# Digital Agendas in the Insurance Industry: The Importance of Comprehensive Approaches

Alexander Bohnert, Albrecht Fritzsche, Shirley Gregor*

This version: August 2018

**Abstract**

With a growing awareness of the potential of innovation provided by digital technology, insurance companies have increasingly adopted digital agendas in their business activities. Our paper studies the relationship between the expression of a digital agenda in annual reports and the business performance of 41 publicly-traded European insurance companies for the time period from 2007 to 2017. Our findings show a positive relationship, which is particularly strong in cases where companies take a comprehensive approach by addressing digital technology both in the context of internal activities within their own organisation and external activities in connection with customers and business partners.

*Keywords*: Digitalisation; firm characteristics; shareholder value; corpus linguistics

## 1. Introduction

Insurance business is coupled to socio-economic change in many different ways. On the one hand, new developments in society and economy affect the demand for insurance. Megatrends such as urbanisation, individualisation, and the ageing society create dynamics in the client markets of insurance companies; climate change, economic instability, and political unrest require alterations to the way how risk is calculated. On the other hand, insurance companies are themselves part of larger socio-economic structures which affect their daily performance. They require qualified personnel, use modern information and communication technologies, and depend on financial products to generate savings. Socio-economic change therefore also has an effect on the way insurance companies perform. In this sense, one can talk about external couplings and internal couplings of insurance business to socio-economic change. While the former concerns markets, customers and offerings, the latter concerns business operations, management and control.

* Alexander Bohnert and Albrecht Fritzsche are at the Friedrich-Alexander University Erlangen-Nürnberg (FAU), School of Business and Economics, Lange Gasse 20, 90403 Nuremberg, Germany, alexander.bohnert@fau.de, albrecht.fritzsche@fau.de; Shirley Gregor is at the Australian National University (ANU) College of Business and Economics, shirley.gregor@anu.edu.au.

Among the many drivers of socio-economic change, digital technology plays a particularly important role. In recent times, mobile, interconnected devices equipped with powerful, miniaturised processors, sensors and actuators have become ubiquitous in daily life. In the years to come, they will permeate human life even more, creating enormous potential for new ways to inform actors, support decision-making and collect data to analyse and predict patterns of behaviour (see, e.g., Lee, 2008; Porter and Heppelmann, 2014; McAfee and Brynjolfsson, 2012). At the same time, existing IT infrastructures enable new forms of commerce, which can lead to innovation through platform-based interaction and systemic value creation (see Lusch and Nambisan, 2015). These changes provide opportunities for insurance companies to enter into a new phase of digital insurance (see, e.g., Nicoletti, 2016).

The potential of digital technology for innovation in the insurance industry is considerable (see Eling and Lehmann, 2018). It includes the implementation of new forms of online marketing and sales activities (see Seitz, 2017), the generation of new business models and value creation processes (see Desyllas and Sako, 2013; Schmidt et al., 2017), and the overall transformation of insurance companies into more agile organisations (see Barkur, Varambally, and Rodrigues, 2007). Given this variety, insurance companies can choose different strategies to approach digital technology. In this paper our research interest is directed at the question of which strategies are most likely to lead to success. In line with the aforementioned considerations, we hypothesise that companies benefit most from digital technology if they use it for a comprehensive approach to innovation, which addresses internal as well as external aspects of change.

The remainder of the paper is structured as follows. Section 2 provides a literature overview leading to hypothesis development. Section 3 describes the data set and outlines the approach to the empirical analysis using a treatment-effects model. Section 4 supplies results including robustness tests, while Section 5 summarises and gives concluding remarks.

## 2. Literature and Hypothesis

### *Digital technologies and innovation*

Understanding the contribution of digital technology to the success of business operations is a central topic in information systems research (see Agarwal and Lucas, 2005; Schryen, 2013). The body of literature in the field is large and covers different approaches with respect to the key constructs, dependent variables and data sources used (see, e.g., Melville, Kraemer, and Gurbaxani, 2004; Kohli and Devaraj, 2003). Recent contributions have emphasised the importance of a broader look at various manifestations of value and their mediating factors (see Kohli and Grover, 2008). Digital technology has to be considered not only as a means of cost reduction, but also as an investment for revenue growth in supporting different functions in the company (see Mithas et al., 2012). Information systems can have a strong impact on organisational agility (see, e.g., Lu and Ramamurty, 2011; Sambamurthy, Bharadwaj, and Grover, 2003). Furthermore, they involve numerous intangible assets related to the implementation and operation of the systems, which affect organisational capabilities in different ways (see, e.g., Saunders and Brynjolfsson, 2016; Mithas, Ramasubbu, and Sambamurthy, 2011).

On a more general level, information systems are drivers of organisational change, as they affect the concepts and operational structures of business practice (see Markus, 2004). Technological determinism is hard to uphold (see Markus and Robey, 1988), as the decisions about how technology is adopted in an organisation result from complex social dynamics (see Boudreau and Robey, 2005; Venkatesh et al., 2003). Nevertheless, technical devices and systemic structures provide an important point of reference for reflection on organisational practice and options for further development (see Orlikowski, 2009), which becomes even more apparent in highly dynamic socio-economic environments where organisational routines and instrumental action are constantly renegotiated (see Pentland et al., 2012; Leonardi, 2011). The transformative power of digital technology (see Lucas et al., 2013; Dhar and Sundararajan, 2007) has recently been documented in various different industries, such as healthcare (see Agarwal et al., 2010), manufacturing (see Brettel et al., 2014), and robotics (see Brynjolfsson and McAfee, 2012).

The notion of digital innovation expands this line of thought towards a new understanding of innovation, following the digitisation of physical artefacts (see Svahn, Henfridsson, and Yoo, 2009). Drawing on the study of multimedia devices, Yoo, Henfridsson, and Lyytinen (2010, p. 725) consider digital innovation as carrying out "new combinations of digital and physical components to produce novel products". New architectural paradigms such as the internet of things (see Atzori, Iera, and Morabito, 2010) and cyber-physical systems (see Lee, 2008) support this process by a stronger association of physical processes with computational events, such that they can be referred to interchangeably (see Gölzer and Fritzsche, 2017). Common examples can be found in the automotive industry, where innovation is increasingly circling around new combinations of physical products and data services in cars and mobility services related to them (see, e.g., Hildebrandt et al., 2015; Hylving and Schultze, 2013).

Fichman, Dos Santos, and Zheng (2014, p. 330) take a broader approach to digital innovation which is associated with any "product, process, or business model that is perceived as new, requires some significant changes on the part of the adopters, and is embodied in or enabled by IT". In a similar way, Nambisan et al. (2017, p. 224) expand the notion of digital innovation to "the creation of (and consequent change in) market offerings, business processes, or models that result from the use of digital technology". Fichman et al. (2014) as well as Nambisan et al. (2017) thus turn the focus towards the added value for customers resulting from applications of digital technology to create new kinds of offerings. Any innovation that relies on the availability of digital technology can accordingly be considered as a digital innovation, no matter how digital technology exerts its influence, as long as it provides a necessary condition for the possibility of the innovation. Engagement in digital innovation has strategic significance for an organisation, because it accompanies decisions to put business operations on a new foundation, rather than just replacing an old tool by a new one.

*Digital innovation and the insurance industry*

Early work on the implications of advanced digital technology for the insurance industry was mainly concerned with new online distribution channels (see Garven, 2002; Dumm and Hoyt, 2003), particularly with respect to their consequences for insurance agents (see Eastman et al., 2002), customer orientation (see Kaiser, 2002) and regulation (see Meyer and Krohm, 1999). While older data processing systems in the companies were seen chiefly as means to increase efficiency, new generations of digital technology are expected to increase market dynamics and competition, due to more transparency and comparability, lower transaction costs and a wider reach of online platforms (see Schulte-Noelle, 2001; Taylor, 2001). As a consequence, possibilities for convergence in financial services are discussed (see Beltratti and Corvino, 2008), as well as implications of formal models of insurance business (see, e.g., Seog, 2009).

Barkur, Varambally, and Rodrigues (2007) emphasise the need for organisational change in the insurance industry to cope with the aforementioned dynamics. The ubiquitous presence of mobile, interconnected devices adds further momentum to this argument, as it enables insurance companies to adopt new business models (see Desyllas and Sako, 2013) and change the types of risks that can be insured against (see, e.g., Gehrke, 2014). Big data analytics allow individual and adaptive calculations of premiums based on information about the insurance holder's behaviour (see McAfee and Brynjolfson, 2012) and risks which have previously not be calculable can now be estimated in ways that make it possible to address them with new types of insurance (see Eling and Schell, 2016). Furthermore, platform-based interactions can complement centralised insurance offerings for specific interest groups (see, e.g., Cole, 2015; Salman, 2014).

Eling and Lehmann (2018) give an overview of current literature related to the digital transformation in the insurance industry. Subtopics include artificial intelligence, big data, the internet of things, blockchain, cloud computing, mobile devices, and various online applications. They show that effects of digital technology cannot be considered in isolated subcategories (see also Nicoletti, 2016; ACORD, 2017). Opportunities for digital innovation span different steps of the value chain and at the same time influence the structure of insurance offerings and their objects of reference. A distinction between internally oriented technical solutions to support business operations and externally oriented solutions for customer interaction can no longer be upheld. To capture the full potential of digital technology for innovation, insurance companies must take a comprehensive approach which includes all their internal and external operations.

While the relevance of the digital transformation as a concept in systems engineering is controversially discussed, experts from industry and academia widely seem to agree on its orientational value in strategic decision making (see, e.g., Legner et al., 2017; Riedl et al., 2017). In combining business- and technology-related topics, this discussion has a strategic orientation (Hess et al., 2016; Matt, Hess, and Benlian, 2015). Digitalisation issues, however, are not always discussed in the same depth. Mertens and Wiener (2018) identify typical structures of a hype cycle in public references to digitalisation, indicating that the term is currently over-used (see also Steininger et al., 2009).

Therefore, in order to distinguish strategic agendas for digital transformation from general mentions of digital technology without further consequence, additional criteria have to be taken into consideration (see Skog, Wimelius, and Sandberg, 2018). Mentions of digitalisation do not necessarily constitute a digital agenda. Digital agendas need to refer to specific operations in a company which define the application context for technology. Based on the literature on the digital transformation of the insurance industry, we distinguish references to internal and external contexts. If both are addressed at the same time, we speak of a comprehensive digital agenda, which can be expected to facilitate digital innovation and give companies a competitive advantage on the market. We express this thought in the following hypothesis: *the existence of a comprehensive digital agenda in a company is positively related to its business performance*.

In following sections, the hypothesis is empirically tested. We study annual reports of insurance companies to assess how companies commit themselves in public to the digital transformation. Our analysis is focused on references to digitalisation topics and the context in which they occur. The context allows us to draw conclusions about the actual digital agendas in the company, which are then set in relation to additional information about their business performance.

## 3. DATA, VARIABLES, AND METHODOLOGY

The European market is particularly interesting for the given research topic, because the infrastructure of the European insurance industry has already reached a high level of maturity. One can therefore expect less interference between effects of the digital transformation and other concurrent change processes than in rapidly developing markets. In line with prior research on the European insurance industry (Bohnert et al., 2018), we use a sample of 41 publicly-traded European insurance companies that covers a sizeable proportion of the insurance market in Europe, representing approximately 60% of gross premiums in the year 2015 (see also Insurance Europe, 2016). We study the development of the companies over the past ten years from 2007 to 2017, for which we retrieved financial data from the Thomson Reuters Datastream. Information on the firms' digitalisation activities was derived from their disclosed annual reports.[1]

*Measuring firm value (dependent variable)*

We use Tobin's *Q* as a proxy for an insurance company's value following prior practice (see, e.g., Hoyt and Liebenberg, 2011; Masli et al., 2011; Bardhan, Krishnan, and Lin, 2013; Bohnert et al., 2018). As stated in Table 1, Tobin's *Q* is calculated as the ratio of the market value of equity plus the book value of liabilities divided by the book value of assets, or equivalently as the ratio of the market value of assets divided by their replacement costs (see, e.g., Hoyt and Liebenberg, 2011). It is held in the finance literature that Tobin's *Q* has several advantages compared to other performance and value measures (see, e.g., Lindenberg and Ross, 1981; Hoyt and Liebenberg, 2011; Lin, Wen, and Yu, 2012).

*Assessing digitalisation activities*

In order to draw conclusions about the digital agendas from annual reports of insurance companies, we use advanced text mining techniques. The firms' annual reports are manually retrieved as PDF documents from the companies' websites and further processed to extract the plain text of the reports, followed by a quantitative text analysis.[2]

---

[1] In order to be able to calculate Tobin's *Q*, we restrict the data set to publicly-traded insurance companies and consider companies that disclose their full annual reports in English for the respective years.

[2] Several approaches were conducted and evaluated; the most suitable results in our case were provided by preprocessing with Ghostscript, plain text extraction via Xpdf, and quantitative text analysis using the programming language R (considered alternatives include, amongst others, PDFBox, RapidMiner, and Tika).

Table 1 provides an overview of the variables used in analysis, with the upper part showing the digitalisation variables.

**Table 1**: Definition of variables

| Variable | Measurement |
|---|---|
| $d_{i,t}^{absolute}$ | Absolute number of occurrences of words containing the strings "digita" or "digiti" (keyword strings *d*) for company *i* in year *t*[3] |
| $w_{i,t}^{absolute}$ | Total number of words in the annual report for company *i* in year *t* (without punctuation and numbers) |
| $d_{i,t}^{relative}$ | $d_{i,t}^{absolute} / w_{i,t}^{absolute} \cdot 100{,}000$, i.e. number of occurrences of words containing the keyword strings *d* for company *i* in year *t* relative to the total number of words in the respective annual report times one hundred thousand |
| $d_{i,t}^{binary}$ | 1 if $d_{i,t}^{absolute} \geq 1$, i.e. at least one occurrence of the keyword strings *d*<br>0 otherwise |
| $d_{i,t}^{c20,e,binary}$ | 1 if at least one occurrence of keyword stems *e* indicating digital external activities focusing on products and sales[4]<br>0 otherwise |
| $d_{i,t}^{c20,i,binary}$ | 1 if at least one occurrence of keyword stems *i* indicating digital internal activities including modelling and management[5]<br>0 otherwise |
| $d_{i,t}^{c20,ei,binary}$ | 1 if $d_{i,t}^{c20,e,binary} = 1$ and $d_{i,t}^{c20,i,binary} = 1$<br>0 otherwise |
| *Q* | (Market value of equity + book value of liabilities) / book value of assets |
| *Size* | Natural logarithm of book value of assets |
| *ROA* | Net income / book value of assets |
| *Leverage* | Book value of liabilities / market value of equity |
| *Dividends* | 1 = Insurer paid dividends (i.e. dividend payments > 0) in the respective year<br>0 = Otherwise |
| *SalesGrowth* | (Sales(t) – sales(t–1)) / sales(t–1) |

*Notes: Quantitative text mining variables are calculated based on the companies' annual reports; financial variables are based on Bohnert et al. (2018) and retrieved from Thomson Reuters Datastream: Market value of equity = market capitalisation (WC08001), book value of liabilities = total assets (WC02999) – total shareholders' equity (WC03995), book value of assets = total assets (WC02999), net income = net income available to common (WC01751), sales = net sales or revenue (WC01001), dividend payments = cash dividends paid total (WC04551), all calculations are done in Euros and converted to Euros if necessary.*

[3] Comprising words such as digital, digitalisation, digitalise, digitalised, digitalising, digitalization, digitalize, digitalized, digitalizing, digitally, digitisation, digitise, digitised, digitising, digitization, and digitize.

[4] Word stems include channel, client, custom, distribut, market, onlin, product, sale, service.

[5] Word stems include board, employe, group, manag, model.

We first count the number of occurrences of words containing the strings “digita” or “digiti” (keyword strings *d* hereafter) for company *i* in year *t* denoted as $d_{i,t}^{absolute}$ comprising any grammatical forms of words referring to digital technology, innovation, and transformation.[6] We next calculate $d_{i,t}^{relative}$ as the ratio of occurrence of these words relative to the overall number of words (without punctuation and numbers) for company *i* in year *t*. In addition to this, we also determine a binary variable $d_{i,t}^{binary}$ that is equal to one in case of $d_{i,t}^{absolute} \geq 1$ and zero otherwise. The occurrence of the keyword strings *d* (measured by the variables $d_{i,t}^{k}$, *k* = absolute, binary, or relative) might be interpreted as an indicator for the general awareness and relevance of the digital transformation in the respective company for a given year. In line with our theoretical considerations, however, we assume that it is too superficial to draw conclusions about digital agendas from simple word counts, as it does not reveal anything about the applications domains for digital technology in the companies. As a consequence, we spent further effort to understand the contexts in which digitisation is addressed.

In doing so, we make use of key word in context (KWIC) concordances as “the most common corpus-linguistic tool currently used” (Gries and Newman, 2013, p. 277). Here, a predefined number of words to the left and to the right of a word of interest (which we define, e.g., as *c*20 as 20 words to both sides of a keyword expression) are extracted from the entire text to further assess the use of the word and get an impression of this word’s immediate context. Since a manual inspection of all concordances is hardly feasible and would further induce a source of subjectivity, we proceed as follows (see also Appendix A.2).

For each concordance, i.e. a certain number of words (e.g. 20 words in case of *c*20) around a word of interest containing our keyword string *d*, we transform these words into word stems.[7] We next determine the most frequent word stems across all concordances and attempt to assign relevant word stems to distinct groups. We build two categories for digitalisation activities with respect to (1) external stakeholders (denoted by *e* hereafter) including word stems such as “channel”, “client”, “custom”, “distribut”, “market”, “online”, “product”, “sale”, “service” and (2) for internal stakeholders (denoted by *i* hereafter) including word stems such as “board”, “employe”, “group”, “manag”, “model”.[8]

While the previous figures $d_{i,t}^{k}$ (*k* = absolute, binary, or relative) allow the general assessment of activities with respect to the digital transformation, we now can further categorise the companies’ engagement as follows. We interpret the occurrence of any word stem of the class “*e*”

6 Note that we explicitly do not use the common word stem “digit” at this point, since it could be misleading.

7 This is done by Porter’s word stemming algorithm via SnowballC (see Bouchet-Valat, 2014).

8 The word stems were gained from a content analysis of a subset of the complete data set. In order to avoid individual bias, key words were chosen by different researchers and then comprehensively discussed before a common agreement on the most suitable items was reached.

("*i*") in the concordance of digitalisation-related words "*d*" as digitalisation activities with respect to external (internal) stakeholders and set the variable $d_{i,t}^{c20,e,binary}$ ($d_{i,t}^{c20,i,binary}$) to 1 and 0 otherwise. The variable $d_{i,t}^{c20,ei,binary}$ (note the "*ei*" in the superscript) is equal to 1 in case of $d_{i,t}^{c20,e,binary} = 1$ and $d_{i,t}^{c20,i,binary} = 1$ (0 otherwise), i.e. in case a company addresses digitalisation in the context of both, external and internal stakeholders, in its annual report.

It is important to keep in mind that this process should be seen as a first objective and reproducible attempt to assess and categorise a firm's engagement with respect to digitalisation based on the annual reports, which, to the best of our knowledge, has not been done to date.

*Further covariates*

In addition to the variables measuring digitalisation engagement, there are further variables that can have an impact on firm value and thus have to be included as covariates in the regression analysis. We use the variables stated in Table 1 (lower part), which are based on Hoyt and Liebenberg (2011) and Bohnert et al. (2018).

*Methodology*

We apply a treatment-effects model to estimate the impact of a company's engagement in digitalisation (binary and endogenous treatment) on its firm value (continuous and dependent variable), which is given by a system of two equations, i.e. the regression equation (denoted as *Q* Equation),

$$Q = f\left(Digital \,|\, Size,\ ROA,\ Leverage,\ Dividends,\ SalesGrowth\right), \tag{1}$$

and the selection equation (denoted as *Digital* Equation)

$$Digital = f\left(Size\right), \tag{2}$$

where the covariates are based on the literature (see, e.g., Bohnert et al., 2018). Since there are firm characteristics that can have an impact on the activities with respect to digitalisation as well as on the firm value directly, we have to deal with endogeneity. In our base case, we set

$$Digital = d_{i,t}^{c20,ei,binary},$$

while we also use $d_{i,t}^{k}$, amongst others, $k$ = "*c*20,*e*,*binary*", "*c*50,*e*,*binary*", "*c*20,*i*,*binary*", "*c*50,*i*,*binary*", "c50,*ei*,*binary*", and "*binary*" in the robustness analysis.

For further details on the treatment-effects model, we refer the reader to the literature (see, e.g., Lee, 1978; Heckman, 1978, 1979; Maddala, 1983; Guo and Fraser, 2010; Hoyt and Liebenberg, 2011; Greene, 2012; Bohnert et al., 2018) and for more technical specifications in our setting see Appendix A.1. See Imbens and Wooldridge (2009) for a comprehensive overview on the statistical analysis of causal effects.

## 4. EMPIRICAL RESULTS

*Descriptive statistics*

We now show descriptive statistics for our full sample set (unbalanced panel) of European insurance companies comprising a total of 393 firm-year observations for the years 2007 to 2017 covering a considerable proportion of the insurance market in Europe.[9] The descriptive summary statistics for all firm-year observations are given in Table 2.

**Table 2**: Summary statistics

| Variable | Mean | Std. Dev. | 1st Quartile | Median | 3rd Quartile |
|---|---|---|---|---|---|
| $d_{i,t}^{absolute}$ | 9.5165 | 22.8342 | 0.0000 | 1.0000 | 9.0000 |
| $d_{i,t}^{relative}$ | 9.7427 | 24.9776 | 0.0000 | 1.2402 | 10.2096 |
| $d_{i,t}^{binary}$ | 0.5929 | 0.4919 | 0.0000 | 1.0000 | 1.0000 |
| $d_{i,t}^{c20,e,binary}$ | 0.4809 | 0.5003 | 0.0000 | 0.0000 | 1.0000 |
| $d_{i,t}^{c20,i,binary}$ | 0.4402 | 0.4970 | 0.0000 | 0.0000 | 1.0000 |
| $d_{i,t}^{c20,ei,binary}$ | 0.3969 | 0.4899 | 0.0000 | 0.0000 | 1.0000 |
| *Q* | 1.0270 | 0.0849 | 0.9867 | 1.0065 | 1.0343 |
| *Size* | 17.8112 | 1.7893 | 16.9304 | 17.9194 | 19.3459 |
| *ROA* | 0.0128 | 0.0172 | 0.0031 | 0.0069 | 0.0170 |
| *Leverage* | 14.2510 | 12.4742 | 5.0538 | 10.7896 | 19.8935 |
| *Dividends* | 0.9644 | 0.1856 | 1.0000 | 1.0000 | 1.0000 |
| *SalesGrowth* | 0.0489 | 0.5272 | -0.0409 | 0.0295 | 0.0903 |

*Notes: Total number of firm-year observations is 393 for a period of 11 years with 30 to 40 yearly observations.*

Table 2 (upper part) shows descriptive statistics for the measures of digitalisation activities. It can be seen that digitalisation is addressed about 10 times on average in every annual report (case for a simple total count $d_{i,t}^{absolute}$). The relatively large standard deviation indicates large differences between the individual firm-year observations that are due to considerable differences over time regarding references to the digital transformation and also with respect to var-

9 By focusing on the European market, we refrain from dealing with market specifics.

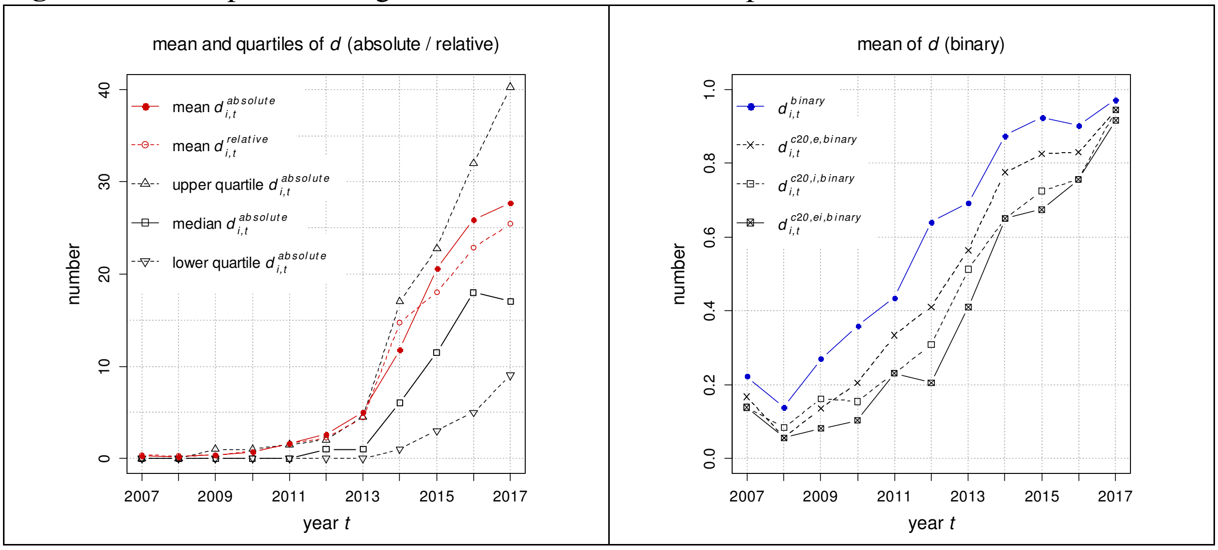
iations for different insurance companies (see Figure 1). The variable $d_{i,t}^{binary}$ shows that digitalisation is addressed at least once on average in only a bit more than half of the individual firm-year observations. This finding is notable, particularly against the background of the considerably increasing importance and omnipresent discussion of digital transformation.

Apart from general references to digitalisation, we further focus on $d_{i,t}^{c20,ei,binary}$ as an indicator for the existence of an actual digital agenda. The variable $d_{i,t}^{c20,ei,binary}$ measures (binary) whether a firm discusses digitalisation in the context of external issues and stakeholders ("*e*"), such as customers, market, or sales, and/or with respect to internal issues and stakeholders ("*i*"), including the board, group, or model(ling). The results show that $d_{i,t}^{c20,ei,binary}$ is equal to 1 in a little fewer than 40% of the firm-year observations (156 firm-year observations with $d_{i,t}^{c20,ei,binary} = 1$ and 237 firm-year observations with $d_{i,t}^{c20,ei,binary} = 0$).

Table 2 (lower part) exhibits the summary statistics for the financial variables, i.e. the dependent (Tobin's *Q*) and the other explanatory variables. It can be seen that the insurers in our sample have *Q*-values that are larger than 1 on average indicating the creation of value (on average) (see, e.g., Lindenberg and Ross, 1981; Bohnert et al., 2018).

We next consider the digitalisation activities of European insurers over time in Figure 1. Figure 1 (left graph) exhibits the developments of the absolute and relative measures over time (for the general case without taking the area of digitalisation activities into account).

**Figure 1**: Development of digitalisation activities of European insurers over time




First, it can be seen that all variables increased considerably over time showing that digitalisation was addressed in more detail in the firms' annual reports and became increasingly relevant to insurers. But the graph further shows that this development did not occur for all insurers to the same extent, which can be seen by taking a look at the quartiles of $d_{i,t}^{absolute}$ (arrow upwards for the upper or 3rd quartile; square for the median or 2nd quartile; and arrow downwards for the lower or 1st quartile), and also in comparison to the mean of $d_{i,t}^{absolute}$ (line with solid red dots). The comparison of the development of the means $d_{i,t}^{absolute}$ (solid red dots) to $d_{i,t}^{relative}$ (red circles) shows a similar development indicating no need for relative measures here.

Figure 1 (right graph) shows the development of the mean of the binary versions of the digitalisation measures. All show a considerable increase over time in general. The only exception is the year 2008, where we can observe a one-year drop in the otherwise clearly increasing trend, which might stem from the financial crisis forcing the financial industry to focus on its core activities. The variable $d_{i,t}^{binary}$ (blue line with solid dots) measures whether an insurer addresses digitalisation at least once in its annual report (1, and 0 otherwise) and shows that only about 20% of the main European insurance companies did so in 2007, whereas in 2017 the situation has changed substantially with the comparable figure of about 90%. The two lower black dashed lines in Figure 1 show that activities with respect to digitalisation in the context of external issues and stakeholders ("*e*"), such as customers, market, or sales ($d_{i,t}^{c20,e,binary}$, black crosses) are discussed a little more often than topics relating to internal issues and stakeholders ("*i*"), including the board, group, or model(ling) ($d_{i,t}^{c20,i,binary}$, black circles). However, both developments are fairly similar.

We focus on the combined variable $d_{i,t}^{c20,ei,binary}$ in the regression analysis. The variable $d_{i,t}^{c20,ei,binary}$ (black crossed circles in Figure 1) depicts the intersection of these two: i.e. it shows insurers that address digitalisation in both the external and internal context. Hence, $d_{i,t}^{c20,ei,binary}$ is the main variable of interest in the subsequent regression analysis (besides Tobin's *Q* for value measurement).

*Regression analysis*

The aim is to statistically assess the impact of a (comprehensive) digital agenda in an insurance company (by means of the disclosure in an annual report) on its firm value. Thus, we now perform a treatment-effects regression analysis with the main results given in Table 3.

We consider *Digital* = $d_{i,t}^{c20,ei,binary}$ as binary and endogenous treatment for Equations (1) and (2) with the dependent (observed and continuous) variable Tobin's *Q* and obtain the regression estimates via full maximum-likelihood using firm-level clustering. The main finding reveals that the variable *Digital* ($d_{i,t}^{c20,ei,binary}$) has a positive and statistically significant (at the 1% level) impact on Tobin's *Q*, i.e. companies that address digitalisation in the context of external issues / stakeholders (such as customers, market, or sales) in addition to internal issues / stakeholders (including the board, group, or model(ling)) exhibit a firm value that is more than 8% higher than for companies that do not (and that are most likely engaged in digitalisation activities to a lower extent) when controlling for relevant covariates and the endogeneity bias.[10]

**Table 3**: Treatment-effects estimates for the value relevance of digitalisation

| Variable | *Digital* Equation (2) | *Q* Equation (1) |
|---|---|---|
| *Digital** | | 0.085942 (0.029323)*** |
| *Size* | 0.175850 (0.060429)*** | -0.001135 (0.005799) |
| *ROA* | | 2.417535 (0.739218)*** |
| *Leverage* | | -0.001104 (0.000476)** |
| *Dividends* | | 0.020304 (0.012614) |
| *SalesGrowth* | | -0.000520 (0.004482) |
| Constant | -3.408263 (1.129320)*** | 0.978326 (0.100810)*** |
| Observations | 393 | |
| Number of clusters (firms) | 41 | |
| Likelihood-ratio test | 4.81** | |
| Wald test | 59.76*** | |

*Notes: The treatment-effects model is fitted via full maximum-likelihood using firm-level clustering; standard errors are given in parentheses with '**' and '***' indicating the level of statistical significance at the 5% and 1% level, respectively.*

It can also be seen that the variable *Size* has a positive and statistically significant (at the 1% level) impact on the digitalisation activities (see *Digital* Equation in Table 3), i.e. the larger the insurer (by means of the book value of assets), the more likely is $d_{i,t}^{c20,ei,binary}$ to be equal to 1 meaning that an insurer is more likely to be active in digitalisation with respect to external and internal issues / stakeholders.

Several robustness checks confirm the main result that European insurance companies with $d_{i,t}^{c20,ei,binary} = 1$ have a higher firm value than companies with $d_{i,t}^{c20,ei,binary} = 0$. The result can be confirmed when dropping data for the year 2008 (as an outlier and year of the financial crisis), which results in a statistically significant (at the 5% level) coefficient (standard error) for $d_{i,t}^{c20,ei,binary}$ of 0.083234 (0.034167). While the treatment-effects analysis also shows positive coefficients for the variables $d_{i,t}^{binary}$ (0.023525), $d_{i,t}^{c20,e,binary}$ (0.025478), $d_{i,t}^{c20,i,binary}$ (0.082637), only the latter has statistical significance (at the 1% level).

The results underline the importance of digital agendas that address internal and external operations comprehensively. General references to digitisation and respective agendas for internal or external operations also seem to be positively related to business performance, but the effects

[10] For $d_{i,t}^{binary}$, the effect seems to be positive as well, but could not be statistically confirmed.

are much less clear than in the case of comprehensive digital agendas. Calculations with other values for the concordances and also the digitalisation variables with lags of one and two years yield similar results, which overall reinforce our previous findings.[11]

## 5. Summary

The results of our study contribute to the ongoing debate around the digital transformation of the insurance industry. While digital technology is widely believed to create huge potential benefit for innovation and business development, relatively little has been said to date about the actual impact of digitisation activities on the business performance of insurance companies. For this reason, we set the focus of our investigation on the relation between the engagement in digital transformation and the firm value of insurance companies. In particular, we look at the role of the digital agendas that guide and direct a companies' digitalisation activities. We draw on extant work in the field of insurance studies as well as information systems research, which allows us to take an interdisciplinary look at digital technology and insurance business. In line with current discussions of digital innovation in the field of information systems research, we identify the need for comprehensive digital agendas to make use of the full potential of digital technology for new directions of growth and development in the insurance industry.

Digital technologies have already found many applications in the context of insurance. In the European market, where the operational infrastructure of the insurance industry is in general very stable, the effects of these applications can be studied rather well. Although the current state of development can only be considered as a first step towards a full digitalisation of the insurance industry, it still seems justified to use it as a basis for the analysis of the relation between digitisation and business performance.

In line with the principles of corpus linguistics, data about digital agendas was collected from the annual reports of publically listed companies, such that personal bias from interviews could be avoided. The annual reports give insight into the official position of the companies regarding the digital transformation and the activities to which they commit themselves publicly. This method provides us with a firm empirical basis to investigate organizations digital agendas and allows the application of standard text mining techniques, which simplify the replication of our approach in subsequent studies.

Our findings reveal that the expression of digital agendas is positively related to business success. Moreover, they show that insurance companies addressing digitalisation activities in their

11 We also calculated $d_{i,t}^{c50,ei,binary}$, $d_{i,t}^{c50,e,binary}$, $d_{i,t}^{c50,i,binary}$, $d_{i,t}^{c100,ei,binary}$, $d_{i,t}^{c100,e,binary}$, and $d_{i,t}^{c100,i,binary}$.

core business as well as activities across organisational boundaries exhibit a firm value measured by Tobin's *Q* that is more than 8% higher than the value for companies that do not. This finding holds true when taking relevant covariates into account and when controlling for endogeneity. While conclusions about unidirectional causality should always be made with caution, these results provide strong evidence that digital agendas and business success are positively related. Furthermore, the results show that the range of applications of digital technology considered in the agendas plays an important role. The strongest and most significant effects are clearly observed for companies with comprehensive digital agendas.

The results of our study have strong implications for insurance companies regarding the treatment of digital technologies. It is not enough to go with the flow and adopt them for specific application cases. Instead, companies need to think more strategically about digitisation and acknowledge its transformative effect on their overall business activities, including internal as well as external operations. From a theoretical point of view, our study shows how the concept of digital innovation can be applied to the insurance industry and which subcategories of digital innovation need further study. Furthermore, we believe that our paper contributes to the development of research methodology in the field with an application of text mining techniques, which are still a rather uncommon means of enquiry but can be expected to play an increasingly important role in the years to come.

## REFERENCES

ACORD (Association for Cooperative Operations Research and Development), 2017, Insurance Digital Maturity Study.

Agarwal, R. and H. C. Lucas, 2005, The Information Systems Identity Crisis: Focusing on High-Visibility and High-Impact Research, *MIS Quarterly,* 29(3): 381–398.

Agarwal, R., G. Gao, C. DesRoches, and A. K. Jha, 2010, Research Commentary: The Digital Transformation of Healthcare: Current Status and the Road Ahead, *Information Systems Research,* 21(4): 796–809.

Atzori, L., A. Iera, and G. Morabito, 2010, The Internet of Things: A Survey, *Computer Networks*, 54(15): 2787–2805.

Bardhan, I., V. Krishnan, and S. Lin, 2013, Research Note – Business Value of Information Technology: Testing the Interaction Effect of IT and R&D on Tobin's Q, *Information Systems Research,* 24(4): 1147–1161.

Barkur, G., K. V. M. Varambally, and L. R. Rodrigues, 2007, Insurance Sector Dynamics: Towards Transformation into Learning Organization, *The Learning Organization,* 14(6): 510–523.

Beltratti, A. and G. Corvino, 2008, Why are Insurance Companies Different? The Limits of Convergence among Financial Institutions, *Geneva Papers on Risk and Insurance—Issues and Practice*, *33*(3): 363–388.

Bohnert, A., N. Gatzert, R. E. Hoyt, and P. Lechner, 2018, The Drivers and Value of Enterprise Risk Management: Evidence from ERM Ratings, *European Journal of Finance,* DOI: 10.1080/1351847X.2018.1514314.

Bouchet-Valat, M., 2014, SnowballC, The Comprehensive R Archive Network (CRAN).

Boudreau, M. C. and D. Robey, 2005, Enacting Integrated Information Technology: A Human Agency Perspective, *Organization Science,* 16(1): 3–18.

Brettel, M., N. Friedrichsen, M. Keller, and M. Rosenberg, 2014, How Virtualizaion, Decentralization and Network Building Change the Manufacturing Landscape: An Industry 4.0 Perspective, *International Journal of Information and Communication Engineering,* 8(1): 37–44.

Brynjolfsson, E. and A. McAfee, 2012, Winning the Race with Ever-Smarter Machines, *Management Review,* 53(2): 53–60.

Cole, S., 2015, Overcoming Barriers to Microinsurance Adoption: Evidence from the Field, *Geneva Papers on Risk and Insurance—Issues and Practice*, 40(4): 720–740.

Desyllas, P. and M. Sako, 2013, Profiting from Business Model Innovation: Evidence from Pay-As-You-Drive Auto Insurance, *Research Policy*, 42(1): 101–116.

Dhar, V. and A. Sundararajan, 2007, Issues and Opinions: Information Technologies in Business: A Blueprint for Education and Research, *Information Systems Research,* 18(2): 125–141.

Dumm, R. E. and R. E. Hoyt, 2003, Insurance Distribution Channels: Markets in Transition, *Journal of Insurance Regulation*, (22)1: 27–47.

Eastman, J. K., A. D. Eastman, and K. L. Eastman, 2002, Insurance Sales Agents and the Internet: The Relationship Between Opinion, Leadership Subjective Knowledge, and Internet Attitudes, *Journal of Marketing Management,* 18: 259–285.

Eling, M. and M. Lehmann, 2018, The Impact of Digitalization on the Insurance Value Chain and the Insurability of Risks, *Geneva Papers on Risk and Insurance—Issues and Practice*, 43(3): 359–396..

Eling, M., and W. Schnell, 2016, What Do We Know about Cyber Risk and Cyber Risk Insurance?, *Journal of Risk Finance*, *17*(5): 474–491.

Fichman, R. G., B. L. Dos Santos, and Z. Zheng, 2014, Digital Innovation as a Fundamental and Powerful Concept in the Information Systems Curriculum, *MIS Quarterly,* 38(2): 329–353.

Garven, J. R., 2002, On the Implication of the Internet for Insurance Markets and Institutions, *Risk Management and Insurance Review,* (5)2: 105–116.

Gehrke, E. (2014). The Insurability Framework Applied to Agricultural Microinsurance: What Do We Know, What Can We Learn? *Geneva Papers on Risk and Insurance—Issues and Practice*, *39*(2): 264–279.

Gölzer, P. and A. Fritzsche, 2017, Data-driven Operations Management: Organisational Implications of the Digital Transformation in Industrial Practice, *Production Planning & Control,* (28)16: 1332–1343.

Greene, W. H., 2012, *Econometric Analysis*, 7th ed., Pearson, Boston.

Gries, S. T. and J. Newman, 2013, Creating and Using Corpora, in R. J. Podesva and D. Sharma, eds., Research Methods in Linguistics, 257–287, Cambridge University Press, Cambridge.

Guo, S. and M. W. Fraser, 2010, Propensity Score Analysis: Statistical Methods and Applications, Sage Publications.

Heckman, J. J., 1978, Dummy Endogenous Variables in a Simultaneous Equation System, *Econometrica*, 46(4): 931–959.

Heckman, J. J., 1979, Sample Selection Bias as a Specification Error, *Econometrica*, 47(1): 153–161.

Hess, T., C. Matt, A. Benlian, and F. Wiesböck, 2016, Options for Formulating a Digital Transformation Strategy, *MIS Quarterly Executive*, 15(2): 123–139.

Hildebrandt, B., A. Hanelt, S. Firk, and L. M. Kolbe, 2015, Entering the Digital Era – The Impact of Digital Technology-related M&As on Business Model Innovations of Automobile OEMs, *36th International Conference on Information Systems,* Forth Worth, 1–21.

Hoyt, R. E. and A. P. Liebenberg, 2011, The Value of Enterprise Risk Management, *Journal of Risk and Insurance*, 78(4): 795–822.

Hylving, L. and U. Schultze, 2013, Evolving the Modular Layered Architecture in Digital Innovation: The Case of the Car's Instrument Cluster, *34th International Conference on Information Systems,* Milan, 1–17.

Imbens G. W. and J. M. Wooldridge, 2009, Recent Developments in the Econometrics of Program Evaluation, *Journal of Economic Literature*, 47(1): 5–86.

Insurance Europe, 2016, European Insurance – Key Facts, www.insuranceeurope.eu, accessed 11/25/2016.

Kaiser, J., 2002, The Customer Shall Lead: E-Business Solutions for the New Insurance Industry, *Geneva Papers on Risk and Insurance—Issues and Practice,* 27(1): 134–145.

Kohli, R. and S. Devaraj, 2003, Measuring Information Technology Payoff: A Meta-Analysis of Structural Variables in Firm-Level Empirical Research, *Information Systems Research,* 14(2): 127–145.

Kohli, R. and V. Grover, 2008, Business Value of IT: An Essay on Expanding Research Directions to Keep up with the Times, *Journal of the Association for Information Systems,* (9)1: 23–39.

Lee, E. A., 2008, Cyber Phyiscal Systems: Design Challenges, *ISORC '08 Proceedings of the 2008 11th IEEE Symposium on Object Oriented Real-Time Distributed Computing, 363–369.*

Lee, L.-F., 1978, Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables, *International Economic Review*, 19(2): 415–433.

Legner C., T. Eymann, T. Hess, C. Matt, T. Böhmann, P. Drews, A. Maedche, N. Urbach, and F. Ahlemann, 2017, Digitalization: Opportunity and Challenge for the Business and Information Systems Engineering Community, *Business & Information Systems Engineering*, 59(4): 301–308.

Leonardi, M. P., 2011, When Flexible Routines meet Flexible Technologies: Affordance, Constraint, and the Imbrication of Human and Material Agencies, *MIS Quarterly,* 35(1): 147–167.

Lin, Y., M.-M. Wen, and J. Yu, 2012, Enterprise Risk Management: Strategic Antecedents, Risk Integration and Performance, *North American Actuarial Journal*, 16(1): 1–28.

Lindenberg, E. B. and S. A. Ross, 1981, Tobin's Q Ratio and Industrial Organization, *The Journal of Business*, 54(1): 1–32.

Lu, Y. and K. R. Ramamurthy, 2011, Understanding the Link between Information Technology Capability and Organizational Agility: An Empirical Examination, *MIS Quarterly,* 35(4): 931–954.

Lucas, H. C., R. Agarwal, E. K. Clemons, O. A. El Sawy, and B. Weber, 2013, Impactful Research on Transformational Information Technology: An Opportunity to Inform New Audiences, *MIS Quarterly,* 37(2): 371–382.

Lusch, R. F. and S. Nambisan, 2015, Service Innovation: A Service-dominant Logic Perspective, *MIS Quarterly,* 39(1): 155–175.

Maddala, G. S., 1983, Limited-Dependent and Qualitative Variables in Econometrics, Cambridge University Press.

Markus, M. L. and D. Robey, 1988, Information Technology and Organizational Change: Causal Structure in Theory and Research, *Management Science,* 34(5): 583–598.

Markus, M. L., 2004, Technochange Management: Using IT to Drive Organizational Change, *Journal of Information Technology*, (19)1: 4–20.

Masli, A., V. J. Richardson, J. M. Sanchez, and R. E. Smith, 2011, The Business Value of IT: A Synthesis and Framework of Archival Research, *Journal of Information Systems*, 25(2): 81–116.

Matt, C., T. Hess, and A. Benlian, 2015, Digital Transformation Strategies, *Business & Information Systems Engineering*, 57(5): 339–343.

McAfee, A. and E. Brynjolffson, 2012, Big Data: The Management Revolution, *Harvard Business Review*, 3–9.

Melville, N., K. Kraemer, and V. Gurbaxani, 2004, Review: Information Technology and Organizational Performance: An Integrative Model of IT Business Value, *MIS Quarterly,* 28(2): 283–322.

Mertens, P., and M. Wiener, 2018, Riding the Digitalization Wave: Toward a Sustainable Nomenclature in Wirtschaftsinformatik, *Business & Information Systems Engineering*, 60(4): 367–372.

Meyer, S. and G. Krohm, 1999, An Overview of Regulatory Developments of Insurance Transactions on the Internet, *Journal of Insurance Regulation,* 17(4): 551–554.

Mithas, S., N. Ramasubbu, and V. Sambamurthy, 2011, How Information Management Capability Influences Firm Performance, *MIS Quarterly,* 35(1): 237–256.

Mithas, S., A. Tafti, I. Bardhan, and J. M. Goh, 2012, Information Technology and Firm Profitability: Mechanisms and Empirical Evidence, *MIS Quarterly,* 36(1): 205–224.

Nambisan, S., K. Lyytinen, A. Majchrzak, and S. Song, 2017, Digital Innovation Management: Reinventing Innovation Management Research in a Digital World, *MIS Quarterly,* 41(1): 224.

Nicoletti, B., 2016, Digital Insurance: Business Innovation in the Post-Crisis Era, *Palgrave MacMillan.*

Orlikowski, W. J., 2009, The Sociomateriality of Organisational Life: Considering Technology in Management Research, *Cambridge Journal of Economics,* 34: 125–141.

Pentland, B. T., M. S. Feldman, M. C. Becker, and P. Liu, 2012, Dynamics of Organizational Routines: A Generative Model, *Journal of Management Studies,* 49(8): 1484–1508.

Porter, M. E. and J. E. Heppelmann, 2014, How Smart Connected Products Are Transforming Competition, *Harvard Business Review,* November, 3–23.

Riedl R., A. Benlian, T. Hess, D. Stelzer, and H. Sikora, 2017, On the Relationship between Information Management and Digitalization, *Business & Information Systems Engineering*, 59(6):475–482

Sambamurthy, V., A. Bharadwaj, and V. Grover, 2003, Shaping Agility through Digital Options: Reconceptualizing the Role of Information Technology in Contemporary Firms, *MIS Quarterly,* 27(2): 237–263.

Salman, S. A., 2014, Contemporary Issues in Takaful (Islamic Insurance), *Asian Social Science*, 10(22): 210–216.

Saunders, A. and E. Brynjolfsson, 2016, Valuing Information Technology Related Intangible Assets, *MIS Quarterly,* 40(1): 83–110.

Schmidt, R., M. Möhring, F. Bär, and A. Zimmermann, 2017, The Impact of Digitization on Information System Design – An Explorative Case Study of Digitization in the Insurance Business, *Business Information Systems Workshops,* 137–149.

Schryen, G., 2013, Revisiting IS Business Value Research: What We Already Know, What We Still Need to Know, and How We Can Get There, *European Journal of Information Systems,* 22(2): 139–169.

Schulte-Noelle, H., 2001, Technological Changes in IT and Their Influence on Insurance: The Change Ahead (I), *Geneva Papers on Risk and Insurance—Issues and Practice*, *26*(1): 83–88.

Seitz, M., 2017, Online Insurance Management among German Farmers, *The Proceedings of the 17th International Joint Conference Central and Eastern Europe in the Changing Business Environment,* 212–220.

Seog S. H., 2009, Insurance Markets with Differential Information, *Journal of Risk and Insurance,* 76(2): 279–294.

Skog, D. A., H. Wimelius, and J. Sandberg, 2018, Digital Disruption, *Business & Information Systems Engineering*, 1–7.

Steininger, K., R. Riedl, F. Roithmayr, and P. Mertens, 2009, Fads and Trends in Business and Information Systems Engineering and Information Systems Research–A Comparative Literature Analysis, *Business & Information Systems Engineering*, 1(6): 411–428.

Svahn, F., O. Henfridsson, and Y. Yoo, 2009, A Threesome Dance of Agency: Mangling the Sociomateriality of Technological Regimes in Digital Innovation, *Proceeding of International Conference on Information Systems,* Phoenix, 1–18.

Taylor, M., 2001, Technological Changes in IT and Their Influence on Insurance: The Change Ahead (II), *Geneva Papers on Risk and Insurance—Issues and Practice*, *26*(1): 89–104.

Venkatesh, V., M. G. Morris, G. B. Davis, and F. D. Davis, 2003, User Acceptance of Information Technology: Toward a Unified View, *MIS Quarterly*, 27(3): 425–478.

Yoo, Y., O. Henfridsson, and K. Lyytinen, 2010, Research Commentary: The New Organizing Logic of Digital Innovation: An Agenda for Information Systems Research, *Information Systems Research,* 21(4): 724–735.

## APPENDIX

### A.1. Treatment-effects model

The treatment-effects model is given by the following two regression equations that are simultaneously estimated via maximum-likelihood. We assume $d_{i,t}^{k} = d_{i,t}^{c20,ei,binary}$ in our base case.[12] The regression equation ("$Q$ Equation") is given by

$$Q_{i,t} = x_{i,t}\beta + d_{i,t}^{k}\delta + \varepsilon_{i,t} \tag{3}$$

and the selection equation ("*Digital* Equation") is defined as

$${}^{*}d_{i,t}^{k} = z_{i,t}\gamma + u_{i,t}\,, \tag{4}$$

where

$$d_{i,t}^{k} = \begin{cases} 1 & \text{if } {}^{*}d_{i,t}^{k} > 0 \\ 0 & \text{otherwise} \end{cases}$$

and error terms $\varepsilon_{i,t}$ and $u_{i,t}$ that are assumed to be normally distributed with a mean vector of zero, variances of $\sigma_{\varepsilon}$ and 1, and a covariance of $\rho$ (see, e.g., Maddala, 1983; Guo and Fraser, 2010; Hoyt and Liebenberg, 2011; Bohnert et al., 2018).

---

[12] In addition to this, we also use, amongst others, $k$ = "*c*20,*e*,*binary*", "*c*50,*e*,*binary*", "*c*20,*i*,*binary*", "*c*50,*i*,*binary*", "*c*50,*ei*,*binary*", and "*binary*" in the robustness analysis.

## A.2. Pseudocode

Repeat for each PDF document, i.e. annual report

- Rename and assign unique identifier (ID) to each PDF document (ID.pdf)
- Prepare document for text extraction, i.e. remove access restrictions (via Ghostscript)
- Extract plain text (via Xpdf) and create a text document with identical ID as the PDF document (ID.txt)

Repeat for each plain text document (for all files ID.txt)

- Create corresponding new (empty) text document for concordances (ID_conc.txt)
- Translate all characters to lower case characters
- Wrap all characters that are not alphabetic characters or digits with one space character before and after
- Replace (multiple) white space characters (including tab keys) and newline with one space character
- Identify occurrences of words containing the strings "digita" or "digiti"
- Repeat for each occurrence
  - Extract corresponding keyword string and certain number of words (string wrapped by space characters) before and after the corresponding keyword, i.e. extract concordance of a given length, e.g. 20 words before and 20 words after the keyword in case of c20
  - Add concordance line to ID_conc.txt
  - Repeat for each concordance document (for all files ID_conc.txt)
    - Remove all characters besides alphabetic characters
    - Remove common words from *common.list*[13]
    - Transform all words into word stems (via Porter's word stemming algorithm)
    - Set variable d_(i,t)^(c20,e,binary) to 1, if at least one of the following word stems can be found: "channel", "client", "custom", "distribut", "market", "online", "product", "sale", "service"
      set variable to 0 otherwise
    - Set variable d_(i,t)^(c20,i,binary) to 1, if at least one of the following word stems can be found: "board", "employe", "group", "manag", "model"
      set variable to 0 otherwise
    - Set variable d_(i,t)^(c20,ei,binary) to 1, if d_(i,t)^(c20,e,binary) == 1 and d_(i,t)^(c20,i,binary) == 1
      set variable to 0 otherwise

End

[13] List of English stop words retrieved from xpo6.com/list-of-english-stop-words.