Safety stock determination in production systems with random yield and positive lead times

Schriftliche Promotionsleistung

zur Erlangung des akademischen Grades

Doctor rerum politicarum

vorgelegt und angenommen

an der Fakultät für Wirtschaftswissenschaft

der Otto-von-Guericke-Universität Magdeburg

Verfasser:	Danja R. Sonntag
Geburtsdatum und -ort:	22.11.1988, Eckernförde
Arbeit eingereicht am:	18.07.2017
Gutachter der schriftlichen Promotionsleistung:	Prof. Dr. Gudrun P. Kiesmüller, Otto-von-Guericke Universität Magdeburg
	Prof. Johan Marklund, PhD, Universität Lund, Schweden
Datum der Disputation:	28.11.2017

Contents

List of Figures

List of Tables

1. Introduction

In recent decades, the production of high-tech products has greatly increased. Within the European Union, the monetary value of high-tech products sold nearly doubled from 170 billion euros to 320 billion euros between 2003 and 2013 (Eurostat, 2014). In such a huge market, companies have to deal with numerous competitors, which makes it necessary to produce high quality products under low costs to guarantee low prices for customers and therefore achieve a competitive advantage. In an environment like the high-tech industry, producing high quality at low prices is not always an easy task. Samsung, for example, has problems producing the curved glass for a new cell phone series. They have to deal with yields of less than 50%, which means that more than half of all glasses manufactured cannot be used in cell phones (McNutt, 2015). Such low yields are the result of very complex production processes—in the Samsung example, the bending of the glass is very difficult—but there are vast differences among various industries and the companies.

One well-known example for random yield problems is the semiconductor industry and the corresponding microchip production. In this sector, the yields can vary tremendously during production and among companies. The chip yields, defined as the fraction of produced microchips which can be sold to customers, can vary between zero and 100% (Leachman & Hodges, 1996). The reasons for yield losses are manifold, as illustrated in Figure 1.1.

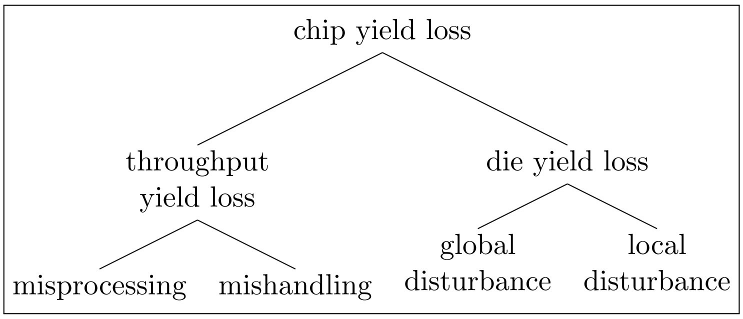


Figure 1.1: Error sources in microchip production (according to yieldWerx, 2016)

On one hand, yield losses can occur due to misprocessing by handlers, or mishandling of the wafers or equipment. Both error sources imply human error and are currently negligible because of automated production systems (yieldWerx, 2016). On the other hand, die yield losses can occur during the production of integrated circuits (ICs) on the wafers. Global disturbances affect all ICs on a wafer due to e.g. variations in the temperature, while local disturbances only affect small parts of the wafer because of e.g. missing connections between two circuit nodes (Khare & Maly, 2012, p. 2f). It is obvious that a global disturbance leads to a yield of zero, while a local disturbance can lead to different yields, depending on the degree of failure. This overview on failure sources in the semiconductor industry illustrates the complexity of production processes in the high-tech industry, which makes it difficult to predict failures.

Besides uncertainties according to the yield between several production runs, the yield can also differ from one company to another, depending on the quality of the production process. For example, Intel has a production yield in the high 90% range, while rivals have yields of 60% to 80% which is a great competitive advantage (Foremski, 2012).

Although many papers on random yield consider the semiconductor industry as an example, random yield occurs in several other industries, such as the agricultural sector as well as the chemical and pharmaceutical industries (Inderfurth & Clemens, 2014). In the agricultural sector the weather determines the quality of the crop. In addition, yield losses occur due to e.g. weeds, fungal diseases, insects or post-harvest losses, which were responsible for 50% of the global harvest losses in 2008 (Statista, 2016).

The example clarifies that the sources of failures are manifold and that, even though production processes have improved over time, failures cannot be avoided.

1.1. Modeling random yield

In the literature, several concepts are known to model yield uncertainty as described above. All the mentioned yield models are able to reflect different situations causing yield uncertainty. An overview can be found in Yano & Lee (1995) and Grosfeld-Nir & Gerchak (2004).

The most common yield model considers *binomial yield* (see e.g. Delfausse & Saltzman, 1966; Sepehri et al., 1986; Parlar & Wang, 1993; Pentico, 1994; Barad & Braha, 1996; Grosfeld-Nir et al., 2000; Sloan, 2004; Ben-Zvi & Grosfeld-Nir, 2007; Inderfurth & Vogelgesang, 2013). In a binomial yield model, each unit has a probability of p to be produced without errors. Thus, the number of good units in a lot of size n is binomially distributed. Binomial yield is appropriate in situations like thermal or chemical treatments e.g. coating, where the probability that one unit is defective is independent of the condition of all other units (Grosfeld-Nir & Gerchak, 2004).

A second common yield model considers *stochastic proportional yield* (see e.g. Bollapragada & Morton, 1999; Henig & Gerchak, 1990; Huh & Nagarajan, 2010; Inderfurth & Vogelgesang, 2013; Güler, 2015; Inderfurth & Kiesmüller, 2015). Stochastic proportional yield is used if a certain quantity of items in a batch is defective due to a limited ability to react to environmental changes or variations in materials (Yano & Lee, 1995). In this case, the output $Y(Q)$ of the production process equals a random fraction $Z \in [0, 1]$ of the input Q: $Y(Q) = Z \cdot Q$ (Shih, 1980).

Another way of modeling yield uncertainty is the use of an *interrupted geometric yield* (see e.g. Zhang & Guu, 1998; Guu & Liou, 1999; Guu & Zhang, 2003; Inderfurth & Vogelgesang, 2013) where a constant probability θ exists that the process falls to an out-of-control state during production. As long as the process is in an in-control state, only error-free products are produced. Once the system falls into an out-of-control state, from that point on, all produced items are defective (Zhang & Guu, 1998). Such a situation occurs when e.g. manufacturing equipment damages due to attrition. In such a production environment, either a preventive maintenance policy before machine breakdown or a corrective maintenance strategy, once the machine falls into an out-of-control state, is required.

The simplest yield model considers *all-or-nothing yield* (see e.g. Tomlin, 2009; Grosfeld-Nir & Gerchak, 2004). All-or-nothing yield leads to only good units in a production batch with probability κ while all items are defective with probability $1 - \kappa$. Such a yield is reasonable in situations of natural strikes, terrorism or disasters e.g. fires, earthquakes or flooding (Tomlin, 2009).

The last common yield model uses *discrete uniform yield* (see e.g. Anily, 1995; Grosfeld-Nir et al., 2000; Inderfurth, 2003), assuming that the quantity of good items is discrete and uniformly distributed. This yield model is used basically in theory, due to its simplicity (Anily, 1995).

Besides these common yield models, others can be found (see Grosfeld-Nir & Gerchak, 1990; Parlar et al., 1995) but will not be discussed in the following.

In all the papers presented below, a stochastic proportional yield model is considered. Interrupted geometric yields as well as all-or-nothing or discrete uniform yields are not suitable for production systems similar to the microchip production. For such production systems, binomial and stochastic proportional yields are applicable, both of which were used by Inderfurth & Vogelgesang (2013) and Inderfurth & Kiesmüller (2015). The main advantage of stochastic proportional yield over binomial yield is that, under stochastic proportional yield, higher yield variances and therefore a higher uncertainty concerning the production output can be modeled. In production systems where the production output can vary between zero and 100% as mentioned by Leachman & Hodges (1996), a high yield variability exists and therefore stochastic proportional yield is more appropriate for modelling such situations.

1.2. Model description

In this work, the focus is on make-to-stock production systems. The production process is followed by a quality control, i.e. the inspection of all produced items in negligible time. The products satisfying the required quality standards are stocked in a warehouse to satisfy stochastic customer demand. Thus, the inspection ensures that only high quality items are sold to customers to guarantee high customer satisfaction. As illustrated in Figure 1.2, defective products - defined as those products not satisfying the quality standards - can be handled in several ways. In the following chapters, defective products are primarily disposed of, which means that they leave the production process (see Chapters 2, 3 and 4). In Chapter 5, defective products can instead be reworked. It is assumed that the rework process can be processed parallel to the production process. After rework, these products are assumed to be in the same condition as immediately perfectly produced items and therefore have the same price. It is also conceivable that defective products are sold as lower quality products at a lower price (see e.g. Gerchak et al., 1996; Hsu & Bassok, 1999) or otherwise utilized. However, these options are not considered in the presented papers.

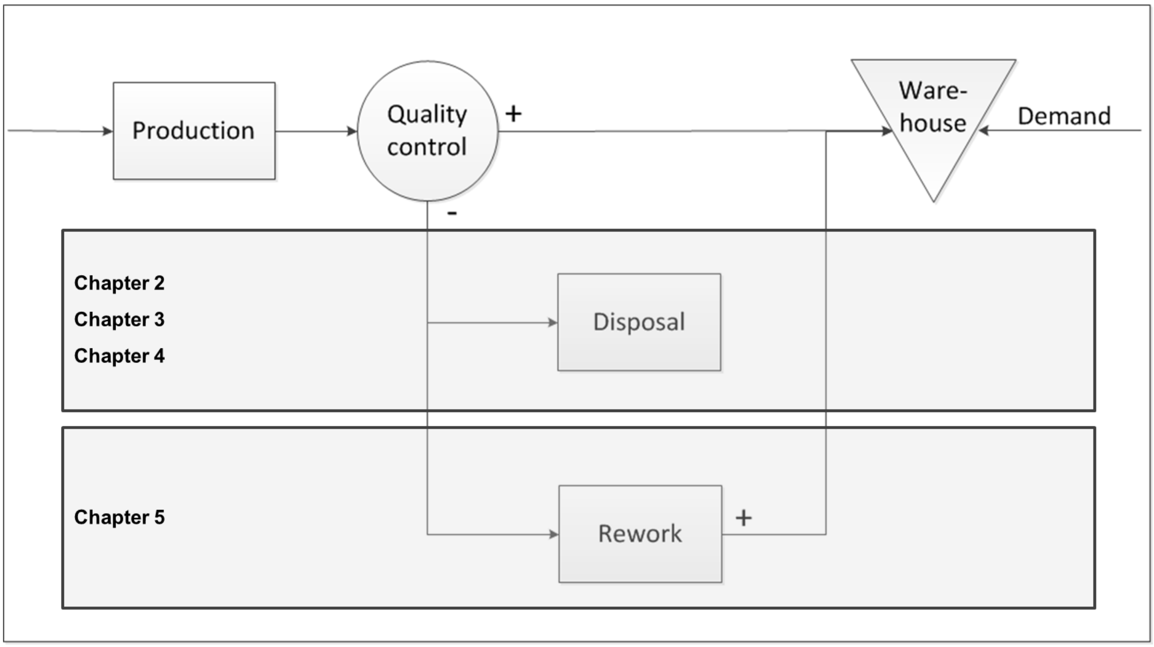


Figure 1.2: Overview of Chapters: Make-to-stock production system with/without rework

To be able to satisfy stochastic customer demand and thus reduce the probability of stock-outs, a planner must ensure that sufficient stock is available. To replenish stock, orders can be placed at the beginning of each period after the delivery of previous orders. Each order initiates the start of an in-house production run. The difficulty is to determine the optimal order quantity, while minimizing average holding and backorder costs per period. Holding costs are charged at the end of a period for each unit in stock whereas backorder costs are charged for each unit backlogged.

In a production environment with random yield, the optimal ordering policy does not have a simple structure (Henig & Gerchak, 1990). On one hand, stochastic demand leads to uncertainty regarding outward stock movements. On the other hand, stochastic supply represented by random yields leads to uncertainty regarding inward stock movements. Huh & Nagarajan (2010) and Bollapragada & Morton (1999) showed that what is called a linear inflation policy performs extremely well under zero lead times. In the case of a linear inflation policy, the order quantity Q_t in the period t equals

$$Q_t = \begin{cases} F(S - X_t) & , \quad X_t < S \\ 0 & , \quad X_t \geq S \end{cases} \tag{1.1}$$

where F is the inflation factor, often defined as the reciprocal of the mean yield (see e.g. Bollapragada & Morton, 1999; Huh & Nagarajan, 2010; Inderfurth & Vogelgesang, 2013; Inderfurth & Kiesmüller, 2015). The inflation factor compensates for expected yield losses during the production process. Note that while the inflation factor takes into consideration average yield losses, variability represented by the variance of the yield is neglected. The order quantity equals the inflated difference between the base stock level S and the actual inventory position X_t if this difference is positive, and zero otherwise. The base stock level equals the mean demand during the lead time plus some safety stock required due to the uncertainty in the system because of demand and yield randomness.

The difficulty in determining the order quantity arises from the definition of the inventory position which is not straightforward under random yield. While the inventory position in a situation without random yield is defined as the stock on hand minus backorders plus all outstanding orders, it changes if not all outstanding orders are delivered to the warehouse due to imperfect production processes. Under random yield, the quantity of good items is uncertain, because a planner does not know how many items will satisfy the quality requirements. Therefore, the inventory position contains only an expectation of items being delivered in future periods. Once the exact quantity of good items and thus the realized yield is known, a planner can update the inventory position. This update is reflected by a "forecast error", which equals the difference between the expected production outcome of an order and its realization. These forecast errors do not exist in the case of zero production times, because ordered items arrive in the warehouse immediately. Therefore, the production outcome can be observed instantaneously and thus the planner can react to higher or lower amounts of perfect items by increasing or decreasing the order quantity of the subsequent order. Because no forecast errors have to be considered for production systems with zero production times, they are much easier to solve and receive more attention in the literature (see e.g. Bollapragada & Morton, 1999; Huh & Nagarajan, 2010; Zipkin, 2000). The focus of this work is on production systems with positive production times.

1.3. Contribution

Table 1.1 categorizes the literature on production systems with random yield regarding two crucial assumptions: the number of production stages and the length of the production time of each stage. It is obvious that a variety of literature exists on make-to-order production systems, in particular with zero production times. All this literature mainly focuses on the question how to set the lot size and thus the number of production runs when yield is random and a given demand quantity has to be fulfilled in total (rigid demand). The authors focus either on optimal or heuristic solution approaches under different yield models and different settings of the model (e.g. different cost structure, rework, downward substitution).

	zero lead time	**positive lead time**
	make-to-order	make-to-order
	Delfausse & Saltzman (1966)	Wang & Gerchak (2000)
	Sepehri et al. (1986)	Hsu et al. (2009)
	Grosfeld-Nir & Gerchak (1990)	
	Anily (1995)	make-to-stock
	Grosfeld-Nir & Gerchak (1996)	(Bollapragada & Morton (1999))
	Zhang & Guu (1998)	Gotzel & Inderfurth (2005)
	Guu & Liou (1999)	Inderfurth & Vogelgesang (2013)
	Grosfeld-Nir et al. (2000)	Inderfurth & Kiesmüller (2015)
single-stage	Guu & Zhang (2003)	**Sonntag & Kiesmüller (2016)**
	Sloan (2004)	**Sonntag & Kiesmüller (in press)**
	Grosfeld-Nir & Gerchak (2004)	
	...	
	make-to-stock	
	Henig & Gerchak (1990)	
	Yano & Lee (1995)	
	Bollapragada & Morton (1999)	
	Zipkin (2000)	
	Huh & Nagarajan (2010)	
	make-to-order	make-to-stock
	Vachanf (1970)	Choi et al. (2008)
	Lee & Yano (1988)	Dettenbach & Thonemann (2015)
	Wein (1992)	**Sonntag & Kiesmüller (2017)**
	Grosfeld-Nir & Ronen (1993)	**Sonntag (2017)**
multi-stage	Pentico (1994)	
	Barad & Braha (1996)	
	Grosfeld-Nir & Gerchak (2002)	
	Grosfeld-Nir & Gerchak (2004)	
	Ben-Zvi & Grosfeld-Nir (2007)	
	...	

Table 1.1: Embedding in existing literature

While a lot of literature on make-to-order production systems under random yield exists, the literature on make-to-stock systems under random yield is very limited and mainly focuses on single-stage systems because they are easier to solve. The contribution of this work is to give solution approaches for imperfect make-to-stock production systems with positive production times.

Examining the existing literature, Bollapragada & Morton (1999) consider random yield problems in particular with zero production times and give only a very brief overview of how to deal with positive production times. Nevertheless, like Inderfurth & Vogelgesang (2013) and Inderfurth & Kiesmüller (2015), they assume that the forecast error can be approximated by a normal distribution due to the central limit theorem (Bollapragada & Morton, 1999). According to Sonntag & Kiesmüller (2016), as presented in Chapter 2, we showed that the assumption of a normally distributed forecast error is not suitable in case of skewed yield. Because, except for Gotzel & Inderfurth (2005), all papers on make-to-stock systems assume that defective items are disposed of, Sonntag & Kiesmüller (in press) consider that defective items are reworked. In addition to having a slightly different model compared to Gotzel & Inderfurth (2005), we also present a different definition of the inventory position, which includes only the relevant information for the actual decision. Beyond that, we also present a solution approach which is very fast, compared to the stochastic dynamic programming of Gotzel & Inderfurth (2005), even for larger problem settings.

The literature on multi-stage make-to-stock production systems with positive production times is even more limited. The only approaches come from Choi et al. (2008) and Dettenbach & Thonemann (2015), who use simulation and stochastic dynamic programming to solve the problem. They are only able to solve small problem sizes, due to these solution approaches. The approach in Sonntag & Kiesmüller (2017) is the first one able to deal with large multi-stage production systems. While Sonntag & Kiesmüller (2017) focus on the question of where to locate intermediate quality inspections to reduce the safety stock level, Sonntag (2017) considers one further option to reduce the safety stock level: under a given budget constraint either additional quality inspections can be introduced or the yield variability can be reduced. Thus, the question is how to invest a given budget to reduce the safety stock.

In the following chapter, a more detailed overview is presented of the papers which form part of this work.

1.4. Bibliography

As already mentioned, the present work considers different problem settings regarding make-to-stock production systems with random demand and random supply and positive production times. An overview is given in Table 1.2.

Chapter	Stages	Defectives	Status	Title
Chapter 2	single-stage	disposed of	published	Sonntag, D., Kiesmüller, G. P. (2016). The shape of the yield and its impact on inventory decisions. 4OR, 14, 405-415.
Chapter 3	multi-stage	disposed of	published	Sonntag, D., Kiesmüller, G. P. (2017). The Influence of Quality Inspections on the Optimal Safety Stock Level. Production and Operations Management, doi: 10.1111/poms.12691.
Chapter 4	multi-stage	disposed of	published	Sonntag, D. (2017). Investing into quality inspections or reducing the yield variability?. Proceedings of the Logistikmanagement 2017, 258-266.
Chapter 5	single-stage	reworked	in press	Sonntag, D., Kiesmüller, G. P. (in press). Disposal versus Rework - Inventory control in a production system with random yield. Accepted for publication in: European Journal of Operational Research.

Table 1.2: List of publications

Although the model is similar in all the papers, each paper considers different research questions which are described in detail in the following text.

Chapter 2 focus on a single-stage production system with positive production times. As already mentioned, the inventory position includes the estimated quantity of goods to be delivered, which leads to a forecast error. Literature focusing on random yield problems with positive production times usually assume that this forecast error is normally distributed (see e.g. Bollapragada & Morton, 1999; Inderfurth & Vogelgesang, 2013). In the presented paper, it is shown that, for skewed yield, the assumption of a normally distributed forecast error leads to poor results. Therefore, the forecast error is approximated by a skewed normal or a generalized extreme value distribution and a Markov chain approach is used to determine the optimal base stock levels. The numerical study reveals that the proposed approaches are excellent and outperform existing ones.

In Chapter 3 the single-stage production system is extended to a multi-stage production system. In such a production system, earlier information about realized yields has a positive effect on cost, because uncertainty about the production outcome decreases. Specifically, production costs as well as holding costs can be reduced by introducing quality inspections, discarding defective items before further production. To achieve the greatest cost savings, it is important to determine the optimal number and positions of these inspections across the production process which, due to several influencing parameters, is not simple. It is shown how the positions of inspection within a production process influence the safety stock level by deriving analytical formulas. The approach is the first to combine decisions about the number and positions of inspections with inventory control strategies in a warehouse. A maximum safety stock reduction of more than 30% can be achieved in the presented examples, which may be even larger depending on the parameter setting. This allows for significant savings for a company like Intel, which reported inventories for finished goods of nearly 1.5 billion dollars in the 2014 Annual Report.

Chapter 4 considers a similar model for the production process to that in Chapter 3, but focuses on the question of how to reduce the required safety stock level and therefore holding cost, resulting in a new optimization model. On the production site, a planner has two opportunities of reducing uncertainty about the production output and therefore the safety stock level. He can implement intermediate quality inspections as in Chapter 3 and/or reduce the yield variance under a given budget constraint. Thus, a planner has to optimize the number and locations of additional inspections, as well as the amount of money spent on each production stage, in order to reduce the yield uncertainty. The problem is solved using a fix-and-optimize solution approach. The results indicate that, depending on the extent of yield variability, either implementing additional inspections or reducing the yield variance is more favorable. Moreover, significant safety stock reductions can be recognized, depending on the current yield variability.

In a production environment where random yield plays a significant role, a decision has to be made how to handle products which do not satisfy the given quality requirements. While, in Chapter 2 to 4 it is assumed that defective items are disposed of, Chapter 5 compares two different strategies for handling defective items: disposal or rework. It is important to note that a planner can make the strategic decision between disposal and rework only once and cannot change it within the planning horizon. It is shown how to determine the optimal base stock level. This is very difficult because of unknown correlations between orders. Subsequently, an optimization model is proposed to support the decision of a planner whether to dispose of or rework defective items. The parameters which directly affect this decision are analyzed within a sensitivity analysis. The analysis indicates that significant cost reductions can be obtained by choosing the best strategy for defective products.

References

Anily, S.. 1995. Single-machine lot-sizing with uniform yields and rigid demands: robustness of the optimal solution. IIE transactions, 27, 634–637.

Barad, M., D. Braha. 1996. Control limits for multi-stage manufacturing processes with binomial yield (single and multiple production runs). Journal of the Operational Research Society, 47, 98–112.

Ben-Zvi, T., A. Grosfeld-Nir. 2007. Serial production systems with random yield and rigid demand: a heuristic. Operations Research Letters, 35, 235–244.

Bollapragada, S., T. E. Morton. 1999. Myopic heuristics for the random yield problem. Operations Research, 47, 713–722.

Choi, H.-c. P., J. D. Blocher, S. Gavirneni. 2008. Value of sharing production yield information in a serial supply chain. Production and Operations Management, 17, 614–625.

Delfausse, J., S. Saltzman. 1966. Values for optimum reject allowances. Naval Research Logistics (NRL), 13, 147–157.

Dettenbach, M., U. W. Thonemann. 2015. The value of real time yield information in multi-stage inventory systems–exact and heuristic approaches. European Journal of Operational Research, 240, 72–83.

Eurostat. 2014. Evolution of total sold production of high-tech products, EU-28, 2003-2013. `http://ec.europa.eu/eurostat/statistics-explained/index.php/File:Evolution_of_total_sold_production_of_high-tech_products,_EU-28,_2003%E2%80%9313.png`. [Online; accessed 25-November-2016].

Foremski, T.. 2012. Intel's battle with ARM is about making its future fabs viable. `http://www.siliconvalleywatcher.com/mt/archives/2012/09/intels_battle_w.php`. [Online; accessed 24-November-2016].

Gerchak, Y., A. Tripathy, K. Wang. 1996. Co-production models with random functionality yields. IIE transactions, 28, 391–403.

Gotzel, C., K. Inderfurth. 2005. Approximate policies for hybrid production and rework systems with stochastic demand and yield. In Operations Research Proceedings 2004 (pp. 41–49). Springer.

Grosfeld-Nir, A., Y. Gerchak. 1990. Multiple lotsizing with random common-cause yield and rigid demand. Operations Research Letters, 9, 383–387.

Grosfeld-Nir, A., Y. Gerchak. 1996. Production to order with random yields: single-stage multiple lotsizing. IIE transactions, 28, 669–676.

Grosfeld-Nir, A., Y. Gerchak. 2002. Multistage production to order with rework capability. Management Science, 48, 652–664.

Grosfeld-Nir, A., Y. Gerchak. 2004. Multiple lotsizing in production to order with random yields: Review of recent advances. Annals of Operations Research, 126, 43–69.

Grosfeld-Nir, A., Y. Gerchak, Q.-M. He. 2000. Manufacturing to order with random yield and costly inspection. Operations research, 48, 761–767.

Grosfeld-Nir, A., B. Ronen. 1993. A single bottleneck system with binomial yields and rigid demand. Management Science, 39, 650–653.

Güler, M. G.. 2015. Coordinating decentralised assembly systems with random yield and random demand. International Journal of Production Research, 53, 886–896.

Guu, S.-M., F.-R. Liou. 1999. An algorithm for the multiple lot sizing problem with rigid demand and interrupted geometric yield. Journal of mathematical analysis and applications, 234, 567–579.

Guu, S.-M., A. X. Zhang. 2003. The finite multiple lot sizing problem with interrupted geometric yield and holding costs. European Journal of Operational Research, 145, 635–644.

Henig, M., Y. Gerchak. 1990. The structure of periodic review policies in the presence of random yield. Operations Research, 38, 634–643.

Hsu, A., Y. Bassok. 1999. Random yield and random demand in a production system with downward substitution. Operations Research, 47, 277–290.

Hsu, H.-M., T.-S. Su, M.-C. Wu, L.-C. Huang. 2009. Multiple lot-sizing decisions with an interrupted geometric yield and variable production time. Computers & Industrial Engineering, 57, 699–706.

Huh, W. T., M. Nagarajan. 2010. Technical note-linear inflation rules for the random yield problem: Analysis and computations. Operations Research, 58, 244–251.

Inderfurth, K.. 2003. Analytical Solution for a Single Period Production Inventory Problem with Uniformly Distributed Yield and Demand. FEMM, Faculty of Economics and Management.

Inderfurth, K., J. Clemens. 2014. Supply chain coordination by risk sharing contracts under random production yield and deterministic demand. OR spectrum, 36, 525–556.

Inderfurth, K., G. P. Kiesmüller. 2015. Exact and heuristic linear-inflation policies for an inventory model with random yield and arbitrary lead times. European Journal of Operational Research, 245, 109–120.

Inderfurth, K., S. Vogelgesang. 2013. Concepts for safty stock determination under stochastic demand and different types of random production yield. European Journal of Operational Research, 224, 293–301.

Khare, J. B., W. Maly. 2012. From contamination to defects, faults and yield loss: simulation and applications volume 5. Springer Science & Business Media.

Leachman, R. C., D. A. Hodges. 1996. Benchmarking semiconductor manufacturing. IEEE Transactions on Semiconductor Manufacturing, 9, 158–169.

Lee, H. L., C. A. Yano. 1988. Production control in multistage systems with variable yield losses. Operations Research, 36, 269–278.

McNutt, C.. 2015. Rumor: Low yield 3D glass forcing Samsung to look for additional suppliers for Galaxy S6 Edge. `http://www.androidheadlines.com/2015/03/rumor-low-yield-3d-glass-forcing-samsung-look-additional-suppliers-galaxy-s6-edge.html`. [Online; accessed 24-November-2016].

Parlar, M., D. Wang. 1993. Diversification under yield randomness in inventory models. European Journal of Operational Research, 66, 52–64.

Parlar, M., Y. Wang, Y. Gerchak. 1995. A periodic review inventory model with markovian supply availability. International Journal of Production Economics, 42, 131–136.

Pentico, D.. 1994. Multistage production systems with random yield: heuristics and optimality. The International Journal of Production Research, 32, 2455–2462.

Sepehri, M., E. A. Silver, C. New. 1986. A heuristic for multiple lot sizing for an order under variable yield. IIE transactions, 18, 63–69.

Shih, W.. 1980. Optimal inventory policies when stockouts result from defective products. International Journal of Production Research, 18, 677–686.

Sloan, T.. 2004. A periodic review production and maintenance model with random demand, deteriorating equipment, and binomial yield. Journal of the Operational Research Society, 55, 647–656.

Sonntag, D.. 2017. Investing into quality inspections or reducing the yield variability? Proceedings of the Logistikmanagement 2017, 258-266.

Sonntag, D., G. P. Kiesmüller. 2016. The shape of the yield and its impact on inventory decisions. 4OR, 14, 405–415.

Sonntag, D., G. P. Kiesmüller. 2017. The influence of quality inspections on the optimal safety stock level. Production and Operations Management, 26, 1284–1298.

Sonntag, D., G. P. Kiesmüller. in press. Disposal versus rework - inventory control in a production system with random yield. Accepted for publication in: EJOR.

Statista. 2016. Ernteverluste durch Schadorganismen in der Landwirtschaft weltweit im Jahr 2008. `http://de.statista.com/statistik/daten/studie/165619/umfrage/ernteverlust-durch-schadorganismen/`. [Online; accessed 24-November-2016].

Tomlin, B.. 2009. Impact of supply learning when suppliers are unreliable. Manufacturing & Service Operations Management, 11, 192–209.

Vachanf, M.. 1970. Determining optimum reject allowances for multistage job-lot manufacturing. AIIE Transactions, 2, 70–77.

Wang, Y., Y. Gerchak. 2000. Input control in a batch production system with lead times, due dates and random yields. European Journal of Operational Research, 126, 371–385.

Wein, A. S.. 1992. Random yield, rework and scrap in a multistage batch manufacturing environment. Operations Research, 40, 551–563.

Yano, C. A., H. L. Lee. 1995. Lot sizing with random yields: A review. Operations Research, 43, 311–334.

yieldWerx. 2016. Conducting yield analysis for semiconductor manufacturing. `http://www.yieldwerx.com/conducting-yield-analysis-semiconductor-manufacturing/`. [Online; accessed 25-November-2016].

Zhang, A. X., S.-m. Guu. 1998. The multiple lot sizing problem with rigid demand and interrupted geometric yield. IIE transactions, 30, 427–431.

Zipkin, P. H.. 2000. Foundations of inventory management.

2. The shape of the yield and its impact on inventory decisions

Sonntag, D., Kiesmüller, G. P. (2016). The shape of the yield and its impact on inventory decisions. 4OR, 14, 405-415.

4OR-Q J Oper Res (2016) 14:405–415
DOI 10.1007/s10288-016-0317-z

RESEARCH PAPER

The shape of the yield and its impact on inventory decisions

Danja Sonntag[1] · **Gudrun P. Kiesmüller**[1]

Received: 8 September 2015 / Revised: 18 April 2016 / Published online: 30 April 2016

Abstract We consider an inventory model with stochastic demand, positive lead time and random yield where ordering decisions are made according to a linear inflation rule. In case of a positive lead times the complexity of such inventory systems increases distinctly. Due to positive lead times, the inventory position contains no longer a term for outstanding orders but the estimated quantity of goods to be delivered after a known positive lead time period, which differ from the realized deliveries. Thus, a forecast error occurs in each period. In previous research this forecast error was assumed to be normally distributed which is not an appropriate assumption in case of symmetric yield. Since yield skewness can't be neglected, we propose to fit a skew normal distribution or a generalized extreme value distribution on the forecast error to account for the yield skewness. A numerical study reveals that the proposed approaches are excellent and outperform existing ones.

Keywords Stochastic inventory model · Random yield · Positive lead time · Yield skewness

Mathematics Subject Classification 90B05 · 60J10

1 Introduction

In many inventory models, e.g. the classical newsvendor model, uncertain demand is assumed. However, these models in the basic form suppose that ordered items are of

✉ Danja Sonntag
danja.sonntag@ovgu.de

1 Chair of Operations Management, Faculty of Economics and Management,
Otto-von-Guericke-University Magdeburg, P.O. Box 4120, 39108 Magdeburg, Germany

perfect quality and that the quantity delivered is equal to the quantity ordered. Thus, the second source of uncertainty—supply quantity uncertainty—is not considered despite the fact that in many industries companies have to deal not only with random demand but also with random supply. The agricultural sector and the related industries for example are to some extent substantially dependent on the not influenceable factor weather (He and Zhang 2008). Furthermore, in electronic fabrication and chemical processes random production yield plays an important role (Yano and Lee 1995).

Beyond a system with random yield we consider a system with positive lead time, which is according to our knowledge only studied by Bollapragada and Morton (1999), Inderfurth and Kiesmüller (2015) and Inderfurth and Vogelgesang (2013). Positive lead times increase complexity distinctly since the information about outstanding orders is uncertain. Thus, the inventory position, which is the basis for the ordering decision, contains a term for the expected yield. Whenever an order is delivered, the inventory position needs to be updated by the difference of the expected yield and the realized yield. Therefore, in every period there exists a forecasting error. In previous literature, Bollapragada and Morton (1999) and Inderfurth and Kiesmüller (2015) as well as Inderfurth and Vogelgesang (2013) assumed that the forecast error is approximately normally distributed due to the central limit theorem (Bollapragada and Morton 1999). Numerical studies show an excellent performance for symmetric demand and yield distributions. However, a symmetric yield is very unlikely especially for a high mean yield, as for example in the agricultural sector (Hennessy 2009). As can be seen in Inderfurth and Kiesmüller (2015), the consideration of an asymmetric yield on the one hand but a symmetric forecast error on the other hand leads to cost 20 % above optimum under positive lead times. This effect is unsurprising bearing in mind that the forecast error somehow depends on the yield. Nonetheless, it is neglected in the literature so far.

We will show that the approximation of the forecast error with a normal distribution fails in case of a skewed yield and investigate two different approaches, able to deal with yield skewness. We suggest fitting a generalized extreme value distribution and a skew normal distribution. Both approximations show excellent results.

The rest of the paper is organized as follows. In Sect. 2, we formulate the mathematical model and outline the markov chain approach to calculate the optimal base stock level. In Sect. 3, we explain the concept of the forecast error in detail as well as the already mentioned generalized extreme value distribution and the skew normal distribution to model the forecast error. Section 4 provides the results of the numerical study followed by a summary and outlook in Sect. 5.

2 Model formulation

We consider a periodic review inventory model where an order of size Q can be placed at the beginning of each period after the delivery of the order placed λ periods before. Subsequently, demand D occurs whereby demand which cannot be satisfied directly from stock is backlogged. We assume that demand is discrete and stochastic with constant parameters over time, and independent and identically distributed (iid) across the periods. At the end of each period inventory holding and backorder costs are charged. Due to e.g. production risks the delivered quantity does not necessarily equal the quantity ordered. Under the assumption of a stochastic proportional yield model the yield is a random multiple of the input, $Y(Q_t) = Z_t \cdot Q_t$, with mean $E[Y(Q_t)] = \mu_Z Q_t$ and variance $VAR[Y(Q_t)] = \sigma_Z^2 Q_t^2$. $Z_t \in [0, 1]$ is denoted as the yield factor in period t with mean μ_Z and variance σ_Z^2. We assume that $(Z_t)_{t=1,2,\ldots}$ is iid and independent on the demand distribution. The lead time λ is deterministic, positive ($\lambda > 1$), and constant over time.

When placing an order, the random yield has to be taken into account. Since the optimal ordering policy has not a simple structure (Henig and Gerchak 1990), a linear inflation policy, which has shown to perform very well (Huh and Nagarajan 2010), is considered. Thus, the order quantity Q_t in period t is determined as

$$Q_t = \begin{cases} F(S - X_t), & X_t < S \\ 0, & X_t \geq S \end{cases} \tag{1}$$

with the yield inflation factor F, the critical stock level S and the inventory position X_t at the beginning of the period before an order is placed. Since only discrete values for the order quantity are permitted, it is rounded to the closest integer value using the following notation: $Q_t = \lfloor F(S - X_t) \rceil$. The inventory position is defined as

$$X_t = I_{t-1} + Y(Q_{t-\lambda}) + \sum_{l=1}^{\lambda-1} \mu_Z Q_{t-l} \tag{2}$$

with the inventory level I_{t-1} at the end of period $t-1$, the delivered quantity $Y(Q_{t-\lambda})$ in period t, and the expected outstanding deliveries, $\sum_{l=1}^{\lambda-1} \mu_Z Q_{t-l}$. We can rewrite Eq. (2) to the following recursive equation for the inventory position in period $t+1$:

$$X_{t+1} = X_t + \mu_Z Q_t - D_t - (\mu_Z Q_{t+1-\lambda} - Y(Q_{t+1-\lambda})). \tag{3}$$

The last term in brackets equals the difference between the expected yield and the actual yield, and can be interpreted as a forecast error. We define the forecast error $R_{t+1-\lambda}$ as follows:

$$R_{t+1-\lambda} = \mu_Z Q_{t+1-\lambda} - Y(Q_{t+1-\lambda}) \tag{4}$$

In order to compute a near optimal critical stock level we model the system as a markov chain similar as in Inderfurth and Kiesmüller (2015). Note, while modelling the system as a markov chain, we ignore effects from previous periods in X_t. Thus, the markov model can only approximate the real system. However, in Inderfurth and Kiesmüller (2015) it is shown that this approximation works well. States are defined as $\Delta_t = X_t - S$, which is positive if the actual yield exceeds the expected yield and negative otherwise. With this definition we get from (1) and (3):

$$\Delta_{t+1} = \begin{cases} \Delta_t + Z \cdot F(-\Delta_t) - (D_t + R_{t+1-\lambda}), & \Delta_t < 0 \\ \Delta_t - (D_t + R_{t+1-\lambda}), & \Delta_t \geq 0 \end{cases} \tag{5}$$

It is obvious that the distribution of the forecast error $R_{t+1-\lambda}$ has an impact on the calculations of the markov chain which clarifies the importance of a good approximation. The markov chain approach is used to determine the stationary distribution of Δ_t and X_t. This information is needed to compute the distribution of the inventory level, $I_{t+\lambda}$, which equals

$$I_{t+\lambda} = X_t + Z \cdot Q_t - \sum_{l=0}^{\lambda-1} R_{t-l} - \sum_{l=0}^{\lambda} D_{t+l}. \tag{6}$$

The forecast error can also be interpreted as a second demand process where demand can be positive as well as negative. Adding this process to the customer demand leads to an adapted demand process, which is defined for general lead times as

$$\eta(\lambda+1) := \sum_{l=0}^{\lambda} D_{t+l} + \sum_{l=0}^{\lambda-1} R_{t-l}. \tag{7}$$

It can be seen that the distribution of the forecast error has again an influence on the computations. Using the information about the stationary distribution of the inventory position and the adapted demand process, a nearly optimal critical stock level S can be determined. For this purpose, the newsvendor like condition in (8) is used. The smallest value of S satisfying this inequality for a given yield factor F minimizes the long run average costs of the system (we refer to Inderfurth and Kiesmüller (2015) for details):

$$\sum_{k=-\infty}^{S-1} P\Big(\eta(\lambda+1) \leq k + \lfloor Z \cdot \lfloor F(S-k) \rceil \rceil\Big) v_k + \sum_{k=S}^{+\infty} P\Big(\eta(\lambda+1) \leq k\Big) v_k \geq \frac{b}{b+h} \tag{8}$$

v_k equals the stationary distribution $v_k = lim_{t\to\infty} P(\Delta_t = k)$.

3 The forecast error

For the newsvendor like condition (8) we need the distribution of the adapted demand process which means, we have to derive the distribution of the sum of the forecast errors, respectively. In Inderfurth and Kiesmüller (2015) and Inderfurth and Vogelgesang (2013) the distribution of the forecast error is approximated by a normal distribution as well as in Bollapragada and Morton (1999), where it is argued that if lead times are sufficiently long the central limit theorem justifies this assumption. Figure 1 shows the frequency distribution of the forecast error for a simulation of the inventory system in ARENA 12.0 (Kelton et al. 1998) for one million periods with the following input parameters: a normally distributed demand with mean $\mu_D = 20$ and coefficient of variation of $\rho_D = 0.1$, a beta distributed unsymmetrical yield with mean $\mu_Z = 0.85$ and coefficient of variation of $\rho_Z = 0.2$, a critical ratio of 0.995 and a lead time of $\lambda = 2$.

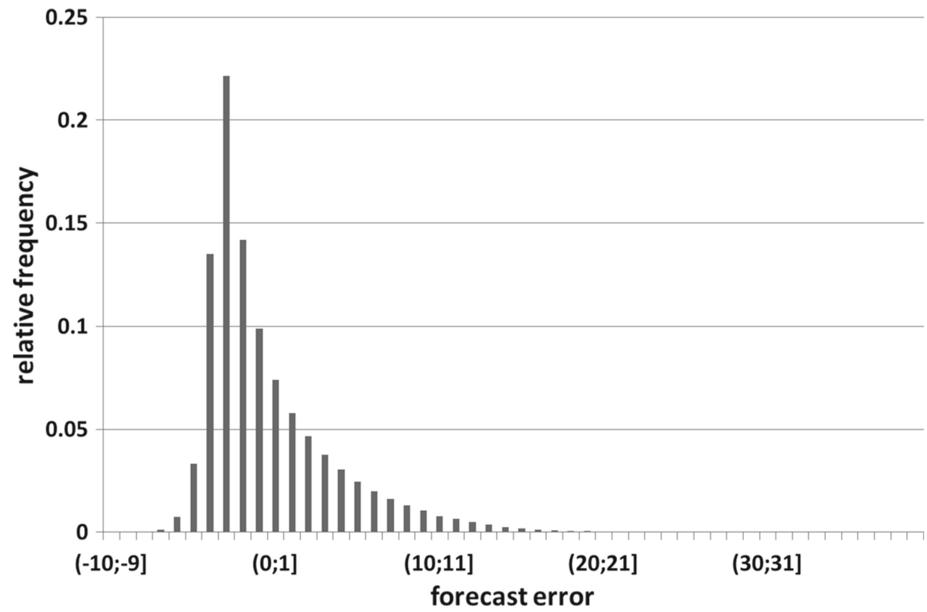


Fig. 1 Relative frequency distribution of the forecast error in case of a symmetrical demand and an unsymmetrical yield

Although the simulated forecast error has a mean of zero, it is obviously not normally distributed, since it is not symmetric, but skewed. For this example the markov chain approach with a normally distributed forecast error leads to a non-optimal critical stock level, resulting in 20 % larger costs compared to the optimum. We can conclude, that using a distribution for the forecast error which is defined only by two central moments is not appropriate. In the following, we investigate two approaches based on three central moments which differ in the distribution for the forecast error.

3.1 Central moments of the forecast error

We start with the derivation of the first three central moments of the forecast error, because in a second step, we fit an appropriate distribution on these moments. It is easy to see that the expectation of R is equal to zero, $E[R] = \mu_R = E(\mu_Z Q - ZQ) = 0$. From Inderfurth and Kiesmüller (2015) we know that the variance $VAR[R] = \sigma_R^2 = VAR[\mu_Z Q - ZQ]$ can be transformed to

$$VAR[R] = \sigma_Z^2(\mu_Q^2 + \sigma_Q^2) \tag{9}$$

with $\mu_Q = \mu_D/\mu_Z$ and $\sigma_Q^2 = [1/\mu_Z^2 \cdot (\sigma_D^2 + \rho_Z^2\mu_D^2)/(1-\rho_Z^2)]$ with the coefficient of variation of the yield factor $\rho_Z = \mu_Z/\sigma_Z$. The skewness of the forecast error equals:

$$\begin{aligned} SWN[R] &= \frac{E[(R-\mu_R)^3]}{\sigma_R^3} \\ &= \frac{E[R^3]}{\sigma_R^3} \end{aligned}$$

$$= \frac{E[(\mu_Z Q - ZQ)^3]}{\sigma_R^3}$$

$$= \frac{E[Q^3] \cdot (-2\mu_Z^3 + 3\mu_Z \mu_{2,Z} - \mu_{3,Z})}{\sigma_R^3} \tag{10}$$

$\mu_{2,Z}$ and $\mu_{3,Z}$ denote the second and third moment of the yield factor Z. From Inderfurth and Kiesmüller (2015) we know that

$$E[Q^3] = \frac{1}{\mu_{3,Z} - 3\mu_Z \sigma_Z^2} \cdot \left[\mu_{3,D} + 3\mu_D \rho_Z^2 \left(\mu_D^2 + \frac{\sigma_D^2 + \rho_Z^2 \mu_D^2}{1 - \rho_Z^2} \right) \right] \tag{11}$$

with the third moment of the demand, $\mu_{3,D}$. Formula (10) and (11) show that the third moment of the yield has an impact on the shape of the forecast error. Thus, it can not ne neglected.

3.2 The generalized extreme value distribution

The generalized extreme value distribution (GEVD) is defined by the location parameter τ, the scale parameter ψ and the shape parameter ξ and has the following probability density function (for more information on the GEVD we refer to Coles (2001), p. 48):

$$f(x; \tau, \psi, \xi) = \frac{1}{\psi} \left[1 + \xi \left(\frac{x - \tau}{\psi} \right) \right]^{(-1/\xi)-1} exp \left\{ - \left[1 + \xi \left(\frac{x - \tau}{\psi} \right) \right]^{-1/\xi} \right\} \tag{12}$$

We fit the distribution of the forecast error on the expectation $E[R]$, the variance $VAR[R]$ and the skewness $SWN[R]$. We first have to solve the following equation to get the shape parameter ξ:

$$SWN[R] = \frac{g_3 - 3g_1 g_2 + 2g_1^3}{(g_2 - g_1^2)^{3/2}} \tag{13}$$

where $g_k = \Gamma(1 - k\xi)$ with the complete gamma function $\Gamma(v)$. In the next step we determine the scale parameter ψ from

$$\psi = \sqrt{VAR[R] \cdot \xi^2 / (g_2 - g_1^2)}. \tag{14}$$

Finally, the location parameter τ is given as

$$\tau = E[R] + \frac{\psi}{\xi}(1 - g_1). \tag{15}$$

3.3 The skew normal distribution

The skew normal distribution is defined by the location parameter ζ, the scale parameter ω and the shape parameter α and has the probability density function (Sartori 2006)

$$f(x;\zeta,\omega,\alpha) = \frac{2}{\omega}\varphi\left(\frac{x-\zeta}{\omega}\right)\Phi\left(\alpha\frac{x-\zeta}{\omega}\right) \tag{16}$$

with the standard normal probability density function φ and the standard normal cumulative distribution function Φ.

To get the shape parameter α, the following equation has to be solved:

$$SWN[R] = \frac{4-\pi}{2}\frac{(\delta\sqrt{2/\pi})^3}{(1-2\delta^2/\pi)^{3/2}} \tag{17}$$

where $\delta = \alpha/\sqrt{1+\alpha^2}$. The scale parameter ω equals

$$\omega = \sqrt{VAR[R]\cdot\frac{1}{1-\frac{2}{\pi}\delta^2}} \tag{18}$$

and the location parameter ζ is defined as

$$\zeta = E[R] - \omega\delta\cdot\sqrt{\frac{2}{\pi}}. \tag{19}$$

3.4 Fitting the distribution

Equations (13) and (17) can easily be solved numerically and also the other parameters of the distribution can be determined such that the GEV or the skew normal distribution equal the first three moments of the forecast error. The convolution of the forecast distribution and the demand distribution in (5) can be determined numerically to obtain the matrix of the transition probabilities of the markov chain.

For the adaptive demand process (7) we directly fit on the sum of the forecast errors $\sum_{l=0}^{\lambda-1} R_{t-l}$, which has mean zero. The variance and the third moment can be computed easily since all terms in the sum are independent and identically distributed. Therefore, we get

$$VAR\left[\sum_{l=0}^{\lambda-1} R_{t-l}\right] = \lambda VAR[R] \quad \text{and} \quad E\left[\left(\sum_{l=0}^{\lambda-1} R_{t-l}\right)^3\right] = \lambda E[R^3]. \tag{20}$$

Thus, if period demand has a distribution where the sum of the random variables stays in the same class, only one convolution has to be computed for the newsvendor condition (8).

4 Numerical results

In a detailed numerical study we investigated the impact of the distribution assumption of the forecast error on the cost performance of the inventory system. We approximated the critical stock level using the markov chain as described in Sect. 2 and determined the corresponding average costs by simulation. We distinguished between the cases where the forecast error is assumed to be normally distributed (N), where a generalized extreme value distribution is fitted (*GEVD*), and a skew normal distribution is fitted (*SN*). We compared the obtained average costs C_{app}, ($app \in \{N, GEVD, SN\}$) with the optimal costs C_{opt} determined by simulation and computed the relative difference for every instance i:

$$\delta_i = \frac{C_{app} - C_{opt}}{C_{opt}} \cdot 100\,\% \quad app \in \{N, GEVD, SN\} \tag{21}$$

Furthermore, we computed the maximum relative difference over N instances as

$$\delta_{max} = max_{i=1,\ldots,N}\,\delta_i \tag{22}$$

and the average relative difference over N instances as

$$\bar{\delta} = \frac{1}{N}\sum_{i=1}^{N}\delta_i. \tag{23}$$

We test our approaches for the same data set as in Inderfurth and Kiesmüller (2015) and we also set the yield inflation factor equal to the reciprocal of the average yield ($F = \frac{1}{\mu_Z}$). We consider on the one hand a normally distributed demand with $\mu_{Dnorm} = 20$ and a coefficient of variation of $\rho_{Dnorm} \in \{0.1, 0.2, 0.3\}$. On the other hand we consider a gamma distributed demand with $\mu_{Dgam} = 20$ and a coefficient of variation of $\rho_{Dgam} \in \{0.1, 0.2, 0.3, 0.5, 0.75\}$ to allow for non-symmetric demand and larger demand uncertainty. We assume that the holding cost parameter equals one ($h = 1$) and chose six different parameters for the backorder costs such that $b/(b + h) \in \{0.85, 0.9, 0.95, 0.97, 0.99, 0.995\}$. Furthermore, we assume a beta distribution for the yield and consider three symmetric distributions with different variability ($\mu_Z = 0.5$, $\rho_Z \in \{0.2, 0.4, 0.5774\}$) and three asymmetric distributions ($\mu_Z = 0.75$, $\rho_Z = 0.2$; $\mu_Z = 0.85$, $\rho_Z = 0.2$; $\mu_Z = 0.85$, $\rho_Z = 0.1$). Further, three different numerical values for the lead time $\lambda \in \{2, 5, 10\}$ are investigated.

Overall we computed 324 instances in case of normally distributed demand and 540 instances in case of gamma distributed demand and measured the impact of the input parameters in the worst case performance (22) and the average performance (23). Table 1 shows the percentage maximum relative deviation from the optimal solution in case of a normally distributed demand.

It can be seen that in case of a symmetrical demand distribution with a low variance and a skewed yield distribution the maximum deviation of the average cost runs up to 20 % if the forecast error is approximated with a normal distribution. This example

Table 1 Maximum relative deviation from optimum (in %) for a normally distributed demand

Parameter	Maximum relative deviation			
	Value	Normal	Skew normal	GEVD
ρ_D	0.1	20.6694	1.0937	2.8142
	0.2	6.6381	0.8579	2.3743
	0.3	1.3230	0.5895	1.5362
(μ_Z, ρ_Z)	(0,85;0,1)	2.3855	0.2537	0.2537
	(0.85;0.2)	20.6694	0.6112	0.1718
	(0.75;0.2)	6.9429	0.2094	2.8142
	(0.5;0.2)	0.3417	0.3417	1.7412
	(0.5;0.4)	0.6386	0.6386	2.1119
	(0.5;0.5774)	1.0937	1.0937	2.8142
λ	2	20.6694	1.0937	2.8142
	5	7.5036	0.3417	1.7412
	10	3.8163	0.3288	0.6914
$b/(b+h)$	0.85	0.5119	0.5119	1.7412
	0.9	0.5613	0.5613	0.5613
	0.95	1.3799	0.3130	0.3130
	0.97	1.9552	0.2662	0.1718
	0.99	7.7290	0.2667	1.1407
	0.995	20.6694	1.0937	2.8142
Total		20.6694	1.0937	2.8142

illustrates that the skewness of the yield cannot be neglected. Approximating the forecast error with a skew normal distribution or with a GEV distribution reduces the maximum relative deviation to acceptable values. It can also be seen in Table 1 that the skew normal distribution outperforms the generalized extreme value distribution, especially in case of symmetric yield. In this situation the skew normal distribution coincides with the normal distribution.

The proposed approaches do not only improve the worst case performance but also the average performance as shown in Table 2.

It can be seen, that independently of the distribution for the forecast error, the maximum as well as the average relative deviation decreases with increasing coefficient of variation of the demand which might be surprising at first glance. The reason for this effect is risk pooling. A higher demand variability leads to a higher safety-stock level. With an increased inventory level, more stock is available for hedging against yield uncertainty. The same effect can be observed for varying lead times since demand uncertainty during the lead time increases.

Beyond that, especially the normal distribution leads to poor results in case of lower lead times. Referring to formula (7), $\eta(\lambda + 1)$ equals the sum of independent random variables. Due to the central limit theorem, $\eta(\lambda + 1)$ can be approximated by a normal random variable for high λ (Bollapragada and Morton 1999). Since $\lambda = 2$ is not

Table 2 Average relative deviation from optimum (in %) for normally distributed demand

Parameter	Average relative deviation			
	Value	Normal	Skew normal	GEVD
ρ_D	0.1	0.7595	0.0572	0.1747
	0.2	0.3149	0.0551	0.1163
	0.3	0.1074	0.0594	0.0756
(μ_Z, ρ_Z)	(0,85;0,1)	0.1324	0.0081	0.0101
	(0.85;0.2)	1.4738	0.0366	0.0200
	(0.75;0.2)	0.4697	0.0113	0.0579
	(0.5;0.2)	0.0245	0.0245	0.1007
	(0.5;0.4)	0.0696	0.0696	0.2222
	(0.5;0.5774)	0.1935	0.1935	0.3221
λ	2	0.6543	0.0856	0.1847
	5	0.3208	0.0375	0.1117
	10	0.2066	0.0487	0.0702
$b/(b+h)$	0.85	0.0501	0.0501	0.1025
	0.9	0.0672	0.0647	0.0783
	0.95	0.0985	0.0239	0.0217
	0.97	0.1836	0.0209	0.0215
	0.99	0.6960	0.0460	0.1499
	0.995	1.2681	0.1379	0.3592
Total		0.3939	0.0573	0.1222

Table 3 Average and maximum relative deviation from optimum (in %) for gamma distributed demand over all instances

	Normal	Skew normal	GEVD
Maximum relative deviation	12.51	1.37	2.90
Average relative deviation	0.21	0.06	0.09

sufficiently high, an approximation of the forecast error with a normal distribution is deficient.

Comparing the different distributions, the results illustrate that the approach based on the skew normal distribution leads to an excellent performance and leads to much better results especially in situations with skew yield and high critical ratios. The GEVD approach also outperforms the approach based on the normal distribution but is not as excellent as the approach based on the skewed normal distribution. The cost differences are a result of different values of the critical stock level. It can be observed that the approaches based on a normal and a generalized extreme value distributed forecast error have a tendency to underestimate the critical stock level. With a skew normal approximation of the forecast error the critical stock level is likely to be overestimated. Since the effect of overestimation is much smaller, because holding costs are usually smaller than backorder costs, the better cost performance can be explained.

Similar observations can be made in case of gamma distributed demand, as shown in Table 3.

5 Summary and outlook

We presented a way to take into account the effect of a skewed yield on the distribution of the forecast error in an appropriate way. The results show, that the skew normal distribution can deal very well with a symmetrical and an unsymmetrical yield factor and thus is a much better approximation for the forecast error than the normal distribution. We can also conclude, that neglecting the skewness of the yield leads to an underestimation of the stocklevel and therefore to more backorders, less service, and larger costs.

For future research, it would be interesting to extend the above approach to more complex systems. In this paper, we assumed that random yield occurs during the lead time, which equals a given number of periods. We can extend this system to a two-stage system where both stages are concerned with random yield. Such a situation exists when e.g. freight is first transported by a container ship and afterwards by a truck. Beyond that, stochastic lead times are not unusual in practice but neglected in our model.

Acknowledgments We thank Rainer Kleber for his valuable comments and discussions.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Bollapragada S, Morton TE (1999) Myopic heuristics for the random yield problem. Oper Res 47(5):713–722

Coles s (2001) An introduction to statistical modeling of extreme values. Springer, London

He Y, Zhang J (2008) Random yield risk sharing in a two-level supply chain. Int J Prod Econ 112(2):769–781

Henig M, Gerchak Y (1990) The structure of periodic review policies in the presence of random yield. Oper Res 38(4):634–643

Hennessy DA (2009) Crop yield skewness and the normal distribution. J Agric Resour Econ 34:34–52

Huh WT, Nagarajan M (2010) Technical note-linear inflation rules for the random yield problem: analysis and computations. Oper Res 58(1):244–251

Inderfurth K, Kiesmüller GP (2015) Exact and heuristic linear-inflation policies for an inventory model with random yield and arbitrary lead times. Eur J Oper Res 245(1):109–120

Inderfurth K, Vogelgesang S (2013) Concepts for safety stock determination under stochastic demand and different types of random production yield. Eur J Oper Res 224(2):293–301

Kelton WD, Sadowski RP, Sadowski DA (1998) Simulation with ARENA. WCB/McGraw-Hill, New York

Sartori N (2006) Bias prevention of maximum likelihood estimates for scalar skew normal and skew t-distributions. J Stat Plan Inference 136(12):4259–4275

Yano CA, Lee HL (1995) Lot sizing with random yields: a review. Oper Res 43(2):311–334

3. The Influence of Quality Inspections on the Optimal Safety Stock Level

Sonntag, D., Kiesmüller, G. P. (2017). The Influence of Quality Inspections on the Optimal Safety Stock Level. Production and Operations Management, 26, 1284-1298.

POMS PRODUCTION AND OPERATIONS MANAGEMENT

PRODUCTION AND OPERATIONS MANAGEMENT POMS

Vol. 0, No. 0, xxxx–xxxx 2017, pp. 1–15
ISSN 1059-1478 | EISSN 1937-5956 | 17 | 00 | 0001
DOI 10.1111/poms.12691

The Influence of Quality Inspections on the Optimal Safety Stock Level

Danja Sonntag, Gudrun P. Kiesmüller
Faculty of Economics and Management, Chair of Operations Management, Otto-von-Guericke-University Magdeburg, P.O. Box 4120, D-39016, Magdeburg, Germany, danja.sonntag@ovgu.de, gudrun.kiesmueller@ovgu.de

Due to yields of less than 50% during the production of curved glass for the displays on their new cell phone series, Samsung has to deal with higher than expected production costs of several million dollars. Where there is random yield, production costs as well as holding costs can be reduced by introducing quality inspections, in which defective items are discarded before further production. To achieve the greatest cost savings, it is important to determine the optimal number and positions of these inspections across the production process which, due to several influencing parameters, is not simple. We show how the positions of inspection within a production process influence the safety stock level that is required to buffer against uncertainties due to demand and yield randomness. Our approach is the first one, combining decisions about the number and positions of inspections with inventory control strategies in a warehouse. We achieve a maximum safety stock reduction of more than 30% in our examples, which can be even larger depending on the parameter setting. For a company like Intel, reporting inventories for finished goods of nearly 1.5 billion dollars in the 2014 annual report, this allows for significant savings.

Key words: stochastic inventory model; random yield; quality inspections; safety stock
History: Received: December 2015; Accepted: December 2016 by Panos Kouvelis, after 1 revision.

1. Introduction

Between 2003 and 2013, the value of high-tech products sold in the European Union rose by over 85% (Eurostat 2016). Due to this increasing number of high-tech products and the challenge of achieving higher quality standards, production processes become more and more complex. Samsung, for example, has problems producing the curved glass for a new cell phone series. Due to low yields of less than 50%, the price for such displays is approximately eight times higher than the former one (McNutt 2015). Samsung ordered around 8.5 million displays in the second quarter of 2015 wherefore several problems arose. On one hand, the production costs increase by millions of dollars if approximately every second item is defective. On the other hand, the low production output leads to difficulties in delivery. By way of comparison, Intel leads the world in the production of microchips especially because their production yields are in the high 90% range, whereas competitors have to deal with yields of 60–80% (Foremski 2012). Taking into consideration that Intel spends around 7 billion dollars for a semiconductor fabrication plant–called "fabs"–high yields are necessary for profitable production and low consumer prices.

Random yields as well as random demand increase the complexity of production systems enormously since the planner can only deal with an expectation of the production outcome. Choi et al. (2008) as well as Dettenbach and Thonemann (2015) showed that early information about realized yields can reduce the uncertainty generated by an unknown number of defective items being produced. The knowledge about the amount of items not fulfilling given quality standards leads to a reduction of the corresponding safety stock level required to achieve a predefined service level for the final product stocked in a warehouse. However, this information is not directly available, with the result that the planner needs to introduce costly quality inspections to obtain it. In the semiconductor industry for example, inspection systems to increase and maintain integrated circuit chip yields cost about 700 million dollars per year (Stokowski and Vaez-Iravani 1998). Depending on the effectiveness of quality inspections, the yield, for example, in microchip production can vary between 0% and 100% (Leachman and Hodges 1996). A trade-off exists between high costs for inspections and the need to ensure high quality throughout the production. On one hand, many quality control stations lead to a direct elimination of all defective items, which reduces uncertainty about the production output and therefore the safety stock level. On the other hand, these inspections are very costly and it might not be economical to order an inspection after each production stage. Therefore, it is important to determine the optimal number and position of these inspections

Please Cite this article in press as: Sonntag, D., G. P. Kiesmüller. The Influence of Quality Inspections on the Optimal Safety Stock Level. *Production and Operations Management* (2017), doi 10.1111/poms.12691

across the production process. This is not easy because several parameters, such as the mean and the variance of the yield and the total production time, influence these decisions.

In this study, we consider multi-stage make-to-stock production systems for a single product. Since studies have shown that product quality has a direct effect on customer satisfaction, purchase intentions, and therefore economic returns (Anderson et al. 1994; Taylor and Baker 1994), we allow only items of perfect quality to be sold to the customer. To ensure that produced products meet the desired quality standards, all units are inspected after the last production stage, before items are stocked in a warehouse, to remove all items not satisfying the quality requirements. Each production stage is related to a positive production time and random yield, which is assumed to be stochastically proportional. Our purpose is to analyze existing production systems with in between quality inspections and afterwards optimize the number as well as the position of these inspections across the production process with the objective of minimizing the overall costs consisting of holding, backorder, production, quality control, and disposal costs.

There exists a variety of literature on the positioning of quality inspections within a production process. Lindsay and Bishop (1964) and Eppen and Hurst Jr (1974) are some of the first authors considering this kind of question. Lindsay and Bishop (1964) used a dynamic programming approach to determine the position of quality control and the optimal percentage of items being checked to reach a contractual given quality level of delivered products. Eppen and Hurst Jr (1974) instead considered imperfect inspections which can lead to good items being rejected and bad items being accepted. They also used dynamic programming to decide whether to inspect products before the next production step. Over the years a lot of literature has come up with problem settings in a similar production environment but with different constraints (e.g., imperfect inspections or reworking of defective items). A comprehensive overview is given by Mandroli et al. (2006) and Shetwan et al. (2011). All literature about the positioning of quality controls has in common that the production quantity is either externally given (e.g., one item or a batch of items with a fixed batch size per period) or affected by demand due to the use of make-to-order production. Peters et al. (1988) and Ben-Daya and Noman (2008) connect "quality-control and inventory policies", but they only consider a one-stage production system with only one quality control inspection before the product enters the warehouse.

In contrast to the existing literature we consider a multi-stage make-to-stock environment in which the warehouse can place an order with the production at the beginning of a period, which determines the production volume. The literature on inventory management reveals that the optimal ordering policy in systems with random yield is very complex (Henig and Gerchak 1990). As a consequence, a linear inflation rule that performs very well (Huh and Nagarajan 2010), is commonly used to take the yield losses into account. A linear inflation policy is defined by two parameters, the pseudo-order-up-to level and the yield inflation factor, whereas we assume that the yield inflation factor is given by the reciprocal of the mean yield of the total system, which is a commonly used assumption (see e.g., Bollapragada and Morton 1999; Henig and Gerchak 1990; Inderfurth and Kiesmüller 2015). Thus, only the pseudo-order-up-to level has to be determined. Contributions in this area are devoted either to the optimal order policy (see Henig and Gerchak 1990; Bollapragada and Morton 1999) or to the optimization of policy parameters of a given policy, for example the linear inflation policy (Huh and Nagarajan 2010). A literature overview was given by Yano and Lee (1995). In this line of research the realized yield can be observed only at the moment of delivery, which means that yields are unknown during the production process.

In contrast, Choi et al. (2008) and Dettenbach and Thonemann (2015) look at multi-stage production with the possibility of observing realized yields between the stages. Thus, they combine the literature on multi-stage serial production systems with the literature on make-to-stock policies and analyze the effect of real-time yield information in a serial supply chain with one manufacturer and one supplier. Choi et al. (2008) compare different alternatives of real-time yield information when the supplier produces the items in two steps. As well as a production time of one period on each production stage, they include a one-period transportation time to the manufacturer in their model. For identical random yield at each production step they show in a simulation study that real-time yield information leads to a reduction in safety stock. The focus of Dettenbach and Thonemann (2015) is different, because they are interested in optimal decisions and therefore present a dynamic programming approach. The optimal decisions can only be determined numerically for very small systems and no structural properties can be obtained. Therefore, they also present two heuristics. The first one is based on an idea of Ehrhardt and Taube (1987) and can lead to poor results since it does not reflect the multi-stage production with random yield on each production stage in an appropriate way. The second heuristic, based on the approach of Huh and Nagarajan (2010), requires simulation since "it is difficult to calculate [...] analytically" (Dettenbach and

Thonemann 2015), and therefore might need considerable computation time.

Choi et al. (2008) and Dettenbach and Thonemann (2015) give very good insights into the effect of real-time yield information on the holding and backorder costs but they do not address the question of which production stage should be followed by a quality control test or how many should be established under a comprehensive cost model.

Our contribution in this study is as follows: (1) For given production systems, we analyze the effect of intermediate quality control on the optimal safety stock level. A steady-state approach leads to a simple newsboy-like equation for the pseudo-order-up-to level and therefore the safety stock. (2) We introduce a new model to optimize the position of one inspection with respect to holding, backorder, production, quality control, and disposal costs. Results show that a safety stock reduction of over 30% is possible, depending on the parameter setting. (3) We determine the optimal number of inspections that minimize overall costs and show that the value of quality control decreases marginally with an increase in the number of inspections. (4) We specify the positioning of the inspection control stations, which reveals the layout of the production process with respect to the inspection strategy. The examples reveal that it is best for inspections to be equally spaced across the production stages.

The remainder of the study is organized as follows: In section 2, we describe the multi-stage production system and formulate the model. In section 3, the complex steady-state approach, which leads to a simple newsboy-like calculation of pseudo-order-up-to levels and safety stocks, is introduced. The formula is used to analyze given production systems with intermediate inspections in section 4. To optimize the number and position of control stations, a comprehensive cost model is introduced in section 5. In section 6, this model is used to show where inspections should be placed over the course of the production stages. We conclude with a summary and suggestions for future research in section 7.

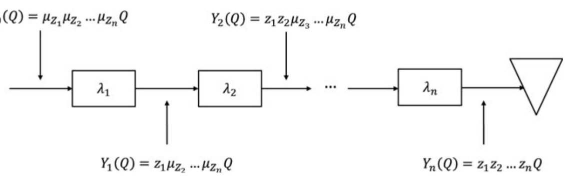
2. Model Formulation

We consider an in-house n-stage serial production system with a warehouse for the final product and a total production time λ for one batch of items. Since no capacity constraints are taken into account the production time is independent of the batch size. We assume that m quality inspections are integrated into the production process ($m \leq n$). They can be placed after any production stage. The inspections perfectly check all incoming items (100% inspection), take a negligible amount of time and are not affected by capacity constraints. Due to perfect inspections all defective items are discarded with certainty and all good items stay in the process. We number the quality inspections in a sequential way such that the last inspection–the inspection of the final product–is denoted with m. After the last inspection, perfect products are stored in the warehouse. We define $\lambda_j (j = 1, 2, \ldots, m)$ as the production time between inspection $j - 1$ and j ($j = 0$ represents the start of the production process). Moreover, we define Λ_j as the cumulated production time until quality control j ($\Lambda_j = \sum_{k=1}^{j} \lambda_k$ and $\Lambda_m = \lambda$). Without loss of generality we can consider an n-stage production system with a quality inspection after each production stage ($m = n$). Note that every n-stage production system with less than n quality inspections ($m < n$) can be modeled as an m-stage production system with m quality inspections by aggregating the stages. Thus, assuming that the number of production stages equals the number of quality inspections does not cause a limitation of the model. The arising n-stage production system is illustrated in Figure 1.

We would like to mention that we formulate the model in a production environment, but each stage can also be interpreted as a transportation process where every transportation stage is related to random yield. One example in a transportation environment, where random yield plays an important role, is the freightage of perishable products (White and Cheong 2012). These products are typically packed in refrigerated trucks or containers whereby it is important that the cool chain is not interrupted. If goods perish during the transportation process, it is important to sort out these goods as soon as possible to avoid wasting even more products. The following formulations are, however, based on a production rather than a transportation process.

The production process suffers from quality problems resulting in an uncertain production output. We assume that these problems can be modeled with a stochastic proportional yield, which means the production output depends on the input Q as follows: $Y(Q) = Z \cdot Q$, where the random variable Z ($Z \in [0, 1]$) denotes the yield factor. If the system is modeled as an n-stage production system, the mean and the variance of the yield for production stage

Figure 1 An n-Stage make to Stock Production System



Please Cite this article in press as: Sonntag, D., G. P. Kiesmüller. The Influence of Quality Inspections on the Optimal Safety Stock Level. *Production and Operations Management* (2017), doi 10.1111/poms.12691

$i = 1, \ldots, n$ equal μ_{Z_i} and $\sigma^2_{Z_i}$ respectively. Thus, the mean yield μ_Z for the whole production system is given as $\prod_{i=1}^{n} \mu_{Z_i}$ and the variance of the total yield σ^2_Z equals $\prod_{i=1}^{n}(\sigma^2_{Z_i} + \mu^2_{Z_i}) - \prod_{i=1}^{n} \mu^2_{Z_i}$.

Since we want to analyze the effect of different positions of quality inspections within the system, we introduce the following random variables $Y_i(Q)$ for $i = 0, 1, 2, \ldots, n$ equal to $\prod_{j=1}^{i} Z_j \prod_{j=i+1}^{n} \mu_{Z_j} Q$ and define $\prod_{i=1}^{0} y_i = 1$. These random variables reflect the fact, that, because quality control is introduced after the production stage i, information about realized yields of all previous stages as well as of stage i becomes available. The planner can use this real-time-yield information when he or she releases new production orders. In advance of any production the realized yields are unknown, so that no information is obtained and the planner has no more than an expectation of the production outcome ($Y_0(Q)$). If, due to quality inspections, yield information becomes available after the first inspection, which means after a production time of λ_1 periods, then the realized yield of the first production stage, z_1, is known and the planner can use this additional information to update his or her expectations of the production outcome. After realization of the first production yield, the expectation of the production outcome $Y_1(Q)$ equals $z_1 \prod_{j=2}^{n} \mu_{Z_j} Q$. This can be done for each production stage until $Y_n(Q)$ equals the real delivery quantity. As mentioned earlier, there is always an inspection after the n-th production stage, such that full yield information is available and only conforming goods are delivered to the warehouse and therefore to the customer.

The production process is used to refill the inventory at the warehouse, which observes stochastic demand. The sequence of events in one period within the warehouse is the following: at the beginning of each period the order placed λ periods before is delivered and the inventory position is updated. Note that realized yields, unknown up to this point, are observed after the final quality inspection which is close to the warehouse due to negligible transportation times between production and the warehouse. Afterwards, a new order is placed, including the yield information of the incoming order, and demand occurs. At the end of the period holding and backorder costs are charged.

Our model is similar to that in Dettenbach and Thonemann (2015) and an extension of the model in Choi et al. (2008), because they consider only two production stages with one period production time and they have allowed only identical yield distributions at both stages.

We assume continuous period demand, independent and identically distributed across the periods and denote the first two moments with (μ_D, σ^2_D). Demand that cannot be satisfied directly from stock is backlogged. Production orders are released periodically by the warehouse according to a linear inflation rule

$$Q_t = \begin{cases} F(S - X_t), & X_t < S \\ 0, & X_t \geq S \end{cases} \tag{1}$$

This ordering policy has shown to have very good performance (Huh and Nagarajan 2010) and it is usually proposed since the optimal ordering policy in the presence of random yield does not possess a simple structure (Henig and Gerchak 1990). The order quantity is the inflated difference between the pseudo-order-up-to level S and the inventory position X_t if this difference is positive, and zero otherwise. The yield inflation factor F is assumed to be the reciprocal of the mean yield over the whole production period ($F = 1/\mu_Z$). In Equation (1) it can be seen that the inventory position has an impact on the order quantity. For periodic order-up-to policies without random yield, the inventory position is defined as the stock on hand minus backorders plus outstanding orders (Tempelmeier 2006, p. 16f). Where there is random yield the order quantity is not necessarily equal to the delivery quantity and thus only limited information is available about the outstanding orders. Therefore, the expected quantities delivered are included in the definition of the inventory position. Under the assumption that n quality inspections exist, which provide information about the realized yields at the corresponding production stage, the inventory position at the beginning of period t after the delivery of the order, placed λ periods before, is defined as

$$\begin{aligned} X_t = {} & I_{t-1} + Y_n(Q_{t-\lambda}) + \sum_{l=1}^{\Lambda_1 - 1} Y_0(Q_{t-l}) \\ & + \sum_{l=\Lambda_1}^{\Lambda_2 - 1} Y_1(Q_{t-l}) + \cdots + \sum_{l=\Lambda_{n-1}}^{\lambda - 1} Y_{n-1}(Q_{t-l}), \end{aligned} \tag{2}$$

where I_{t-1} is the inventory level at the end of period $t - 1$ and $Y_n(Q_{t-\lambda})$ the delivered quantity of the order placed in period $t - \lambda$. The following terms are related to the outstanding orders. For example, $\sum_{l=1}^{\Lambda_1 - 1} Y_0(Q_{t-l})$ describes the outstanding orders that have not passed the first production stage, which means that no realized yields are known. If the total production time is zero or one period it is obvious that the terms for the outstanding orders cancel out since we define $\sum_{i=1}^{0} y_i = 0$. This explains why neglecting positive production times reduce the complexity of the system enormously: there is no uncertainty about the production outcome when making an ordering decision.

Please Cite this article in press as: Sonntag, D., G. P. Kiesmüller. The Influence of Quality Inspections on the Optimal Safety Stock Level. *Production and Operations Management* (2017), doi 10.1111/poms.12691

Based on the the definition in Equation (2) a recursive equation for the inventory position can be obtained:

$$\begin{aligned}X_{t+1} =& X_t + Y_0(Q_t) - D_t - (Y_0(Q_{t+1-\lambda_1}) - Y_1(Q_{t+1-\lambda_1}))\\ &- (Y_1(Q_{t+1-\Lambda_2}) - Y_2(Q_{t+1-\Lambda_2}))\\ &- \cdots - (Y_{n-1}(Q_{t+1-\lambda}) - Y_n(Q_{t+1-\lambda})).\end{aligned} \tag{3}$$

The inventory position in period $t + 1$ before an order is placed equals the inventory position at the beginning of the previous period plus the expected delivery of the order placed in period t, $Y_0(Q_t)$, minus the demand in period t, D_t. The rest of the terms reflect updates of the inventory position. More precisely, whenever the realization of a yield is known due to a quality inspection, this information is included in the inventory position and replaces the information about the expected yield. Thus, these terms ensure that the forecasts about the expected deliveries are updated whenever we get information about the realized yields within the production process (compare with Figure 1). These forecast errors R_i are defined as

$$\begin{aligned}R_{i,t+1-\Lambda_i} =& Y_{i-1}(Q_{t+1-\Lambda_i}) - Y_i(Q_{t+1-\Lambda_i})\\ =& \prod_{j=1}^{i-1} Z_j \prod_{l=i}^{n} \mu_{Z_l} Q_{t+1-\Lambda_i}\\ &- \prod_{j=1}^{i} Z_j \prod_{l=i+1}^{n} \mu_{Z_l} Q_{t+1-\Lambda_i}, \quad i = 1, \ldots, n\end{aligned} \tag{4}$$

and also play an important role when the distribution of the inventory level is determined.

In a periodic order-up-to policy without random yield, the inventory level is given by the inventory position after ordering, which equals the order-up-to level and includes all incoming quantities minus the outgoing quantities during the risk period (Axsäter 2007, p. 69). In the presence of random yield, the incoming quantities may differ from the quantities included in the inventory position, modeled by the forecast error as defined in Equation (4). This leads to the following equation for the inventory level at the end of period $t + \lambda$:

$$\begin{aligned}I_{t+\lambda} =& X_t + \prod_{i=1}^{n} \mu_{Z_i} Q_t - \sum_{l=0}^{\lambda} D_{t+l}\\ &- \left(\sum_{l=0}^{\lambda_1 - 1} R_{1,t-l} + \sum_{l=0}^{\lambda_1+\lambda_2-1} R_{2,t-l} + \cdots + \sum_{l=0}^{\lambda-1} R_{n,t-l} \right)\\ =& X_t + \mu_Z Q_t - \sum_{l=0}^{\lambda} D_{t+l} - \sum_{i=1}^{n} \sum_{l=0}^{\Lambda_i - 1} R_{i,t-l}.\end{aligned} \tag{5}$$

We can also interpret the forecast error as an additional demand, which can be positive as well as negative, and define an adapted demand process as

$$\eta(\lambda + 1) = \sum_{l=0}^{\lambda} D_{t+l} + \sum_{i=1}^{n} \sum_{l=0}^{\Lambda_i - 1} R_{i,t-l}. \tag{6}$$

As a result, we get the following equation for the inventory level, which has a structure similar to that of the case of no random yield.

$$I_{t+\lambda} = X_t + \mu_Z Q_t - \eta(\lambda + 1). \tag{7}$$

Our first purpose is to analyze given production systems with given positions of quality inspections. This means that we cannot influence quality inspection costs, production costs or disposal costs. Relevant costs are only holding and backorder costs which can be controlled by the amount of safety stock stored in the warehouse. Holding and backorder costs are charged at the end of a period based on the average inventory level. If h denotes the unit holding and b the unit backorder costs, we obtain

$$C(S) = hE[(I(S))^+] + bE[(-I(S))^+], \tag{8}$$

where $(X)^+$ is defined as $\max\{0, X\}$. It can be seen that the costs depend on the pseudo-order-up-to level which is related to the amount of safety stock.

If we denote with $\varphi_I(x)$ the density of the inventory level as given in Equation (7), the average cost for a given pseudo-order-up-to level can be computed as

$$C(S) = h \int_0^{\infty} x\varphi_I(x)dx - b \int_{-\infty}^{0} x\varphi_I(x)dx. \tag{9}$$

The density of the inventory level remains to be determined.

3. The Optimal Safety Stock for a Given Production Layout

In previous literature, the pseudo-order-up-to level in a multi-stage serial production system was determined using simulation (Choi et al. 2008; Dettenbach and Thonemann 2015). The disadvantage of this solution method is that it leads to high computation times and is therefore not suitable for practical application. We instead use a heuristic steady-state approach which is shown to perform very well by Inderfurth and Kiesmüller (2015) for a single-stage production system. The steady-state approach assumes a linearization of the function for the order quantity in Equation (1), which means that the possibility that X_t is larger or equal S is neglected (Zipkin 2000) and the order quantity Q_t equals $F(S - X_t)$. Under this assumption, closed-form expressions for the first two moments of the inventory level can be obtained and a reasonable distribution can be fitted. In Inderfurth and Kiesmüller (2015) it is shown, that a normal distribution leads to excellent results in symmetric situations and in case of non-symmetric yield or demand a mirrored gamma distribution is appropriate. This approach is easy to implement in a spreadsheet with low computation times and leads to very good results.

Since we neglect any kind of inventory between production stages, the total holding and backorder costs at the end of a period as given in Equation (9) are of a structure similar to that of the single-stage system, and we obtain for the pseudo-order-up-to level the following expression (see Inderfurth and Kiesmüller 2015):

$$S = (\lambda + 1)\mu_D + \Phi^{-1}\left(\frac{b}{b+h}\right)\sigma_I. \tag{10}$$

$\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution and σ_I is the standard deviation of the inventory level. Similar to the classical newsvendor model, the pseudo-order-up-to level equals the mean demand during the risk period plus some safety stock. It is obvious that real-time yield information can just have an influence on the safety stock due to a reduction of the variance of the inventory level. To determine the influence of quality inspections on the safety stock, the standard deviation of the inventory level is required.

LEMMA 1. *Under a strictly linear control rule, the variance of the inventory level is given as:*

$$\sigma_I^2 = (\lambda + 1)\sigma_D^2 + \sum_{i=1}^{n} \Lambda_i \sigma_{R_i}^2. \tag{11}$$

PROOF 1. In case of a strictly linear control rule the order quantity Q_t equals $F(S - X_t)$. Since we have assumed that the yield inflation factor equals the reciprocal of the average yield, for the inventory position after ordering we get:

$$X_t + \mu_Z Q_t = X_t + \mu_Z F(S - X_t) = S. \tag{12}$$

Therefore, Equation (11) follows directly from Equation (5). □

Formula (11) shows that the variance of the inventory level is influenced by both the variance of the demand and by the variances of the forecast errors. As defined in section 2, Λ_j reflects the cumulated production time until quality inspection j which equals the time until forecast error R_i is known ($\forall i, \quad i = 1, \ldots, n$).

We illustrate the influence of production times with a small example where we consider a two-stage production system with one intermediate quality inspection. In this environment, the variance of the inventory level equals: $\sigma_I^2 = (\lambda + 1)\sigma_D^2 + \lambda_1\sigma_{R_1}^2 + (\lambda_1 + \lambda_2)\sigma_{R_2}^2$. In the following, we show how the variances of the forecast errors depend on the position of quality inspections implicitly given by the moments of the yield factors Z_i ($i = 1, \ldots, n$).

LEMMA 2. *Let n denote the number of stages of the production system. Then, under a strictly linear control rule, the variance of the forecast error for stage i ($i \leq n$) is given as:*

$$\sigma_{R_i}^2 = (\sigma_Q^2 + \mu_Q^2)\left[\prod_{j=1}^{i-1}(\sigma_{Z_j}^2 + \mu_{Z_j}^2) \cdot \prod_{l=i+1}^{n} \mu_{Z_l}^2 \cdot \sigma_{Z_i}^2\right]. \tag{13}$$

The proof of the lemma is given in Appendix A.

We illustrate the result for an example with a three-stage production system and obtain the following variances for the forecast errors.

$$\begin{aligned}
\sigma_{R_1}^2 &= (\sigma_Q^2 + \mu_Q^2)[\sigma_{Z_1}^2\mu_{Z_2}^2\mu_{Z_3}^2] \\
\sigma_{R_2}^2 &= (\sigma_Q^2 + \mu_Q^2)[\sigma_{Z_2}^2\mu_{Z_1}^2\mu_{Z_3}^2 + \sigma_{Z_2}^2\sigma_{Z_1}^2\mu_{Z_3}^2] \\
\sigma_{R_3}^2 &= (\sigma_Q^2 + \mu_Q^2)[\sigma_{Z_3}^2\mu_{Z_1}^2\mu_{Z_2}^2 + \sigma_{Z_3}^2\sigma_{Z_2}^2\mu_{Z_1}^2 + \sigma_{Z_3}^2\sigma_{Z_1}^2\mu_{Z_2}^2 \\
&\quad + \sigma_{Z_3}^2\sigma_{Z_2}^2\sigma_{Z_1}^2]
\end{aligned}.$$

The example confirms that the variances of the forecast errors can be easily calculated, depending on the information on the mean and the variance of the yield for each production stage and the mean and the variance of the order quantity. The following lemma shows how to calculate the variance of the order quantity.

LEMMA 3. *Under a strictly linear control rule with F equal to $1/\mu_Z$, the variance of the order quantity in an n-stage production system is given as:*

$$\sigma_Q^2 = \frac{(\rho_Z^2\mu_D^2 + \sigma_D^2)}{\mu_Z^2 - \sigma_Z^2}, \tag{14}$$

where ρ_Z equals the coefficient of variation of the total yield as the ratio of the standard deviation σ_Z and the mean μ_Z.

The proof can be found in Appendix B.

In the long run, the mean order quantity has to equal the inflated mean of the adapted demand defined in Equation (6). Thus, $E[Q]$ equals $F \cdot (\sum_{i=1}^{n} E[R_i] + E[D])$. Since the means of the forecast errors are equal to zero, the mean order quantity equals the inflated mean demand $F\mu_D$. The results for the mean and the variance of the order quantity show that the ordering process is independent of the forecast errors and the production times and therefore independent of the position of quality inspections. Thus, real-time yield information has no impact upon this. It only has an impact on the variance of the inventory level and therefore on the safety stock.

The above formulae show that the effect of real-time yield information obtained by quality inspections and analyzed by Choi et al. (2008) and Dettenbach and Thonemann (2015) via simulation can be easily calculated in a spreadsheet.

As posed in the contribution of this study, our first purpose is to analyze given production systems with intermediate quality control. In such a given environment the position of the inspection cannot be changed, which results in given production costs, quality control costs and disposal costs. Nevertheless, the planner in the warehouse can influence the amount of holding and backorder costs by making the best choice for the safety stock level. In the following, we show how formula (11) can be used to determine the optimal safety stock level for a given production system, which minimizes the average holding and backorder costs.

4. The Impact of the Position of Inspections on the Safety Stock

In the following, we compare given production systems which contain one intermediate inspection. We analyze the impact of its position on the optimal safety stock, using the approximate steady-state approach. The safety stock level can be obtained directly from Equation (10) and equals $\Phi^{-1}(\frac{b}{b+h})\sigma_I$.

We fix the total production time λ and model the system as an n-stage production system with a production time of one period at each stage and a fixed mean yield on every production stage and consider different scenarios with respect to yield uncertainty. We distinguish between constant yield variance across all production stages (1), decreasing (2) or increasing (3) yield variance over all stages. With these three cases, we represent balanced production systems as well as production systems that have to deal with higher discard variations at the beginning or at the end of the production process. A process with increasing yield variance can be found in microchip production. At the end of the production process the manufacturer checks whether the chips are working or whether there was a production fault in the electrical circuits. As mentioned in the beginning of the study, the yield in microchip production can vary between 0% and 100% (Leachman and Hodges 1996). An example of a decreasing yield variance is the processing of agricultural products. These products enter production in large amounts and are afterwards cleaned and sorted before further production. Depending on the quality of the harvest, the yield at the beginning of the production process differs. Note that these are only two examples of possible parameter constellations. The steady-state approach presented is able to deal with all other conceivable yield settings across the production process as well. For the examples in this study, with constant, decreasing or increasing yield, it is important to remark that although the yield variability over the production stages is different in all three systems, the total yield for the whole production process as well as the mean yield remains the same over all scenarios, making the examples comparable.

We define a skewness factor δ ($\delta \le 1$) which reflects the degree of skewness of the yield variability across the system. For δ equal to one no skewness exists, resulting in a balanced production with constant yield variance across all production stages. The yield variances of the systems will be calculated as follows. Considering the production system with decreasing yield variance across the stages first, for a given yield variance in the first stage σ^2_{dec,Z_1} the yield variance for all following production stages σ^2_{dec,Z_i} equals $\sigma^2_{dec,Z_{i-1}} \cdot \delta^2$ ($\forall i = 2, \ldots, n$). Where there is increasing yield variability the variance of the yield looks the same but is mirrored: $\sigma^2_{inc,Z_i} = \sigma^2_{dec,n+1-i}$, $\forall i = 1, \ldots, n$. As mentioned, the overall yield variability σ^2_Z stays the same. Thus, the variance of the yield in the symmetric case σ^2_{sym,Z_i} equals $\sqrt[n]{\sigma^2_Z + \mu^{2n}_{Z_i}} - \mu^2_{Z_i}$, $\forall i = 1, \ldots, n$ with σ^2_Z equal to $\prod_{i=1}^n(\sigma^2_{dec,Z_i} + \mu^2_{Z_i}) - \prod_{i=1}^n \mu^2_{Z_i}$.

Remember that up to now the position of the intermediate quality inspection is not a decision criterion. We assume that the production process with the control station is given such that only the determination of the safety stock level implies a potential for optimization to the planner. The following parameter setting is used to analyze the amount of safety stock required when quality inspections are given at different positions in the process.

The total production time λ equals ten periods and we model the system as a ten-stage production system (n = 10) with a production time of one time period for each stage. We consider ten different system configurations based on the position of the quality inspection resulting in: $(\lambda_1, \lambda_2) = (1, 9), (2, 8), \ldots, (9, 1)$.

We allow for one intermediate inspection after stage one to nine which meas that λ_1 and λ_2 can equal 1, 2, …, 9 and 9, 8, …, 1. The mean demand μ_D and the corresponding coefficient of variation ρ_D equal 20 and 0.1. The results of Choi et al. (2008) and Dettenbach and Thonemann (2015) show that the value of

Please Cite this article in press as: Sonntag, D., G. P. Kiesmüller. The Influence of Quality Inspections on the Optimal Safety Stock Level. *Production and Operations Management* (2017), doi 10.1111/poms.12691

Table 1 Yield Scenarios

Stage i	1	2	3	4	5	6	7	8	9	10
μ_{Z_i}	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
σ_{sym,Z_i}	0.0109	0.0109	0.0109	0.0109	0.0109	0.0109	0.0109	0.0109	0.0109	0.0109
σ_{dec,Z_i}	0.0564	0.0276	0.0135	0.0066	0.0033	0.0016	0.0008	0.0004	0.0002	0.0001
σ_{inc,Z_i}	0.0001	0.0002	0.0004	0.0008	0.0016	0.0033	0.0066	0.0135	0.0276	0.0564

real-time yield information decreases with increasing demand variability. In case of a larger demand variability a higher amount of safety stock is required to buffer against demand uncertainty. Therefore, more stock is made available to hedge against yield uncertainty. As a result, the effect of introducing quality control is the largest for small demand variability.

Service level α equals 0.99 which leads to a backorder costs parameter b of $\frac{\alpha}{1-\alpha} \cdot h$ equal to 99 for given holding costs h per item on stock of one unit. Choi et al. (2008) and Dettenbach and Thonemann (2015) reveal, that the service level has nearly no impact on the value of real-time yield information as long as it is appropriately high (>0.8). The mean yield μ_{Z_i} per production stage i equals 0.95 ($\forall i = 1, \ldots, n$) and the coefficient of variation of the yield ρ_{dec,Z_1} on the first production stage is assigned to be 0.25. The skewness parameter δ equals 0.7. Table 1 shows how the variance of the yield in the three production systems differs across the production stages given δ. It is evident that the variance of the yield decreases/increases exponentially when there is asymmetric yield variability. As the skewness parameter δ decreases, the distribution of the yield variability across the stages steepens.

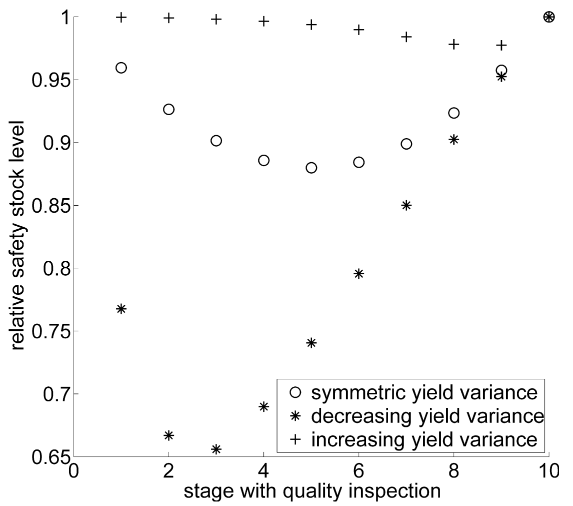
In the following, we determine the optimal safety stock level according to Equation (10) for the three yield scenarios with fixed mean yield and varying yield variance across the stages, when there is only one additional inspection before the final inspection. We compute the optimal fraction of safety stock needed to ensure an alpha service level of 99% relative to the benchmark with no intermediate control for different scenarios. The numerical results are depicted in Figure 2.

It is obvious that the non-optimal position of an inspection substantially increases the required safety stock, especially when there is decreasing yield variability. This phenomenon can be explained by the utility of real-time yield information generated by quality inspections after different stages in the production process. If early information about realized yields is obtained, the forecast errors defined in Equation (4) are known at early production stages, which allows an update of the inventory position and therefore a reduction of the safety stock level. Due to decreasing yield variability over all stages, yield becomes less uncertain at later production stages and to some extent more predictable. Since real-time yield information has no value if the production output is 100% certain, the value of real-time yield information and therefore of quality inspections is greatest at early stages, with a relatively high corresponding yield variability.

When there is increasing yield variability, quality control after an early stage has a limited effect on the options for safety stock reduction. This is because, in early stages the number of defective items is roughly equal to the mean, which is not critical for the warehouse because it is predictable with near certainty. Since real-time yield information has the greatest utility when the forecast errors are significantly different from zero, quality inspections are more valuable after later stages, when the yield variability is high. Nevertheless, the problem arises that after the 10th stage a final quality inspection exists which means that real-time yield information generated after stage nine would have been available after the 10th stage anyway, which reduces the utility of the information enormously.

If yield variability is constant across all production stages, a quality control measure is most valuable if it lies in the middle of the production process. Notice that early quality control leads to early information about realized yields, but the information is generated out of a small number of stages and therefore limited. A late quality control process in turn leads to a lot

Figure 2 Fraction of Safety Stock Relative to no Additional Inspection



more information about realized yields, but is less valuable since after the 10th stage the final inspection takes place. As a result, it is obvious that the greatest benefits are obtained from an inspection in the middle of the production process, because real-time yield information has been generated from a certain number of stages but there are still stages left before the final quality control.

It is important to keep in mind that the safety stock level can be reduced enormously for a given service level depending on the moments of the yield across the stages and the position of the quality inspection. For a company like Intel, reporting inventories for finished goods of nearly 1.5 billion dollars in the 2014 annual report, a significant safety stock reduction can lead to savings of several million dollars.

After analyzing different production systems with one fixed intermediate quality inspection, our next aim is to determine the optimal position for one or more intermediate control stations in an environment where there has thus been only a final inspection. If the control station is not fixed, production costs, costs for additional quality control, and disposal costs are relevant to holding and backorder costs. Therefore, in the following, we investigate a comprehensive cost model.

5. Optimization Model

In the previous section, we showed how to determine the optimal safety stock level, which influences the amount of holding and backorder costs, for given production systems with a fixed position for the intermediate quality control. In the continued analysis, we focus on the optimization of the number and the positions of the possible inspections. In this situation, not only holding and backorder costs have to be taken into account, but also production costs, costs for quality inspections and disposal costs.

Before we look into more detail on the cost parameters, we will focus on the position of quality inspections within the production process. In the following, the planner can decide after which stage $i = 1, \ldots, n-1$ he or she wants to place quality inspections in addition to the final one after stage n. Let $x = (x_i)_{i=1,\ldots,n}$ be a binary row vector of length n with

$$x_i = \begin{cases} 1, & \text{if a quality control exists after stage } i \\ 0, & \text{otherwise.} \end{cases} \tag{15}$$

If we consider a production time of λ equal to ten periods and model the system as a ten-stage production system with a production time of one unit at each stage, then x_i could look like (0, 0, 1, 0, 0, 0, 1, 0, 1, 1). This would mean that there are quality inspections after stage three, seven, nine and ten (final inspection). In section 2, we introduced $\lambda_j\,(j = 1, \ldots, m)$ reflecting the production time between quality inspection $j-1$ and j. Thus, we would get λ_1, λ_2, λ_3, and λ_4 equal to 3, 4, 2 and 1, which reveals that using x_i leads to the same model as presented in Figure 1. Beyond the definition in (15), we define q_i as $max\{x_k \cdot k | k = 1, \ldots, i-1\}$ $\forall i = 2, \ldots, n$ with q_1 equal to zero. q_i determines after which stage the latest quality inspection before stage i took place ($q_i > 0$) or if there has been non so far ($q_i = 0$).

Based on the position of quality inspections represented by x, the following costs are considered: Variable production costs $P(x)$ occur for each stage of production and each item being produced. We neglect fixed production costs which might arise whenever the production process has to be set up, since we assume the production of only a single product and therefore no product changes within production. At the end of the production process, quality control exists to ensure that only products fulfilling the desired quality standards are stored in the warehouse. Beyond this quality control, the planner can decide whether additional quality inspections should be introduced after other stages. We measure the fixed and variable costs of each quality control. Fixed costs $F(x)$ occur whenever a batch of items, in our case one order, is inspected and variable costs $V(x)$ are charged for each item being checked. Therefore, the costs for quality control depend on the batch size. For every non-conforming item, disposal costs $G(x)$ are taken into account. Within the warehouse holding and backorder costs $C(x)$ are charged. Holding costs occur for each item and period on stock and backorder costs at the end of each period for unsatisfied demand.

Thus, we get the following total cost function $TC(x)$ which has to be minimized depending on the position of quality inspections represented by x:

$$TC(x) = P(x) + F(x) + V(x) + G(x) + C(x). \tag{16}$$

Variable production costs $P(x)$ are charged with parameter p_i for each item produced on production stage i. If no quality control–except the final one–exists, on average the mean order quantity μ_Q is produced on every production stage. If an additional inspection is introduced, all defective items produced so far are sorted out and are no longer part of the production process. Thus, after an inspection, the production costs decrease. Since under a strictly linear control rule an order is placed in every period, production costs occur in each period for all stages of production (with a production time of one time unit on each stage). The average variable production costs per period can be determined by

$$P(x) = \mu_Q \cdot \left[\sum_{i=1}^{n} p_i \cdot \prod_{k=1}^{q_i} \mu_{Z_k} \right], \tag{17}$$

with $\prod_{i=1}^{0} y_i = 1$. Since a quality inspection sorts out all defective items, the production costs are reduced after each control station represented by $\prod_{k=1}^{q_i} \mu_{Z_k}$.

The costs for a quality control are divided into fixed and variable costs. The fixed costs $F(x)$ occur for each batch of items independent of the batch size if an inspection takes place after stage i with the cost parameter f_i:

$$F(x) = \sum_{i=1}^{n} x_i \cdot f_i. \tag{18}$$

Variable inspection costs $V(x)$ can be calculated similarly to the production costs, with the difference that they occur only if there is quality control after production stage i. Variable inspection costs equal

$$V(x) = \mu_Q \cdot \left[\sum_{i=1}^{n} v_i \cdot x_i \cdot \prod_{k=1}^{q_i} \mu_{Z_k} \right], \tag{19}$$

with the cost parameter v_i for each item being checked after stage i.

If an item does not meet the quality requirements, it cannot be sold to the customer and therefore is not stored in the warehouse. The quality control located after the whole production process ensures that all defective items are discarded. If prior inspections exist, the defective items are sorted out earlier. Average disposal costs $G(x)$ occur for each item that is discarded from production with the cost parameter g_i:

$$G(x) = \mu_Q \cdot \left[\sum_{i=1}^{n} g_i \cdot x_i \cdot \left(1 - \prod_{k=1}^{i} \mu_{Z_k} \right) \right]. \tag{20}$$

The holding and backorder costs $C(S)$ have been defined in (9) to be dependent on the optimal pseudo-order-up-to level. The pseudo-order-up-to level S depends on the variance of the inventory level, which in turn depends on x_i as follows:

$$\sigma_I^2(x) = (\lambda + 1)\sigma_D^2 + \sum_{i=1}^{n} \left[\sum_{l=1}^{i} \Lambda_l \cdot \sum_{k=q_i+1}^{i} \sigma_{R_k}^2 \cdot x_i \right]. \tag{21}$$

To determine the optimal number and positions of intermediate quality inspections, a planner has to solve the following model:

$$\min_x TC(x) = P(x) + F(x) + V(x) + G(x) + C(x) \tag{22}$$

$$= \mu_Q \cdot \left[\sum_{i=1}^{n} p_i \cdot \prod_{k=1}^{q_i} \mu_{Z_k} \right] \tag{23}$$

$$+ \sum_{i=1}^{n} x_i \cdot f_i \tag{24}$$

$$+ \mu_Q \cdot \left[\sum_{i=1}^{n} v_i \cdot x_i \cdot \prod_{k=1}^{q_i} \mu_{Z_k} \right] \tag{25}$$

$$+ \mu_Q \cdot \left[\sum_{i=1}^{n} g_i \cdot x_i \cdot \left(1 - \prod_{k=1}^{i} \mu_{Z_k} \right) \right] \tag{26}$$

$$+ h \int_0^{\infty} y \varphi_I(y) dy - b \int_{-\infty}^{0} y \varphi_I(y) dy \tag{27}$$

subject to

$$x_n = 1 \tag{28}$$

$$x_i \in \{0, 1\} \; \forall i = 1, \ldots, n. \tag{29}$$

The constraint in (28) reflects that there is always a final quality control after the nth production stage.

6. Numerical Analysis

In the following, we optimize the position of one intermediate quality control and afterwards determine the optimal number and position of several inspections.

6.1. Optimal Position of One Intermediate Control Test

In section 4, we determined the optimal safety stock level for a given production system with one intermediate quality control. Now, we consider a production system where only a final quality control test exists, whereas the planner has the possibility to place one additional intermediate inspection. Due to a limited budget, it may not be possible to place more control stations. The question is, after which stage this additional inspection should be placed in order to generate the largest benefits. Unlike the model in section 4, not only holding and backorder costs have to be optimized, but also production costs, inspection costs, and disposal costs. Thus, the cost model presented in equation (22) is considered with the additional restriction:

$$\sum_{i=1}^{n-1} x_i = 1. \tag{30}$$

Since the yield plays an important role when answering the question of where to place inspections, we assume different shapes in the yield variance as already shown in Table 1. Furthermore, we assume the same parameter setting as in section 4: we consider λ equal to 10 production periods, a mean demand μ_D and a corresponding coefficient of variation ρ_D of 20 and 0.1, a service level α of 0.99, a mean yield μ_{Z_i} per production stage of 0.95 ($\forall i = 1, \ldots, n$) and a coefficient of variation of the yield on the first production stage for decreasing yield variance ρ_{dec,Z_1} of 0.25. We assume the following setting for the cost parameters: production costs p_i equal to 2 ($\forall i = 1, \ldots, n$), fixed costs f_i for inspecting one batch of items of 5 ($\forall i = 1, \ldots, n$), variable inspection costs v_i for one item after stage i of 0.1 times p_i which equals 0.2 ($\forall i = 1, \ldots, n$) and disposal costs g_i of 0 ($\forall i = 1, \ldots, n$). The costs for quality inspections are a fraction of the production costs (see e.g., Schiffauerova and Thomson 2006). This assumption leads to notably lower costs for inspections than for production costs. We neglect disposal costs since the final quality inspection leads to a 100% disposal of all defective items. Disposal costs play a role only if the costs differ from one stage to another since waste separation becomes more difficult at a later stage of production.

Despite the above parameter setting, we need an assumption about the distribution of the inventory level to determine holding and backorder costs (formula (9)). Like Inderfurth and Kiesmüller (2015) we assume a normally distributed inventory level which leads to the following formula:

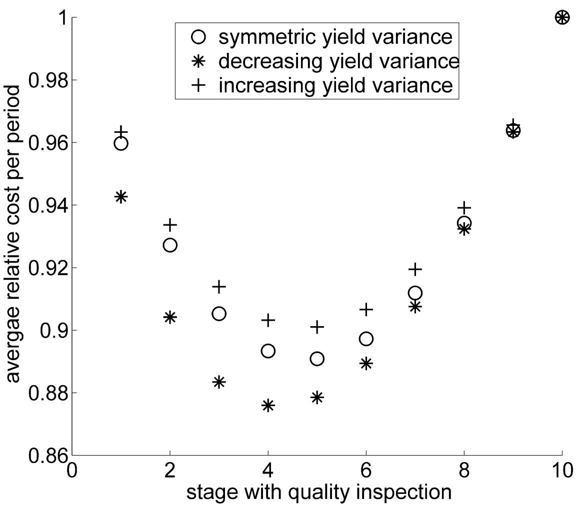
$$C(x) = (h+b)\left[\sigma_I(x)\cdot\varphi\left(\frac{-\mu_I}{\sigma_I(x)}\right) + \mu_I\cdot\left(1-\Phi\left(\frac{-\mu_I}{\sigma_I(x)}\right)\right)\right] - b\cdot\mu_I. \tag{31}$$

We run a complete enumeration to determine the optimal position of one intermediate control. Negligible computation times make this possible, using the steady-state approach for determining the optimal safety stock level and therefore holding and backorder costs.

Figure 3 shows the average cost computed with formulae (24)–(29) for different shapes of the yield variance across the production stages.

It is clear that the effect of different yield scenarios vanishes distinctly when all cost parameters are involved (compare with Figure 2). This is because of the production costs, which are the main cost driver beside the holding and backorder costs. Each item that is produced at one stage generates costs. Unlike the possibility of reducing the safety stock, the possibility of reducing production costs is heavily influenced by the mean yield. Every defective item that is discarded reduces production costs at later stages, whereas only defective items differing from the mean

Figure 3 Influence of the Inspection Station on the Average Cost



lead to changing safety stock levels. Changing the mean yield strongly influences the optimal position of the intermediate inspection. Figure 4 shows how the optimal position of the quality control station changes as the mean yield varies. If the mean yield is very high ($\mu_{Z_i} = 0.99, \forall i = 1, \ldots, n$), which is desirable for an economically sound company, the production costs can be influenced only marginally and holding and backorder costs become the dominant factor (see also Figure 2). On the other hand, if the mean yield is relatively low ($\mu_{Z_i} = 0.5, \forall i = 1, \ldots, n$) the production costs dominate all other cost parameters. In this case, the allocation of the yield variance across the stages has no influence on the total costs, which is why it is best to place an inspection after early production stages.

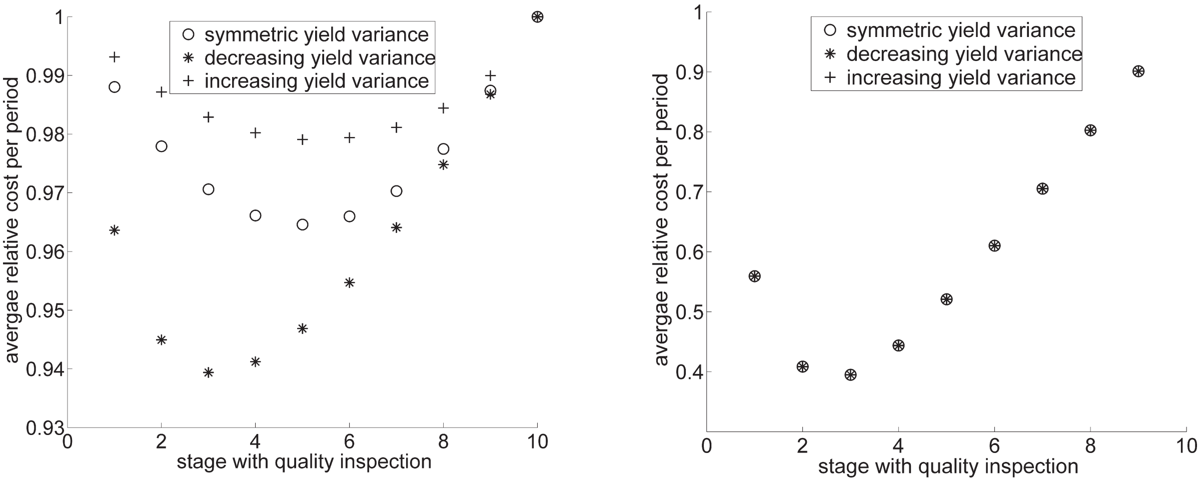
To sum up, for a low mean yield such as in the Samsung example, production costs are the main cost driver and thus one intermediate inspection should be placed after the first stages of production. For a high mean yield such as the one in the Intel example, the holding and backorder costs, and therefore the safety stock level, determine the optimal position of the inspection. Since the safety stock level is strongly dependent on the shape of the yield variance across the production stages, the position of an inspection should be determined with respect to the minimum safety stock level where there is a high mean yield.

6.2. Optimal Number and Position of Multiple Quality Inspections

After answering the question of the optimal position of one intermediate quality control, we focus on the optimal quantity and position of several inspections. This analysis can work to support decisions in cases

Please Cite this article in press as: Sonntag, D., G. P. Kiesmüller. The Influence of Quality Inspections on the Optimal Safety Stock Level. *Production and Operations Management* (2017), doi 10.1111/poms.12691

Figure 4 Influence of the Mean Yield Rate on the Optimal Inspection Station



(a) $\mu_{Z_i} = 0.99, \; \forall i = 1, \ldots, n$ (b) $\mu_{Z_i} = 0.50, \; \forall i = 1, \ldots, n$

where a given number of inspections has to be located optimally.

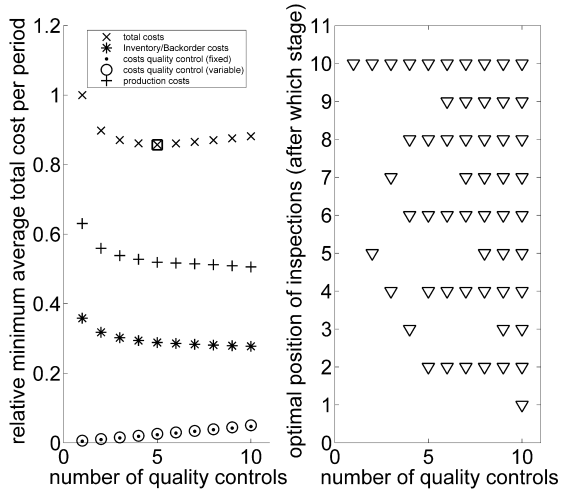
We use the same parameter setting as in the setting with one intermediate control despite the coefficient of variation of the yield ρ_{Z_i} which equals 0.25 ($\forall i = 1, \ldots, n$) since we consider only production systems with constant yield across the production stages.

We used a complete enumeration to calculate the total costs for all possible combinations of control stations and their positions. Since the final quality control is fixed, this leads to $\sum_{i=0}^{n-1} \frac{(n-1)!}{i!(n-1-i)!}$ equal to 512 scenarios if a maximum of $n - 1$ equal to nine quality controls can be placed. For every number of installed inspections, we determined the one that minimizes the total costs. This solution is presented in Figure 5.

The first diagram reflects the minimum total costs per period under the assumption that only a given number of inspections can be introduced. If there is only one control station, this inspection equals the final one after stage ten, and provides a benchmark. The square marks the point where the total costs are minimized over all possible solutions. The second diagram shows after which production stages the control stations should be placed to determine the minimum total costs for the corresponding number of inspections.

It is clear that it is best to introduce five quality controls and space them equally over the production stages. When deciding about the number of control stations it is obvious that the marginal utility of an additional quality control decreases until the optimum is reached. For more inspections, the running costs increase slightly. Nevertheless, the reduction in

Figure 5 Optimal Number and Position of Quality Inspections



costs between three, four, and five quality inspections is very small whereas the marginal utility for less inspections is significantly higher.

The optimal number of control stations is highly dependent on the input parameters. Changing the cost parameters or the mean/variance of the yield leads to different results. Nevertheless, for a limited number of inspections, it is best to place them equally across the production stages. After the fifth inspection is placed, the parameter setting determines whether further inspections are located between existing control stations in the beginning or at the end of the production process.

Please Cite this article in press as: Sonntag, D., G. P. Kiesmüller. The Influence of Quality Inspections on the Optimal Safety Stock Level. *Production and Operations Management* (2017), doi 10.1111/poms.12691

It is important to be aware of the fact that every additional quality inspection leads to the reallocation of all existing inspections. Thus, it is not possible to simply add additional inspections at later points in time. If we assume that production systems work with high mean yields, which should be the target for an economically sound company, the decision about the number and position of quality inspections should be driven primarily by holding and backorder costs and therefore by the amount of safety stock. This makes clear the need for an appropriate inventory management.

7. Summary and Outlook

We examined an in-house multi-stage serial production system with random yield and a warehouse for the final product. Within the warehouse, a linear inflation policy was used to determine the order quantities. Our contribution was to (1) analyze certain given production systems with intermediate quality control and we show how the optimal safety stock can be computed, (2) optimize the position of one inspection, (3) determine the optimal number and (4) the optimal position of quality inspections across the production stages.

The results are as follows: (1) the safety stock level heavily depends on the shape of the yield across the production stages. In the case of decreasing yield, quality inspections are most valuable. The examples show a maximum safety stock reduction of over 30% which can be even larger depending on the parameter setting. Where there is increasing yield variability across the stages, the position of one intermediate inspection has almost no effect on the safety stock. For symmetric yield variability, the safety stock level can be reduced by placing an inspection in the middle of the production process. (2) The effect of the safety stock reduction vanishes when production costs and the costs for quality inspections are taken into account in addition to holding and backorder costs, since production costs are the main cost driver. Nevertheless, as the mean production yield increases, holding and backorder costs become dominant and determine the placement of an inspection. (3) The marginal cost reduction of every additional worthwhile quality inspection decreases rapidly. (4) In general, for symmetric yields, it is best to locate control stations equally spaced across the production stages to minimize overall costs.

Since the steady-state approach is a powerful tool when it comes to the calculation of safety stocks, it can be easily used to answer other research questions concerning multi-stage production systems with random yield. We modeled an n-stage production system where inspection stations can be placed after each production stage. However, in some situations, this might not be possible due to the design of the production process. Think of two working steps which are closely related because they are processed on one machine. In such a case, it might not be reasonable to place an inspection between the stages, thus interrupting production. To see how the number of possible positions influences the required safety stock level and the costs, μ_{Z_i} and $\sigma^2_{Z_i}$ have to be adjusted (for all $i = 1, 2, \ldots, n$). Beyond that, the analysis is similar to the numerical analysis in this study.

If defective products remain in the production process and continue to be produced even though they are defective, production costs increase. In this study, we have introduced quality inspections as an opportunity to reduce production as well as overall costs. A second option for cost reduction is increasing the yield. Bohn and Terwiesch (1999) mention that a 1% increase in yields can lead to a 6% increase in gross revenue. Thus, a decision maker with limited budget has to choose between the implementation of inspection stations and the improvement of yields within the system. Both approaches can lead to an overall cost reduction. Our approach, together with the cost model we have presented, is able to determine, what situations warrant the introduction of quality inspections and when it is better to concentrate on the reduction of yield losses. If one reduces yield losses, the question remains where within the system it is best to do so. There is also the option of reducing yield losses as well as introducing quality inspections, in which case it would be interesting to determine whether the decisions affect each other and if so, how.

Despite the issues mentioned above, concerning further analysis of the production system, using the steady-state approach, future research should take into consideration that inspections are not necessarily perfect. Sometimes defective items are not discarded although they have been checked through an inspection; conversely perfectly produced items can be discarded by mistake.

Acknowledgments

The authors thank the department editor, the senior editor, and the two reviewers for their valuable feedback on the article. The authors also thank Srinagesh Gavirneni from Samuel Curtis Johnson Graduate School of Management (Cornell University) for his valuable comments and suggestions which helped to improve the quality of the study.

Appendix A: Proof of Lemma 2

In order to improve readability we will use $E[X]$ for the expectation of a random variable and $VAR[X]$ for its variance.

The proof of Lemma 2 is based on an induction and we start with the case $n = 2$. Then the forecast error for the first stage ($i = 1$) according to Equation (4) is given as

$$\begin{aligned} R_{1,t+1-\Lambda_1} =& Y_0(Q_{t+1-\Lambda_1}) - Y_1(Q_{t+1-\Lambda_1}) \\ =& \mu_{Z_1}\mu_{Z_2}Q_{t+1-\Lambda_1} - Z_1\mu_{Z_2}Q_{t+1-\Lambda_1} \end{aligned}. \tag{A.1}$$

This yields the following variance in a steady-state:

$$\begin{aligned} VAR[R_1] &= VAR[\mu_{Z_1}\mu_{Z_2}Q - Z_1\mu_{Z_2}Q] \\ &= E[(\mu_{Z_1}\mu_{Z_2}Q - Z_1\mu_{Z_2}Q)^2] \\ &\quad - E[\mu_{Z_1}\mu_{Z_2}Q - Z_1\mu_{Z_2}Q]^2 \\ &= E[Q^2]E[\mu_{Z_1}^2\mu_{Z_2}^2 - 2\mu_{Z_1}\mu_{Z_2}Z_1\mu_{Z_2} + Z_1^2\mu_{Z_2}^2] \\ &\quad - E[Q]^2(\mu_{Z_1}\mu_{Z_2} - \mu_{Z_1}\mu_{Z_2})^2 \\ &= E[Q^2](E[Z_1^2]\mu_{Z_2}^2 - \mu_{Z_1}^2\mu_{Z_2}^2) \\ &= (VAR[Q] + E[Q]^2)VAR[Z_1]\mu_{Z_2}^2 \\ &= (\sigma_Q^2 + \mu_Q^2)\left[\prod_{j=1}^{1-1}(\sigma_{Z_j}^2 + \mu_{Z_j}^2)\cdot \prod_{l=1+1}^{2}\mu_{Z_l}^2\cdot\sigma_{Z_i}^2\right] \end{aligned}. \tag{A.2}$$

For the forecast error at the second stage we obtain:

$$\begin{aligned} R_{2,t+1-\Lambda_2} =& Y_1(Q_{t+1-\Lambda_2}) - Y_2(Q_{t+1-\Lambda_2}) \\ =& Z_1\mu_{Z_2}Q_{t+1-\Lambda_2} - Z_1Z_2Q_{t+1-\Lambda_2} \end{aligned}. \tag{A.3}$$

This leads to:

$$\begin{aligned} VAR[R_2] =& VAR[Z_1\mu_{Z_2}Q - Z_1Z_2Q] \\ =& E[(Q(Z_1\mu_{Z_2} - Z_1Z_2))^2] \\ &- E[(Q(Z_1\mu_{Z_2} - Z_1Z_2))]^2 \end{aligned}. \tag{A.4}$$

Since the yield is iid we get

$$\begin{aligned} VAR[R_2] &= E[Q^2(Z_1\mu_{Z_2} - Z_1Z_2)^2] \\ &= E[Q^2]E[Z_1^2](\mu_{Z_2}^2 - 2\mu_{Z_2}^2 + E[Z_2^2]) \\ &= (\sigma_Q^2 + \mu_Q^2)(\sigma_{Z_1}^2 + \mu_{Z_1}^2)\sigma_{Z_2}^2 \\ &= (\sigma_Q^2 + \mu_Q^2)\left[\prod_{j=1}^{2-1}(\sigma_{Z_j}^2 + \mu_{Z_j}^2)\cdot \prod_{l=2+1}^{2}\mu_{Z_l}^2\cdot\sigma_{Z_2}^2\right] \end{aligned}. \tag{A.5}$$

which means the statement in Lemma 2 is true for $n = 1, 2$. Let us now assume that Equation (4) is true for an arbitrary n and i, $i = 1, \ldots, n$. We will show that it then follows that it is also true for $n + 1$. To make clear how many stages are included in the system considered, we denote this number as a superscript.

$$\begin{aligned} VAR[R_i^{(n+1)}] &= VAR[Y_{i-1}^{(n+1)}(Q) - Y_i^{(n+1)}(Q)] \\ &= VAR\left[\prod_{j=1}^{i-1}Z_j\prod_{l=i}^{n+1}\mu_{Z_l}Q - \prod_{j=1}^{i}Z_j\prod_{l=i+1}^{n+1}\mu_{Z_l}Q\right] \\ &= E\left[Q^2\left(\prod_{j=1}^{i-1}Z_j\prod_{l=i}^{n}\mu_{Z_l}\mu_{Z_{n+1}}\prod_{j=1}^{i}Z_j\prod_{l=i+1}^{n}\mu_{Z_l}\mu_{Z_{n+1}}\right)^2\right] \\ &\quad - E\left[\prod_{j=1}^{i-1}Z_j\prod_{l=i}^{n}\mu_{Z_l}\mu_{Z_{n+1}}Q - \prod_{j=1}^{i}Z_j\prod_{l=i+1}^{n}\mu_{Z_l}\mu_{Z_{n+1}}Q\right]^2 \\ &= \mu_{Z_{n+1}}^2\left\{E\left[Q^2\left(\prod_{j=1}^{i-1}Z_j\prod_{l=i}^{n}\mu_{Z_l} - \prod_{j=1}^{i}Z_j\prod_{l=i+1}^{n}\mu_{Z_l}\right)^2\right]\right. \\ &\quad \left. - E\left[\prod_{j=1}^{i-1}Z_j\prod_{l=i}^{n}\mu_{Z_l}Q - \prod_{j=1}^{i}Z_j\prod_{l=i+1}^{n}\mu_{Z_l}Q\right]^2\right\} \\ &= \mu_{Z_{n+1}}^2 VAR\left[R_i^{(n)}\right] \\ &= \mu_{Z_{n+1}}^2(\sigma_Q^2 + \mu_Q^2)\left[\prod_{j=1}^{i-1}(\sigma_{Z_j}^2 + \mu_{Z_j}^2)\cdot \prod_{l=i+1}^{n}\mu_{Z_l}^2\cdot\sigma_{Z_i}^2\right] \\ &= (\sigma_Q^2 + \mu_Q^2)\left[\prod_{j=1}^{i-1}(\sigma_{Z_j}^2 + \mu_{Z_j}^2)\cdot \prod_{l=i+1}^{n+1}\mu_{Z_l}^2\cdot\sigma_{Z_i}^2\right] \end{aligned}. \tag{A.6}$$

Appendix B. Proof of Lemma 3

Since we assume a steady-state, the variance of the order quantity equals the variance of the inflated adapted demand, as defined in Equation (6). Thus, for the variance of the order quantity we get:

$$VAR[Q] = F^2 VAR[\sum_{i=1}^{n} R_i] + F^2 VAR[D]. \tag{B.1}$$

The variance of the sum of the forecast errors is

$$\begin{aligned} VAR\left[\sum_{i=1}^{n}R_i\right] =& E\left[\left(\sum_{i=1}^{n}R_i\right)^2\right] - \left(E\left[\sum_{i=1}^{n}R_i\right]\right)^2 \\ =& E\left[\left(\sum_{i=1}^{n}R_i\right)^2\right] \\ =& \sigma_Z^2\cdot[VAR[Q] + E[Q]^2] \\ =& \sigma_Z^2 VAR[Q] + \mu_D^2\rho_Z^2 \end{aligned} \tag{B.2}$$

From Equations (B.1) to (B.2) we get:

$$\begin{aligned} VAR[Q] &= F^2 VAR\left[\sum_{i=1}^{n}R_i\right] + F^2 VAR[D] \\ \sigma_Q^2 &= F^2(\sigma_Z^2\sigma_Q^2 + \mu_D^2\rho_Z^2) + F^2\sigma_D^2 \\ &= \frac{F^2(\rho_Z^2\mu_D^2 + \sigma_D^2)}{1 - \rho_Z^2}. \end{aligned} \tag{B.3}$$

Please Cite this article in press as: Sonntag, D., G. P. Kiesmüller. The Influence of Quality Inspections on the Optimal Safety Stock Level. *Production and Operations Management* (2017), doi 10.1111/poms.12691

whereby ρ_Z^2 equals the squared coefficient of variation equal to σ^2/μ_Z^2 with

$$\begin{aligned} VAR\left[\prod_{i=1}^{n} Z_i\right] &= E\left[\left(\prod_{i=1}^{n} Z_i\right)^2\right] - \left(E\left[\left(\prod_{i=1}^{n} Z_i\right)\right]\right)^2 \\ &= E\left[\prod_{i=1}^{n} Z_i^2\right] - \left(\prod_{i=1}^{n}\left(E[Z_i]\right)\right)^2 \\ &= \prod_{i=1}^{n}(VAR[Z_i] + (E[Z_i])^2) - \prod_{i=1}^{n}(E[Z_i])^2 \\ &= \prod_{i=1}^{n}(\sigma_{Z_i}^2 + \mu_{Z_i}^2) - \prod_{i=1}^{n}\mu_{Z_i}^2 \end{aligned} \qquad (B.4)$$

References

Anderson, E. W., C. Fornell, D. R. Lehmann. 1994. Customer satisfaction, market share, and profitability: Findings from sweden. *J. Market.* **58**, 53–66.

Axsäter, S. 2007. *Inventory Control Volume 90*. Springer Science & Business Media, New York.

Ben-Daya, M., S. M. Noman. 2008. Integrated inventory and inspection policies for stochastic demand. *Eur. J. Oper. Res.* **185**, 159–169.

Bohn, R. E., C. Terwiesch. 1999. The economics of yield-driven processes. *J. Oper. Manag.* **18**, 41–59.

Bollapragada, S., T. E. Morton. 1999. Myopic heuristics for the random yield problem. *Oper. Res.* **47**, 713–722.

Choi, H.-C. P., J. D. Blocher, S. Gavirneni. 2008. Value of sharing production yield information in a serial supply chain. *Prod. Oper. Manag.* **17**, 614–625.

Dettenbach, M., U. W. Thonemann. 2015. The value of real time yield information in multi-stage inventory systems-exact and heuristic approaches. *Eur. J. Oper. Res.* **240**, 72–83.

Ehrhardt, R., L. Taube. 1987. An inventory model with random replenishment quantities. *Int. J. Prod. Res.* **25**, 1795–1803.

Eppen, G. D., E. G. Hurst Jr. 1974. Optimal location of inspection stations in a multistage production process. *Management Sci.* **20**, 1194–1200.

Eurostat. 2016. Production and international trade in high-tech products. Available at http://ec.europa.eu/eurostat/statistics-explained/index.php/Production_and_international_trade_in_high-tech_products (accessed date December 13, 2016).

Foremski, T. 2012. Intel's battle with ARM is about making its future fabs viable. Available at http://www.siliconvalleywatcher.com/mt/archives/2012/09/intels_battle_w.php (accessed date December 13, 2016).

Henig, M., Y. Gerchak. 1990. The structure of periodic review policies in the presence of random yield. *Oper. Res.* **38**, 634–643.

Huh, W. T., M. Nagarajan. 2010. Technical note-linear inflation rules for the random yield problem: Analysis and computations. *Oper. Res.* **58**, 244–251.

Inderfurth, K., G. P. Kiesmüller. 2015. Exact and heuristic linear-inflation policies for an inventory model with random yield and arbitrary lead times. *Eur. J. Oper. Res.* **245**, 109–120.

Leachman, R. C., D. A. Hodges. 1996. Benchmarking semiconductor manufacturing. *IEEE Trans. Semicond. Manuf.* **9**, 158–169.

Lindsay, G. F., A. B. Bishop. 1964. Allocation of screening inspection effort-a dynamic-programming approach. *Management Sci.* **10**, 342–352.

Mandroli, S. S., A. K. Shrivastava, Y. Ding. 2006. A survey of inspection strategy and sensor distribution studies in discrete-part manufacturing processes. *IIE Trans.* **38**, 309–328.

McNutt, C. 2015. Rumor: Low yield 3D glass forcing Samsung to look for additional suppliers for Galaxy S6 Edge. Available at http://www.androidheadlines.com/2015/03/rumor-low-yield-3d-glass-forcing-samsung-look-additional-suppliers-galaxy-s6-edge.html (accessed date December 13, 2016).

Peters, M. H., H. Schneider, K. Tang. 1988. Joint determination of optimal inventory and quality control policy. *Management Sci.* **34**, 991–1004.

Schiffauerova, A., V. Thomson. 2006. A review of research on cost of quality models and best practices. *Int. J. Qual. Reliab. Manag.* **23**, 647–669.

Shetwan, A. G., V. I. Vitanov, B. Tjahjono. 2011. Allocation of quality control stations in multistage manufacturing systems. *Comput. Ind. Eng.* **60**, 473–484.

Stokowski, S., M. Vaez-Iravani. 1998. Wafer inspection technology challenges for ulsi manufacturing. *AIP Conf. Proc.* **449**, 405–418.

Taylor, S. A., T. L. Baker. 1994. An assessment of the relationship between service quality and customer satisfaction in the formation of consumers' purchase intentions. *J. Retail.* **70**, 163–178.

Tempelmeier, H. 2006. *Inventory Management in Supply Networks: Problems, Models, Solutions*. Books on Demand, Norderstedt.

White, C. C., T. Cheong. 2012. In-transit perishable product inspection. *Logist. Transp. Rev.* **48**, 310–330.

Yano, C. A., H. L. Lee. 1995. Lot sizing with random yields: A review. *Oper. Res.* **43**, 311–334.

Zipkin, P. H. 2000. *Foundations of Inventory Management*. McGraw-Hill, New York.

Please Cite this article in press as: Sonntag, D., G. P. Kiesmüller. The Influence of Quality Inspections on the Optimal Safety Stock Level. *Production and Operations Management* (2017), doi 10.1111/poms.12691

4. Investing into quality inspections or reducing the yield variability?

Sonntag, D. (2017). Investing into quality inspections or reducing the yield variability?. Proceedings of the Logistikmanagement 2017, 258-266.

Investing into quality inspections or reducing the yield variability?

Sonntag, Danja, Otto-von-Guericke-University Magdeburg, Faculty of Economics and Management, Chair of Operations Management, Danja.Sonntag@ovgu.de

In this paper, a multi-stage make-to-stock production system is considered. Due to imperfect production, each production stage produces a random fraction of defective items. A quality inspection, screening all produced products, exists at the end of the production process. While defective items are disposed of, items conforming to quality standards are stored in a warehouse to satisfy stochastic customer demand. Random demands as well as random yields make it necessary to hold some safety stock to reduce the probability of stock-outs. On the production site a planner has two opportunities to reduce uncertainty about the production output and therefore the safety stock level. He can implement intermediate quality inspections and/or reduce the yield variance under a given budget constraint. Thus, he has to optimize the number and locations of additional inspections as well as the amount of money spend on each production stage to reduce the yield uncertainty. The problem is solved using a fix-and-optimize solution approach. The results indicate that depending on the extent of yield variability either implementing additional inspections or reducing the yield variance is more favorable. Moreover, significant safety stock reductions can be recognized depending on the investment strategy and the current yield variability.

Keywords: quality inspection, yield reduction, safety stock

1 Introduction

Nowadays product quality becomes more and more important due to high customer requirements. In addition with significantly more complex production processes, especially the high-tech industry has to deal with partially high yield losses. Yield losses occur if produced items do not correspond to given quality standards. Samsung, for example, has problems producing the curved glass for a new cell phone series because of low yields of less than 50 % (McNutt, 2015). But not every industry and not every company has to deal with such low yields; Intel for example has a production yield in the high 90 % range (Foremski, 2012). To ensure that only high quality products are sold to the customers, a quality inspection exists at the end of the production process checking all produced items. The products meeting the quality requirements are stored in a warehouse to satisfy stochastic customer demand whereas defective products are disposed of.

Besides the uncertain supply there is also uncertainty about the demand. Because of these uncertainties, costly safety stock is required to minimize the probability of stock outs. A planner has two opportunities to reduce the safety stock level by capital investments: On one hand, he can decrease the yield variability through investments into modern production technology, which leads to a realized production outcome that differs less from the expected one. Thus, a planner can improve the production process itself. On the other hand, he can introduce further quality inspections, which results in earlier information about realized yields. In this case, the production process remains the same without reducing the number of defectives being produced but the information system is improved, which enables the planner to react to unexpectedly low or high yield losses at early production stages. Both strategies as well as a combination of them reduce the uncertainty in the system and therefore the safety stock level. For a company like Intel, reporting inventories for finished goods of more than 1.5 billion dollars in the 2016 annual report (Intel, 2017), a safety stock reduction can lead to significant savings.

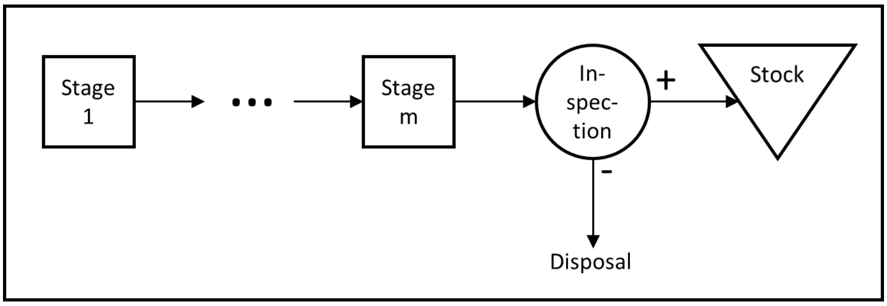
Within this paper, the objective is to minimize the required amount of safety stock in the warehouse by determining the number and location of additional intermediate quality inspections and the amount of yield variance reduction on each production stage under a given budget constraint. The presented research is closely related to the literature on random yield problems (see Henig / Gerchak, 1990; Yano / Lee, 1995; Huh / Nagarajan, 2010), especially those focusing on positive production times (Bollapragada /Morton, 1999; Choi et al., 2008; Inderfurth / Vogelgesang, 2013; Inderfurth / Kiesmüller, 2015; Dettenbach / Thoneman, 2015; Sonntag / Kiesmüller, 2017). Positive production times in combination with random yield increase the complexity of the system tremendously because the inventory position contains only an expectation about outstanding orders and is therefore uncertain (Sonntag / Kiesmüller, 2016). Because of this complexity, existing literature, analyzing the effect of an investment to reduce yield variability, consider much simpler production models e.g. without positive production times (see e.g. Lin / Hou, 2005; Kulkarni, 2008). Nonetheless, there is no literature combining a yield variability reduction and additional quality inspections under stochastic demand in make-to-stock environments.
The problem is solved by applying a fix-and-optimize solution approach, which is suitable to solve mixed-integer programs as given in this paper (for more details about the fix-and-optimize approach see Helber / Sahling, 2010). In a first step, all possible sets of additional inspections under a given budget constraint are determined. Afterwards, for each of these fixed sets, the allocation of the remaining budget to production stages is optimized with respect to minimizing average holding and backorder cost per period. To determine on which stage the remaining budget should be invested and therefore the yield variability be reduced, a nonlinear program is solved.
The remainder of the paper is organized as follows: In section 2, the model is presented followed by the solution approach in section 3. Subsequently, results of a numerical

analysis are shown and some managerial insights are given in section 4. The paper concludes with a summary in section 5.

2 Model formulation

The considered production system, consisting of m production stages, a quality inspection after the last production stage and a warehouse for the final product, is presented in Figure 1.

Figure 1: Production system



Because of imperfect production, defective items can be produced during each production stage. These products remain in the process until the final quality control at the end of production, which inspects all produced items regarding their required quality standards. The quality inspection eliminates all defective items with certainty. While defective products are disposed of and leave the production system, perfectly produced items are stored in a warehouse to satisfy stochastic customer demand. The warehouse can replenish stock periodically at the beginning of each period e.g. one day using an adjusted base stock policy. Hereinafter, the described situation illustrated in Figure 1 is used as a benchmark. In the following, the whole system will be described in more detail.

The stochastic customer demand, which has to be satisfied, is modelled with a continuous random variable in each period. Demand across the periods is independent and identically distributed with mean μ_D and standard deviation σ_D and demand which cannot be satisfied directly from stock is backlogged. For the production process we assume that each production stage m involves a positive production time λ_i ($\forall i = 1, \dots, m$) and random yield, which is assumed to be stochastically proportional with mean $\mu_{Z,i}$ and standard deviation $\sigma_{Z,i}$ ($\forall i = 1, \dots, m$). In a stochastic proportional yield model the production output Q depends on the input as follows: $Y(Q) = Z \cdot Q$, where Z ($Z \in [0,1]$) is a random variable called yield factor. The system has a total production time of λ periods ($\lambda = \sum_{i=1}^{m} \lambda_i$) as well as a total yield factor with mean μ_Z and standard deviation σ_Z equal to $\prod_{i=1}^{m} \mu_{Z,i}$ and $\prod_{i=1}^{m}(\sigma_{Z,i}^2 + \mu_{Z,i}^2) - \prod_{i=1}^{m} \mu_{Z,i}^2$, respectively

(Sonntag / Kiesmüller, 2017). The sequence of events in one period is as follows: at the beginning of each period the order placed λ periods before arrives, the realized yield is observed and the inventory position is updated. Based on the new value of the inventory position an order is placed. During the period, demand occurs and at the end of the period holding and backorder cost are charged based on the inventory level.
To replenish stock a production of quantity Q is released periodically according to a linear inflation policy. A linear inflation policy shows a very good performance (Huh / Nagarajan, 2010) and is used because the optimal ordering policy under random yield does not possess a simple structure (Henig / Gerchak, 1990). The ordering policy is an adapted form of a classical base stock policy whereas the order quantity Q_t in period t equals

$$Q_t = \begin{cases} F(S - IP_t) & , IP_t < S \\ 0 & , IP_t \geq S \end{cases} \tag{1}$$

S is the base stock level and IP_t the inventory position at the beginning of period t.
F is called the yield inflation factor usually set to $1/\mu_Z$ (e.g. Huh / Nagarajan, 2010; Inderfurth / Vogelgesang, 2013; Inderfurth / Kiesmüller, 2015). The yield inflation factor compensates for the expected yield losses occurring during the production process. For the benchmark scenario in Figure 1, the inventory position IP_t before ordering is defined as

$$IP_t = IL_{t-1} + Z_{t-\lambda} Q_{t-\lambda} + \sum_{l=1}^{\lambda-1} \mu_Z Q_{t-l} \tag{2}$$

according to Inderfurth / Kiesmüller (2015). IL_{t-1} is the inventory level at the end of period $t-1$. The second term reflects the order placed λ periods before, which is delivered in period t. The moment an order arrives in the warehouse, the corresponding realized yield $Z_{t-\lambda}$ can be observed. The last term equals the expected yield of all outstanding orders. Since no information about realized yields is available before or during production, an expectation μ_Z about the yield factors is included in the inventory position.
The inventory level $IL_{t+\lambda}$ at the end of period $t+\lambda$ is given as

$$IL_{t+\lambda} = IP_t + \mu_Z Q_t - \sum_{l=0}^{\lambda} D_{t+l} - \sum_{l=0}^{\lambda-1} R_{t-l} \tag{3}$$

with the period demand D_t and the forecast error R_t defined as $\mu_Z Q_t - Z_t Q_t$. The forecast error replaces the expected yield by the realized yield once the products pass the quality inspection and thus realized yields can be observed. This update is necessary because when placing an order only the expected yield is included in the inventory position ($\sum_{l=1}^{\lambda-1} \mu_Z Q_{t-l}$). When the planner observes how many good items leave the production process, he can replace the estimated outcome with the realized one. This update is expressed by the forecast error R_t (for more details see Inderfurth / Kiesmüller, 2015 or Sonntag / Kiesmüller, 2016). Due to uncertainties in the inventory position, production systems with positive production times are very complex and therefore difficult to analyze.

As mentioned before, a planner has two opportunities to reduce the uncertainty in the system. On one hand, he can introduce further quality inspections to improve the information system and therefore get earlier information about realized yields. On the other hand, he can reduce the yield variability and therefore improve the production process.
Let $\boldsymbol{n}$ denote a binary vector of size $m-1$, reflecting if an additional quality inspection exists after production stage i $(i=1,\dots,m-1)$ or not. The implementation of each inspection involves cost of cqc_i $(\forall i=1,\dots,m-1)$ if the inspection is located after stage i $(i=1,\dots,m-1)$. Besides, $\boldsymbol{v}$ is a binary vector of size m, containing the amount invested into yield variability reduction on each production stage i $(i=1,\dots,m)$. Reducing the yield variability for stage i $(i=1,\dots,m)$ from $\sigma_{Z,i}$ to $\sigma^*_{Z,i}$ imply cost of $1/c_i$ defined as the fraction of yield standard deviation reduction per monetary unit increase in investment. The relation between the yield standard deviation and the capital investment in the yield standard deviation reduction $\xi(\sigma^*_{Z,i})$ is defined according to Lin / Hou (2005):

$$\xi\left(\sigma^*_{Z,i}\right) = c_i \cdot \ln\left(\frac{\sigma_{Z,i}}{\sigma^*_{Z,i}}\right), \qquad 0 < \sigma^*_{Z,i} \leq \sigma_{Z,i} \tag{4}$$

Thus, $\boldsymbol{v}$ equals $\left(\xi\left(\sigma^*_{Z,1}\right), \xi\left(\sigma^*_{Z,2}\right), \dots, \xi\left(\sigma^*_{Z,m}\right)\right)$.

The investment into further quality inspections as well as into yield variability reduction is limited due to a given budget B. The objective is to determine the investment strategy minimizing holding and backorder cost C, which is similar to minimizing the required safety stock level. The total average holding and backorder cost depend on the number of additional inspections represented by $\boldsymbol{n}$, the reduction of the yield variability reflected by $\boldsymbol{v}$ and the base stock level S. The total average holding and backorder cost per period C can be calculated based on the expected inventory level as follows:

$$C(S,\boldsymbol{n},\boldsymbol{v}) = hE\left[\left(IL(S,\boldsymbol{n},\boldsymbol{v})\right)^+\right] + bE\left[\left(-IL(S,\boldsymbol{n},\boldsymbol{v})\right)^+\right] \tag{5}$$

$(M)^+$ is defined as *max{0,M}* whereas h and b denote the unit holding and unit backorder cost, respectively. Thus, the optimization model is defined as:

$$\min C(S, \boldsymbol{n}, \boldsymbol{v}) \tag{6}$$
$$s.t. \sum_{i=1}^{m-1} cqc_i \cdot \boldsymbol{n}_i + \sum_{i=1}^{m} \boldsymbol{v}_i \leq B$$

In the next section, the solution approach to determine the best combination of *S*, $\boldsymbol{n}$ and $\boldsymbol{v}$ is presented.

3 Solution approach

Because of the complexity of the model due to uncertain yield and demand in combination with positive production times, Inderfurth / Kiesmüller (2015) developed a heuristic solution approach to determine the base stock level and therefore the required safety stock for a production system like the benchmark scenario. The approach is easy to implement in a spreadsheet with low computation times even for large problem sizes, leading to excellent results. One assumption is that the system is in a steady state and that the order quantity is linearized, such that Q_t always equals $F(S - IP_t)$. Furthermore, the distribution of the inventory level IL_t is approximated by a normal distribution with the first two moments μ_{IL} and σ_{IL}. In this case, the base stock level *S* can be calculated as

$$S = (\lambda + 1)\mu_D + \Phi^{-1}\left(\frac{b}{b+h}\right)\sigma_{IL} \tag{7}$$

where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution. The base stock level is composed of the average demand during the risk period $(\lambda + 1)\mu_D$ and the safety stock level $\Phi^{-1}\left(\frac{b}{b+h}\right)\sigma_{IL}$. The variance of the inventory level σ_{IL}^2 follows directly from equation (1) and (3) and equals $(\lambda + 1)\sigma_D^2 + \lambda\sigma_R^2$. The variance of the forecast error σ_R^2 can be calculated using the definition of the forecast error and equals $\sigma_R^2 = (\sigma_Q^2 + \mu_Q^2)\sigma_Z^2$. According to Inderfurth / Kiesmüller (2015), the mean μ_Q and the variance σ_Q^2 of the order quantity can be calculated as μ_D/μ_Z and $(\rho_Z^2\mu_D^2 + \sigma_D^2)/(\mu_Z^2 - \sigma_Z^2)$ where ρ_Z is the coefficient of variation of the yield factor defined as σ_Z/μ_Z. Using all these formulae the optimal base stock level and therefore the optimal safety stock level can be determined easily. If further quality inspections are introduced, the above formulae have to be extended to incorporate earlier information about realized yields, which requires some calculation effort. The results can be found in Sonntag / Kiesmüller, 2017.

To determine the cost minimizing investment strategy, the best set of $\boldsymbol{n}$, $\boldsymbol{v}$ and *S*–the number and position of additional inspections, the yield variability reduction on each production stage and the base stock level–has to be determined. In this paper, a fix-and-optimize solution approach is used to determine these parameters. For a fixed set of inspections $\boldsymbol{n}$, the reduction of yield variability $\boldsymbol{v}$ and the base stock level S are optimized. The investment in yield variability reduction is optimized solving a nonlinear program using the Matlab function *fmincon* whereas the base stock level is calculated using formula (7). After the optimization over all $\boldsymbol{n}$, the best combination of $\boldsymbol{n}$, $\boldsymbol{v}$ and S regarding average holding and backorder cost per period is determined out of the set of solutions.

As an example consider a production system consisting of three production stages and a budget of 10 monetary units. If one inspection costs 6 units, only one inspection additional to the final one can be implemented and $\boldsymbol{n}$ could equal (0,1), meaning that an additional inspection exists after stage 2. Other possible solutions for $\boldsymbol{n}$ are (0,0) and (1,0), which will not be considered in the following. Note, that after the final stage an inspection always exists wherefore it is not part of $\boldsymbol{n}$. The remaining budget equals 4 monetary units and will be used to reduce the yield variability. Thus, $\boldsymbol{v}$ could e.g. equal (0.5,1.5,2) meaning that 0.5 monetary units are used to reduce the yield variability during the first, 1.5 during the second and 2 during the third production stage.

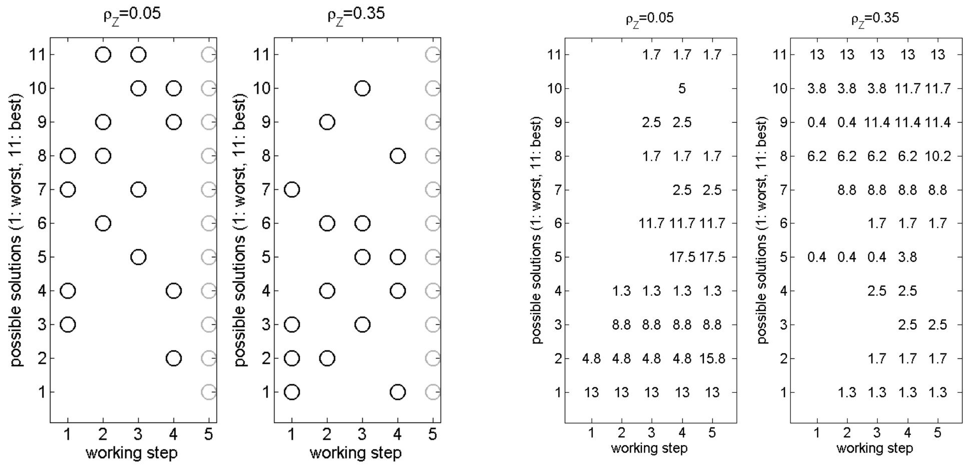
4 Numerical analysis

To give some managerial insights, a production system consisting of five production stages is examined. The demand is normally distributed with mean 20 and a coefficient of variation of 0.1 while holding and backorder cost are charged with one dollar and 99 dollars, respectively. This cost structure leads to a cost ratio $b/(b+h)$ of 0.99. All stages have a mean yield of 95% and coefficients of variation of either 0.05 or 0.35 to model a production system with low as well as with high uncertainty in the yield. The planner has an available budget of 65,000 dollars whereas each additional quality inspection costs 30,000 dollars and the yield standard deviation of each production stage can be reduced by 0.01 for 1,000 dollars investment ($c_i = 100{,}000\ \forall i = 1, \dots, m$). It is obvious that a maximum of two additional inspections can be introduced.

Figure 2a shows after which production stages quality inspections can be located additionally to the final inspection, while Figure 2b shows how the remaining budget (in 1,000 dollars) is allocated to the five production stages to minimize the required safety stock. At the top of the figures the best solution as a combination of $\boldsymbol{n}$, $\boldsymbol{v}$ and S is presented whereas the following solutions are sorted in ascending order according to the required safety stock level.

It can be seen that for low yield variability ($\rho_Z = 0.05$) it is reasonable to introduce as many inspection stations as possible and spend just little money on reducing the yield variability. Instead for higher yield uncertainty ($\rho_Z = 0.35$) it is better to place fewer inspections and therefore invest the budget in yield uncertainty reduction. The reason for this phenomenon is that a reduction of the yield variability is not very valuable if the yield variance is already very low. Instead, if the yield variability is very high even small reductions lead to large effects according to the required safety stock level.

Figure 2: Locations of inspections (a) and investment in yield variance reduction at each working step in 1,000 dollars (b)

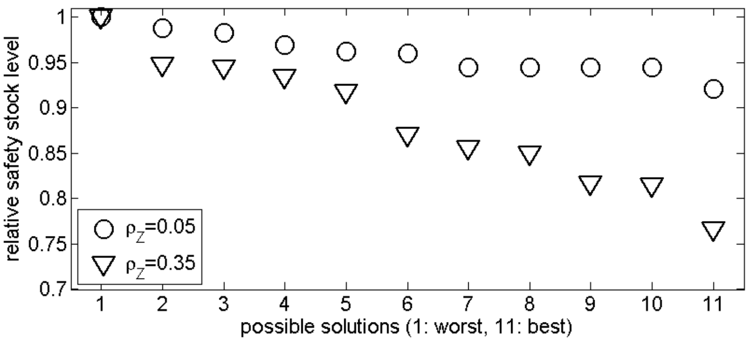


(a) (b)

A closer look at Figure 2b indicates that the budget is particularly invested at later production stages. To explain this phenomenon the second best solution for a coefficient of variation of the yield of 0.35 is used where 3,800 dollars are invested at the first three and 11,700 dollars at the last two stages. Figure 2a shows that for this solution an additional inspection is located after the third production stage, which means that all uncertain yields of the first three stages become known after this inspection and all defective items are disposed of before the fourth production stage. While the uncertainty about the production outcome of the first three stages is eliminated, the uncertainty about the yield of the last two stages still exists. A high yield uncertainty for later production stages require a larger amount of safety stock because the uncertainty cannot be eliminated significantly before finishing the whole production. Thus, reducing the yield variability at the last two production stages is more valuable and therefore most of the budget is invested there.

Figure 3 summarizes the results from Figure 2 by showing the corresponding safety stock level, which minimizes average holding and backorder cost as in formula (5). The horizontal axis shows all optimal solutions regarding $\boldsymbol{v}$ and S for given $\boldsymbol{n}$ in descending order according to the required safety stock level. On the vertical axis the relative safety stock level compared to the worst solutions in case of a coefficient of variation of the yield of 0.05 and 0.35 is presented. Thus, the safety stock level for the worst solution equals 100%.

Figure 3: Minimum safety stock level for all sets of additional inspections and optimal investments in yield variability reduction



The best investment strategy as a combination of additional inspections and a reduced yield variability can reduce the required safety stock level by 8% for low yield variability and nearly 25% for high yield variability compared to the worst strategy.

5 Summary

In this paper, a production system producing a random fraction of defective items due to imperfect production processes is considered. Supply as well as demand uncertainty require large safety stock in a warehouse resulting in high holding cost. It is assumed that a planner has two opportunities to reduce the safety stock level. He can implement additional quality inspections and/or reduce the variability of the yield.

It is shown how to determine the best combination of quality inspections and a reduction in the yield variability under a given budget constraint. The results emphasize that in production systems with low yield variability money should be invested into further quality inspections whereas in production systems with high yield variability it is best to use the budget to reduce the yield variance rather than investing into further inspections. Applying the best budget allocation strategy can lead to significant reductions in the required safety stock level and therefore holding and backorder cost.

References

Bollapragada, S. / Morton, T. E. (1999), Myopic heuristics for the random yield problem, in: Operations Research, Vol. 47, No. 5, pp. 713-722.

Choi, H. P. / Blocker, J. D. / Gavirneni, S. (2008), Value of Sharing Production Yield Information in a Serial Supply Chain, in: Production and Operations Management, Vol. 17, No. 6, 614-625.

Dettenbach, M. / Thonemann, U. W. (2015). The value of real time yield information in multi-stage inventory systems – exact and heuristic approaches, in: European Journal of Operational Research, Vol. 240, Issue 1, 72-83.

Foremski, T. (2012), Intel's battle with ARM is about making its future fabs viable, in: http://www.siliconvalleywatcher.com/mt/archives/2012/09/intels_battle_w.php (18.04.2017)

Helber, S. / Sahling, F. (2010), A fix-and-optimize approach for the multi-level capacitated lot sizing problem, in: International Journal of Production Economics, Vol. 123, Issue 2, pp. 247-256.
Henig, M. / Gerchak, Y. (1990), The structure of periodic review policies in the presence of random yield, in: Operations Research, Vol. 38, No. 4, pp. 634–643.
Huh, W. T. / Nagarajan, M. (2010), Technical note-linear inflation rules for the random yield problem: Analysis and computations, in: Operations Research, Vol. 58, No. 1, pp. 244–251.
Inderfurth, K. / Kiesmüller, G. P. (2015), Exact and heuristic linear-inflation policies for an inventory model with random yield and arbitrary lead times, in: European Journal of Operational Research, Vol. 245, No. 1, pp. 109–120.
Inderfurth, K. / Vogelgesang, S. (2013), Concepts for safety stock determination under stochastic demand and different types of random production yield, in: European Journal of Operational Research, Vol. 224, No. 2, pp. 293-301.
Intel (2017), 2016 Annual report, in: https://s21.q4cdn.com/600692695/files/doc_financials/2016/annual/Intel_2016_Annual_Report.pdf (18.04.1017)
Kulkarni, S. S. (2008), Loss-based quality costs and inventory planning: General models and insights, in: European Journal of Operational Research, Vol. 188, No. 2, pp. 428–449.
Lin, L. / Hou, K. (2005), An inventory system with investment to reduce yield variability and set-up cost, in: Journal of the Operational Research Society, Vol. 56, No. 1, pp. 67–74.
McNutt, C. (2015), Rumor: Low yield 3D glass forcing Samsung to look for additional supplier for Galaxy S6 Edge, in: http://www.androidheadlines.com/2015/03/rumor-low-yield-3d-glass-forcing-samsung-look-additional-suppliers-galaxy-s6-edge.html (18.04.2017)
Sonntag, D. / Kiesmüller, G. P. (2016), The shape of the yield and its impact on inventory decisions, in: 4OR, Vol. 14, No. 4, pp. 405-415.
Sonntag, D. / Kiesmüller, G. P. (2017), The influence of quality inspections on the optimal safety stock level, in: Production and Operations Management, doi 10.1111/poms.12691.
Yano, C. A. /Lee, H. L. (1995), Lot sizing with random yields: A review, in: Operations Research, Vol. 43, No. 2, pp. 311-334.

5. Disposal versus Rework - Inventory control in a production system with random yield

Sonntag, D., Kiesmüller, G. P. (in press). Disposal versus Rework - Inventory control in a production system with random yield. Accepted for publication in: European Journal of Operational Research.

European Journal of Operational Research 000 (2017) 1–12

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Production, Manufacturing and Logistics

Disposal versus rework – Inventory control in a production system with random yield

Danja Sonntag*, Gudrun P. Kiesmüller

Faculty of Economics and Management, Chair of Operations Management, Otto-von-Guericke-University Magdeburg, P.O. Box 4120, D-39106, Magdeburg, Germany

ARTICLE INFO

Article history:
Received 9 December 2016
Accepted 9 November 2017
Available online xxx

Keywords:
Inventory
Stochastic inventory model
Random yield
Rework
Disposal

ABSTRACT

In a production environment where random yield plays a fairly significant role, a decision has to be made on how to handle products that do not satisfy given quality requirements. We consider a single-stage production system with a positive production time and random yield. To ensure that only high quality items are sold to the customer, a post-production quality control system has been put in place. We compare two different strategies for defective items: disposal or rework. Disposal is possible without any time delay whereas the rework process requires a positive rework time. While disposed-of items leave the production process, reworked products stay in the process and are assumed to be as good as products that are perfect when they are initially produced. The end products are stored in a warehouse to satisfy stochastic demand. We show how to determine the optimal base-stock level, which is very difficult because of unknown covariances between orders. Subsequently, an optimization model is proposed to support the planner's decision on which strategy to choose when it comes to whether to dispose of or rework defective items. By means of a sensitivity analysis we show which parameters directly affect this decision and give managerial insights. The analysis indicates that significant cost reductions can be obtained by choosing the best strategy for defective products.

1. Introduction

Customer service plays an especially important role in highly competitive markets where dissatisfaction about e.g., product quality leads to a loss of the customer's goodwill resulting in the customer selection of a new vendor. Thus, the plan must be to sell only high-quality products. A problem arises when the production process is not perfect, such that random yield losses occur. Random production yield is a common problem in the high-tech industry with complex production processes. For example, in the production of microchips, yields differ between 60% and the high 90% range depending on the manufacturer (Foremski, 2012). A second example is the production of curved glass for the display of a new cell phone series, where Samsung has to deal with yields down to less than 50% (McNutt, 2015). In such an environment, where sometimes more than every second item is defective, it is obvious that random production yield cannot be neglected. To guarantee that only high-quality products are sold to the customers, a quality control inspecting 100% of all produced items is required. The items passing the inspection are stocked in a warehouse to serve stochastic customer demand. This begs the question, how to handle all the defective products which should not be sold to the customer due to poor quality. Several opportunities arise: the products can either be scrapped (see e.g., Yano & Lee, 1995; Huh & Nagarajan, 2010; Inderfurth & Kiesmüller, 2015; Sonntag & Kiesmüller, 2017), sold as lower quality products for a lower price (see e.g., Gerchak, Tripathy, & Wang, 1996; Hsu & Bassok, 1999), reworked (see e.g., Wein, 1992; Grosfeld-Nir & Gerchak, 2004) or used otherwise.

In this paper, we study the strategic choice between scrapping and reworking which means that the planner can decide between these opportunities only at the beginning of the planning horizon. We assume that reworked items satisfy all quality requirements and are as good as items that are well made from the start and can be sold for the same price (see e.g., Inderfurth, Lindner, & Rachaniotis, 2005; Gotzel & Inderfurth, 2005; Buscher & Lindner, 2007). Reworking defective products to raise their quality might be desirable for a company for several reasons: First of all, rework might be reasonable for economic reasons. This is the case when defective products are of substantial value because of expensive input materials, e.g., in the high tech industry (Buscher & Lindner, 2007; Inderfurth et al., 2005) or when the time and cost for rework are

* Corresponding author.
E-mail address: danja.sonntag@ovgu.de (D. Sonntag).

https://doi.org/10.1016/j.ejor.2017.11.019
0377-2217/

Please cite this article as: D. Sonntag, G.P. Kiesmüller, Disposal versus rework – Inventory control in a production system with random yield, European Journal of Operational Research (2017), https://doi.org/10.1016/j.ejor.2017.11.019

lower than for the initial production of new items. Second, new legislation might force companies to reduce waste (Teunter & Flapper, 2003). Third, ecological aspects are gaining more and more attention and therefore influence the waste policy and the image of the company (Inderfurth et al., 2005). The image of a company has an influence on customer satisfaction and therefore on sales and the equity, which provide a competitive advantage especially in highly competitive markets where it is difficult to differentiate between the products (Chen, 2010; Teunter & Flapper, 2003). Flapper, Fransoo, Broekmeulen, and Inderfurth (2002) give an overview of industries mentioned in the literature where rework plays an important role for at least one of the specified reasons (e.g., the semiconductor and pharmaceutical industry).

We consider a single-stage production system producing batches of items with a known and constant production time independent of the batch size, and stochastic proportional yield, which leads to a random number of defective items. In a stochastic proportional yield model, the yield is a random multiple of the input (Henig & Gerchak, 1990). Subsequent to the production process, a quality control system is in place, inspecting all items with no time delay. Items satisfying the quality requirements are stocked in a warehouse to satisfy incoming stochastic customer demand, whereas defective items are either entirely disposed of or reworked. Note that once the planner has chosen one of these strategies, he cannot change it in the near future. The rework process – like the production process – corresponds to a known and constant rework time but is performed on a different machine. Thus, the production and the rework process require different resources. The rework process brings all defective items in a condition equal to that of perfectly produced products such that they can be stored in the warehouse as well. The warehouse has to initiate the production of a batch of items with varying lot size periodically to replenish stock.

The literature includes work on imperfect production systems where defective products are either scraped or reworked – totally or partially. Yano and Lee (1995) give a literature overview of different problem settings and approaches to solving imperfect production environments where yield losses are disposed of. In the following discussion, we will amplify two different streams of literature: make-to-stock models under random yield with disposal of defective items and make-to-order models under random yield with rework of defective items.

The literature on make-to-stock production systems, combining random yield settings with inventory control strategies, can be divided into two groups: production time zero or one (e.g., Henig & Gerchak, 1990; Bollapragada & Morton, 1999; Huh & Nagarajan, 2010) and arbitrary positive production times (Dettenbach & Thonemann, 2015; Inderfurth & Kiesmüller, 2015; Inderfurth & Vogelgesang, 2013; Sonntag & Kiesmüller, 2017). Production times of zero or one period (called zero production time in the following discussion) substantially reduce the complexity of the problem since no uncertainties of outstanding orders have to be considered, which means that the inventory position used to determine the order quantity is known. For zero production times, Henig and Gerchak (1990) and Bollapragada and Morton (1999) focus on the optimal order policy whereas Huh and Nagarajan (2010) concentrate on the optimization of the policy parameters in case of a linear inflation rule. A linear inflation rule is the commonly used heuristic order policy under random production yield since the optimal ordering policy is very complex (Henig & Gerchak, 1990).

Positive production times involve an uncertainty in the inventory position which makes the problem far more complex. Therefore, only a few authors have considered positive production times (e.g., Inderfurth & Vogelgesang, 2013; Dettenbach & Thonemann, 2015; Inderfurth & Kiesmüller, 2015; Sonntag & Kiesmüller, 2017). Inderfurth and Vogelgesang (2013) present concepts to determine safety stocks under different types of yield randomness. Dettenbach and Thonemann (2015) take into consideration multi-stage production systems with the aim of determining the location of quality inspections to obtain real-time yield information which reduces the required safety stock. They use dynamic programming for small and medium-sized problems and two heuristic approaches for larger problems. One of the heuristics is based on an idea of Ehrhardt and Taube (1987) and can lead to poor results depending on the parameter setting. The second heuristic is based on an idea of Huh and Nagarajan (2010) and leads to very good results but requires simulation since "it is difficult to calculate […] analytically" (Dettenbach & Thonemann, 2015). Since the optimal order policy is difficult to determine and dynamic programming as well as simulation might involve high computation times, Inderfurth and Kiesmüller (2015) developed a new heuristic solution method. The so-called steady-state approach leads to very good results and is easy to implement in a spreadsheet. While Inderfurth and Kiesmüller (2015) present a single-stage production system, Sonntag and Kiesmüller (2017) extend the approach to analyze multi-stage production systems with in-between quality inspections. All these papers have in common that defective items are disposed of and therefore leave the production system.

Unlike the limited literature on make-to-stock production systems with random yield and disposal of defectives, there exists a variety of literature on make-to-order systems where defective items are reworked. Most researchers investigate production and rework processes on the same machine and thus have to solve a scheduling problem (e.g., So & Tang (1995)) or a lot-sizing problem (see e.g., Liu & Yang, 1996; Teunter & Flapper, 2003; Inderfurth, Kovalyov, Ng, & Werner, 2007; Grosfeld-Nir & Gerchak, 2002; Wein, 1992). Jaber and Khan (2010) extend the production system with learning curves for the time to produce and rework items.

There is only one paper, by Gotzel and Inderfurth (2005), where an inventory-control policy in a production environment with random yield and stochastic demand is studied in which defective items are reworked. While Gotzel and Inderfurth (2005) assume that defective items can be temporarily stocked before rework, we consider a situation where defective items have to be reworked immediately. As an example, consider the steel industry where steel coming out of the furnace is inspected for quality. If it does not satisfy the given quality standards it has to be returned to a furnace. This reworking in a furnace is much faster when the steel is still hot and has had no time to cool down. In such a situation intermediate stock points are not favorable for economic reasons. Other examples can be found in industries such as the chemical, pharmaceutical or food industries, where items cannot be stored to wait for rework. Further, it is sometimes the case that there is no storage space available for items waiting for rework or that the company wants to reduce work in process inventory.

Allowing only one stockpoint in the system has consequences for the analysis, because no decision has to be made on the amount to be reworked. The rework quantity is dependent only on the production output, and therefore on the order quantity of a previous period and the realized yield. The number of defective items increases as the order quantity increases, thus the order quantity in a previous period determines the amount to be reworked in a later period. If a large number of reworked items arrive at the warehouse, the order quantity in the actual period can be reduced. Therefore, the actual order quantity depends on previous order quantities and thus covariances occur.

Aside from the fact that our model includes only one stockpoint we define the inventory position differently from Gotzel and Inderfurth (2005) and include only information about orders arriving during the risk period and thus only relevant information for the actual decision.

The contribution of this paper is as follows: (1) We show how to determine the base-stock level in a production environment where defective products are not disposed of but reworked and thus stay within the system. The induced covariances cannot be calculated easily and thus we propose an approximation. In a detailed numerical study we illustrate the excellent performance of our approximation. (2) We introduce a mathematical model representing production, quality control, disposal or rework as well as holding and backorder costs. The model can be used as a decision support tool for a planner when he or she has to decide if defective items should be disposed of or reworked. We show which parameters have an influence on this decision and give an idea on how robust the decision of the planner is regarding changes in the environment or the cost parameters. With the derived model we also gain some managerial insights.

The remainder of the paper is organized as follows: In Section 2, we describe the multi-stage production system and formulate the model. In Section 3, the steady-state approach is introduced. Since covariances between orders occur, which are not easy to calculate but cannot be neglected, we present an approximation and analyze its accuracy in Section 4. In Section 5, we present a mathematical model considering above-mentioned cost parameters, and analyze the effect of changes within the input parameters of the production system (5.3) and the cost parameters (5.4) on the decision whether to rework or dispose of defective items. Based on this analysis, we formulate managerial insights in Section 5.5. We conclude with a summary and suggestions for future research in Section 6.

2. Model formulation

We consider a single-stage production system producing batches of one single product with a constant production time of L_P periods per batch ($L_P > 0$). The production time can be independent of the batch size in, for example, the chemical industry where processing times are often independent of the amount being produced. Further, in the context of an MRP planning system, planned lead times are assumed to be constant, even though some variability exists, in order to enable coordinated decisions.

Due to deficiencies in the production, not all produced items are of perfect quality. Since it is not desirable to sell products of lower quality to the customers, a quality inspection subsequent to the production process is established. The inspection station checks the quality properties of all produced items with no time lag. We would like to note that the time for an inspection can be included in the production time since all items pass quality control. Thus, neglecting any delay for the inspection does not reflect a limitation in the model.

Items that satisfy the quality requirements are stocked in a warehouse to serve incoming stochastic customer demand. We assume that the demand across periods is independent and identically distributed (iid) and backlogged if it cannot be satisfied directly from stock. Inderfurth and Kiesmüller (2015) as well as Dettenbach and Thonemann (2015) and Sonntag and Kiesmüller (2017) analyzed single or multi-stage production systems where defective items are scrapped. In contrast to these contributions, we focus on a situation where defective items are reworked. The reworking process – like the production process – requires a rework time of L_R time units ($L_R > 0$) whereas the rework time can be either smaller, equal to or larger than the production time. After rework these products are stocked in a warehouse with the same quality as items that were perfect when first produced. Note that the rework process proceeds on a different machine from the one used in the production process, which means that different resources are required and they do not interfere with each other.

Fig. 1 illustrates the whole model composed of a production and a rework process, a quality-control process and a warehouse for the final product. The sequence of events in one period is as follows: First, the good items of the order placed L_P periods before as well as the reworked items of the order placed $L_P + L_R$ periods before are delivered. Subsequently, a new order is placed and demand occurs. At the end of the period, inventory holding and backorder costs are charged based on the inventory level.

We apply a stochastic proportional yield model which is commonly used to describe random yield due to an imperfect production process (Yano & Lee, 1995). In a stochastic proportional yield model the output $Y(Q)$ of the production process equals a positive fraction Z of the input Q such that $Y(Q) = Z \cdot Q$. In our model, the input for production is determined by the order quantity Q requested by the warehouse to refill stock and ensure that demand can be satisfied. $Z \in [0, 1]$ is a random variable called the yield factor with mean μ_Z and variance σ_Z^2 and is iid across the periods and independent of the demand distribution.

As already mentioned, at the beginning of each period, after receiving the batch of a prior order, the warehouse has to determine the required order quantity in order to minimize average holding and backorder costs. We consider a periodic review base-stock policy with a review period of one time unit and a base-stock level S. Such a policy is used because the optimal ordering policy for production systems with random yield does not possess a simple structure, even for production times equal to zero (Henig & Gerchak, 1990). In this paper, positive instead of zero production times are considered, which increases the complexity of the system and makes it even more difficult to determine the optimal policy structure. Therefore, we propose a heuristic ordering policy, which is easy to implement. The periodic review base-stock policy is such a candidate, and it is also optimal if there is no uncertainty in the yield. In such a situation the order quantity in period t equals the difference between the base-stock level S and the actual inventory position IP_t (Tempelmeier, 2006): $Q_t = S - IP_t$, where the inventory position is defined as the physical stock on hand minus backorders plus the outstanding order quantities.

In case of random production yield and a positive production lead time it is not reasonable to include the outstanding order quantities, because the production output is uncertain. Therefore, we have to define the inventory position differently and suggest the use of the expected amount to be delivered instead of the outstanding order quantities, as long as we do not have information about the realized yield. This means that for all orders in the production process the yield is unknown and the expected amount to be delivered is included in the inventory position, while for all orders in the rework process realized yield is known and the estimates can be updated. Further, we only include information about the orders that will be delivered during the risk period (see also Kiesmüller, 2003). The risk period equals $L_P + 1$ review periods, because between review periods it is not possible to influence the amount of incoming items, either from production or from rework. Altogether, we get the following definition of the inventory position IP_t at the beginning of period t before ordering:

$$IP_t = \begin{cases} IL_{t-1} + Z_{t-L_P} Q_{t-L_P} + \sum_{l=1}^{L_P-1} \mu_Z Q_{t-l} + \sum_{l=L_P}^{L_P+L_R-1} (1 - Z_{t-l}) Q_{t-l} \\ + (1 - Z_{t-L_P-L_R}) Q_{t-L_P-L_R} + \sum_{l=L_R}^{L_P-1} (1 - \mu_Z) Q_{t-l} & , \ L_P \geq L_R \\ IL_{t-1} + Z_{t-L_P} Q_{t-L_P} + \sum_{l=1}^{L_P-1} \mu_Z Q_{t-l} + \sum_{l=L_R}^{L_P+L_R-1} (1 - Z_{t-l}) Q_{t-l} \\ + (1 - Z_{t-L_P-L_R}) Q_{t-L_P-L_R} & , \ L_P < L_R \end{cases} \quad (1)$$

Note that the realized yield is modeled with a random variable (Z_t) because all possible values have to be considered. However,

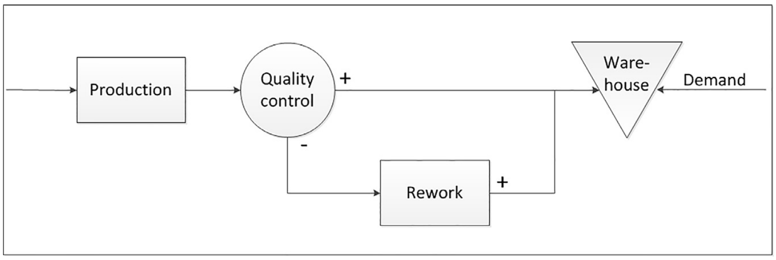


Fig. 1. A single-stage make-to-stock production system with random yield and rework.

if the model is applied in practice the random variables have to be replaced by the observed realized values. For example, if the planner has to decide on the order quantity, he uses the actual values for the yield (z_t), the demand (d_t) and the outstanding order quantities (q_t).

It is clear that we also have to distinguish between a situation where the production time is larger than or equal to the rework time, and one where the production time is smaller than the rework time. For the first case ($L_P \geq L_R > 0$), the inventory position equals the inventory level IL_{t-1} at the end of the previous period plus the sum of the following components: The first term $Z_{t-L_P}Q_{t-L_P}$ represents the delivered number of good items of the order placed in period $t-L_P$, because in the moment when production is finished, the yield $Z_{t-L_P}Q_{t-L_P}$ is known. The second term $\sum_{l=1}^{L_P-1}\mu_Z Q_{t-l}$ represents all orders still in production such that the yield is not known and therefore the expected amount to be delivered after production has to be estimated. The following term $\sum_{l=L_P}^{L_P+L_R-1}(1-Z_{t-l})Q_{t-l}$ equals outstanding quantities within the rework process, where yield is known because the production process for these items has already been completed. $(1-Z_{t-L_P-L_R})Q_{t-L_P-L_R}$ equals the number of delivered reworked items of the order placed L_P+L_R periods before. The last term $\sum_{l=L_R}^{L_P-1}(1-\mu_Z)Q_{t-l}$ is related to the orders in production and represents an estimate of the number of units which will not satisfy the quality requirements and therefore have to be reworked.

It is important to note that while all outstanding orders within the production process are included in the inventory position ($\sum_{l=1}^{L_P-1}\mu_Z Q_{t-l}$), only a part of the outstanding orders in the rework process is taken into account ($\sum_{l=L_R}^{L_P-1}(1-\mu_Z)Q_{t-l}$), because we consider only orders that arrive during the risk period. Thus, in contrast to Gotzel and Inderfurth (2005), who include all outstanding orders, we only include orders which will be delivered to the warehouse during the risk period of L_P+1 periods.

In cases where the rework time exceeds the production time ($L_R > L_P > 0$), the inventory position consists of the same elements with one difference: the last term, the defective quantities thus so far unknown, which will enter the rework process in future periods, does not appear. If the production time is smaller than the rework time, no order exists where the quantity of defective units is not yet known but will be delivered during the risk period of length L_P+1.

We can merge some terms in (1) and reformulate the inventory position as

$$IP_t = \begin{cases} IL_{t-1} + \sum\limits_{l=L_P+1}^{L_P+L_R}(1-Z_{t-l})Q_{t-l} + \sum\limits_{l=L_R}^{L_P} Q_{t-l} + \sum\limits_{l=1}^{L_R-1}\mu_Z Q_{t-l} & , L_P \geq L_R \\ IL_{t-1} + \sum\limits_{l=L_R}^{L_P+L_R}(1-Z_{t-l})Q_{t-l} + Z_{t-L_P}Q_{t-L_P} + \sum\limits_{l=1}^{L_P-1}\mu_Z Q_{t-l} & , L_P < L_R \end{cases} \tag{2}$$

with $\sum_{i=a}^{b} x_i = 0$ for $b < a$.

Since the inventory position includes the expected number of units to be delivered, the moment when the realized yield is observed the inventory position has to be updated as follows.

$$IP_{t+1} = IP_t - D_t - (\mu_Z - Z_{t+1-L_P})Q_{t+1-L_P} \tag{3}$$

It can happen that the realized yield (Z_{t+1-L_P}) is much larger than expected (μ_Z), which increases the inventory position without placing an order. In extreme cases, which occur vary rarely, it is possible that the inventory position before ordering is already above the base-stock level. Therefore, the ordering policy in case of random yield has to be adjusted as follows:

$$Q_t = \begin{cases} S - IP_t, & IP_t < S \\ 0, & IP_t \geq S \end{cases} \tag{4}$$

Note that we do not apply a linear inflation policy as often used in the random yield literature with disposal of defective items, because all ordered units will arrive. There is only a difference in the observed lead time, because some of the units have to be reworked. Inflation factors are necessary if units not satisfy quality requirements are disposed of.

The inventory position in formula (2) is not only used to determine the order quantity, but it is also used in our model to determine the inventory level IL_{t+L_p} at the end of period $t+L_P$:

$$IL_{t+L_p} = IP_t - \sum_{l=0}^{L_P} D_{t+l} + \mu_Z Q_t - R_t - \sum_{l=1}^{min\{L_P,L_R\}-1} R_{t-l} \tag{5}$$

The term $\sum_{l=0}^{L_p} D_{t+l}$ equals the demand that occurs between the beginning of period t and the end of period $t+L_P$. The second term reflects the estimated amount to be delivered in period $t+L_p$ from the order, placed in period t, where the yield is unknown when production starts. The next terms are related to the updates of the inventory position as shown in Eq. (3) which are done in each period, when yield is realized. We call the difference between expected and realized yield for a given order quantity Q_t the forecast error, which is defined as

$$R_t = \mu_Z Q_t - Z_t Q_t. \tag{6}$$

It is obvious that the inventory level as given in (5) is a function of the base-stock level S and hence we denote it as $IL(S)$ in the following discussion. The higher the base-stock level S, the higher the stock-on-hand and the lower the backorder quantities and vice versa. Our objective is to determine a base-stock level S, which minimizes the average holding and backorder cost $C(S)$:

$$C(S) = hE[(IL(S))^+] + bE[(-IL(S))^+] \tag{7}$$

where h denotes the unit holding and b the unit backorder costs, and $(M)^+$ is defined as $max\{0, M\}$.

Determining the optimal base-stock level in case of random yield is not easy in the presence of positive production times, even without reworking defective items. For production systems where imperfect products are scraped instead of being reworked, previous approaches in the literature involve high computation times due to the application of Markov chains (Dettenbach & Thonemann, 2015; Inderfurth & Kiesmüller, 2015), simulation (Dettenbach & Thonemann, 2015; Inderfurth & Kiesmüller, 2015) or stochastic dynamic programming (Gotzel, 2010). Therefore, Inderfurth and Kiesmüller (2015) introduced an approximate steady-state approach for a single-stage production system where defective items are disposed of. The performance of the approach has been shown to be excellent while it is easy to implement in a spreadsheet. Because of the excellent performance of the approach and the absence of efficient solution methods for production systems where defective items are reworked, we adapt the idea for the production system described above.

As a starting point for this approach, formula (7) can be rewritten as

$$C(S) = h\int_0^{\infty} x\varphi_{IL}(x)dx - b\int_{-\infty}^{0} x\varphi_{IL}(x)dx \tag{8}$$

where φ_{IL} reflects the probability density function of the inventory level IL defined in (5). To determine the base-stock level S which minimizes the average holding and backorder cost, the distribution of the inventory level IL with the density function φ_{IL} is required.

3. Determining the base-stock level

To calculate the average cost for a given base-stock level, Inderfurth and Kiesmüller (2015) showed that, for symmetric demand distributions, a normal distribution with mean μ_{IL} and variance σ_{IL}^2 is a suitable approximation of the inventory level. For asymmetric demand distributions Inderfurth and Kiesmüller (2015) as well as Sonntag and Kiesmüller (2016) showed that other distribution functions for modelling the inventory level are suitable. In this paper, we will not focus on asymmetric demand distribution because the analysis and the insights are similar.

For a normally distributed inventory level, the optimal base-stock level is given by the following newsboy equation (Inderfurth & Kiesmüller, 2015):

$$P(IL \geq 0) = \frac{b}{b+h} \tag{9}$$

We will fit a normal distribution on the first two moments of the inventory level, which means we need to derive expressions for the moments. The mean inventory level μ_{IL} can be determined directly from (5) and (4).

Lemma 1. *Under a strictly linear control rule (which means: $Q_t = S - IP_t$), the mean inventory level μ_{IL} in a production system with positive production and rework times is given as:*

$$\mu_{IL} = S - (L_P + 1)\mu_D - (1 - \mu_Z)\mu_Q \tag{10}$$

where μ_D and μ_Q reflect the mean demand and the mean order quantity, respectively.

Proof. For the proof see Appendix A. □

In order to derive an expression for the variance of the inventory level, we need the moments of the forecast error as defined in (6). While the mean of the forecast error equals zero ($\mu_R = E[R_t] = E[\mu_Z Q_t - Z_t Q_t] = 0$), the variance does not. Sonntag and Kiesmüller (2017) demonstrate that the following equation holds:

$$\sigma_R^2 = (\sigma_Q^2 + \mu_Q^2)\sigma_Z^2 \tag{11}$$

Knowing the first two moments of the forecast error, the second central moment of the inventory level – the variance – can be determined.

Lemma 2. *Under a strictly linear control rule, the variance of the inventory level σ_{IL}^2 in a production system with positive production and rework times can be calculated as*

$$\sigma_{IL}^2 = (L_P + 1)\sigma_D^2 + L_R\sigma_R^2 + (1 - \mu_Z)^2\sigma_Q^2 \tag{12}$$

where σ_D^2, σ_R^2 and σ_Q^2 reflect the variances of the period demand, the forecast error and the order quantity.

Proof. For the proof see Appendix B. □

It is clear that the mean and the variance of the order quantity are required to determine the mean and the variance of the inventory level. The mean order quantity equals the mean demand ($\mu_Q = \mu_D$) since in the long run all demands have to be satisfied. To determine the variance of the order quantity, σ_Q^2, a recursive equation for the order quantity can be obtained:

$$Q_t = \begin{cases} D_{t-1}, & L_P \geq L_R = 1 \\ D_{t-1} + (1-\mu_Z)Q_{t-1} - (1-\mu_Z)Q_{t-L_R}, & L_P \geq L_R > 1 \\ D_{t-1} + (1-\mu_Z)Q_{t-1} - (1-Z_{t-L_R})Q_{t-L_R} + (\mu_Z - Z_{t-L_P})Q_{t-L_P}, & L_R > L_P > 0 \end{cases} \tag{13}$$

For the proof see Appendix C.

It is clear that for rework times larger than one time unit, the order quantity Q_t in period t depends on the order quantities in previous periods. To explain this phenomenon, remember that yield is random and estimates are included in the inventory position for the amount that is delivered L_P periods after production starts and the quantity delivered after $L_P + L_R$ periods when the rework is finished. When yield is realized these estimates are updated. If in one period the realized yield of the production process is larger than expected, more items than expected arrive in the warehouse and fewer items have to be reworked. Therefore the items arrive earlier than expected. This means that the order quantity in one period depends on the order quantities of previous periods and covariances occur when the variance of the order quantity has to be determined.

It is obvious that, if rework time is one period, the variance of the order quantity equals the variance of the demand: $\sigma_Q^2 = \sigma_D^2$. Moreover, for larger rework times the variance of the order quantity depends on the covariances of orders as follows.

$$VAR[Q_t] = \begin{cases} VAR[D_{t-1}] + VAR[(1-\mu_Z)Q_{t-1}] + VAR[(1-\mu_Z)Q_{t-L_R}] & \\ -2(1-\mu_Z)^2 COV[Q_{t-1}, Q_{t-L_R}], & L_P \geq L_R > 1 \\ VAR[D_{t+1}] + VAR[(1-\mu_Z)Q_{t-1}] & \\ +VAR[(1-Z_{t-L_R})Q_{t-LR}] + VAR[(\mu_Z - Z_{t-L_P})Q_{t-L_P}] & \\ -2COV[(1-\mu_Z)Q_{t-1}, (1-Z_{t-L_R})Q_{t-L_R}] & \\ +2COV[(1-\mu_Z)Q_{t-1}, (\mu_Z - Z_{t-L_P})Q_{t-L_P}] & \\ -2COV[(1-Z_{t-L_R})Q_{t-LR}, (\mu_Z - Z_{t-L_P})Q_{t-L_P}], & L_R > L_P > 0 \end{cases} \tag{14}$$

To determine the covariances, we need to know corresponding joint probability distribution of the two random variables considered. These joint distributions are unknown and therefore the covariances cannot be calculated easily. In the following discussion, we show how the covariances between order quantities can be approximated to get a good estimate. The numerical study reveals that the performance of the approximation is excellent.

4. Approximation of the covariances between orders

For all the subsequent analyses we focus on rework times that are smaller than or equal to the production times ($1 < L_R \leq L_P$). Nevertheless, an analysis similar to the one for production times exceeding rework times can be adapted for the opposite relation.

We approximate the covariance between order quantities by using the recursive equation of the order quantity as given in (13). For every L_R, the recursive equation of the order quantity is plugged into the formula for the covariance in (14). Unfortunately, we end up with different formulae for the covariances for different values of L_R (for details see Appendix D).

Table 1
Covariance factors.

L_R	A_{L_R}
4	$[-(1-\mu_Z)\sum_{k=0}^{\infty}(-(1-\mu_Z))^k]$
5	$[-(1-\mu_Z)]$
6	$[-(1-\mu_Z)-(1-\mu_Z)^3\cdot\sum_{k=0}^{\infty}(-(1-\mu_Z))^k\cdot f_{k+1}]$
7	$[-(1-\mu_Z)\cdot\sum_{k=0}^{\infty}((1-\mu_Z)^2)^k]$
8	$[-(1-\mu_Z)-(1-\mu_Z)^3-2(1-\mu_Z)^5+(1-\mu_Z)^6-5(1-\mu_Z)^7\cdots]$
9	$[-(1-\mu_Z)-(1-\mu_Z)\cdot\sum_{k=1}^{\infty}2^{k-1}((1-\mu_Z)^2)^k]$
10	$[-(1-\mu_Z)-(1-\mu_Z)^3-2(1-\mu_Z)^5-5(1-\mu_Z)^7\cdots]$
⋮	⋮

The covariances, and therefore the variance of the order quantity, can be calculated nearly exactly for rework times of two and three periods. For rework times larger than three, the covariances can be approximated very well by $(\sigma_Q^2-\sigma_D^2)\cdot A_{L_R}$, whereas A_{L_R} depends on the mean yield μ_Z in different ways for each L_R (for details see Appendix E).

Lemma 3. *Under a strictly linear control rule, the variance of the order quantity σ_Q^2 in a production system with positive production and rework times ($L_P \geq L_R > 0$) can be approximated as*

$$\sigma_Q^2 \approx \begin{cases} \sigma_D^2, & L_R = 1 \\ \frac{\sigma_D^2}{1-2(1-\mu_Z)^2\left[1-\frac{1}{2-\mu_Z}\right]}, & L_R = 2 \\ \frac{\sigma_D^2}{1-2(1-\mu_Z)^2}, & L_R = 3 \\ \frac{\sigma_D^2(1+2(1-\mu_Z)^2\cdot A_{L_R}}{1-2(1-\mu_Z)^2+2(1-\mu_Z)^2\cdot A_{L_R}}, & L_R > 3 \end{cases} \tag{15}$$

with the factors A_{L_R} as given in Table 1 and f_k denoting the Fibonacci numbers starting with $f_0 = 0,\ f_1 = 1$ and $f_k = f_{k-1} + f_{k-2}$ ($\forall k \geq 2$).

Proof. Formula (15) directly follows from (14). For details on Table 1 see Appendix E. □

Before we make further use of the above formulae within the solution approach, it is important to validate the quality of the approximation of the variance of the order quantity as given in Eq. (15).

The aim of this study is to investigate the performance of the approximations of the covariances relating to the optimal base-stock levels and costs. We analyze several instances with respect to the effect of different production and rework times and different coefficients of variation for both demand and yield factor under two different cost ratios $b/(b+h)$. We compare the results of the steady-state approach with the optimal solution determined by simulation. The parameter setting is presented in Table 2:

The critical ratio $b/(b+h)$ equals 0.9 or 0.95. The demand parameters are similar as in Inderfurth and Kiesmüller (2015). Demand is normally distributed with a fixed mean ($\mu_D = 20$) whereas the coefficient of variation varies between 0.1 and 0.3.

Table 2
Numerical values of the input parameters.

	$L_P = 5$	$L_P = 10$
$b/(b+h)$	{0.9; 0.95}	
μ_D	20	
ρ_D	{0.1; 0.2; 0.3}	
μ_Z	{0.5}; {0.8;0.9}	
ρ_Z	{0.1; 0.2; 0.3; 0.4; 0.5};	{0.1;0.2;0.3}
L_R	{1; 3; 4; 5}	{1; 5; 9; 10}

If higher coefficients of variation for the demand are requested, a gamma distribution is suitable which is not considered in the following discussion. Furthermore, we assume a symmetric ($\mu_Z = 0.5$) as well as an asymmetric ($\mu_Z = \{0.8; 0.9\}$) beta distributed yield factor. Using these mean yields we are able to model productions where on average half of all products are defective, as in the Samsung example in the introduction, or situations where fewer yield losses occur (e.g., Intel's microchip production Foremski, 2012). In the symmetric case, we allow for values for the coefficient of variation of the yield between 0.1 and 0.5 whereas for the asymmetric case, the coefficient of variation varies between 0.1 and 0.3.

We distinguish between a production time L_P of five and ten periods. Depending on the production length, we consider a system, where the rework time is very short ($L_R = 1$), half of the length of the production time ($L_R = 3$ for $L_P = 5$ and $L_R = 5$ for $L_P = 10$), just one period smaller ($L_R = 4$ for $L_P = 5$ and $L_R = 9$ for $L_P = 10$) or of equal length as the production time ($L_R = L_P = 5$ and $L_P = L_R = 10$). In total we analyze 240 instances of symmetric yield and 288 instances of asymmetric yield.

For the steady-state approach the base-stock levels were calculated solving formula (8) with respect to S, whereas the base-stock level was rounded up to the next integer. Therefore, only discrete values occur for S. We compare the base-stock levels with the optimal solution determined via simulation by increasing S stepwise by one unit until the minimum costs are reached. This procedure is possible because the cost function is convex in S (Huh & Nagarajan, 2010). Each simulation run represented 5000 periods with a 1000-period warm-up phase. To guarantee high accuracy, a sequential sampling procedure was used where the number of simulation runs was determined such that the half width of the 95% confidence interval of the average cost per period was smaller than 0.5% of the corresponding sample average. The simulation-based optimal base-stock level is the one minimizing the simulated cost based on formula (8).

Tables 3 and 4 show the results for symmetric and asymmetric yield. The first column gives the number of instances within the full factorial design in which the base-stock level of the steady-state approach is equal to the optimum S_{Sim} determined via simulation. The second column ($S_{Sim} +/- 1$) indicates that the base-stock level is one unit below or above the optimum. The third column reveals that the deviation from optimum is larger than one unit.

For symmetric yields in Table 3 the approach leads to excellent results independent of the production time L_P. We would like to mention that in only six instances the base-stock level is underestimated, which is worse than overestimating it because one unit backordered is more expensive than one unit of additional inventory.

For asymmetric yields in Table 4, the results are similar even though the number of instances with a deviation from the simulated solution increased. It is obvious that, for a cost ratio $b/(b+h)$ of 0.9, the approach leads to very good results for short as well as for long production times – independent of all other parameter settings. For a cost ratio of 0.95, deviations from optimum can be greater than one unit with a maximum of four. The results differ only slightly for increased production time. Nevertheless, since the base-stock level increases for longer production times, a higher absolute variation from the optimal solution has only small effects, especially when it comes to cost.

Therefore, after showing the effect of the approximations on the base-stock level, we analyzed the effect on the corresponding average inventory holding and backorder costs. We simulated the average cost C^* for the base-stock levels S_{SS} calculated with the steady-state approach and compared the results with the minimum average cost C^*_{Sim} obtained by simulation. The percentage

Table 3
Quality of base-stock level under symmetric yield.

		$L_P = 5$			$L_P = 10$				
		S_{Sim}	$S_{Sim} +/- 1$	Larger	S_{Sim}	$S_{Sim} +/- 1$	Larger		
L_R	1	19	11	0	17	13	0	1	L_R
	3	19	11	0	20	10	0	5	
	4	17	13	0	24	6	0	9	
	5	17	13	0	12	16	2	10	
ρ_D	0.1	22	18	0	28	12	0	0.1	ρ_D
	0.2	24	16	0	26	14	0	0.2	
	0.3	26	14	0	19	19	2	0.3	
$b/(b+h)$	0.90	32	28	0	33	26	1	0.90	$b/(b+h)$
	0.95	40	20	0	40	19	1	0.95	
ρ_Z	0.1	14	10	0	16	8	0	0.1	ρ_Z
	0.2	19	5	0	16	8	0	0.2	
	0.3	12	12	0	12	11	1	0.3	
	0.4	13	11	0	14	9	1	0.4	
	0.5	14	10	0	15	9	0	0.5	

Table 4
Quality of base-stock level under asymmetric yield.

		$L_P = 5$			$L_P = 10$				
		S_{Sim}	$S_{Sim} +/- 1$	Larger	S_{Sim}	$S_{Sim} +/- 1$	Larger		
L_R	1	19	16	1	21	15	0	1	L_R
	3	21	13	2	23	10	3	5	
	4	22	12	2	18	14	4	9	
	5	17	15	4	20	13	3	10	
ρ_D	0.1	29	14	5	27	16	5	0.1	ρ_D
	0.2	24	21	3	27	18	3	0.2	
	0.3	26	21	1	28	18	2	0.3	
$b/(b+h)$	0.90	45	27	0	47	25	0	0.90	$b/(b+h)$
	0.95	34	29	9	35	27	10	0.95	
μ_Z	0.8	38	33	1	39	32	1	0.8	μ_Z
	0.9	41	23	8	43	20	9	0.9	
ρ_Z	0.1	21	27	0	27	21	0	0.1	ρ_Z
	0.2	33	15	0	32	15	1	0.2	
	0.3	25	14	9	23	16	9	0.3	

cost difference of instance i was then calculated as

$$\delta_i = \frac{C^*(S_{SS}) - C^*_{Sim}}{C^*_{Sim}} \cdot 100\% \tag{16}$$

and the maximum relative difference of N instances was computed as

$$\delta_{max} = max_{i=1,\ldots,N}\delta_i \tag{17}$$

and the average relative difference of N instances as

$$\bar{\delta} = \frac{1}{N}\sum_{i=1}^{N}\delta_i. \tag{18}$$

Table 5 shows the average and maximum percentage cost deviation from the optimal solution for a production time of five and ten periods under symmetric yield ($\mu_Z = 0.5$).

The calculations reveal that the approximation of the covariance and the variance as given in (15) shows excellent performs especially for high production time. The results for an asymmetric yield are similar when it comes to an optimal solution with an average percentage cost deviation of 0.15% for a production time of five periods and 0.09% for a production time of ten periods, the maximum deviation over all instances equalling 2.5% and 1.4%, respectively.

Due to the satisfying results, formula (15) can be used to approximate the variance although the covariances are unknown. Note that the higher the mean yield the lower the influence of the covariances because they are multiplied with the term $(1-\mu_Z)^2$

Table 5
Average and maximum percentage deviation from optimal costs for $L_P = 5$ and $L_P = 10$.

$L_P = 5$				$L_P = 10$			
		$\bar{\delta}$	δ_{max}	$\bar{\delta}$	δ_{max}		
L_R	1	0.08	0.64	0.08	0.58	1	L_R
	3	0.10	0.72	0.04	0.30	5	
	4	0.16	1.22	0.02	0.18	9	
	5	0.09	0.60	0.06	0.34	10	
ρ_D	0.1	0.21	1.22	0.06	0.58	0.1	ρ_D
	0.2	0.07	0.45	0.03	0.36	0.2	
	0.3	0.04	0.24	0.06	0.34	0.3	
$b/(b+h)$	0.90	0.12	1.22	0.05	0.58	0.90	$b/(b+h)$
	0.95	0.09	1.12	0.05	0.42	0.95	
ρ_Z	0.1	0.17	1.22	0.09	0.42	0.1	ρ_Z
	0.2	0.04	0.45	0.02	0.18	0.2	
	0.3	0.15	0.72	0.07	0.39	0.3	
	0.4	0.10	0.45	0.05	0.58	0.4	
	0.5	0.07	0.64	0.02	0.12	0.5	

(compare formula (14)) and the approximation itself also consists of several such terms.

5. Disposal versus rework

Since we are able to compute the optimal base-stock policy, we can compare two strategies for defective products. The first strategy results in the disposal of all defective items whereas the

Table 6
Formulae for rework or disposal.

	With rework ($i = wR$)	With disposal ($i = nR$)
$\mu_{Q,\,i}$	μ_D	μ_D / μ_Z
$\sigma^2_{Q,i}$	$\sigma^2_D, \quad L_R = 1$	$(\rho^2_Z \mu^2_D + \sigma^2_D)/(\mu^2_Z - \sigma^2_Z)$
	$\frac{\sigma^2_D}{1-2(1-\mu_Z)^2\left[1-\frac{1}{2-\mu_Z}\right]}, \quad L_R = 2$	
	$\frac{\sigma^2_D}{1-2(1-\mu_Z)^2}, \quad L_R = 3$	
	$\frac{\sigma^2_D(1+2(1-\mu_Z)^2 \cdot A_{L_R}}{1-2(1-\mu_Z)^2+2(1-\mu_Z)^2 \cdot A_{L_R}}, \quad L_R > 1$	
$\sigma^2_{R,i}$	$(\sigma^2_{Q,wR} + \mu^2_{Q,wR})\sigma^2_Z$	$(\sigma^2_{Q,nR} + \mu^2_{Q,nR})\sigma^2_Z$
$\sigma^2_{IL,i}$	$(L_P + 1)\sigma^2_D + L_R \sigma^2_{R,wR} + (1-\mu_Z)^2 \sigma^2_{Q,wR}$	$(L_P + 1)\sigma^2_D + L_P \sigma^2_{R,nR}$
S_i	$(L_P + 1)\mu_D + (1-\mu_Z)\mu_{Q,wR}\Phi^{-1}(b/(b+h))\sigma_{IL,wR}$	$(L_P + 1)\mu_D + \Phi^{-1}(b/(b+h))\sigma_{IL,nR}$
$\mu_{IL,\,i}$	$S_{wR} - (L_P + 1)\mu_D - (1-\mu_Z)\mu_{Q,wR}$	$S_{nR} - (L_P + 1)\mu_D$

second strategy considers that all defective items are reworked. Note that whether to dispose of or rework imperfect products is a one-time decision. Once the planner has made a decision as to which strategy to choose, this decision cannot be changed, e.g., from batch to batch.

First, we focus on the differences in the model and therefore in the formulae depending on what has been decided about what to do with imperfect products. Afterwards a mathematical model incorporating different cost parameters, e.g., production and inspection cost, is presented. This model is used in the numerical analysis in Section 5.2 to examine how sensitive the decision concerning disposal or rework is to various input and cost parameters.

In a production system where defective products are disposed of, not all ordered items arrive at the warehouse. Unlike the model for rework, in the case of disposal, defective items leave the production system. Therefore, an ordering policy as presented in (4) is not suitable because in every period fewer products than required are received. Therefore, a linear inflation policy, which has been shown to perform very well (Huh & Nagarajan, 2010), is commonly used. The order quantity in this case equals:

$$Q_t = \begin{cases} F(S - IP_t), & IP_t < S \\ 0, & IP_t \geq S \end{cases} \tag{19}$$

F is called the linear inflation factor and is often defined as the reciprocal of the mean yield: $F = 1/\mu_Z$ (see e.g., Bollapragada & Morton, 1999; Huh & Nagarajan, 2010; Inderfurth & Vogelgesang, 2013; Inderfurth & Kiesmüller, 2015). The yield inflation factor takes into account that defective items are disposed of, which reduces the output of the production process. It compensates for fewer items with better quality.

Using the definition of the order quantity in (19), the steady-state formulae for the case with disposal rather than rework are required, which were derived by Inderfurth and Kiesmüller (2015) and Sonntag and Kiesmüller (2017). In Table 6, we summarize the formulae for the cases with rework (wR) and with no rework (nR).

These formulae show, that the variance of the order quantity differs a lot. Since the variance of the order quantity influences the variance of the forecast error as well as the variance of the inventory level, it has a large effect on the base-stock level S.

As already mentioned, we approximate the inventory level with a normal distribution. For a normally distributed inventory level, the average inventory holding and backorder cost can be calculated as in Sonntag and Kiesmüller (2017):

$$H(S_i, i) = (h+b)\left[\sigma_{IL,i} \cdot \varphi\left(\frac{-\mu_{IL,i}}{\sigma_{IL,i}}\right) + \mu_{IL,i} \cdot \left(1 - \Phi\left(\frac{-\mu_{IL,i}}{\sigma_{IL,i}}\right)\right)\right] - b \cdot \mu_{IL,i} \quad, \; i \in \{wR, nR\} \tag{20}$$

with the corresponding mean and variance of the inventory level as presented in Table 6.

5.1. Mathematical model

To support the decision on whether a rework station should be integrated or defective items should be scrapped, a cost model considering production and quality control costs, possible rework or disposal costs as well as holding and backorder costs is introduced to calculate the average cost per period. For simplicity, we introduce the binary variable X, which indicates whether defective products are reworked or disposed of:

$$X = \begin{cases} 1, & rework \\ 0, & disposal \end{cases} \tag{21}$$

We consider variable production costs P per period, charged with p for every produced item ($p \geq 0$). On average, the production volume equals the mean order quantity $(\mu_{Q,i})_{i \in \{wR,nR\}}$ in each period. Thus, we get

$$P(i) = p \cdot L_P \cdot \mu_{Q,i}, \quad i \in \{wR, nR\}. \tag{22}$$

We do not consider fixed production costs because only a single product is produced and thus no set-up costs are required to initialize the machine in advance of each production run. Concerning quality control costs, we neglect fixed costs for implementing such a control station because these costs are not relevant for the decision as to whether products should be scrapped or reworked. A quality control process exists in both cases. Variable quality control costs A are charged with parameter a for each item produced ($a \geq 0$):

$$A(i) = a \cdot \mu_{Q,i}, \quad i \in \{wR, nR\} \tag{23}$$

If defective products are reworked, variable costs $R(i)$ occur with parameter r ($r \geq 0$)for every defective item ($(1-\mu_Z)\mu_{Q,wR}$):

$$R(i) = X \cdot r \cdot L_R \cdot (1-\mu_Z)\mu_{Q,i}, \quad i = wR \tag{24}$$

Instead, if defective products are scrapped, disposal costs g ($g \geq 0$) are charged for each defective and thus disposed of item ($(1-\mu_Z)\mu_{Q,nR}$):

$$G(i) = (1 - X) \cdot g \cdot (1-\mu_Z)\mu_{Q,i}, \quad i = nR \tag{25}$$

Finally, inventory holding and backorder costs H are charged as in formula (20) with $h \geq 0$ and $b \geq 0$:

$$H(S_i, i) = h \int_0^\infty x\varphi_{IL,i} dx - b \int_{-\infty}^0 x\varphi_{IL,i} dx, \quad i \in \{wR, nR\} \tag{26}$$

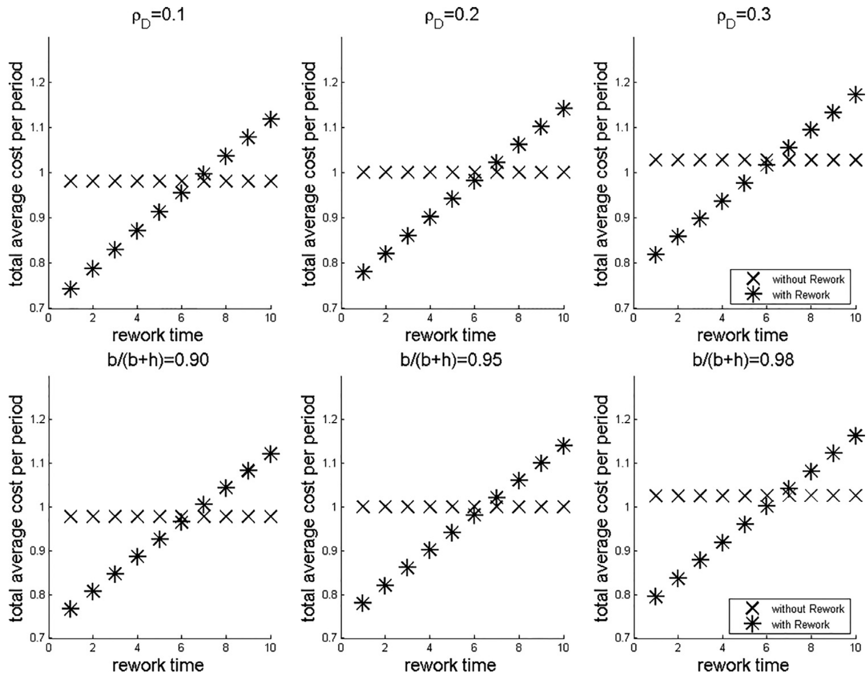


Fig. 2. Effect of input parameter variations (ρ_D and $b/(b+h)$) on total cost.

Summarizing all the different cost terms, we get the following cost function for $i \in \{wR, nR\}$:

$$\begin{aligned} TC(S_i, i) &= P(i) + A(i) + R(wR) + G(nR) + H(S_i, i) \\ &= p \cdot L_P \cdot \mu_{Q,i} \\ &\quad + a \cdot \mu_{Q,i} \\ &\quad + X \cdot r \cdot L_R \cdot (1-\mu_Z)\mu_{Q,wR} \\ &\quad + (1-X) \cdot g \cdot (1-\mu_Z)\mu_{Q,nR} \\ &\quad + h\int_0^{\infty} x\varphi_{IL,i}dx - b\int_{-\infty}^{0} x\varphi_{IL,i}dx \end{aligned} \tag{27}$$

In the next section, we analyze different production systems where all defective products are either scrapped or reworked.

5.2. Numerical analysis

To analyze the effect of different input parameters on the decision concerning whether defective items should be scrapped or reworked, we use one example as a benchmark and change one parameter after another. This analysis indicates which input parameters are critical and therefore should be accorded greater attention. As a benchmark we consider a production system with a production time L_P of ten periods and a rework time L_R which can vary between one and ten periods. The mean demand μ_D and the corresponding coefficient of variation are 20 and 0.2, respectively. The cost ratio $b/(b+h)$ is set to 0.95. The mean yield μ_Z equals 0.8 with a coefficient of variation ρ_Z of 0.3, which indicates that yield losses are not negligible but improvable.

The setting for the cost parameters is as follows: variable production cost p are charged with one unit for each produced item, quality control cost a equal the production cost. If defective items are disposed of, cost g of two units per item occur. Instead, if defective products are reworked, variable cost r of three units per item are charged. Thus, we get a rework to disposal cost ratio of 1.5 and a production to rework cost ratio of 1/3.

In the following section, we run a sensitivity analysis to illustrate which parameters are critical and should therefore receive more attention than others.

5.3. Variations in the input of the production environment

First, we change the values of the coefficients of variation of the demand, ρ_D, as well as the mean μ_Z and the coefficients of variation ρ_Z^2 of the yield. Additionally, we look at the effect of changes in the cost ratio $b/(b+h)$, which follows from service level agreements with the customers. Note that while one of the parameters is changed, all the other parameters are fixed. Obviously, changes in demand can occur over time due to the addition of new customers or varying demand quantities of existing customers. Changes in the yield arise from an improved production system, producing less defective items. The service level is agreed by contract with the customer. There are situations conceivable where a customers willingness to pay for high service increases and therefore he signs a contract, which guarantees a higher service level.

The considered scenarios are as follows: $\rho_D \in \{0.1,\ 0.2,\ 0.3\}$, $b/(b+h) \in \{0.90, 0.95, 0.98\}$, $\mu_Z \in \{0.7,\ 0.8,\ 0.9\}$ and $\rho_Z \in \{0.1,\ 0.3,\ 0.5\}$. Thus, we analyze the effect of single parameter changes – either a decrease or an increase – compared with the benchmark scenario. To make the results comparable, we calculated the total average cost per period relative to the cost of the production system without rework in the benchmark scenario.

Fig. 2 shows the results for a variation of the demand variability and the cost ratio $b/(b+h)$. It can be seen that demand variations as well as variations in the cost ratio $b/(b+h)$ have almost no effect on the decision whether to dispose of or rework defective products. The point of indifference between both strategies always lies between a rework time of six and seven periods and thus depends entirely on the length of rework times. Because of the slope of the cost when reworking defectives, the decision whether to rework or scrap defective items becomes even more important

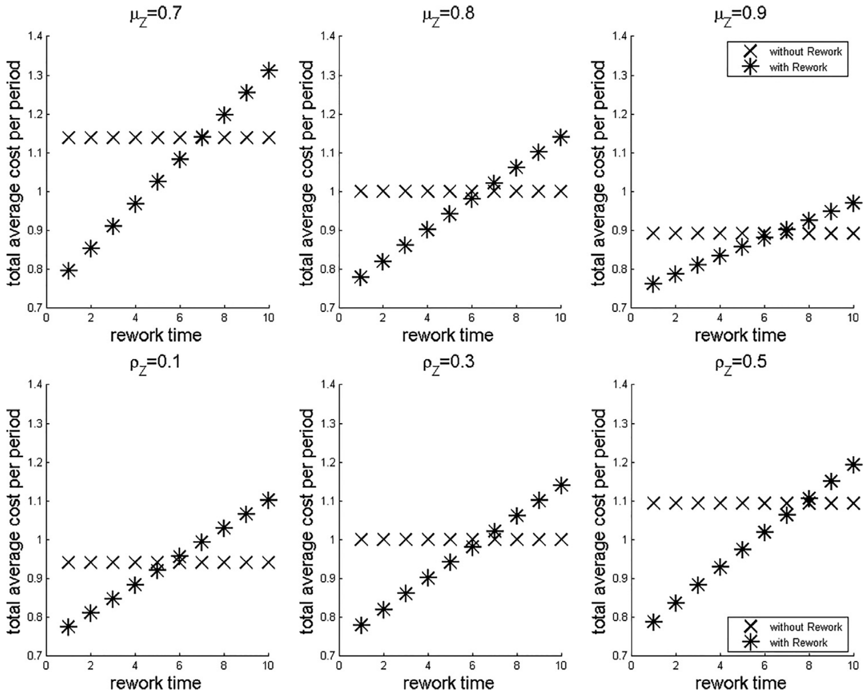


Fig. 3. Effect of input parameter variations (μ_Z and ρ_Z) on total cost.

because a wrong decision leads to a high amount of additional cost.

Fig. 3 shows the results for a variation of the mean and the coefficient of variation of the yield.

If the results are compared with the results in Fig. 2, it is obvious that the yield parameters have a greater impact. On one hand, an increasing mean yield goes in line with decreasing yield losses, which reduces total cost. On the other hand, the higher the mean yield, the less valuable is a costly rework process, represented by a decreasing slope. In other words, the greater the yield, the smaller the difference between the cost of reworking and of disposing of defective items.

As well as the mean yield, the coefficient of variation of the yield also has a large effect. We can see that, compared with the other parameters where the decision whether to rework or not was independent of the parameter setting, the variation of the yield forces the decision of a planner. For a coefficient of variation of the yield of 0.1, the point of indifference lies between a rework time of five and six periods: for a coefficient of variation of 0.2, between a rework time of six and seven periods; and for a coefficient of variation of 0.3, between a rework time of seven and eight periods. Thus, with higher yield variability reworking remains the best strategy even for larger rework times. High yield variability makes it difficult to estimate the yield losses. Thus, the probability of stock-outs during the risk period increases. The shorter the rework time, the earlier initially imperfect items arrive in the warehouse, reducing the probability that customer demand cannot be fulfilled.

The analysis illustrates the fact that the demand variability as well as the target cost ratio $b/(b+h)$ should not affect the decision of a planner on whether to scrap or rework defective items. In other words, the model is robust against variations in the demand and changes in the required cost ratio. Changes in these parameters affect only total cost. The planner should instead decide based on the mean and the variance of the yield whether it is worthwhile to rework defective items or not. While the point where the planner is undecided between rework and disposal is not affected by the demand parameters, the cost ratio or the mean yield, this point changes for different coefficients of variation of the yield. In this case, the decision on how to handle imperfect items depends heavily on the ratio between production time and rework time.

5.4. Variations in costs

We now focus on the robustness of the decision whether to rework or not if cost parameters change. Specifically, we change the ratio of rework and disposal cost r/g, the ratio of production and rework cost p/r and quality control cost a. The cost parameters may increase if the products become more and more complex over time due to new functionality. On the other hand, the cost parameters may decrease due to learning effects and improvements in the production, rework or quality control processes.

For all three cost parameters we analyze three scenarios as in the previous section:

$r/g \in \{1.25,\ 1.5,\ 1.75\}$, $p/r \in \{1/6,\ 1/3,\ 1/2\}$ and $a \in \{0.5 \cdot p,\ p,\ 1.5 \cdot p\}$. Fig. 4 shows the results.

Of course, the cost ratio of rework and disposal has no effect on total cost for the model where defective products are disposed of. However, with increasing cost ratio the slope of the total cost function increases if the products are reworked. The decision whether to dispose of or rework defective items is also affected by the rework cost parameter. The higher the rework cost per item and period relative to the disposal cost, the lower the rework time L_R has to be to make rework profitable. The longer the rework times, the higher the cost during the reworking process for bringing imperfect products to a condition equivalent to that of items that were produced correctly in the first place.

If we look at the production to rework cost ratio, reworking becomes more favorable with increasing production costs even if the rework time equals the production time. The reason for this is that

Please cite this article as: D. Sonntag, G.P. Kiesmüller, Disposal versus rework – Inventory control in a production system with random yield, European Journal of Operational Research (2017), https://doi.org/10.1016/j.ejor.2017.11.019

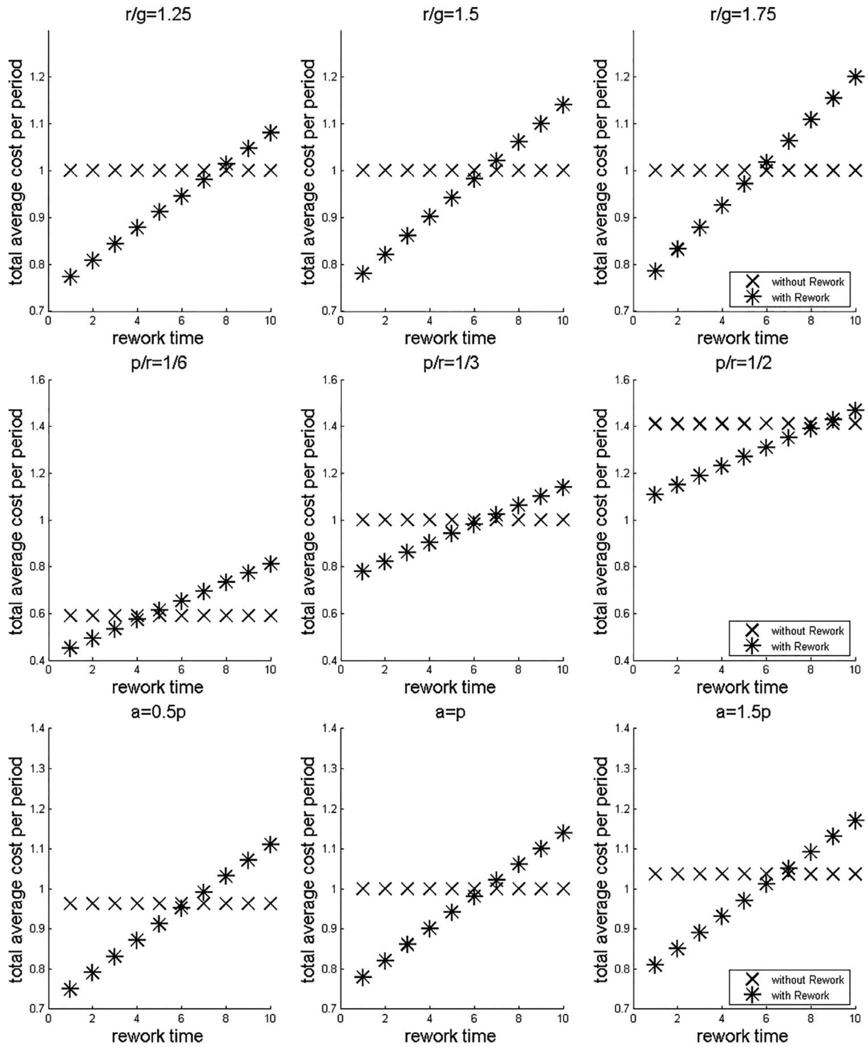


Fig. 4. Effect of cost parameter variations on total cost.

if all defective items are scrapped, these products have to be produced again, which becomes more expensive with increasing production costs. If production becomes expensive, the length of the rework time becomes more and more negligible to a point where reworking is always cheaper than producing defective items again.

Next, a change in quality control cost is analyzed. It is obvious that this parameter has only small effects. Only the total costs increase because of the higher cost parameter. The small effect arises due to small changes in the cost parameter. For larger variations an effect is recognizable in such a way that reworking becomes more and more favorable with an increasing quality control cost per item. This effect results from the assumption that reworked products are of perfect quality without any further inspection.

With these results in mind, a planner should focus mainly on production and rework costs when he has to decide whether to scrap or rework defective items. The higher the production cost relative to rework cost, the more profitable is rework.

5.5. Managerial insights

Whether to dispose or rework defective items is a difficult decision for a planner due to lots of different influencing parameters. It is obvious that reworking becomes more favorable for shorter rework times and lower corresponding costs compared to production and disposal. Nevertheless, a planner should not only focus on these effects but should be aware that the decision between rework and disposal has wide-ranging consequences especially regarding the required safety stock. Reducing the safety stock is important especially in situations where limited storage capacity is available. Although rework is in some situations more costly than disposal, the safety stock level under rework is always below the safety stock level under disposal. The reason for this is that under rework former defective items enter the warehouse after rework at a certain time wherefore these quantities – different from disposal – do not have to be reordered.

Even though the safety stock level is lower under rework, the difference compared to disposal of defectives highly depends on the parameter setting. As an example consider the demand variability, which has no effect on the question whether disposal or rework is favorable, but a high effect on the safety stock level. The difference in safety stock between rework and disposal decreases with increasing demand variability, which means rework is even more favorable for low demand variability. This is not intuitive at the first moment, because one would expect that rework becomes more favorable for increasing uncertainty in the system. The effect can be explained in two ways: First, under low demand variability the uncertainty due to random yield is dominant. This uncertainty is lower if defective items are reworked and therefore do not leave the system and do not have to be reordered. Second, under low demand variability less safety stock is available to hedge against demand uncertainty. Thus, there is less stock available which can also be used to hedge against yield uncertainty. Indeed, there is a pooling effect which reduces the safety stock level under high demand and yield uncertainty.

6. Summary and Outlook

We studied a single-stage production system with stochastic proportional yield, which results in random yield losses in each period. Since only items of perfect quality are stored in the warehouse to satisfy stochastic customer demand, defective products are either disposed of or reworked. We assumed that the reworking process converts all defective items into products that satisfy the required quality standards.

Our contribution was (1) to show how to determine the base-stock level minimizing average inventory holding and backorder cost in a production environment where defective products are not disposed of but reworked, and (2) to develop a mathematical model to be used as a decision-making support for the planner when it comes to the question of whether defective items should be disposed of or reworked.

The results are as follows: (1) the adaptation of the steady-state approach to a situation where defective products are not disposed of but reworked is not easy. The reworking process results in covariances between orders, which are difficult to calculate exactly because the joint distribution is unknown. We presented an approximation of the covariances depending on the reworking times. A numerical study confirmed that the approximation works very well. In 319 of 528 instances the approximation leads to the optimal solution as determined by the simulation. For all other 209 instances the average deviation equals approximately 0.10%. Over all 528 instances we get an average deviation from optimum of 0.10% with a standard deviation of 0.23%, which is excellent.

(2) We introduced a mathematical model addressing production, quality control, rework and inventory holding, and backorder cost. Based on this model, we analyzed the effect of varying parameter settings. The results show that the demand variation as well as the cost ratio $b/(b+h)$ has nearly no effect on cost and on the decision whether to dispose of or rework defective items. Thus, the model is robust for these parameters. On the other hand, the mean and the coefficient of variation of the yield have an enormous effect. The higher the mean yield, the less valuable is reworking because only a few items are of imperfect quality. The higher the coefficient of variation of the yield, the more valuable is a rework. Concerning a change in the cost parameters, the ratio of production and rework cost is the main determiner of whether to dispose of or rework a defective item.

Future research should focus on production systems where the planner can decide in each period if he or she wants to rework defective items or dispose of them. A mixture of both strategies is conceivable, where some products are scrapped and some are reworked. Furthermore, the reworking process like the production process might be imperfect, which means that either good products would be classified as defective or imperfect products would stay in the system by mistake. In such situations it might be necessary to place several inspection stations in tandem.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.ejor.2017.11.019

References

Bollapragada, S., & Morton, T. E. (1999). Myopic heuristics for the random yield problem. *Operations Research, 47*(5), 713–722.
Buscher, U., & Lindner, G. (2007). Optimizing a production system with rework and equal sized batch shipments. *Computers & Operations Research, 34*(2), 515–535.
Chen, Y.-S. (2010). The drivers of green brand equity: Green brand image, green satisfaction, and green trust. *Journal of Business Ethics, 93*(2), 307–319.
Dettenbach, M., & Thonemann, U. W. (2015). The value of real time yield information in multi-stage inventory systems – exact and heuristic approaches. *European Journal of Operational Research, 240*(1), 72–83.
Ehrhardt, R., & Taube, L. (1987). An inventory model with random replenishment quantities. *International Journal of Production Research, 25*(12), 1795–1803.
Flapper, S. D. P., Fransoo, J. C., Broekmeulen, R. A., & Inderfurth, K. (2002). Planning and control of rework in the process industries: A review. *Production Planning & Control, 13*(1), 26–34.
Foremski, T. (2012). Intel's battle with ARM is about making its future fabs viable. http://www.siliconvalleywatcher.com/mt/archives/2012/09/intels_battle_w.php (Accessed 18-April-2017).
Gerchak, Y., Tripathy, A., & Wang, K. (1996). Co-production models with random functionality yields. *IIE Transactions, 28*(5), 391–403.
Gotzel, C. (2010). Materialplanung bei internen produktkreisläufen. In *Mrp zur materialplanung für kreislaufprozesse* (pp. 101–234). Springer.
Gotzel, C., & Inderfurth, K. (2005). Approximate policies for hybrid production and rework systems with stochastic demand and yield. In *Operations research proceedings* (pp. 41–49). Springer.
Grosfeld-Nir, A., & Gerchak, Y. (2002). Multistage production to order with rework capability. *Management Science, 48*(5), 652–664.
Grosfeld-Nir, A., & Gerchak, Y. (2004). Multiple lotsizing in production to order with random yields: Review of recent advances. *Annals of Operations Research, 126*(1–4), 43–69.
Henig, M., & Gerchak, Y. (1990). The structure of periodic review policies in the presence of random yield. *Operations Research, 38*(4), 634–643.
Hsu, A., & Bassok, Y. (1999). Random yield and random demand in a production system with downward substitution. *Operations Research, 47*(2), 277–290.
Huh, W. T., & Nagarajan, M. (2010). Technical note – linear inflation rules for the random yield problem: Analysis and computations. *Operations research, 58*(1), 244–251.
Inderfurth, K., & Kiesmüller, G. P. (2015). Exact and heuristic linear-inflation policies for an inventory model with random yield and arbitrary lead times. *European Journal of Operational Research, 245*(1), 109–120.
Inderfurth, K., Kovalyov, M. Y., Ng, C., & Werner, F. (2007). Cost minimizing scheduling of work and rework processes on a single facility under deterioration of reworkables. *International Journal of Production Economics, 105*(2), 345–356.
Inderfurth, K., Lindner, G., & Rachaniotis, N. (2005). Lot sizing in a production system with rework and product deterioration. *International Journal of Production Research, 43*(7), 1355–1374.
Inderfurth, K., & Vogelgesang, S. (2013). Concepts for safety stock determination under stochastic demand and different types of random production yield. *European Journal of Operational Research, 224*(2), 293–301.
Jaber, M. Y., & Khan, M. (2010). Managing yield by lot splitting in a serial production line with learning, rework and scrap. *International Journal of Production Economics, 124*(1), 32–39.
Kiesmüller, G. P. (2003). A new approach for controlling a hybrid stochastic manufacturing/ remanufacturing system with inventories and different leadtimes. *European Journal of Operational Research, 147*(1), 62–71.
Liu, J. J., & Yang, P. (1996). Optimal lot-sizing in an imperfect production system with homogeneous reworkable jobs. *European Journal of Operational Research, 91*(3), 517–527.
McNutt, C. (2015). Rumor: Low yield 3D glass forcing Samsung to look for additional suppliers for Galaxy S6 Edge. http://www.androidheadlines.com/2015/03/rumor-low-yield-3d-glass-forcing-samsung-look-additional-suppliers-galaxy-s6-edge.html (Accessed 18-April-2017).
So, K. C., & Tang, C. S. (1995). Optimal operating policy for a bottleneck with random rework. *Management Science, 41*(4), 620–636.
Sonntag, D., & Kiesmüller, G. P. (2016). The shape of the yield and its impact on inventory decisions. *4OR, 14*(4), 405–415.
Sonntag, D., & Kiesmüller, G. P. (2017). The influence of quality inspections on the optimal safety stock level. *Production and Operations Management, 26*(7), 1284–1298.
Tempelmeier, H. (2006). *Inventory management in supply networks: Problems, models, solutions*. Norderstedt: Books on Demand.
Teunter, R. H., & Flapper, S. D. P. (2003). Lot-sizing for a single-stage single-product production system with rework of perishable production defectives. *OR Spectrum, 25*(1), 85–96.
Wein, A. S. (1992). Random yield, rework and scrap in a multistage batch manufacturing environment. *Operations Research, 40*(3), 551–563.
Yano, C. A., & Lee, H. L. (1995). Lot sizing with random yields: A review. *Operations Research, 43*(2), 311–334.

Supplementary Materials

Appendix A

$$\begin{aligned}
E[IL_{t+L_p}] &= E[IP_t - \sum_{l=0}^{L_P} D_{t+l} + \mu_Z Q_t - R_t - \sum_{l=1}^{min\{L_P,L_R\}-1} R_{t-l}] \\
&= E[(S - Q_t) - \sum_{l=0}^{L_P} D_{t+l} + Z_t Q_t - \sum_{l=1}^{min\{L_P,L_R\}-1} R_{t-l}] \\
&= E[S - (1 - Z_t)Q_t - \sum_{l=0}^{L_P} D_{t+l} - \sum_{l=1}^{min\{L_P,L_R\}-1} R_{t-l}] \\
&= S - (1 - \mu_Z)\mu_Q - (L_P + 1)\mu_D - (min\{L_P, L_R\} - 1)\mu_R \\
&= S - (L_P + 1)\mu_D - (1 - \mu_Z)\mu_Q
\end{aligned} \tag{28}$$

Appendix B

$$\begin{aligned}
VAR[IL_{t+L_p}] &= VAR[IP_t - \sum_{l=0}^{L_P} D_{t+l} + \mu_Z Q_t - R_t - \sum_{l=1}^{min\{L_P,L_R\}-1} R_{t-l}] \\
&= VAR[(S - Q_t) - \sum_{l=0}^{L_P} D_{t+l} + Z_t Q_t - \sum_{l=1}^{min\{L_P,L_R\}-1} R_{t-l}] \\
&= VAR[S - (1 - Z_t)Q_t - \sum_{l=0}^{L_P} D_{t+l} - \sum_{l=1}^{min\{L_P,L_R\}-1} R_{t-l}] \\
&= VAR[S] + VAR[(1 - Z_t)Q_t] + VAR[\sum_{l=0}^{L_P} D_{t+l}] + VAR[\sum_{l=1}^{min\{L_P,L_R\}-1} R_{t-l}] \\
&= VAR[(1 - Z_t)Q_t] + (L_P + 1)VAR[D_{t+l}] + (min\{L_P, L_R\} - 1)VAR[R_{t-l}] \\
&= VAR[(1 - Z_t)Q_t] + (L_P + 1)\sigma_D^2 + (min\{L_P, L_R\} - 1)\sigma_R^2
\end{aligned} \tag{29}$$

with $VAR[(1 - Z_t)Q_t]$ equal to:

$$\begin{aligned}
VAR[(1 - Z_t)Q_t] &= E[(1 - Z_t)^2 Q_t^2] - E[(1 - Z_t)Q_t]^2 \\
&= E[(1 - 2Z_t + Z_t^2)Q_t^2] - E[(1 - Z_t)Q_t]^2 \\
&= (1 - 2\mu_Z + E[Z_t^2])E[Q_t^2] - [(1 - \mu_Z)\mu_Q]^2 \\
&= (1 - 2\mu_Z + (\mu_Z^2 + \sigma_Z^2))(\mu_Q^2 + \sigma_Q^2) - (1 - \mu_Z)^2\mu_Q^2 \\
&= ((1 - \mu_Z)^2 + \sigma_Z^2)(\mu_Q^2 + \sigma_Q^2) - (1 - \mu_Z)^2\mu_Q^2 \\
&= \sigma_Q^2(1 - \mu_Z)^2 + (\mu_Q^2 + \sigma_Q^2)\sigma_Z^2 \\
&= \sigma_Q^2(1 - \mu_Z)^2 + \sigma_R^2
\end{aligned} \tag{30}$$

Thus, we get for the variance of the inventory level

$$\begin{aligned}
VAR[IL_{t+L_p}] &= VAR[(1 - Z_t)Q_t] + (L_P + 1)\sigma_D^2 + (min\{L_P, L_R\} - 1)\sigma_R^2 \\
&= \sigma_Q^2(1 - \mu_Z)^2 + \sigma_R^2 + (L_P + 1)\sigma_D^2 + (min\{L_P, L_R\} - 1)\sigma_R^2 \\
&= (L_P + 1)\sigma_D^2 + min\{L_P, L_R\}\sigma_R^2 + (1 - \mu_Z)^2\sigma_Q^2
\end{aligned} \tag{31}$$

Appendix C

In case that the production time exceeds the rework time ($L_P \geq L_R$), the order quantity can be calculated using the definition of the inventory position in (2):

$$\begin{aligned} Q_t &= S - IP_t \\ &= S - \left[IL_{t-1} + \sum_{l=L_P+1}^{L_P+L_R} (1 - Z_{t-l}) Q_{t-l} + \sum_{l=L_R}^{L_P} Q_{t-l} + \sum_{l=1}^{L_R-1} \mu_Z Q_{t-l} \right] \end{aligned} \tag{32}$$

The inventory level IL_{t-1} at the end of period $t-1$ equals the inventory level at the end of period $t-2$ less the incoming customer demand D_{t-1} between the end of period $t-2$ and $t-1$, while stock increases due to the arrival of an order leaving production and one leaving rework:

$$IL_{t-1} = IL_{t-2} - D_{t-1} + (1 - Z_{t-1-L_P-L_R}) Q_{t-1-L_P-L_R} + Z_{t-1-L_P} Q_{t-1-L_P} \tag{33}$$

$(1 - Z_{t-1-L_P-L_R}) Q_{t-1-L_P-L_R}$ equals the amount leaving the rework process. The underlying order was placed in period $t - 1 - L_P - L_R$. The last term $Z_{t-1-L_P} Q_{t-1-L_P}$ reflects the fraction of good items leaving production after L_P periods wherefore the underlying order was placed in period $t - 1 - L_P$.

Using the recursive definition of the inventory level in (33), the order quantity in (32) can be rewritten as

$$\begin{aligned} Q_t &= S - [IL_{t-2} - D_{t-1} + (1 - Z_{t-1-L_P-L_R}) Q_{t-1-L_P-L_R} + Z_{t-1-L_P} Q_{t-1-L_P}] \\ &\quad - \left[\sum_{l=L_P+1}^{L_P+L_R} (1 - Z_{t-l}) Q_{t-l} + \sum_{l=L_R}^{L_P} Q_{t-l} + \sum_{l=1}^{L_R-1} \mu_Z Q_{t-l} \right]. \end{aligned} \tag{34}$$

Using equation (32), the order quantity in period $t-1$ equals

$$\begin{aligned} Q_{t-1} &= S - IP_{t-1} \\ &= S - \left[IL_{t-2} + \sum_{l=L_P+1}^{L_P+L_R} (1 - Z_{t-1-l}) Q_{t-1-l} + \sum_{l=L_R}^{L_P} Q_{t-1-l} + \sum_{l=1}^{L_R-1} \mu_Z Q_{t-1-l} \right]. \end{aligned} \tag{35}$$

We transform equation 35 to determine an equation for $S - I_{t-2}$ in formula (34):

$$S - IL_{t-2} = Q_{t-1} + \sum_{l=L_P+1}^{L_P+L_R} (1 - Z_{t-1-l}) Q_{t-1-l} + \sum_{l=L_R}^{L_P} Q_{t-1-l} + \sum_{l=1}^{L_R-1} \mu_Z Q_{t-1-l} \tag{36}$$

$S - IL_{t-2}$ is plugged in formula (34):

$$\begin{aligned} Q_t &= S - IL_{t-2} - [-D_{t-1} + (1 - Z_{t-1-L_P-L_R}) Q_{t-1-L_P-L_R} + Z_{t-1-L_P} Q_{t-1-L_P}] \\ &\quad - \left[\sum_{l=L_P+1}^{L_P+L_R} (1 - Z_{t-l}) Q_{t-l} + \sum_{l=L_R}^{L_P} Q_{t-l} + \sum_{l=1}^{L_R-1} \mu_Z Q_{t-l} \right] \end{aligned}$$

$$
\begin{aligned}
&= Q_{t-1} + \sum_{l=L_P+1}^{L_P+L_R} (1 - Z_{t-1-l}) Q_{t-1-l} + \sum_{l=L_R}^{L_P} Q_{t-1-l} + \sum_{l=1}^{L_R-1} \mu_Z Q_{t-1-l} \\
&\quad + D_{t-1} - (1 - Z_{t-1-L_P-L_R}) Q_{t-1-L_P-L_R} - Z_{t-1-L_P} Q_{t-1-L_P} \\
&\quad - \sum_{l=L_P+1}^{L_P+L_R} (1 - Z_{t-l}) Q_{t-l} - \sum_{l=L_R}^{L_P} Q_{t-l} - \sum_{l=1}^{L_R-1} \mu_Z Q_{t-l} \\
&= Q_{t-1} + \sum_{l=L_P+1}^{L_P+L_R} (1 - Z_{t-1-l}) Q_{t-1-l} \\
&\quad - \left(\sum_{l=L_P+2}^{L_P+L_R+1} (1 - Z_{t-l}) Q_{t-l} + (1 - Z_{t-L_P-1}) Q_{t-L_P-1} - (1 - Z_{t-L_P-L_R-1}) Q_{t-L_P-L_R-1} \right) \\
&\quad + \sum_{l=L_R}^{L_P} Q_{t-1-l} - \left(\sum_{l=L_R+1}^{L_P+1} Q_{t-l} + Q_{t-L_R} - Q_{t-L_P-1} \right) \\
&\quad + \sum_{l=1}^{L_R-1} \mu_Z Q_{t-1-l} - \left(\sum_{l=2}^{L_R} \mu_Z Q_{t-l} + \mu_Z Q_{t-1} - \mu_Z Q_{t-L_R} \right) \\
&\quad + D_{t-1} - (1 - Z_{t-1-L_P-L_R}) Q_{t-1-L_P-L_R} - Z_{t-1-L_P} Q_{t-1-L_P}
\end{aligned} \tag{37}
$$

Simplifying this equation leads to

$$
\begin{aligned}
Q_t &= Q_{t-1} + - ((1 - Z_{t-L_P-1}) Q_{t-L_P-1} - (1 - Z_{t-L_P-L_R-1}) Q_{t-L_P-L_R-1}) \\
&\quad - (Q_{t-L_R} - Q_{t-L_P-1}) - (\mu_Z Q_{t-1} - \mu_Z Q_{t-L_R}) \\
&\quad + D_{t-1} - (1 - Z_{t-1-L_P-L_R}) Q_{t-1-L_P-L_R} - Z_{t-1-L_P} Q_{t-1-L_P} \\
&= D_{t-1} + (1 - \mu_Z) Q_{t-1} \\
&\quad - (1 - Z_{t-L_P-1}) Q_{t-L_P-1} - Z_{t-1-L_P} Q_{t-1-L_P} + Q_{t-L_P-1} \qquad (38) \\
&\quad (Q_{t-L_R} + \mu_Z Q_{t-L_R} \\
&= D_{t-1} + (1 - \mu_Z) Q_{t-1} - (1 - \mu_Z) Q_{t-L_R} \qquad (39)
\end{aligned}
$$

Thus, the order quantity in period t equals the demand of the previous period plus two additional terms. It is obvious, that for $L_R = 1$ the order quantity equals the demand of the previous period: $Q_t = D_{t-1}$. The term $+(1 - \mu_Z) Q_{t-1}$ compensates for the items of the order in the previous period, which are expected to be defective and thus will move to the rework process. The last term arises through the forecast error which adjusts the amount of items if new information about realized yields become known.

For the case, that the production time is smaller than the rework time ($L_P < L_R$), the procedure is similar:

$$
\begin{aligned}
Q_t &= S - IP_t \\
&= S - \left[IL_{t-1} + Z_{t-L_P} Q_{t-L_P} + \sum_{l=1}^{L_P-1} \mu_Z Q_{t-l} + \sum_{l=L_R}^{L_P+L_R} (1 - Z_{t-l}) Q_{t-l} \right]
\end{aligned} \tag{40}
$$

with

$$
IL_{t-1} = IL_{t-2} - D_{t-1} + (1 - Z_{t-1-L_P-L_R}) Q_{t-1-L_P-L_R} + Z_{t-1-L_P} Q_{t-1-L_P} \tag{41}
$$

and

$$Q_{t-1} = S - \left[IL_{t-2} + Z_{t-1-L_P} Q_{t-1-L_P} + \sum_{l=1}^{L_P-1} \mu_Z Q_{t-1-l} + \sum_{l=L_R}^{L_P+L_R} (1 - Z_{t-1-l}) Q_{t-1-l} \right] \quad (42)$$

Thus,

$$S - IL_{t-2} = Q_{t-1} + Z_{t-1-L_P} Q_{t-1-L_P} + \sum_{l=1}^{L_P-1} \mu_Z Q_{t-1-l} + \sum_{l=L_R}^{L_P+L_R} (1 - Z_{t-1-l}) Q_{t-1-l} \quad (43)$$

which will be plugged in (40) under consideration of (41):

$$\begin{aligned}
Q_t &= S - [IL_{t-2} - D_{t-1} + (1 - Z_{t-1-L_P-L_R}) Q_{t-1-L_P-L_R} + Z_{t-1-L_P} Q_{t-1-L_P}] \\
&\quad - \left[Z_{t-L_P} Q_{t-L_P} + \sum_{l=1}^{L_P-1} \mu_Z Q_{t-l} + \sum_{l=L_R}^{L_P+L_R} (1 - Z_{t-l}) Q_{t-l} \right] \\
&= Q_{t-1} + Z_{t-1-L_P} Q_{t-1-L_P} + \sum_{l=1}^{L_P-1} \mu_Z Q_{t-1-l} + \sum_{l=L_R}^{L_P+L_R} (1 - Z_{t-1-l}) Q_{t-1-l} \\
&\quad + D_{t-1} - (1 - Z_{t-1-L_P-L_R}) Q_{t-1-L_P-L_R} - Z_{t-1-L_P} Q_{t-1-L_P} \\
&\quad - Z_{t-L_P} Q_{t-L_P} - \sum_{l=1}^{L_P-1} \mu_Z Q_{t-l} - \sum_{l=L_R}^{L_P+L_R} (1 - Z_{t-l}) Q_{t-l} \qquad (44)
\end{aligned}$$

The order quantity in period t then equals:

$$\begin{aligned}
Q_t &= Q_{t-1} + D_{t-1} + \sum_{l=L_R}^{L_P+L_R} (1 - Z_{t-1-l}) Q_{t-1-l} \\
&\quad - \left[\sum_{l=L_R+1}^{L_P+L_R+1} (1 - Z_{t-l}) Q_{t-l} + (1 - Z_{t-L_R}) Q_{t-L_R} - (1 - Z_{t-L_P-L_R-1}) Q_{t-L_P-L_R-1} \right] \\
&\quad + \sum_{l=1}^{L_P-1} \mu_Z Q_{t-1-l} - \left[\sum_{l=2}^{L_P} \mu_Z Q_{t-l} + \mu_Z Q_{t-1} - \mu_Z Q_{t-L_P} \right] \\
&\quad - (1 - Z_{t-1-L_P-L_R}) Q_{t-1-L_P-L_R} - Z_{t-L_P} Q_{t-L_P} \\
&= Q_{t-1} + D_{t-1} \\
&\quad - [(1 - Z_{t-L_R}) Q_{t-L_R} - (1 - Z_{t-L_P-L_R-1}) Q_{t-L_P-L_R-1}] \\
&\quad - [\mu_Z Q_{t-1} - \mu_Z Q_{t-L_P}] \\
&\quad - (1 - Z_{t-1-L_P-L_R}) Q_{t-1-L_P-L_R} - Z_{t-L_P} Q_{t-L_P} \\
&= D_{t-1} + (1 - \mu_Z) Q_{t-1} - (1 - Z_{t-L_R}) Q_{t-L_R} + (\mu_Z - Z_{t-L_P}) Q_{t-L_P} \qquad (45)
\end{aligned}$$

Appendix D

1. *Calculation of $COV[Q_{t-1}, Q_{t-LR}]$ for $LR = 2$*

$$\begin{aligned}
COV[Q_{t-1}, Q_{t-2}] &= COV[(1-\mu_Z)Q_{t-2} - (1-\mu_Z)Q_{t-3}, Q_{t-2}] \\
&= (1-\mu_Z)[\sigma_Q^2 - COV[Q_{t-2}, Q_{t-3}]] \\
&= (1-\mu_Z)[\sigma_Q^2 - COV[(1-\mu_Z)Q_{t-3} - (1-\mu_Z)Q_{t-4}, Q_{t-3}]] \\
&= (1-\mu_Z)[\sigma_Q^2 - ((1-\mu_Z)\sigma_Q^2 - (1-\mu_Z)COV[Q_{t-3}, Q_{t-4}])] \\
&= (1-\mu_Z)\sigma_Q^2[1 - (1-\mu_Z) + (1-\mu_Z)COV[Q_{t-3}, Q_{t-4}]] \\
&\dots \\
&= (1-\mu_Z)\sigma_Q^2 \sum_{k=0}^{\infty} (-(1-\mu_Z))^k \\
&\approx \sigma_Q^2 \frac{1-\mu_Z}{2-\mu_Z}
\end{aligned}$$

2. *Calculation of $COV[Q_{t-1}, Q_{t-LR}]$ for $LR = 3$*

$$\begin{aligned}
COV[Q_{t-1}, Q_{t-3}] &= COV[(1-\mu_Z)Q_{t-2} - (1-\mu_Z)Q_{t-4}, Q_{t-3}] \\
&= (1-\mu_Z)(COV[Q_{t-2}, Q_{t-3}] - COV[Q_{t-3}, Q_{t-4}] \\
&= 0
\end{aligned}$$

3. *Calculation of $COV[Q_{t-1}, Q_{t-LR}]$ for $LR = 4$*

$$\begin{aligned}
COV[Q_{t-1}, Q_{t-4}] &= (1-\mu_Z)COV[Q_{t-2}, Q_{t-4}] - (1-\mu_Z)COV[Q_{t-4}, Q_{t-5}] \\
&= lag1 \cdot [-(1-\mu_Z)] \\
&\quad +(1-\mu_Z)COV[Q_{t-2}, Q_{t-4}]
\end{aligned}$$

$$\begin{aligned}
COV[Q_{t-1}, Q_{t-4}] &= lag1 \cdot [-(1-\mu_Z)] \\
&\quad +(1-\mu_Z)^2 COV[Q_{t-3}, Q_{t-4}] - (1-\mu_Z)^2 COV[Q_{t-4}, Q_{t-6}] \\
&= lag1 \cdot \left[-(1-\mu_Z) + (1-\mu_Z)^2\right] \\
&\quad -(1-\mu_Z)^2 COV[Q_{t-4}, Q_{t-6}] \\
&= lag1 \cdot \left[-(1-\mu_Z) + (1-\mu_Z)^2\right] \\
&\quad -(1-\mu_Z)^3 COV[Q_{t-5}, Q_{t-6}] + (1-\mu_Z)^3 COV[Q_{t-6}, Q_{t-8}] \\
&= lag1 \cdot \left[-(1-\mu_Z) + (1-\mu_Z)^2 - (1-\mu_Z)^3\right] \\
&\quad +(1-\mu_Z)^3 COV[Q_{t-6}, Q_{t-8}] \\
&\dots \\
&= lag1 \cdot \left[-(1-\mu_Z) \sum_{k=0}^{\infty} (-(1-\mu_Z))^k\right] \\
&\approx lag1 \cdot \frac{-(1-\mu_Z)}{1-(-(1-\mu_Z))} \\
&\approx lag1 \cdot \frac{-(1-\mu_Z)}{2-\mu_Z}
\end{aligned}$$

4. *Calculation of $COV[Q_{t-1}, Q_{t-LR}]$ for $LR = 5$*

$$\begin{aligned}
COV[Q_{t-1}, Q_{t-5}] &= (1-\mu_Z)COV[Q_{t-2}, Q_{t-5}] - (1-\mu_Z)COV[Q_{t-5}, Q_{t-6}] \\
&= lag1 \cdot [-(1-\mu_Z)] \\
&\quad +(1-\mu_Z)COV[Q_{t-2}, Q_{t-5}] \\
&= lag1 \cdot [-(1-\mu_Z)] \\
&\quad +(1-\mu_Z)^2 COV[Q_{t-3}, Q_{t-5}] - (1-\mu_Z)^2 COV[Q_{t-5}, Q_{t-7}] \\
&= lag1 \cdot [-(1-\mu_Z)]
\end{aligned}$$

5. *Calculation of $COV[Q_{t-1}, Q_{t-LR}]$ for $LR = 6$*

$$\begin{aligned}
COV[Q_{t-1}, Q_{t-6}] &= (1-\mu_Z)COV[Q_{t-2}, Q_{t-6}] - (1-\mu_Z)COV[Q_{t-6}, Q_{t-7}] \\
&= (1-\mu_Z)^2 COV[Q_{t-3}, Q_{t-6}] - (1-\mu_Z)^2 COV[Q_{t-6}, Q_{t-8}] \\
&\quad -(1-\mu_Z)COV[Q_{t-6}, Q_{t-7}] \\
&= (1-\mu_Z)^3 COV[Q_{t-4}, Q_{t-6}] - (1-\mu_Z)^3 COV[Q_{t-6}, Q_{t-9}] \\
&\quad -(1-\mu_Z)^3 COV[Q_{t-7}, Q_{t-8}] + (1-\mu_Z)^3 COV[Q_{t-8}, Q_{t-12}] \\
&\quad -(1-\mu_Z)COV[Q_{t-6}, Q_{t-7}] \\
&= lag1 \cdot \left[-(1-\mu_Z) - (1-\mu_Z)^3\right] + (1-\mu_Z)^3 COV[Q_{t-8}, Q_{t-12}] \\
&\quad +(1-\mu_Z)^3 COV[Q_{t-4}, Q_{t-6}] - (1-\mu_Z)^3 COV[Q_{t-6}, Q_{t-9}] \\
&= lag1 \cdot \left[-(1-\mu_Z) - (1-\mu_Z)^3 + (1-\mu_Z)^4\right] \\
&\quad -(1-\mu_Z)^4 COV[Q_{t-6}, Q_{t-10}] \\
&\quad -(1-\mu_Z)^4 COV[Q_{t-7}, Q_{t-9}] + (1-\mu_Z)^4 COV[Q_{t-9}, Q_{t-12}] \\
&\quad +(1-\mu_Z)^4 COV[Q_{t-9}, Q_{t-12}] - (1-\mu_Z)^4 COV[Q_{t-12}, Q_{t-14}] \\
&= lag1 \cdot \left[-(1-\mu_Z) - (1-\mu_Z)^3 + (1-\mu_Z)^4\right] \\
&\quad -(1-\mu_Z)^4 COV[Q_{t-6}, Q_{t-10}] - (1-\mu_Z)^4 COV[Q_{t-7}, Q_{t-9}] \\
&\quad +2(1-\mu_Z)^4 COV[Q_{t-9}, Q_{t-12}] - (1-\mu_Z)^4 COV[Q_{t-12}, Q_{t-14}] \\
&= lag1 \cdot \left[-(1-\mu_Z) - (1-\mu_Z)^3 + (1-\mu_Z)^4 - 2(1-\mu_Z)^5\right] \\
&\quad -(1-\mu_Z)^5 COV[Q_{t-7}, Q_{t-10}] + (1-\mu_Z)^5 COV[Q_{t-9}, Q_{t-13}] \\
&\quad +3(1-\mu_Z)^5 COV[Q_{t-10}, Q_{t-12}] - 2(1-\mu_Z)^5 COV[Q_{t-12}, Q_{t-15}] \\
&\quad +(1-\mu_Z)^5 COV[Q_{t-14}, Q_{t-18}] \\
&= lag1 \cdot \left[-(1-\mu_Z) - (1-\mu_Z)^3 + (1-\mu_Z)^4 - 2(1-\mu_Z)^5 + 3(1-\mu_Z)^6\right] \\
&\quad -(1-\mu_Z)^6 COV[Q_{t-8}, Q_{t-10}] + 2(1-\mu_Z)^6 COV[Q_{t-10}, Q_{t-13}] \\
&\quad -3(1-\mu_Z)^6 COV[Q_{t-12}, Q_{t-16}] - 3(1-\mu_Z)^6 COV[Q_{t-13}, Q_{t-15}] \\
&\quad +3(1-\mu_Z)^6 COV[Q_{t-15}, Q_{t-18}] - (1-\mu_Z)^6 COV[Q_{t-18}, Q_{t-20}]
\end{aligned}$$

$$\begin{aligned}
COV[Q_{t-1}, Q_{t-6}] &= lag1 \cdot [-(1-\mu_Z) - (1-\mu_Z)^3 + (1-\mu_Z)^4 - 2(1-\mu_Z)^5 + 3(1-\mu_Z)^6 \\
&\quad -5(1-\mu_Z)^7] \\
&\quad +(1-\mu_Z)^7 COV[Q_{t-10}, Q_{t-14}] + 2(1-\mu_Z)^7 COV[Q_{t-11}, Q_{t-13}] \\
&\quad -5(1-\mu_Z)^7 COV[Q_{t-13}, Q_{t-16}] + 3(1-\mu_Z)^7 COV[Q_{t-15}, Q_{t-19}] \\
&\quad +6(1-\mu_Z)^7 COV[Q_{t-16}, Q_{t-18}] - 3(1-\mu_Z)^7 COV[Q_{t-18}, Q_{t-21}] \\
&\quad +(1-\mu_Z)^7 COV[Q_{t-20}, Q_{t-24}] \\
&= lag1 \cdot [-(1-\mu_Z) - (1-\mu_Z)^3 + (1-\mu_Z)^4 - 2(1-\mu_Z)^5 + 3(1-\mu_Z)^6 \\
&\quad -5(1-\mu_Z)^7 \\
&\quad +8(1-\mu_Z)^8] \\
&\quad +(1-\mu_Z)^8 COV[Q_{t-11}, Q_{t-14}] - 2(1-\mu_Z)^8 COV[Q_{t-13}, Q_{t-17}] \\
&\quad -6(1-\mu_Z)^8 COV[Q_{t-14}, Q_{t-16}] + 8(1-\mu_Z)^8 COV[Q_{t-16}, Q_{t-19}] \\
&\quad -6(1-\mu_Z)^8 COV[Q_{t-18}, Q_{t-22}] - 6(1-\mu_Z)^8 COV[Q_{t-19}, Q_{t-21}] \\
&\quad +4(1-\mu_Z)^8 COV[Q_{t-21}, Q_{t-24}] - (1-\mu_Z)^8 COV[Q_{t-24}, Q_{t-26}] \\
&= lag1 \cdot [-(1-\mu_Z) - (1-\mu_Z)^3 + (1-\mu_Z)^4 - 2(1-\mu_Z)^5 + 3(1-\mu_Z)^6 \\
&\quad -5(1-\mu_Z)^7 \\
&\quad +8(1-\mu_Z)^8 - 13(1-\mu_Z)^9] \\
&\quad +(1-\mu_Z)^9 COV[Q_{t-12}, Q_{t-14}] - 3(1-\mu_Z)^9 COV[Q_{t-14}, Q_{t-17}] \\
&\quad +6(1-\mu_Z)^9 COV[Q_{t-16}, Q_{t-20}] + 10(1-\mu_Z)^9 COV[Q_{t-17}, Q_{t-19}]
\end{aligned}$$

$$
\begin{aligned}
& -14(1-\mu_Z)^9 COV[Q_{t-19}, Q_{t-22}] + 6(1-\mu_Z)^9 COV[Q_{t-21}, Q_{t-25}] \\
& +10(1-\mu_Z)^9 COV[Q_{t-22}, Q_{t-24}] - 4(1-\mu_Z)^9 COV[Q_{t-24}, Q_{t-27}] \\
& +(1-\mu_Z)^9 COV[Q_{t-26}, Q_{t-30}] \\
\dots & \\
= \; & lag1 \cdot \left[-(1-\mu_Z) - (1-\mu_Z)^3 \cdot \sum_{k=0}^{\infty} (-(1-\mu_Z))^k \cdot f_k \right]
\end{aligned}
$$

where f_k equals the Fibonacci numbers starting with $f_0 = 0$, $f_1 = 1$ and $f_k = f_{k-1} + f_{k-2}$ ($\forall k \geq 2$).

6. *Calculation of* $COV[Q_{t-1}, Q_{t-LR}]$ *for* $LR = 7$

$$
\begin{aligned}
COV[Q_{t-1}, Q_{t-7}] &= (1-\mu_Z) COV[Q_{t-2}, Q_{t-7}] - (1-\mu_Z) COV[Q_{t-7}, Q_{t-8}] \\
&= lag1 \cdot [-(1-\mu_Z)] \\
&\quad +(1-\mu_Z)^2 COV[Q_{t-3}, Q_{t-7}] - (1-\mu_Z)^2 COV[Q_{t-7}, Q_{t-9}] \\
&= lag1 \cdot [-(1-\mu_Z)] \\
&\quad -(1-\mu_Z)^3 COV[Q_{t-8}, Q_{t-9}] + (1-\mu_Z)^3 COV[Q_{t-9}, Q_{t-14}] \\
&= lag1 \cdot \left[-(1-\mu_Z) - (1-\mu_Z)^3 \right] \\
&\quad +(1-\mu_Z)^3 COV[Q_{t-9}, Q_{t-14}] \\
&= lag1 \cdot \left[-(1-\mu_Z) - (1-\mu_Z)^3 \right] \\
&\quad +(1-\mu_Z)^4 COV[Q_{t-10}, Q_{t-14}] - (1-\mu_Z)^4 COV[Q_{t-14}, Q_{t-16}] \\
&= lag1 \cdot \left[-(1-\mu_Z) - (1-\mu_Z)^3 \right] \\
&\quad -(1-\mu_Z)^5 COV[Q_{t-15}, Q_{t-16}] + (1-\mu_Z)^5 COV[Q_{t-16}, Q_{t-21}] \\
&= lag1 \cdot \left[-(1-\mu_Z) - (1-\mu_Z)^3 - (1-\mu_Z)^5 \right] \\
&\quad +(1-\mu_Z)^6 COV[Q_{t-17}, Q_{t-21}] - (1-\mu_Z)^6 COV[Q_{t-21}, Q_{t-23}] \\
&= lag1 \cdot \left[-(1-\mu_Z) - (1-\mu_Z)^3 - (1-\mu_Z)^5 \right] \\
&\quad +(1-\mu_Z)^7 COV[Q_{t-18}, Q_{t-21}] - (1-\mu_Z)^7 COV[Q_{t-21}, Q_{t-24}] \\
&\quad -(1-\mu_Z)^7 COV[Q_{t-22}, Q_{t-23}] + (1-\mu_Z)^7 COV[Q_{t-23}, Q_{t-28}] \\
&= lag1 \cdot \left[-(1-\mu_Z) - (1-\mu_Z)^3 - (1-\mu_Z)^5 - (1-\mu_Z)^7 \right] \\
&\quad +(1-\mu_Z)^7 COV[Q_{t-23}, Q_{t-28}]
\end{aligned}
$$

$$
\begin{aligned}
COV[Q_{t-1}, Q_{t-7}] &= lag1 \cdot \left[-(1-\mu_Z) \cdot \sum_{k=0}^{\infty} ((1-\mu_Z)^2)^k \right] \\
&\approx lag1 \cdot \left[-(1-\mu_Z) \cdot \frac{1}{1-(1-\mu_Z)^2} \right] \\
&\approx lag1 \cdot \left[\frac{-(1-\mu_Z)}{\mu_Z(2-\mu_Z)} \right]
\end{aligned}
$$

7. *Calculation of* $COV[Q_{t-1}, Q_{t-LR}]$ *for* $LR = 8$

$$
\begin{aligned}
COV[Q_{t-1}, Q_{t-8}] &= (1-\mu_Z) COV[Q_{t-2}, Q_{t-8}] - (1-\mu_Z) COV[Q_{t-8}, Q_{t-9}] \\
&= lag1 \cdot [-(1-\mu_Z)] \\
&\quad +(1-\mu_Z)^2 COV[Q_{t-3}, Q_{t-8}] - (1-\mu_Z)^2 COV[Q_{t-8}, Q_{t-10}] \\
&= lag1 \cdot [-(1-\mu_Z)] \\
&\quad +(1-\mu_Z)^3 COV[Q_{t-4}, Q_{t-8}] - (1-\mu_Z)^3 COV[Q_{t-8}, Q_{t-11}] \\
&\quad -(1-\mu_Z)^3 COV[Q_{t-9}, Q_{t-10}] + (1-\mu_Z)^3 COV[Q_{t-10}, Q_{t-16}] \\
&= lag1 \cdot \left[-(1-\mu_Z) - (1-\mu_Z)^3 \right]
\end{aligned}
$$

$$
\begin{aligned}
&+(1-\mu_Z)^4COV[Q_{t-5},Q_{t-8}]-(1-\mu_Z)^4COV[Q_{t-8},Q_{t-12}]\\
&-(1-\mu_Z)^4COV[Q_{t-9},Q_{t-11}]+(1-\mu_Z)^4COV[Q_{t-11},Q_{t-16}]\\
&+(1-\mu_Z)^4COV[Q_{t-11},Q_{t-16}]-(1-\mu_Z)^4COV[Q_{t-16},Q_{t-18}]\\
=\;& lag1\cdot\left[-(1-\mu_Z)-(1-\mu_Z)^3-2(1-\mu_Z)^5\right]\\
&+(1-\mu_Z)^5COV[Q_{t-6},Q_{t-8}]-(1-\mu_Z)^5COV[Q_{t-8},Q_{t-13}]\\
&-(1-\mu_Z)^5COV[Q_{t-9},Q_{t-12}]+(1-\mu_Z)^5COV[Q_{t-11},Q_{t-17}]\\
&+3(1-\mu_Z)^5COV[Q_{t-12},Q_{t-16}]-2(1-\mu_Z)^5COV[Q_{t-16},Q_{t-19}]\\
&+(1-\mu_Z)^5COV[Q_{t-18},Q_{t-24}]
\end{aligned}
$$

$$
\begin{aligned}
COV[Q_{t-1},Q_{t-8}] =\;& lag1\cdot\left[-(1-\mu_Z)-(1-\mu_Z)^3-2(1-\mu_Z)^5+(1-\mu_Z)^6\right]\\
&-(1-\mu_Z)^6COV[Q_{t-8},Q_{t-14}]-(1-\mu_Z)^6COV[Q_{t-9},Q_{t-13}]\\
&-(1-\mu_Z)^6COV[Q_{t-10},Q_{t-12}]+2(1-\mu_Z)^6COV[Q_{t-12},Q_{t-17}]\\
&+4(1-\mu_Z)^6COV[Q_{t-13},Q_{t-16}]-3(1-\mu_Z)^6COV[Q_{t-16},Q_{t-20}]\\
&-3(1-\mu_Z)^6COV[Q_{t-17},Q_{t-19}]+3(1-\mu_Z)^6COV[Q_{t-19},Q_{t-24}]\\
&-(1-\mu_Z)^6COV[Q_{t-24},Q_{t-26}]\\
=\;& lag1\cdot\left[-(1-\mu_Z)-(1-\mu_Z)^3-2(1-\mu_Z)^5+(1-\mu_Z)^6-5(1-\mu_Z)^7\right]\\
&-(1-\mu_Z)^7COV[Q_{t-9},Q_{t-14}]-(1-\mu_Z)^7COV[Q_{t-10},Q_{t-13}]\\
&+(1-\mu_Z)^7COV[Q_{t-12},Q_{t-18}]+3(1-\mu_Z)^7COV[Q_{t-13},Q_{t-17}]\\
&+5(1-\mu_Z)^7COV[Q_{t-14},Q_{t-16}]-(1-\mu_Z)^7COV[Q_{t-16},Q_{t-21}]\\
&-5(1-\mu_Z)^7COV[Q_{t-17},Q_{t-20}]+3(1-\mu_Z)^7COV[Q_{t-19},Q_{t-25}]\\
&+6(1-\mu_Z)^7COV[Q_{t-20},Q_{t-24}]-3(1-\mu_Z)^7COV[Q_{t-24},Q_{t-27}]\\
&+(1-\mu_Z)^7COV[Q_{t-26},Q_{t-32}]\\
=\;& lag1\cdot[-(1-\mu_Z)-(1-\mu_Z)^3-2(1-\mu_Z)^5+(1-\mu_Z)^6-5(1-\mu_Z)^7\\
&+5(1-\mu_Z)^8]\\
&-(1-\mu_Z)^8COV[Q_{t-10},Q_{t-14}]-(1-\mu_Z)^8COV[Q_{t-11},Q_{t-13}]\\
&+2(1-\mu_Z)^8COV[Q_{t-13},Q_{t-18}]+4(1-\mu_Z)^8COV[Q_{t-14},Q_{t-17}]\\
&-5(1-\mu_Z)^8COV[Q_{t-16},Q_{t-22}]-4(1-\mu_Z)^8COV[Q_{t-17},Q_{t-21}]\\
&-6(1-\mu_Z)^8COV[Q_{t-18},Q_{t-20}]+8(1-\mu_Z)^8COV[Q_{t-20},Q_{t-25}]\\
&+7(1-\mu_Z)^8COV[Q_{t-21},Q_{t-24}]-6(1-\mu_Z)^8COV[Q_{t-24},Q_{t-28}]\\
&-6(1-\mu_Z)^8COV[Q_{t-25},Q_{t-27}]+4(1-\mu_Z)^8COV[Q_{t-27},Q_{t-32}]\\
&-(1-\mu_Z)^8COV[Q_{t-32},Q_{t-34}]
\end{aligned}
$$

$$
\begin{aligned}
COV[Q_{t-1},Q_{t-8}] =\;& lag1\cdot[-(1-\mu_Z)-(1-\mu_Z)^3-2(1-\mu_Z)^5+(1-\mu_Z)^6-5(1-\mu_Z)^7\\
&+5(1-\mu_Z)^8-14(1-\mu_Z)^9]\\
&-(1-\mu_Z)^9COV[Q_{t-11},Q_{t-14}]+(1-\mu_Z)^9COV[Q_{t-13},Q_{t-19}]\\
&+3(1-\mu_Z)^9COV[Q_{t-14},Q_{t-18}]+4(1-\mu_Z)^9COV[Q_{t-15},Q_{t-17}]\\
&-9(1-\mu_Z)^9COV[Q_{t-17},Q_{t-22}]-6(1-\mu_Z)^9COV[Q_{t-18},Q_{t-21}]\\
&+6(1-\mu_Z)^9COV[Q_{t-20},Q_{t-26}]+12(1-\mu_Z)^9COV[Q_{t-21},Q_{t-25}]\\
&+12(1-\mu_Z)^9COV[Q_{t-22},Q_{t-24}]-7(1-\mu_Z)^9COV[Q_{t-24},Q_{t-29}]\\
&-14(1-\mu_Z)^9COV[Q_{t-25},Q_{t-28}]+6(1-\mu_Z)^9COV[Q_{t-27},Q_{t-33}]\\
&+10(1-\mu_Z)^9COV[Q_{t-28},Q_{t-32}]-4(1-\mu_Z)^9COV[Q_{t-32},Q_{t-35}]\\
&+(1-\mu_Z)^9COV[Q_{t-34},Q_{t-40}]\\
=\;& lag1\cdot[-(1-\mu_Z)-(1-\mu_Z)^3-2(1-\mu_Z)^5+(1-\mu_Z)^6-5(1-\mu_Z)^7\\
&+5(1-\mu_Z)^8-14(1-\mu_Z)^9+16(1-\mu_Z)^{10}]
\end{aligned}
$$

$$
\begin{aligned}
&-(1-\mu_Z)^{10}COV[Q_{t-12},Q_{t-14}]+2(1-\mu_Z)^{10}COV[Q_{t-14},Q_{t-19}]\\
&+3(1-\mu_Z)^{10}COV[Q_{t-15},Q_{t-18}]-4(1-\mu_Z)^{10}COV[Q_{t-17},Q_{t-23}]\\
&-12(1-\mu_Z)^{10}COV[Q_{t-18},Q_{t-22}]-7(1-\mu_Z)^{10}COV[Q_{t-19},Q_{t-21}]\\
&+12(1-\mu_Z)^{10}COV[Q_{t-21},Q_{t-26}]+21(1-\mu_Z)^{10}COV[Q_{t-22},Q_{t-25}]\\
&-12(1-\mu_Z)^{10}COV[Q_{t-24},Q_{t-30}]-19(1-\mu_Z)^{10}COV[Q_{t-25},Q_{t-29}]\\
&-20(1-\mu_Z)^{10}COV[Q_{t-26},Q_{t-28}]+20(1-\mu_Z)^{10}COV[Q_{t-28},Q_{t-33}]\\
&+17(1-\mu_Z)^{10}COV[Q_{t-29},Q_{t-32}]-10(1-\mu_Z)^{10}COV[Q_{t-32},Q_{t-36}]\\
&-10(1-\mu_Z)^{10}COV[Q_{t-33},Q_{t-35}]+5(1-\mu_Z)^{10}COV[Q_{t-35},Q_{t-40}]\\
&-(1-\mu_Z)^{10}COV[Q_{t-40},Q_{t-42}]
\end{aligned}
$$

8. *Calculation of $COV[Q_{t-1},Q_{t-LR}]$ for $LR=9$*

$$
\begin{aligned}
COV[Q_{t-1},Q_{t-9}] &= (1-\mu_Z)COV[Q_{t-2},Q_{t-9}]-(1-\mu_Z)COV[Q_{t-9},Q_{t-10}]\\
&= lag1\cdot[-(1-\mu_Z)]+(1-\mu_Z)COV[Q_{t-2},Q_{t-9}]\\
&= lag1\cdot[-(1-\mu_Z)]\\
&\quad +(1-\mu_Z)^2COV[Q_{t-3},Q_{t-9}]-(1-\mu_Z)^2COV[Q_{t-9},Q_{t-11}]\\
&= lag1\cdot\left[-(1-\mu_Z)-(1-\mu_Z)^3\right]+(1-\mu_Z)^3COV[Q_{t-11},Q_{t-18}]\\
&\quad +(1-\mu_Z)^3COV[Q_{t-4},Q_{t-9}]-(1-\mu_Z)^3COV[Q_{t-9},Q_{t-12}]\\
&= lag1\cdot\left[-(1-\mu_Z)-(1-\mu_Z)^3\right]\\
&\quad -(1-\mu_Z)^4COV[Q_{t-10},Q_{t-12}]+2(1-\mu_Z)^4COV[Q_{t-12},Q_{t-18}]\\
&\quad -(1-\mu_Z)^4COV[Q_{t-18},Q_{t-20}]\\
&= lag1\cdot\left[-(1-\mu_Z)-(1-\mu_Z)^3-2(1-\mu_Z)^5\right]\\
&\quad +(1-\mu_Z)^5COV[Q_{t-12},Q_{t-19}]+2(1-\mu_Z)^5COV[Q_{t-13},Q_{t-18}]\\
&\quad -2(1-\mu_Z)^5COV[Q_{t-18},Q_{t-21}]+(1-\mu_Z)^5COV[Q_{t-20},Q_{t-27}]\\
&= lag1\cdot\left[-(1-\mu_Z)-(1-\mu_Z)^3-2(1-\mu_Z)^5\right]\\
&\quad +(1-\mu_Z)^6COV[Q_{t-13},Q_{t-19}]-3(1-\mu_Z)^6COV[Q_{t-19},Q_{t-21}]\\
&\quad +3(1-\mu_Z)^6COV[Q_{t-21},Q_{t-27}]-(1-\mu_Z)^6COV[Q_{t-27},Q_{t-29}]\\
&= lag1\cdot\left[-(1-\mu_Z)-(1-\mu_Z)^3-2(1-\mu_Z)^5-4(1-\mu_Z)^7\right]\\
&\quad +(1-\mu_Z)^7COV[Q_{t-14},Q_{t-19}]-(1-\mu_Z)^7COV[Q_{t-19},Q_{t-22}]\\
&\quad +3(1-\mu_Z)^7COV[Q_{t-21},Q_{t-28}]+3(1-\mu_Z)^7COV[Q_{t-22},Q_{t-27}]\\
&\quad -3(1-\mu_Z)^7COV[Q_{t-27},Q_{t-30}]+(1-\mu_Z)^7COV[Q_{t-29},Q_{t-36}]\\
&= lag1\cdot\left[-(1-\mu_Z)-(1-\mu_Z)^3-2(1-\mu_Z)^5-4(1-\mu_Z)^7\right]\\
&\quad -(1-\mu_Z)^8COV[Q_{t-20},Q_{t-22}]+4(1-\mu_Z)^8COV[Q_{t-22},Q_{t-28}]\\
&\quad -6(1-\mu_Z)^8COV[Q_{t-28},Q_{t-30}]+4(1-\mu_Z)^8COV[Q_{t-30},Q_{t-36}]\\
&\quad -(1-\mu_Z)^8COV[Q_{t-36},Q_{t-38}]
\end{aligned}
$$

$$
\begin{aligned}
COV[Q_{t-1},Q_{t-9}] &= lag1\cdot\left[-(1-\mu_Z)-(1-\mu_Z)^3-2(1-\mu_Z)^5-4(1-\mu_Z)^7-8(1-\mu_Z)^9\right]\\
&\quad +(1-\mu_Z)^9COV[Q_{t-22},Q_{t-29}]+4(1-\mu_Z)^9COV[Q_{t-23},Q_{t-28}]\\
&\quad -4(1-\mu_Z)^9COV[Q_{t-28},Q_{t-31}]+6(1-\mu_Z)^9COV[Q_{t-30},Q_{t-37}]\\
&\quad +4(1-\mu_Z)^9COV[Q_{t-31},Q_{t-36}]-4(1-\mu_Z)^9COV[Q_{t-36},Q_{t-39}]\\
&\quad +(1-\mu_Z)^9COV[Q_{t-38},Q_{t-45}]\\
&= lag1\cdot\left[-(1-\mu_Z)-(1-\mu_Z)^3-2(1-\mu_Z)^5-4(1-\mu_Z)^7-8(1-\mu_Z)^9\right]\\
&\quad +1(1-\mu_Z)^{10}COV[Q_{t-23},Q_{t-29}]+4(1-\mu_Z)^{10}COV[Q_{t-24},Q_{t-28}]\\
&\quad -5(1-\mu_Z)^{10}COV[Q_{t-29},Q_{t-31}]+10(1-\mu_Z)^{10}COV[Q_{t-31},Q_{t-37}]\\
&\quad +4(1-\mu_Z)^{10}COV[Q_{t-32},Q_{t-36}]-10(1-\mu_Z)^{10}COV[Q_{t-37},Q_{t-39}]
\end{aligned}
$$

$$
\begin{aligned}
& -4(1-\mu_Z)^{10}COV[Q_{t-28},Q_{t-32}]+5(1-\mu_Z)^{10}COV[Q_{t-39},Q_{t-45}] \\
& -4(1-\mu_Z)^{10}COV[Q_{t-36},Q_{t-40}]-(1-\mu_Z)^{10}COV[Q_{t-45},Q_{t-47}] \\
= \;& lag1\cdot[-(1-\mu_Z)-(1-\mu_Z)^3-2(1-\mu_Z)^5-4(1-\mu_Z)^7-8(1-\mu_Z)^9 \\
& -16(1-\mu_Z)^{11}] \\
& +1(1-\mu_Z)^{11}COV[Q_{t-24},Q_{t-29}]+4(1-\mu_Z)^{11}COV[Q_{t-25},Q_{t-28}] \\
& -4(1-\mu_Z)^{11}COV[Q_{t-28},Q_{t-33}]-5(1-\mu_Z)^{11}COV[Q_{t-29},Q_{t-32}] \\
& +5(1-\mu_Z)^{11}COV[Q_{t-31},Q_{t-38}]+14(1-\mu_Z)^{11}COV[Q_{t-32},Q_{t-37}] \\
& +4(1-\mu_Z)^{11}COV[Q_{t-33},Q_{t-36}]-4(1-\mu_Z)^{11}COV[Q_{t-36},Q_{t-41}] \\
& -14(1-\mu_Z)^{11}COV[Q_{t-37},Q_{t-40}]+10(1-\mu_Z)^{11}COV[Q_{t-39},Q_{t-46}] \\
& +9(1-\mu_Z)^{11}COV[Q_{t-40},Q_{t-45}]-5(1-\mu_Z)^{11}COV[Q_{t-45},Q_{t-48}] \\
& +(1-\mu_Z)^{11}COV[Q_{t-47},Q_{t-54}] \\
\ldots & \\
= \;& lag1\cdot\left[-(1-\mu_Z)-(1-\mu_Z)\cdot\sum_{k=1}^{\infty}2^{k-1}((1-\mu_Z)^2)^k\right]
\end{aligned}
$$

9. *Calculation of $COV[Q_{t-1},Q_{t-LR}]$ for $LR=10$*

$$
\begin{aligned}
COV[Q_{t-1},Q_{t-10}] = \;& (1-\mu_Z)COV[Q_{t-2},Q_{t-10}]-(1-\mu_Z)COV[Q_{t-10},Q_{t-11}] \\
= \;& lag1\cdot[-(1-\mu_Z)] \\
& +(1-\mu_Z)COV[Q_{t-2},Q_{t-10}] \\
= \;& lag1\cdot[-(1-\mu_Z)] \\
& +(1-\mu_Z)^2COV[Q_{t-3},Q_{t-10}]-(1-\mu_Z)^2COV[Q_{t-10},Q_{t-12}] \\
= \;& lag1\cdot\left[-(1-\mu_Z)-(1-\mu_Z)^3\right] \\
& +(1-\mu_Z)^3COV[Q_{t-4},Q_{t-10}] \\
& -(1-\mu_Z)^3COV[Q_{t-10},Q_{t-13}] \\
& +(1-\mu_Z)^3COV[Q_{t-12},Q_{t-20}] \\
= \;& lag1\cdot\left[-(1-\mu_Z)-(1-\mu_Z)^3\right] \\
& +(1-\mu_Z)^4COV[Q_{t-5},Q_{t-10}]-(1-\mu_Z)^4COV[Q_{t-10},Q_{t-14}] \\
& -(1-\mu_Z)^4COV[Q_{t-11},Q_{t-13}]+2(1-\mu_Z)^4COV[Q_{t-13},Q_{t-20}] \\
& -(1-\mu_Z)^4COV[Q_{t-20},Q_{t-22}] \\
= \;& lag1\cdot\left[-(1-\mu_Z)-(1-\mu_Z)^3-2(1-\mu_Z)^5\right] \\
& +(1-\mu_Z)^5COV[Q_{t-6},Q_{t-10}]-(1-\mu_Z)^5COV[Q_{t-10},Q_{t-15}] \\
& -(1-\mu_Z)^5COV[Q_{t-11},Q_{t-14}]+(1-\mu_Z)^5COV[Q_{t-13},Q_{t-21}] \\
& +3(1-\mu_Z)^5COV[Q_{t-14},Q_{t-20}]-2(1-\mu_Z)^5COV[Q_{t-20},Q_{t-23}] \\
& +(1-\mu_Z)^5COV[Q_{t-22},Q_{t-30}]
\end{aligned}
$$

$$
\begin{aligned}
COV[Q_{t-1},Q_{t-10}] = \;& lag1\cdot\left[-(1-\mu_Z)-(1-\mu_Z)^3-2(1-\mu_Z)^5\right] \\
& +(1-\mu_Z)^6COV[Q_{t-7},Q_{t-10}]-(1-\mu_Z)^6COV[Q_{t-10},Q_{t-16}] \\
& -(1-\mu_Z)^6COV[Q_{t-11},Q_{t-15}]-(1-\mu_Z)^6COV[Q_{t-12},Q_{t-14}] \\
& +2(1-\mu_Z)^6COV[Q_{t-14},Q_{t-21}]+4(1-\mu_Z)^6COV[Q_{t-15},Q_{t-20}] \\
& -3(1-\mu_Z)^6COV[Q_{t-20},Q_{t-24}]-3(1-\mu_Z)^6COV[Q_{t-21},Q_{t-23}] \\
& +3(1-\mu_Z)^6COV[Q_{t-23},Q_{t-30}]-(1-\mu_Z)^6COV[Q_{t-30},Q_{t-32}] \\
= \;& lag1\cdot\left[-(1-\mu_Z)-(1-\mu_Z)^3-2(1-\mu_Z)^5-5(1-\mu_Z)^7\right] \\
& +(1-\mu_Z)^7COV[Q_{t-8},Q_{t-10}]-(1-\mu_Z)^7COV[Q_{t-10},Q_{t-17}]
\end{aligned}
$$

$$
\begin{aligned}
& -(1-\mu_Z)^7 COV[Q_{t-11}, Q_{t-16}] - (1-\mu_Z)^7 COV[Q_{t-12}, Q_{t-15}] \\
& +(1-\mu_Z)^7 COV[Q_{t-14}, Q_{t-22}] + 3(1-\mu_Z)^7 COV[Q_{t-15}, Q_{t-21}] \\
& +5(1-\mu_Z)^7 COV[Q_{t-16}, Q_{t-20}] - 4(1-\mu_Z)^7 COV[Q_{t-20}, Q_{t-25}] \\
& -5(1-\mu_Z)^7 COV[Q_{t-21}, Q_{t-24}] + 3(1-\mu_Z)^7 COV[Q_{t-23}, Q_{t-31}] \\
& +6(1-\mu_Z)^7 COV[Q_{t-24}, Q_{t-30}] - 3(1-\mu_Z)^7 COV[Q_{t-30}, Q_{t-33}] \\
& +(1-\mu_Z)^7 COV[Q_{t-32}, Q_{t-40}] \\
= \; & lag1 \cdot \left[-(1-\mu_Z) - (1-\mu_Z)^3 - 2(1-\mu_Z)^5 - 5(1-\mu_Z)^7 + (1-\mu_Z)^8\right] \\
& -(1-\mu_Z)^8 COV[Q_{t-10}, Q_{t-18}] - (1-\mu_Z)^8 COV[Q_{t-11}, Q_{t-17}] \\
& -(1-\mu_Z)^8 COV[Q_{t-12}, Q_{t-16}] - (1-\mu_Z)^8 COV[Q_{t-13}, Q_{t-15}] \\
& +2(1-\mu_Z)^8 COV[Q_{t-15}, Q_{t-22}] + 4(1-\mu_Z)^8 COV[Q_{t-16}, Q_{t-21}] \\
& +6(1-\mu_Z)^8 COV[Q_{t-17}, Q_{t-20}] - 5(1-\mu_Z)^8 COV[Q_{t-20}, Q_{t-26}] \\
& -7(1-\mu_Z)^8 COV[Q_{t-21}, Q_{t-25}] - 6(1-\mu_Z)^8 COV[Q_{t-22}, Q_{t-24}] \\
& +8(1-\mu_Z)^8 COV[Q_{t-24}, Q_{t-31}] + 10(1-\mu_Z)^8 COV[Q_{t-25}, Q_{t-30}] \\
& -6(1-\mu_Z)^8 COV[Q_{t-30}, Q_{t-34}] - 6(1-\mu_Z)^8 COV[Q_{t-31}, Q_{t-33}] \\
& +4(1-\mu_Z)^8 COV[Q_{t-33}, Q_{t-40}] - (1-\mu_Z)^8 COV[Q_{t-40}, Q_{t-42}]
\end{aligned}
$$

$$
\begin{aligned}
COV[Q_{t-1}, Q_{t-10}] \; = \; & lag1 \cdot [-(1-\mu_Z) - (1-\mu_Z)^3 - 2(1-\mu_Z)^5 - 5(1-\mu_Z)^7 + (1-\mu_Z)^8 \\
& -14(1-\mu_Z)^9] \\
& -(1-\mu_Z)^9 COV[Q_{t-11}, Q_{t-18}] - (1-\mu_Z)^9 COV[Q_{t-12}, Q_{t-17}] \\
& -(1-\mu_Z)^9 COV[Q_{t-13}, Q_{t-16}] + (1-\mu_Z)^9 COV[Q_{t-15}, Q_{t-23}] \\
& +3(1-\mu_Z)^9 COV[Q_{t-16}, Q_{t-22}] + 5(1-\mu_Z)^9 COV[Q_{t-17}, Q_{t-21}] \\
& +7(1-\mu_Z)^9 COV[Q_{t-18}, Q_{t-20}] - 6(1-\mu_Z)^9 COV[Q_{t-20}, Q_{t-27}] \\
& -9(1-\mu_Z)^9 COV[Q_{t-21}, Q_{t-26}] - 9(1-\mu_Z)^9 COV[Q_{t-22}, Q_{t-25}] \\
& +6(1-\mu_Z)^9 COV[Q_{t-24}, Q_{t-32}] + 15(1-\mu_Z)^9 COV[Q_{t-25}, Q_{t-31}] \\
& +15(1-\mu_Z)^9 COV[Q_{t-26}, Q_{t-30}] - 10(1-\mu_Z)^9 COV[Q_{t-30}, Q_{t-35}] \\
& -14(1-\mu_Z)^9 COV[Q_{t-31}, Q_{t-34}] + 6(1-\mu_Z)^9 COV[Q_{t-33}, Q_{t-41}] \\
& +10(1-\mu_Z)^9 COV[Q_{t-34}, Q_{t-40}] - 4(1-\mu_Z)^9 COV[Q_{t-40}, Q_{t-43}] \\
& +(1-\mu_Z)^9 COV[Q_{t-42}, Q_{t-50}] \\
= \; & lag1 \cdot [-(1-\mu_Z) - (1-\mu_Z)^3 - 2(1-\mu_Z)^5 - 5(1-\mu_Z)^7 \\
& +(1-\mu_Z)^8 - 14(1-\mu_Z)^9 + 7(1-\mu_Z)^{10}] \\
& -(1-\mu_Z)^{10} COV[Q_{t-12}, Q_{t-18}] - (1-\mu_Z)^{10} COV[Q_{t-13}, Q_{t-17}] \\
& -(1-\mu_Z)^{10} COV[Q_{t-14}, Q_{t-16}] + 2(1-\mu_Z)^{10} COV[Q_{t-16}, Q_{t-23}] \\
& +4(1-\mu_Z)^{10} COV[Q_{t-17}, Q_{t-22}] + 6(1-\mu_Z)^{10} COV[Q_{t-18}, Q_{t-21}] \\
& -7(1-\mu_Z)^{10} COV[Q_{t-20}, Q_{t-28}] - 11(1-\mu_Z)^{10} COV[Q_{t-21}, Q_{t-27}] \\
& -12(1-\mu_Z)^{10} COV[Q_{t-22}, Q_{t-26}] - 10(1-\mu_Z)^{10} COV[Q_{t-23}, Q_{t-25}] \\
& +15(1-\mu_Z)^{10} COV[Q_{t-25}, Q_{t-32}] + 24(1-\mu_Z)^{10} COV[Q_{t-26}, Q_{t-31}] \\
& +21(1-\mu_Z)^{10} COV[Q_{t-27}, Q_{t-30}] - 15(1-\mu_Z)^{10} COV[Q_{t-30}, Q_{t-36}] \\
& -25(1-\mu_Z)^{10} COV[Q_{t-31}, Q_{t-35}] - 20(1-\mu_Z)^{10} COV[Q_{t-32}, Q_{t-34}] \\
& +20(1-\mu_Z)^{10} COV[Q_{t-34}, Q_{t-41}] + 20(1-\mu_Z)^{10} COV[Q_{t-35}, Q_{t-40}] \\
& -10(1-\mu_Z)^{10} COV[Q_{t-40}, Q_{t-44}] - 10(1-\mu_Z)^{10} COV[Q_{t-41}, Q_{t-43}] \\
& +5(1-\mu_Z)^{10} COV[Q_{t-43}, Q_{t-50}] - (1-\mu_Z)^{10} COV[Q_{t-50}, Q_{t-52}]
\end{aligned}
$$

Appendix E

We approximate the covariance of the order quantity by using the recursive equation of the order quantity in formula (45): $Q_t = D_{t-1} + (1-\mu_Z)Q_{t-1} - (1-\mu_Z)Q_{t-L_R}$. For every L_R, the recursive equation for the order quantity is plugged in the formula for the covariance. For example, if the rework time equals two periods, the covariance can be reformulated as: $COV[Q_{t-1}, Q_{t-2}] = COV[D_{t-2} + (1-\mu_Z)Q_{t-2} - (1-\mu_Z)Q_{t-3}, Q_{t-2}]$. Repeating this procedure, we end up with: $COV[Q_{t-1}, Q_{t-2}] = \sigma_Q^2(1-\mu_Z)\sum_{k=0}^{\infty}(-(1-\mu_Z))^k$. The same procedure follows for all $COV[Q_{t-1}, Q_{t-L_R}]$ with respect to all L_R taken into consideration e.g. in a numerical analysis. For L_R equal to two and three, it is possible to derive exact formulae for the covariances (see Appendix D). For larger rework times, this is not possible. Therefore, at first we express the covariance for a given L_R as a function of the first lag ($lag1_{L_R}$) varying in L_R. $lag1_{L_R}$ is defined as $VAR[Q_{t-L_R-i}, Q_{t-L_R-i-1}]$ for all $i \in \mathbb{Z}_{\geq 0}$ in a steady-state system.

Lemma 1. *Under a strictly linear control rule, the covariance $COV[Q_{t-1}, Q_{t-L_R}]$ of the order quantities Q_{t-1} and Q_{t-L_R} in a production system with positive production and rework times ($L_P \geq L_R > 1$) can be approximated as*

L_R	Covariances	
2	$\sigma_Q^2(1-\mu_Z)\sum_{k=0}^{+\infty}(-(1-\mu_Z))^k$	$\approx \sigma_Q^2 \cdot [(1-\mu_Z)/(2-\mu_Z)]$
3	0	
4	$lag1_{L_R} \cdot [-(1-\mu_Z)\sum_{k=0}^{\infty}(-(1-\mu_Z))^k]$	$\approx lag1_{L_R} \cdot [-(1-\mu_Z)/(2-\mu_Z)]$
5	$lag1_{L_R} \cdot [-(1-\mu_Z)]$	
6	$lag1_{L_R} \cdot [-(1-\mu_Z) - (1-\mu_Z)^3 \cdot \sum_{k=0}^{\infty}(-(1-\mu_Z))^k \cdot f_{k+1}]$	
7	$lag1_{L_R} \cdot [-(1-\mu_Z) \cdot \sum_{k=0}^{\infty}((1-\mu_Z)^2)^k]$	$\approx lag1_{L_R} \cdot [-(1-\mu_Z)/(\mu_Z(2-\mu_Z))]$
8	$lag1_{L_R} \cdot [-(1-\mu_Z) - (1-\mu_Z)^3 - 2(1-\mu_Z)^5 + (1-\mu_Z)^6 - 5(1-\mu_Z)^7 \cdots]$	
9	$lag1_{L_R} \cdot [-(1-\mu_Z) - (1-\mu_Z) \cdot \sum_{k=1}^{\infty} 2^{k-1}((1-\mu_Z)^2)^k]$	
10	$lag1_{L_R} \cdot [-(1-\mu_Z) - (1-\mu_Z)^3 - 2(1-\mu_Z)^5 - 5(1-\mu_Z)^7 \cdots]$	

Table 5:

where f_k equals the Fibonacci numbers starting with $f_0 = 0$, $f_1 = 1$ and $f_k = f_{k-1} + f_{k-2}$ ($\forall k \geq 2$).

For the proof see Appendix D.

Table 5 shows that the covariance can be calculated nearly exactly for rework times of two and three periods—just the geometric series for a rework time of two periods is an approximation. For a rework time of eight and ten periods the formula for the covariance indicates no structure. Therefore, the formula will be cut after $(1-\mu_Z)^{10}$ in all following analyses. For rework times larger than three, the covariances depend on $lag1_{L_R}$. An extensive study has shown that $lag1_{L_R}$ can be approximated very well by $\sigma_Q^2 - \sigma_D^2$. Using this approximation, the covariance of the order quantity for $L_R > 3$ can be written as $(\sigma_Q^2 - \sigma_D^2) \cdot A_{L_R}$.