nonparametric argument (Section 4.4) |
|--|--|--|---|---|
| Valid argument | Yes | Yes | Yes | Yes |
| Functional relation | Yes | Yes | Yes | Yes |
| Common sense evidence | L,K incomplete L,K subjective | L,K incomplete L,K subjective | L,K incomplete L,K subjective | L,K incomplete L,K subjective |
| Universal argument | Yes, but refuted | No | No | No |
| Decidable Argument | Yes | Only in AIC for (s,y) _[1,1976] . | Only in AIC for (s,y) _[1,1976] . | Only in likelihood |
| Decidable about L _T →K _{T+1} w.r.t. {lt,kt,kt+1} | No | Only in AIC for (\$,y) _[1,1976] . | Only in AIC for (\$,y) _[1,1976] . | Only in likelihood for (s,y)[1,1976]. |

Table 21 Definite inference precision of the candidate arguments

The reliability engineering argument appeared to be somewhat superfluous in this discussion. For the dichotomous sample, it reconciled with the maintenance prognostic argument and for the cardinal samples, it appeared to delimit the amount of admissible evidence rather than to reduce controversy about the presumptions in the maintenance prognostic argument.

The nonparametric argument was geared to *existential* claims about the likelihood of a presumption of independence. The nonparametric argument was most restrictive about the composition of the sample and only the sample $(s,y)_{[1,1976]}$ appeared to consist of sufficient replications. Table 16 showed that today's maintenance policy compliance S_T is unlikely to prima facie cause tomorrow's functionality Y_{T+1} but we also sought for

long-term effects. In Section 5.4.2, we confirmed that maintenance policy compliance $S_T=0$ associates with longer remaining uptimes and shorter remaining downtimes. Provided that the current functionality Y_T is known, this association would also imply a prima facie causality between maintenance policy compliance and functionality.

Furthermore, the nonparametric argument is appreciable for (*i*) the computational efficiency, i.e. not all candidate models should be assessed as required for a maintenance policy validation by the maintenance prognostic argument (Table 11), (*ii*) its potential to compare different samples at a reduced information set $V=\{l_t,k_{t+1}\}$ like we did in Section 5.4.3, (*iii*) its ability to quantify its ambiguity about a prima facie causality by a likelihood whereas the AIC and AIC_c scores in Section 5.3.2 only compared models on expected information loss when approximating functionality K_{T+1} , (*iv*) its ability to both existentially confirm and existentially refute the prima facie causality in Equation 41 and (*v*) its more compelling causal interpretation upon an existential confirmation of the prima facie causality in Equation 41, whereas the maintenance prognostic argument and the reliability engineering argument would then require an additional test for independence between L_T and K_T.

Therefore, the nonparametric argument became our argument of choice in the case study.

6.1.2 Our contribution regarding the argument selection

The challenge of the argument selection was:

To achieve inference precision by choosing an adequate argument.

This iterative quest for an argument may be seen as just another instantiation in the choice of some prognostic or diagnostic argument. The candidate arguments in this work covered various approaches like analytical redundancy by the maintenance optimisation argument, a model based approach by the maintenance prognostic argument, a history based approach by the nonparametric argument and some hybrid reliability engineering argument. The influence of categorical and cardinal samples on the choice of an argument is similarly well explored by conventional diagnostics and prognostics. Still, we consciously sought for a *causal* argument whereas diagnostic and prognostic conventions often effectively reason about physical symptoms. This work may be seen as an initial step to explore causal inferences in a maintenance decision making context. As the amount of recording routines that could support maintenance decision making seem to be growing, causal inferences under observational research constructs that seem well explored in other fields could potentially become important here. Finally, we critically compared the candidate arguments on inference precision to arrive at an argument of choice in a specific case. This practice of empirical science is transferrable to other cases in principle. In summary:

We have implemented a maintenance policy validation by a causal argument and a sample from a realistic case study at an improved inference precision.

6.2 Choice of an operationalisation

Inference precision heavily relies on the evidence we are willing to accept as true. To mitigate controversy, we sought for common sense. To validate a presumption regarding causality of maintenance policy compliance and functionality, we needed common sense about three aspects:

- Maintenance policy compliance;
- Functionality;
- Causality.

In this section, we will summarise and discuss our achievements regarding the operationalisation of these three aspects.

6.2.1 Main findings

The literature review on normative decision theory did not appear to be very encouraging for policy validations since the assessment of an individual's preference is problematic. Decisions to carry out maintenance usually happen in some organisation which is a choice to collaborate. To enable group members to align with the organisation's goals, maintenance performance indicators appear to be helpful. We therefore proposed to depart from this common sense about the organisation's goals of a maintenance policy.

Although an intuition exists that leading performance indicators (maintenance policy compliance) cause lagging performance indicators (functionality), we found that conventional maintenance scorecards do not really accommodate causal inferences. This might explain why we did not find any attempt to do so in the literature.

Our major and partly unresolved concern is the operationalisation of causality. Given the constraint on an observational research, we were unable to find a common sense operationalisation of causality. Still, we posited common sense about three causality principles that we tried to address in the candidate arguments:

- The past and present may cause the future, but the future cannot cause the past. This causality principle has been addressed by collecting a time series sample $(l,k)_{[1,t]}$ and presumption P5 that asserts that future functionality K_{T+1} could never have caused current maintenance policy compliance L_T .
- All causal relationships remain constant in direction throughout time. This causality principle has only been addressed by the split sample validation in Section 5.5.3. Although the second causality principle was not directly problematic, we clearly revealed some non-stationarity concerns.
- A cause comprises unique information about the effect that is not available otherwise. This causality principle could not be addressed by the obvious choice of random assignment of treatments. We therefore resorted to a modest notion of prima facie causality that we enhanced with quantifiers. These quantifiers appeared to be very useful in comparing the inference precision of a maintenance policy validation by the candidate arguments and samples.

The causality principles are meant to hold *universally*. We have similarly been explicit about the *universal* but controversial presumptions of the candidate arguments that were

needed to make a *universal* claim from a spatiotemporally constrained sample $(l,k)_{[1,t]}$. Except for the maintenance optimisation argument that was essentially not a causal argument, the other candidate arguments only allowed for *existential* claims regarding prima facie causality.

The recording routines of the case study were not predominantly affected by background variables like software conversions or reorganisations. The case study was therefore quite advanced in terms of influences from background variables. Furthermore, we ignored some lagging indicators on cost control, on resource allocation and on health, safety and environment in the case study. By just confining to functionality, we have been incomplete about the case organisation's goals of the maintenance policy. To alleviate this concern, we implemented a removal factor on the maintenance policy compliance variable which, in hindsight, would not have altered our result.

In the case study, we did not recognise calibration issues in the recordings of the physical variable (output) which represented functionality. The objective to maximise output did not seem to be controversial either. The maintenance policy compliance variable was built on a queue of delayed maintenance. We noticed some ambiguity in the perception of a delay since an individual field expert did not always agree with the case organisation's definition. We just chose to follow the case organisation's convention. Furthermore, we found some symptoms of human factor bias in the maintenance policy compliance recordings of unknown magnitude.

Finally, we tested for a prima facie causality that did not capture all information in the sample $(l,k)_{[1,t]}$ of the case study. A split sample test confirmed some non-stationary drifts that allude to effects from background variables which potentially reduce the inferred prima facie causality to spurious (Figure 5). Still, the split sample validation was not really problematic for the second causality principle which asserts that a causality should remain constant in direction throughout time.

6.2.2 Our contribution regarding the operationalisation

The challenge of an operationalisation was:

To achieve inference precision by establishing common sense about the evidence.

This work addressed the challenge to observe the effects of decisions in an organisation's recording routines. Maintenance policies appeared to be suitable for an empirical validation because they trigger decisions at a high rate and because the abundant policy violations are typically also recorded. An attempt to use violations to get a glimpse into the counterfactual reality appears to be new from a normative decision theoretical perspective. We realise that maintenance cases allow for this research construct that is atypical in normative decision theory.

Common sense is not enforceable. So, we merely critically compared the case organisation's common sense about its subjective goals of a maintenance policy with literature conventions and field expert judgement. However, it ultimately remains a

matter of personal taste to accept or reject our proposed operationalisation of maintenance policy compliance and functionality.

We pioneered a modest notion of prima facie causality in a maintenance decision making context. We added quantifiers to Granger's (1980) original definition of a prima facie causality to be more precise about its assessment. These quantifiers revealed that the sample space of the information set V is decisive for the assessment of a prima facie causality, as we conceptually showed in Figure 7. In Section 5.4.2, we existentially confirmed a prima facie causality between maintenance policy compliance and prospective functionality, provided that the current functionality is known. This delimitation of the *possible* values in the information set V by common sense may similarly enhance our ability to observe and meaningfully use prima facie causalities in other cases as well. In summary:

We have implemented a maintenance policy validation based on evidence about policy violations that appears to be new from a normative decision theoretical perspective.

6.3 Choice of a sampling procedure

The sampling procedure determines inference precision by both its influence on the acceptance of causality and the efficiency of collecting evidence. In this section, we will survey our achievements regarding the choice of the sampling procedure.

6.3.1 Main findings

We found that precision of causal inferences highly benefits from experimental research that allows for random assignment of treatments. Since we confined ourselves to evidence about recording routines, the resulting sampling procedure is not optimal. Nevertheless, many organisations apparently have loads of recording routines that potentially contain valuable knowledge. The choice of this observational research construct therefore contributes to the practical applicability of the causal inference.

A survey revealed that conventional maintenance performance indicators are typically geared to show posterior satisfaction of goals. For decisions that only influence the yet to be observed future, predictive capabilities would be desirable. We have not found any organisation that inferred predictive models from its maintenance performance indicators.

We found that conventional maintenance performance indicators do not really allow for causal inferences that enable maintenance performance predictions. We derived and implemented some construction rules to adjust them and we revealed a major improvement in our predictive capabilities using a typical realistic case study. So, the case organisation (and other organisations with similar performance indicators) could potentially learn more from its recording routines.

We also compared a dichotomous sample $(s,y)_{[1,1976]}$ with a cardinal sample $(d,q)_{[1,1977]}$. Although the dichotomous sample is a less precise quantification, we showed its merits in terms of sampling efficiency and in terms of tenable presumptions for the candidate causal arguments. In Section 5.4.3, we compared the inference precision of the maintenance policy validation by the nonparametric argument and the candidate samples $(d,\bar{y})_{[1,70]}$, $(d,q)_{[1,1977]}$ and $(s,y)_{[1,1976]}$. Clearly, the sample $(s,y)_{[1,1976]}$ yielded the highest inference precision, i.e. lowest p-values, while enabling claims about an interval rather than just one-step-ahead. Therefore, the sample $(s,y)_{[1,1976]}$ became our sample of choice.

6.3.2 *Our contribution regarding the sampling procedure*

The challenge of the selection of a sampling procedure was:

To achieve inference precision by composing a suitable sample, given the constraint on an observational research.

We pioneered the redesigning of a maintenance scorecard which allows for causal inferences that are essential for maintenance performance predictions. We therefore put forward some construction rules that (*i*) comply with the practice of multiple criteria decision making (*ii*) respect the subjective goals to be measured by the maintenance performance indicators and (*iii*) are straightforwardly implementable on typical recording routines. We have constructed two alternative samples in the case study that respected these construction rules and we showed that both of these alternative samples outperformed the case organisation's convention to maintenance performance is rather typical, it is expected that other organisations may similarly benefit from the proposed construction rules. In summary:

We have implemented alternatives for conventional maintenance performance indicators that enable more precise causal inferences in the case study.

7 Conclusion

7.1 Maintenance is unjustifiable; an improved inference

The proposition "maintenance is unjustifiable" obstructs the justification of maintenance. In this work, we were looking for an improved inference that would enhance the justifiability of maintenance. We stuck to a quest for an inference that is more precise about the proposition:

Maintenance policy compliance causes functionality

We have implemented a policy validation which relies on policy compliance that appeared to be unconventional from a normative decision theoretical perspective where "choosing" and "doing" coincide and correspond. A maintenance policy appeared to be suitable for such a validation since it triggers decisions at a high rate and the abundant maintenance policy violations are typically also recorded. Policy violations may reveal the counterfactual reality which maintenance policy compliance intends to avoid in the first place.

As anticipated, we failed to be precise in the operationalisation, the sampling and the argument:

- Although we *operationalised* maintenance policy compliance and functionality by common sense and we pioneered a modest notion of prima facie causality in a maintenance decision making context, we failed to be complete in our operationalisation and we recognised some concerns regarding background variables that are problematic for causal explanations of the observed associations.
- Although we demonstrated how the *sampling* of maintenance performance indicators could better enable causal inferences that are essential in deciding about the future, we failed to apply a well-designed experimental research that would have been more compelling for a causality. This sampling constraint applies to many operating organisations that typically only have recording routines about maintenance performance available.
- Although we constructed several causal *arguments*, we resorted to just an existential claim regarding some presumed prima facie causality.

In a practical case study, we showed that maintenance policy compliance S_T prima facie causes functionality $Y_{[T+1,T+x]}$ with respect to the information set $V=\{s_t, y_t, y_{[t+1,t+x]}\}$ provided that the current functionality Y_T is known. Then, maintenance policy compliance prima facie causes longer remaining uptimes and shorter remaining downtimes. Since the case organisation's approach to maintenance performance indicators appears to be conventional, inference precision may similarly improve in other cases. However, the magnitude of the background variables appeared to be rather limited in the present case study, which might be less ideal in other cases. The final conclusion is thus that a maintenance policy validation by the proposed approach is very difficult if the only evidence available is from an organisation's recording routines. This makes it difficult to obtain explicit justification. However, the proposed approach showed how to improve the associated inference precision in a specific case study.

7.2 Practical implications

In this work, we merely tried to better approximate reality by an increased inference precision. We did not, for example, intervene in the course of operations of the organisation in the case study. This observer perspective, which seems unconventional in research on maintenance, obviously raises questions on its practical implications.

Maintenance originates from decisions that can only influence a yet to be observed future. Maintenance policy assessments therefore predominantly rely on expert judgement about the prospective future. We take the viewpoint that this prospective future should materialise in order to retain maintenance policy assessments which are meaningful to practitioners and empirical scientists.

This work revealed that conventional maintenance performance indicators typically do not sufficiently capture the variations that allow us to learn about the system behaviour. We proposed and implemented some construction rules for maintenance performance indicators that enabled us to reveal prima facie causalities from recording routines.

We *firstly* eliminated redundancy to avoid dependencies that are definitional rather than causal. To avoid longitudinal redundancy, we suggested instantaneously observable variables that allow for any convenient sampling rate. We *secondly* proposed to increase the sampling rate to enable a reconstruction of the original signals. This gave insight into the perturbations that are informative for causality. We *finally* balanced completeness with efficiency to arrive at a meaningful but tractable inference.

Although these construction rules appear to be straightforwardly implementable on recording routines, they are often violated in the practice of maintenance performance measurement. We therefore argue that organisations could potentially learn more about the causal effects of their decisions. Eventually, by validating some formal argument, like we did in this work, or by simply asking: "Where does this peak come from?".

7.3 Further research

The approach of this research may be regarded as a case of satisficing. The conclusion therefore only existentially holds with respect to the options that we considered for the operationalisation, the sampling and the argument:

- Regarding the *operationalisation*, we could have considered another incomplete representation of the case organisation's common sense or we could have challenged it.

- Regarding the *sampling*, we could have extended the sample in various ways and we could have considered a sampling procedure that better approximates a random assignment of maintenance policy compliance treatments.
- Regarding the *argument*, we could have considered arguments that rely on neural networks or genetic algorithms and we could have explored some hybrid argument that merges the inferences of the arguments we did consider.

The remainder of this section will just outline some applications of this work that we are currently working on.

7.3.1 Predictive maintenance performance

We are currently working on the case of an organisation that is quite mature in its prognostic health management capabilities. Prognostic health management triggers decisions to carry out maintenance based on observable evidence. This could, in a similar way, enhance observing maintenance policy violations.

The organisation still uses conventional maintenance performance indicators which are hard to associate with health indicators. We propose that we can improve associating prognostic health information with maintenance performance indicators.

For this endeavour, the present work gave us the experience to implement causal inferences on maintenance performance indicators. Our approach to inference precision allows us compare alternative inferences that predict functionality.

7.3.2 Data driven decision support

We are currently working on the case of an organisation that is reconsidering its maintenance performance indicators. This reconsideration has been triggered by an upgrading of the computerised maintenance management system. This investment should better support maintenance decision making, which implies a need to apply causal inferences to maintenance recordings. This work turned out to be a demonstrator for the construction of maintenance performance indicators.

The organisation operates a fleet that is being exposed to varying operating conditions affecting maintenance. We are running pilot projects to investigate whether our adjusted maintenance performance indicators are predictable from operating conditions.

It is too early to claim successful causal inferences, but the instantaneously observable performance indicators that are sampled at a high rate seem to contribute to data quality. They provide immediate feedback to various stakeholders while coming up with qualitative explanations for perturbations. This side effect may mitigate the effects from background variables which were problematic for this research.

References

Abernethy, R. (2006). *The New Weibull Handbook; reliability and statistical analysis for predicting life, safety, survivability, risk, cost and warranty claims.* North Palm Beach: Robert B. Abernethy.

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle. *Second international symposium on information theory* (pp. 267-281). Budapest: Akademiai Kiado.

- ARMP-7. (2008). Nato R&M terminology applicable to ARMPs. Brussels: NATO.
- Aven, T. (2011). Interpretations of alternative uncertainty representations in a reliability and risk analysis context. *Reliability Engineering and System Safety*, *96*(3), 353-360.
- Baker, R., & Wang, W. (1991). Determining the delay time distribution of faults in repairable machinery from failure data. *IMA Journal of Mathematics Applied in Business and Industry*, 3(4), 259-282.
- Bandyopadhyay, P., & Forster, M. (2011). Philosophy of statistics. In D. Gabbay, P. Thagard, & J. Woods, *Handbook of the philosophy of science*. Elsevier.
- Banjevic, D. (2009). Remaining useful life in theory and practice. Metrika, 69(2-3), 337-349.
- Bernoulli, D. (transl. 1954). Exposition of a new theory on the measurement of risk. *Econometrika*, 22, 23-36.
- Beshears, J., Choi, J., Laibson, D., & Madrian, B. (2008). How are preferences revealed? *Journal of Public Economics*, 92(8), 1787-1794.
- Birolini, A. (2007). Reliability engineering; theory and practice. Springer Verlag.
- Blanchard, B. (2004). Logistics engineering and management. Prentice Hall.
- Bobbio, A., Franceschinis, G., Gaeta, R., & Portinale, L. (2003). Parametric fault tree for the dependability analysis of redundant systems and its high level Petri net semantics. *IEEE Transactions on Software Engineering*, 29(3), 270-287.
- Bouissou, M., & Bon, J. (2003). A new formalism that combines advantages of fault trees and Markov models: Boolean logic driven Markov processes. *Reliability Engineering and System Safety*, 82(2), 149-163.
- Box, G., Jenkins, G., & Reinsel, G. (2008). Time series analysis. Hoboken: John Wiley & Sons.
- Broome, J. (2004). Ethics out of economics. Cambridge: Cambridge University Press.
- Bucci, P., Kirschenbaum, J., Mangan, A., Aldemir, T., Smith, C., & Wood, T. (2008). Construction of event-tree/fault tree models from a Markov approach to dynamic system reliability. *Reliability Engineering and System Safety*, 93(11), 1616-1627.
- Burnham, K., & Anderson, D. (2010). *Model selection and multi model inference; a practical information theoretic approach.* New York: Springer Verlag.
- CEN. (2001). EN 13306; Maintenance terminology. Brussels: European Committee for Standardisation.
- CEN. (2007). *EN 15341; Maintenance- maintenance key performance indicators*. Brussels: European Committee for Standardisation.
- Cheung, J.-Y., & Stephanopoulos, G. (1990). Representation of process trends-part I. A formal representation framework. *Computers & Chemical Engineering*, *14*(4), 495-510.
- Chiang, L., Russell, E., & Braatz, R. (2001). *Fault Detection and Diagnosis in Industrial Systems*. London: Springer-Verlag.
- Christer, A. (1973). Innovatory decision making. In K. Bowen, & W. D.J., The role and effectiveness of theories of decision in practice (pp. 368-377). Hodder and Stoughton.
- Christer, A., Wang, W., Baker, R., & J.M., S. (1995). Modelling maintenance practice of production plant using the delay time concept. *IMA Journal of Management Mathematics*, 6(1), 67-83.
- Coolen, F., Coolen-Schrijner, P., & Yan, K. (2002). Nonparametric predictive inference in reliability. *Reliability Engineering & System Safety*, 78(2), 185-193.
- Copi, I., & Cohen, C. (2009). Introduction to logic. Pearson Education Inc.
- Cox, D. (1972). Regression models and life-tables (with discussion). Journal of the Royal Statistical Society, 187-220.
- Crespo Marquez, A. (2007). The maintenance management framework; models and methods for complex systems maintenance. Springer Verlag.
- Dekker, R. (1996). Applications of maintenance optimisation models: a review and analysis. *Reliability Engineering & System Safety*, 51(3), 229-240.
- Dekker, R., & Scarf, P. (1998). On the impact of optimisation models in maintenance decision making; the state of the art. *Reliability Engineering & System Safety*, 60(2), 111-119.

- Dekker, R., Wildeman, R., & Duyn Schouten, F. v. (1997). A review of multi-component maintenance models with economic dependence. *Mathematical Methods of Operations Research*, 45(3), 411-435.
- DiStefano, S., & Puliafito, A. (2008). Dependability evaluation with dynamic reliability block diagrams and dynamic fault trees. *IEEE Transactions on Dependable and Secure Computing*, 6(1), 4-17.
- Doyle, K. (2002). On the application of stochastic models in nuclear power plant maintenance. *European Journal of Operational Research*, 154(3), 673-690.
- Drucker, P. (1954). The practice of management. New York: Harper & Row.
- EFNMS; SMRP. (2011). Global maintenance and reliability indicators; fitting pieces together.
- Finetti, B. d. (2008). *Philosophical lectures on probability; collected, edited and annotated by Alberto Mura.* Springer Verlag.
- Foucher, B., Boullié, J., Meslet, B., & Das, D. (2002). A review of reliability prediction methods for electronic devices. *Microelectronics Reliability*, 42(8), 1155-1162.
- Frank, P., Ding, S., & Marcu, T. (2000). Model based fault diagnosis in technical processes. Transactions of the Institution of Measurement and Control, 22(1), 57-101.
- Gao, D., Wu, C., Zhang, B., & Ma, X. (2010). Signed Directed Graph and Qualitative Trend Analysis Based Fault Diagnosis in Chemical Industry. *Chinese Journal of Chemical Engineering*, 18(2), 265-276.
- Gauch, H. (2002). Scientific method in practice. Cambridge University Press.
- Gibbard, A., & Harper, W. (1981). Counterfactuals and two kinds of expected utility. In W. Harper, R. Stalnaker, & G. Pearce, *Ifs: conditionals, belief, decision, chance and time* (pp. 153-190). Dordrecht: Reidel Publishing Company.
- Gilb, T. (2006). Competitive engineering; a handbook for systems engineering, requirements engineering and software engineering using planguage. Elsevier.
- Gorjian, N., Ma, L., Mittinty, M., Yarlagadda, P., & Sun, Y. (2010). A review on reliability models in reliability analysis. In *Engineering asset lifecycle management* (pp. 369-384). London: Springer.
- Granger, C. (1980). Testing for causality, a personal viewpoint. *Journal of Economic Dynamics and Control*, 2, 329-352.
- Greenland, S. (2011). The logic and philosophy of causal inference: a statistical perspective. In P. Bandyopadhyay, & M. Forster, *Handbook of philosophy of science Volume 7: Philosophy of statistics* (pp. 813-830). Elsevier.
- Gruene, T. (2006). The problems of testing preference axioms with revealed preference theory. Analyse & Kritik, 26(2), 382-397.
- Haarman, M., & Delahay, G. (2004). Value Driven Maintenance; New faith in maintenance. Dordrecht: Mainnovation.
- Halpern, J. (2005). Reasoning about uncertainty. MIT-Press.
- Hammond, K., Steward, T., Brehmer, B., & Steinmann, D. (1975). Social judgment theory. In M. Kaplan, & S. Schwartz, *Human judgment and decision processes* (pp. 271-307). New York: Academic Press.
- Hitchcock, C. (2013). What is the "cause" in causal decision theory? Erkenntnis, 78(1), 129-146.
- Hogarth, R. (1980). Judgment and choice: the psychology of decision. New York: John Wiley & Sons.
- Horenbeek, A. v., Pintelon, L., & Muchiri, P. (2011). Maintenance optimization models and criteria. International Journal of System Assurance Engineering and Management, 1(3), 189-200.
- Hume, D. (1739). A Treatise of human nature; being an attempt to introduce the experimental method of reasoning into moral subjects.
- Hurvich, C., & Tsai, C. (1989). Regression and time series selection in small samples. *Biometrika*, 76(2), 297-307.
- IEC. (1990). *IEC 60050-191; International Electrotechnical Vocabulary, chapter 191: Dependability and quality of service.* Geneva: International Electrotechnical Commission.
- Isermann, R. (2005). Model based fault-detection and diagnosis status and applications. Annual Reviews in Control, 29(1), 71-85.
- Jardine, A., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical systems and signal processing*, 20(7), 1483-1510.
- Jeffrey, R. (1974). Frameworks for preference. In M. Balch, D. McFadden, & S. Wu, *Essays of economic behaviour under uncertainty* (pp. 74-79). Amsterdam: North Holland.
- Jones, J. (2007). Supportability engineering handbook: implementation measurement and management. Sole Press.

- Kahneman, D., & Tversky, A. (1979). Prospect theory: an analysis of decision under risk. *Econometrica*, 263-291.
- Kaplan, E., & Meier, P. (1958). Nonparametric estimation for incomplete observations. Journal of the American Statistical Association, 53(282), 457-481.
- Kaplan, R., & Norton, D. (1996). The balanced scorecard; translating strategy into action. Boston: Harvard Business School Press.
- Keeney, R., & Raiffa, H. (1976). *Decisions with multiple objectives: preference and value tradeoffs*. New York: John Wiley and Sons.
- Kelly, A. (2006). Strategic maintenance planning. Butterworth Heinemann.
- Kleer, J. d., & Brown, J. (1984). A qualitative physics based on confluences. *Artificial Intelligence*, 24(1-3), 7-83.
- Lakatos, I. (1976). *Proofs and refutations; the logic of mathematical discovery*. Cambridge: Cambridge University Press.
- Lewis, D. (1981). Causal decision theory. Australasian Journal of Philosophy, 59(1), 5-30.
- Lewis, D. (2000). Causation as influence. Journal of Philosophy, 97(4), 182-197.
- Lin, C., & Yang, M. (2009). Improved p-value tests for comparing two independent binomial proportions. *Communications in Statistics-Simulation and Computation*, 38(1), 78-91.
- Little, I. (1949). A reformulation of the theory of consumer's behavior. *Oxford Economic Papers*, 90-99. Little, J. (1961). A proof of the queuing formula L=IW. *Operations Research*, 9(3), 383-387.
- Lorenz, E. (1972). Predictability; does the flap of a butterfly's wings in Brazil set off a tornado in Texas? Annual Meeting of the American Association for the Advancement of Science. Washington DC.
- MacGregor, J., & Cinar, A. (2012). Monitoring, fault diagnosis, fault-tolerant control and optimization: data driven models. *Computers & Chemical Engineering*, 47, 111-120.
- Moubray, J. (2004). Reliability Centred Maintenance. Elsevier.
- Muchiri, P., Pintelon, L., Gelders, L., & Martin, H. (2011). Development of maintenance function performance measurement framework and indicators. *International Journal of production* economics, 131(1), 295-302.
- Nakajima, S. (1988). Introduction to TPM: Total Productive Maintenance (preventative maintenance series). Productivity Press.
- Neumann, J. v., & Morgenstern, O. (1944). *Games and economic behaviour*. Princeton: Princeton University Press.
- Neyman, J. (1923). On the application of probability theory to agricultural experiments. Essay on principles (with discussion). *Statistical Science*, *5*, 465-480.
- Nicolai, R., & Dekker, R. (2008). Optimal Maintenance of Multi-component systems: a review. In K. Kobbacy, & D. Murthy, *Complex System Maintenance Handbook* (pp. 263-286). London: Springer Verlag.
- Nowlan, F., & Heap, H. (1978). Reliability Centered Maintenance. Dolby Access Press.
- Nozick, R. (1969). Newcomb's problem and two principles of choice. In N. Rescher, *Essays in honor of Carl G. Hempel; a tribute on the occasion of his sixty fifth birthday* (pp. 114-146). Dordrecht: D. Reidel Publishing Company.
- NVDO. (2011). *NVDO onderhoudskompas; visie, trends en ontwikkelingen in de Nederlandse onderhoudsmarkt*. Amsterdam: Industrielings.
- Pareto, V. (transl. 1971). Manual of political economy. New York: Augustus M. Kelley Publishers.
- Parry, G. (1996). The characterisation of uncertainty in probabilistic risk assessments of complex systems. *Reliability Engineering & System Safety*, 54(2), 119-126.
- Pearl, J. (2000). Causality: models, reasoning, and inference. New-York: Cambridge University Press.
- Pearl, J. (2010). An introduction to causal inference. International Journal of Biostatistics, 6(2).
- Peng, Y., Dong, M., & Zuo, M. (2010). Current status of machine prognostics in condition based maintenance: a review. *International Journal of Advanced Manufacturing Technology*, 50(1-4), 297-313.
- Pham, H., & Wang, H. (1996). Imperfect Maintenance. *European Journal of Operational Research*, 94(3), 425-438.
- Popper, K. (2002). The logic of scientific discovery. London and New York: Routledge Classics.
- Puterman, M. (2005). *Markov Decision Processes; discrete stochastic dynamic programming*. Hoboken: Wiley Interscience.
- Rauzy, A. (2008). Binary Decision Diagrams for Reliability Studies. In K. Misra, Handbook of performability engineering (pp. 381-396). London: Springer.

- Rehage, D., & Carl, U. V. (2005). Redundancy management of fault tolerant aircraft system architecturesreliability synthesis and analysis of degraded system states. *Aerospace Science and Technology*, 9(4), 337-347.
- Remenyte-Prescott, R., & Andrews, J. (2009). An efficient real-time method of analysis for non-coherent fault trees. *Quality and Reliability Engineering International*, 25(2), 129-150.
- Rich, S., & Venkatasubramanian, V. (1987). Model-based reasoning in diagnostic expert systems for chemical process plants. *Computers & Chemical Engineering*, 11(2), 111-122.
- Robertson, S., & Robertson, J. (2006). Mastering the requirements process. Addison Wesley.
- Rubin, D. (1974). Estimating causal effects of treatments in randomized and non-randomized studies. *Journal of Educational Psychology*, 66(5), 688-701.
- Samuelson, P. (1938). A note on the pure theory of consumer's behaviour. *Economica*, 5(17), 61-71.
- Savage, L. (1954). The foundations of statistics. New York: John Wiley & Sons.
- Scarf, P. (1997). On the application of mathematical models in maintenance. European Journal of Operational Research, 99(3), 493-506.
- Schüller, J. (1997). *CPR12E. Methods for determining and processing probabilities.* SDU Uitgevers. Sen, A. (1973). Behaviour and the concept of preference. *Economica*, 40(159), 241-259.
- Sharma, A., Yadava, G., & Deshmukh, S. (2011). A literature review and future perspectives on maintenance optimization. *Journal of Quality in Maintenance Engineering*, 17(1), 5-25.
- Si, X., Wang, W., Hu, C., & Zhou, D. (2011). Remaining useful life estimation; a review on the statistical data driven approaches. *European Journal of Operational Research*, 213(1), 1-14.
- Sikorska, J., Hodkiewicz, M., & Ma, L. (2011). Prognostic modelling options for remaining useful life estimation by industry. *Mechanical Systems & Signal Processing*, 25(5), 1803-1836.
- Simon, H. (1955). A behavioral model of rational choice. The Quarterly Journal of Economics, 99-118.
- Simon, H. (1956). Rational choice and the structure of the environment. *Psychological Review*, 63(2), 129-138.
- Simon, H. (1978). Rationality as process and as product of thought. *The American Economic Review*, 68(2), 1-16.
- Simon, H. (1997). Administrative behavior; a study of decision making processes in administrative organizations. New York: The Free Press.
- Spirtes, P., Glymour, C., & Scheines, R. (2000). *Causation, Prediction and search*. Cambridge (Ma): The MIT Press.
- Stidham, S. (1974). A last word on L=lW. Operations Research, 22(2), 417-421.
- Suh, N. (2001). Axiomatic design; advances and applications. New York: Oxford University Press.
- Suppes, P. (1970). A probabilistic theory of causality. Amsterdam: North-Holland Publishing.
- Vaurio, J. (1998). An implicit method for incorporating common-cause failures in system analysis. *IEEE Transactions on Reliability*, 47(2), 173-180.
- Venkatasubramanian, V., Rengaswamy, R., & Kavuri, S. (2003). A review of process fault detection and diagnosis part II: Qualitative models and search strategies. *Computers & Chemical Engineering*, 27(3), 313-326.
- Venkatasubramanian, V., Rengaswamy, R., Kavuri, S., & Yin, K. (2003). A review of process fault detection and diagnosis part III: Process history based methods. *Computers & Chemical Engineering*, 27(3), 327-346.
- Venkatasubramanian, V., Rengaswamy, R., Yin, K., & Kavuri, S. (2003). A review of process fault isolation and diagnosis Part I: Quantitative model-based methods. *Computers & Chemical Engineering*, 27(3), 293-311.
- Vesely, W. (2002). Fault tree handbook with aerospace applications. Version 1.1. NASA office of safety and mission assurance.
- Villez, K., Rosen, C., Anctil, F., & Vanrolleghem, P. (2013). Qualitative Representation of trends (QRT): Extended method for identification of consecutive inflection points. *Computers & Chemical Engineering*, 48, 187-199.
- Walter, M., Siegle, M., & Bode, A. (2008). OpenSESAME- the simple but extensive, structured availability modeling environment. *Reliability Engineering & System Safety*, 93(6), 857-873.
- Wang, H. (2002). A survey of maintenance policies of deteriorating systems. European Journal of Operational Research, 139(3), 469-489.
- Wang, W. (2012). An overview of the recent advances in delay-time-based maintenance modelling. *Reliability Engineering & System Safety*, 106, 165-178.
- Weber, A., & Thomas, R. (2005). *Key performance indicators; measuring and managing the maintenance function*. Burlington, Ontario: Ivara corporation.
- Wright, S. (1921). Correlation and causation. Journal of Argicultural Research, 20(7), 557-585.

- Xu, H., Xing, L., & Robidoux, R. (2009). DRBD: Dynamic Reliability Block Diagrams for System Reliability Modelling. *International Journal of Computers and Applications*, *31*(2), 132-141.
- Yuge, T., & Yanagi, S. (2008). Quantitative analysis of a fault tree with priority AND gates. *Reliability Engineering & System Safety*, 93(111), 1577-1583.
- Zio, E. (2009). Reliability Engineering: old problems and new challenges. *Reliability Engineering & System Safety*, 94(2), 125-141.
- Zorriassatine, F., & Tannock, J. (1998). A review of neural networks for statistical process control. *Journal of Intelligent Manufacturing*, 9(3), 209-224.

List of publications

Conference contributions

- Rijsdijk, C. 2016. "Big data is not big enough to justify maintenance". *Marine maintenance world expo conference*. Amsterdam.
- Rijsdijk, C. and Tinga, T. 2016. "Predicting maintenance performance". 22nd HDO International conference. Sibenik.
- Rijsdijk, C. et.al. 2016. "Modelling failure behaviour; a case study". *IMA World Maintenance Forum*. Lugano.

Rijsdijk, C. 2015. "Enhance decision supportive capabilities of maintenance performance indicators". Global forum on total maintenance and operational excellence. Berlin.

Rijsdijk, C. 2015. "Data driven decision support". *Marine maintenance world expo conference*. Antwerp. Rijsdijk, C. and Tinga, T. 2015. "Enabling maintenance performance prediction by improving

performance indicators". *Proceedings of the 25th European safety and reliability conference*. Zürich. Rijsdijk, C. 2015. "Data driven decision support". *4th World congress of MIRCE mechanics*. Exeter.

- Rijsdijk, C. 2014. "Predicting from performance indicators". *Enterprise Asset Management 2014*. Copenhagen.
- Rijsdijk, C. and Tinga, T. 2014. "The effect of maintenance policy violations". *Euromaintenance conference proceedings*. Helsinki.
- Rijsdijk, C. 2013. "The effects of timely maintenance". *Conference proceedings of the 2nd world maintenance forum*. Lugano.
- Rijsdijk, C. 2013. "The effects of maintenance decision making". 2nd World congress of MIRCE mechanics. Exeter.
- Rijsdijk, C. 2013. "Availability predictions". Maintworld. Vol.1.
- Rijsdijk, C. 2012. "Practical availability predictions". 22nd MIRCE international symposium. Exeter.
- Rijsdijk, C. and Ghobbar, A. 2012. "An attempt to quantitatively underpin the rationality of maintenance support". *Air transport and operations; proceedings of the third international air transport and operations symposium.* Edited by R. Curran et al. IOS Press. Amsterdam.
- Rijsdijk, C. and Ghobbar, A. 2012. "A test of the importance of supportability". *Euromaintenance conference proceedings*. Edited by B. Vasic et.al. DOTS. Belgrade.
- Rijsdijk, C. and Ghobbar, A. 2011. "A quest for a dependency between equipment quality and supportability". *Maintenance 2011*. Edited by V. Cervenka. CSPU. Liblice.
- Rijsdijk, C. and Ghobbar, A. 2011. "Election of a hypothesis asserting that supportability implies quality". *Air transport and operations; proceedings of the second international air transport and operations symposium.* Edited by R. Curran et al. IOS-Press. Amsterdam.
- Rijsdijk, C. and Ghobbar, A. 2010. "A proposal for a supportability measure". *Proceedings conference day 2010*. Edited by H. Jodlbauer. Shaker Verlag. Aachen.
- Rijsdijk, C. and Ghobbar, A. 2010. "Measure and predict supportability". *Euromaintenance conference proceedings*. Edited by F. Santini. AIMAN. Milano.

Journal papers

Rijsdijk, C. and Tinga, T. 2016. "Observing the effect of a policy; a maintenance case". *Journal of Quality in Maintenance Engineering*. 22(3). 277-301. DOI 10.1108/JQME-10-2014-0055

