Patience is Key: The Time it Takes to see Benefits from Continuous Auditing

Marc Eulerich  
University of Duisburg-Essen  
marc.eulerich@uni-due.de

Benjamin Fligge  
University of Duisburg-Essen  
benjamin.fligge@uni-due.de

Vanessa Lopez Kasper  
University of Duisburg-Essen  
vanessa.lopez-kasper@uni-due.de

David A. Wood  
Brigham Young University  
davidwood@byu.edu

Running Head: Benefits from Continuous Auditing

Marc Eulerich, University of Duisburg-Essen, Mercator School of Management, Duisburg, Germany; Benjamin Fligge, University of Duisburg-Essen, Mercator School of Management, Duisburg, Germany; Vanessa Lopez Kasper, University of Duisburg-Essen, Mercator School of Management, Duisburg, Germany; David A. Wood, Brigham Young University, Marriott School of Business, School of Accountancy, Provo, UT, USA

Data Availability: The data used in this study cannot be made publicly available due to confidentiality agreements with the participating organization.

JEL Classifications: M42, G32

Keywords: Continuous Auditing; Key Risk Indicator; internal auditing
Patience is Key: The Time it Takes to see Benefits from Continuous Auditing

Abstract
Despite research showing numerous benefits to continuous auditing, uptake by internal audit functions has been quite slow. Using field data from a multinational company, we study two possible reasons for the slow uptake of continuous auditing—the time it takes for continuous auditing to result in measurable reductions in audit risks and that not every type of risk is equally likely to improve from continuous auditing. Results suggest it takes three years before observing significant risk reductions from implementing continuous auditing. We also find that the benefits of implementing continuous auditing vary by risk factor, ranging in improvement from 0 percent to 51.6 percent for each additional year of continuous auditing use. These findings are important for internal auditors to understand to have a realistic expectation of the benefits and their limitations, as well as the timetable for realizing those benefits, when adopting continuous auditing.

I. INTRODUCTION
Continuous auditing (CA) was proposed decades ago to improve assurance quality (Groomer and Murthy 1989; Vasarhelyi and Halper 1991). It has many theoretical benefits and research has already demonstrated several of them (reviewed in section II). Yet, despite these benefits, the adoption of CA in the field of internal auditing remains relatively low (German IIA 2023; Polizzi and Scannella 2023). In this study, we examine two reasons to partially explain why this may be the case: how long it takes to observe benefits from implementing CA, and what types of risks are most and least likely to improve from CA.

We analyze the effects of CA on various key risk indicators (KRIs). “A KRI is an operational or financial variable that provides a reliable basis for estimating the likelihood and the severity of one or more operational risk events” (Scandizzo 2005, 235) and should be a warning system for future actions (Scandizzo 2005; Tupa, Simota, and Steiner 2017). Internal auditors use KRIs to monitor the risks affecting an organization. Based on KRIs, they inform the respective stakeholders about problems, so the stakeholders can take measures to manage, avert, or minimize risks. Internal auditing also relies on KRIs to identify areas in need of additional testing.
For our analysis we use a field-study setting. The studied organization, which provided us with data from their CA system, is a multinational company that is among the largest food retailers in the world. The company’s internal audit function (IAF) implemented CA in 2015 and provided us with data on the analyzed KRI. Specifically, we analyzed data from 20 KRI for 41 entities operating in different countries.

The studied company uses a self-designed CA system, which conducts automated tests on a quarterly basis. The results are consolidated at an entity-level and analyzed by internal auditing. All findings are reported to the superior of the audited area and trigger further actions in accordance with the risk assessment.

We analyze the effect of CA on 20 KRI between 2015 and 2020. Our initial results confirm that CA is associated with an improvement in the firm’s overall risk level. Yet, we only observe statistically significant effects starting three years after the implementation of CA. We find that KRI continue to improve in the fourth year and then level off, not fully mitigating all risk, but at a level that is substantially improved from when CA was implemented. Thus, one key reason why CA may not be heavily adopted is that it takes a significant amount of time—three years—before observing benefits from CA.

We also analyze each KRI separately and find significant improvement in 12 KRI, while 8 KRI show no significant improvement despite the use of CA. Our exploratory analyses suggest that one reason for why KRI do not improve is that the risks were sufficiently mitigated before the CA was implemented.

We contribute to practice and research in several ways. First, our findings are important for practitioners in that we show the need for patience when implementing CA.

---

1 Even though CA theory talks about near real-time testing (Vasarhelyi and Halper 1991), it practically depends on the available database. In addition, near real-time implementation is sometimes not cost-effective and it may be sufficient to audit in frequent cycles (Chan and Vasarhelyi 2011). Practical studies report that CA frequency varies from daily, monthly or quarterly to even semiannually and sometimes varies across different CA indicators within one CA system (Codesso et al. 2020; de Freitas, Codesso, and Rodrigues Augusto 2020).
The benefits of implementation will likely take some time before they are observed. Second, we answer the call of Brown, Wong and Baldwin (2007) for research that quantifies the actual benefits of CA, as we are the first to analyze the effects of an organizational CA system. Anecdotal evidence suggests that practitioners struggle to “sell” the benefits of CA to management. Our results document the improvement of multiple KRI that enable practitioners to determine the value added by CA, addressing the call for research on what types of business value are delivered by CA technologies, and the timing and sustainability of these impacts over time (Rikhardsson and Dull 2016). Third, we provide evidence in response to the call of Campbell, Epstein, and Martinez-Jerez (2011) for more research on the relationship between learning and control. In particular, we find that CA can improve various KRI over time, which adds value to the organization.

II. LITERATURE REVIEW AND THEORY DEVELOPMENT

CA is defined as “data flowing through the system [that] are monitored and analyzed continuously…using a set of auditor defined rules. Exceptions…will trigger alarms which are intended to call the auditor’s attention” (Vasarhelyi and Halper 1991, 114). Its aim is to enhance the overall visibility of risk and performance to the organization through the effective use of technology (KPMG 2009). Terms such as continuous monitoring, continuous assurance, real-time auditing, or online auditing are often used synonymously and refer to similar concepts (Alles, Kogan, and Vasarhelyi 2002; Brown, Wong, and Baldwin 2007). All these approaches rely on comparable information technologies and aim to improve internal control systems and transparency of business processes.

CA is expected to enhance monitoring because it provides more feedback, and more timely feedback, than traditional auditing. Prior research shows that CA is associated with

---

2 Following prior research (e.g., Brown, Wong, and Baldwin 2007; Eulerich and Kalinichenko 2018), we thus do not differentiate among the terms and use “CA” generically.

3 Timely feedback is more specific as the details of the situation are more apparent than at a later time (Michaelsen and Schultheiss 1989). While a delay in the feedback process can make the result causally
reductions in risk exposure and improvements in internal controls (Codesso et al. 2020; de Freitas, Codesso, and Augusto 2020). More specifically CA has been found to avoid undue costs (de Freitas, Codesso, and Augusto 2020), increase reliance of external auditors on internal audit work (Davidson, Desai, and Gerard 2013; Malaescu and Sutton 2015), reduce fraud (Gonzalez and Hoffmann 2018), and investors’ firm risk estimates (El-Masry and Reck 2008). Additionally, CA increases the perception of the value created or added by the department in use (Eulerich, Georgie, Schmidt 2020; Rikhardsson and Dull 2016). On the other hand, it is also found to increase non-compliant behavior (Eulerich, Lopez Kasper, Sofla 2020) and there are concerns it can decrease critical thinking and increase alert immunization (Rikhardsson and Dull 2016).

Given these arguments and prior research findings, the question posed by influential internal audit thought leader Norman Marks remains, “Why is the move to continuous auditing so slow?” For example, recent survey data of 597 IAFs in Europe shows that 70.8 percent of respondents do not used CA and 49.9 percent of respondents have no intention to implement CA (German IIA 2023). If the research findings are so strong, what is prohibiting IAFs from embracing this technology?

The findings of Eulerich et al. (2023) shed light on why technology may be slowly adopted by IAFs and what can be done to increase adoption rates. Eulerich et al. (2023) find that many IAFs do not adopt technology-based audit techniques because of the difficulty in observing quantifiable benefits in a timely manner. Furthermore, internal auditors often struggle to quantify the benefits of new technologies and they look to other companies or ambiguous (Argote, Lee, and Park 2021; Levitt and March 1988), implying that more timely feedback is of higher quality. Delayed feedback has even found to impair team performance (Diehl and Sterman 1995) and to hinder employees from understanding the effectiveness of process improvement initiatives (Repenning and Sterman 2002). Further, Christ et al. (2012) find that employees receiving immediate feedback perform better than employees exposed to delayed feedback or no feedback.

See https://normanmarks.wordpress.com/2010/01/14/why-is-the-move-to-continuous-auditing-so-slow/.

For additional research on technology adoption see discussions in the following papers, Applebaum, Kogan, and Vasarhelyi (2018), Christ et al. (2021), Kreiger, Drews, and Veltz (2021), and Li et al. (2018).
research to help them understand the benefits of technology adoption. Thus, research that can help quantify the benefits of new technologies can be persuasive in increasing adoption.

To this end, we study how long it takes to observe benefits from the implementation of a CA system. The specific benefit we examine is how CA reduces risks within an organization, given internal audit’s focus on risk reduction as part of their work (Carcello et al. 2020; Christ et al. 2021; Jiang, Messier, and Wood 2020; Prawitt, Sharp, and Wood 2012). We are not aware of theory that predicts exactly when CA will reduce risk, suggesting the importance of documenting how fast risk reduction can incur from CA.

The second aspect of CA that we study is the types of risks that CA is most effective in reducing. As recognized by risk management frameworks such as COSO and COSO-ERM, risks can relate to many different aspects of an organization (Chan, Chen, and Liu 2021; Janvrin et al. 2012; Klamm and Watson 2009). CA can be designed to report on many different risks, but it is not clear which types of risks CA will be effective in addressing. Thus, we add a descriptive analysis of what types of risks CA effectively addresses and which risks it does not, elaborating on the reasons for these (non) improvements.

III. DATA AND METHODOLOGY

For this paper, we analyze data from a multinational company that is among the largest food retailers in the world. For the fiscal year 2020, the company reports more than $100 billion in revenues and approximately 500,000 employees working in 34 countries and 41 different entities.

CA System

The company’s CA system was implemented in 2015 and initially covered 29 KRI s. All KRI s were selected by the IAF and considered in the CA system if adequate and complete data is available and can be extracted efficiently. Furthermore, the consideration of all KRI s was discussed with other internal assurance providers regarding their appropriateness and the
extent to which they are already monitored by other measures. Over time, the IAF considered the creation of new KRI s, modifying existing KRI s, and removing KRI s that were not deemed unhelpful. At the end of 2020, the CA system covered 26 different KRI s.

The CA system consolidates the data at the entity-level and analyzes it automatically. Results are interpreted by the IAF and reported to the CFO of the respective entity quarterly, whereby the format of reporting depends on the findings. This represents a drastic improvement in audit timeliness as previously, the audit cycle to analyze each KRI was six to eight years. In case of a low level of risk or no abnormalities, a report is sent to the CFO of the respective entity. For a medium level of risk, the report is also sent to the responsible management and internal auditors discuss the findings with management. For a high level of risk, internal auditing immediately conducts an on-site audit after the distribution of the report.

**Key Risk Indicators**

We analyze the development of several KRI s starting with the implementation of CA in 2015 until 2020. Between 2015 and 2020 a total 57 KRI s were monitored by the IAF. We omit 4 KRI s that do not show any variation during this period. Additionally, we exclude 32 KRI s which did not have sufficient data (usually because of only being implemented for a short time), and 1 KRI that captures firm size rather than a risk. Our final sample consists of 20 KRI s. We discuss each KRI in Appendix A.

To examine the effect of CA on the overall risk in the organization, we calculate a combined KRI score \( (Overall\_KRI) \). To calculate \( Overall\_KRI \), we compute the mean of all

\[\text{Overall}_{-} \text{KRI} = \frac{\sum_{i=1}^{n} \text{KRI}_i}{n}\]

---

6 Although these KRI s do not show variation, the IAF continued to monitor them given the level of risk impacting the company should one of these KRI s show any variation.

7 We exclude the number of super-users for each entity from our analyses as this KRI is strongly correlated with the size of the entity.
$z$-transformed KRI$s$ for a given period of time the KRI has been implemented in the CA system at a specific entity.  

**Research Models**

To explore our research questions, we regress $Overall\_KRI$ on our measure for the length of CA use:

$$Overall\_KRI_{i,\text{CA}} = \beta_0 + \beta_1 CA\_Years_{i,t} + \epsilon_{i,t} \quad (1)$$

The variable of interest, $CA\_Years$, measures the number of years the respective KRI$s$ were analyzed by the CA system. We expect to observe a negative coefficient on $CA\_Years$, suggesting that the more years the company uses CA, the lower the overall risk in the organization. As $Overall\_KRI$ summarizes various KRI$s$ from various years that were analyzed for the same length of time by the CA system, thus it is only possible to control for entity-specific factors that are constant over time, instead of for time specific factors. Hence, we include entity fixed-effects and cluster standard errors on the entity-level.

To provide further insights into the effects of CA, we regress each individual KRI on $CA\_Years$ and include several control variables at the entity level, as follows:

$$KRI_{i,t} = \beta_0 + \beta_1 CA\_Years_{i,t} + \beta_2 GDP_{i,t} + \beta_3 Unemployment_{i,t} + \beta_4 Complexity_{i,t} + \epsilon_{i,t} \quad (2)$$

The dependent variable in equation (2) is $KRI$ and represents one of 20 KRI$s$. Each KRI is calculated as the average quarterly value for entity $i$ in year $t$. This allows us to reduce the influence of random fluctuations among quarters. We include entity fixed-effects to control for entity-specific factors that are constant over time. Additionally, we include the gross domestic product ($GDP$) and unemployment rate ($Unemployment$) to account for the time-variant economic conditions within the entity’s country. Finally, we include the number of KRI$s$ reported ($Complexity$) to control for the possibility that too much data from the CA

---

8 We note that we observe the same results if we transform each KRI to an interval ranging from 0 (lowest value) to 100 (highest value) and calculate the combined KRI score as the average of these transformed KRI$s$. 

Electronic copy available at: https://ssrn.com/abstract=4454724
system may overwhelm the internal auditors and impair risk improvements. The standard errors in our models are clustered on entity-level.

IV. Results

Descriptive Statistics

Figure 1 illustrates the Overall_KRI relative to the amount of time CA has been employed. On average, z-transformed KRIs have a value of 0.100 in the first year of CA usage. While we observe a similar risk level in the second year (0.102), Overall_KRI decreases to 0.019 in the third year of CA usage. This decrease continues until the fifth year (-0.132), but is followed by a slight increase in the overall risk level in the sixth year (-0.064).

Table 1 provides descriptive statistics on the individual KRIs. Consistent with our expectations most (16 of 20 KRIs) have decreased the longer the CA has been employed. These results indicate that specific KRIs are more (less) likely to improve from CA usage. We also provide a test in Table 1 of whether the KRI values in the first year of implementation of CA are different from zero. If the KRI’s are not different than zero in period 1, we would not expect that CA can reduce the risk of these variables to a significant degree.

Empirical Results

Table 2 summarizes our empirical results for estimating the effect of CA on Overall_KRI. The first column shows that the number of years CA is implemented (CA-Years) is associated with the overall KRI measure (p < 0.01). As we expected, the use of CA is associated with a reduction in risk.

To understand how long it takes to see the benefits, the second column splits the CA variable into dummy variables for each year (Year 1 is omitted and is thus the comparison

---

9 We also gather information on the entity size and entity age. While the inclusion of both factors yields similar results, we lose roughly half of the entity-years due to missing values. Thus, we decide not to include these variables in our main models but note that our results are robust to this sensitivity analysis.
group for p-values). The results show that there are no significant improvements in risk in year 2, but that from year 3 forward, there are significant differences in risk relative to year 1. When we perform F-tests to compare differences for each year, we find that risk in year 3 is significantly lower than in year 2 (p-value < 0.05), in year 4 it is significantly lower than in year 3 (p-value < 0.01), but that in year 5 and year 6 it is no different than the previous year (p-values > 0.10). Thus, column 2 shows that, on average, it takes 3 years to see improvement in risk after adopting CA and that it continues to improve for two more years before it then levels off.\textsuperscript{10}

We test whether CA has the same effect across each of the different KRIs and report the results in Table 3. We find CA to have a significant and negative effect on 12 KRIs. These variables are more amenable to economic interpretation than the z-score in Table 2. We present economic significance by computing the percentage improvement for each KRI for each additional year of CA use. We find that each of the 12 variables in Panel A improved by 13.0 percent to 51.6 percent. This suggests that the results are practically important as CA substantially reduces risk.\textsuperscript{11}

We present the KRIs that are not associated with improvement over time in Table 3, Panel B. We find CA has no effect on the number of bookings into closed accounting periods \textit{(Closed\_Period)}, usage of suspense accounts for existing creditors \textit{(CPD\_Receipts)}, the value and percentage of cash discount losses \textit{(CashDiscount\_Loss, CashDiscount\_Pct)}, the number of logins with an SAP-user holding infinite rights \textit{(SAP\_All\_Login)}, the number of unauthorized payment block removal or modification \textit{(Payment\_Block)}, the number of failed automated receipt transfers \textit{(ERP\_Transfer)}, and the number of open items on difference accounts \textit{(Difference\_Account)}.

\textsuperscript{10} When we test for nonlinear learning effects (Hodge, Hopkins, and Wood 2010), we find a cubed term best explains the pattern in our data.

\textsuperscript{11} In Appendix B, we also test whether CA improves each KRI at the same pace. Results suggest differences in the speed to which CA improves different KRIs.
We conduct additional testing to shed light on why these KRI s did not show improvement. As previously explained, KRI s that are already very low, may not show improvement as there is no room to show statistical improvement. Consistent with this explanation, we find that three KRI s (i.e., Closed_Period, CPD_Receipts, Difference_Account) are not statistically different from zero in the first period measured (see Table 1); thus, CA is very unlikely to show statistically significant improvements for these variables.

For four of the remaining five variables (i.e., CashDiscount_Loss, CashDiscount_Pct, Payment_Block, and ERP_Transfer), we note that the pattern in Table 1, Panel B shows that the risk is decreasing for each of the variables, it was just not statistically significant. The company’s internal auditors state that one possible reason for not seeing statistically significant differences is that the company had many different risks to focus on, and these ones may not deemed as critical for the entities to address. For example, ERP_Transfer is related to automation of transferring receipts to the ERP system. While the automation saves time, if something is not automatically transferred, it can be manually transferred which would hurt efficiency but not necessarily increase the risk of malfeasance in the system. If the organization is busy trying to handle other tasks, it may just be that they are not focused on these lower-level risks. Additionally, discussions with the company’s internal auditors reveal that improvements for some KRI s require enormous efforts that are expected to exceed the potential benefits.

V. LIMITATIONS AND CONCLUSION

We find that CA takes on average, three years, to see overall benefits in terms of reductions in risk to adopting CA. Also, while CA does improve the majority of risks measured, it is not associated with reducing all types of risks.
Our study is subject to limitations. Although we analyze real-world data of a large multinational company, we cannot rule out that company-wide factors affect our results and whether our results will generalize to other organizations. Nevertheless, we focus on a very rich, longitudinal dataset that allows for analyzing a problem that prior research could not analyze for lack of data. Thus, while our results add to the literature, additional work is necessary to test for generalizability. Second, we are unable to provide empirical answers for why CA works for some KRIIs but not others. This suggests an important avenue for future research. Research that can empirically identify what types of risks are best mitigated by CA will allow practitioners to better target CA to appropriate risks, saving time and money and likely showing a better payoff for adopting these technologies.

We encourage additional research into CA. Given the theoretical benefits of this technology, more research should study why adoption is slow and how to increase adoption. We specifically encourage future research to help design CA systems that can be implemented at a lower cost to increase adoption of this technology.
REFERENCES


Electronic copy available at: https://ssrn.com/abstract=4454724


FIGURE 1
Overall KRI improvement

This figure presents the association between average z-transformed KRI scores and the length of CA usage.