Measuring dynamic capabilities in new ventures: exploring strategic change in US green goods manufacturing using website data

Sanjay K. Arora¹ · Yin Li² · Jan Youtie³ · Philip Shapira^{4,5}

Published online: 22 August 2019
© The Author(s) 2019

Abstract

Entrepreneurial scholarship suggests that a small firm's ability to grow is a function of its capacity to sense and respond to changes in the market as well as the broader environment for the firm's goods and services. Developing detailed measures of internal capabilities at a large scale, however, is often hampered by limitations in the availability of data from conventional sources, low survey response rates and panel attrition. The emergence of new information sources, including big data sets derived from the online activities of firms, coupled with advanced computational approaches, raises fresh analytical possibilities. In this exploratory study, we turn to freely accessible website data to gauge internal capabilities, specifically for market sensing and responding. To operationalize the construct of seizing, the paper uses an application of topic modeling, a text mining approach commonly used in computer science, on archived website data from the Wayback Machine for two time periods, 2008–2009 and 2010–2011, to explain sales growth for green goods enterprises in two later time periods, from 2010 to 2012. We find an endogenous inverse U-shaped relationship exists between market seizing and sales growth. Increasing levels of focus on a firm's local geographic area also predict sales growth. We consider these findings in light of the practitioner literature on firm agility and pivoting and discuss opportunities for future work using website data to study entrepreneurship and the strategic management of innovation.

Keywords Dynamic capabilities · Entrepreneurship · SMEs · Big data · Website analytics · Text mining

JEL Classification C18 · C81 · D22 · L21 · L26 · O32

✉ Philip Shapira
pshapira@manchester.ac.uk

¹ Ernst & Young, LLP, Washington, DC 20005, USA

² School of International Relations and Public Affairs, Fudan University, Shanghai 200433, China

³ Enterprise Innovation Institute, Georgia Institute of Technology, Atlanta, GA 30308, USA

⁴ Manchester Institute of Innovation Research, Alliance Manchester Business School, University of Manchester, Manchester M15 6PB, UK

⁵ School of Public Policy, Georgia Institute of Technology, Atlanta, GA 30332, USA

1 Introduction

Organizational agility is increasingly seen by researchers as a source of competitive advantage as firms alter their capabilities and business models to better respond to market opportunities and threats (Roberts and Grover 2012; Teece et al. 2016). Agility is essentially an outcome of enacting effective strategic change at the right time to achieve beneficial performance outcomes. According to the dynamic capability literature, agility involves the continuing ability to sense new market conditions, adapt or seize upon opportunities, and alter strategy (Teece 2007; Teece et al. 2016).

While both large and small firms can be agile, small firms are often better able to quickly respond to market needs and changing environments than their larger counterparts (Arora and Gambardella 1994; Rothwell 1989). They have fewer well-defined routines than large firms, and entrepreneurial leadership may alter the course of a firm's direction more readily than in firms with complex bureaucracies (Covin and Slevin 1988). While the contribution of small firms to economic growth is sometimes overpromoted (c.f., Atkinson and Lind 2018), small firms are often considered catalysts of disruptive innovation (Kassicieh et al. 2002). Investigating relationships between agility and innovation outcomes in small firms is thus valuable from both management and policy perspectives. Yet studying small firm agility and change is particularly difficult for a number of reasons, including low survey participation rates and panel attrition, lack of representation in databases, and other limitations in traditional social science research designs involving primary data collection (Bartholomew and Smith 2006; Baruch 1999; Laaksonen and Peltoniemi 2016; March and Sutton 1997).

Our study develops a non-traditional approach to measuring the evolution of firm capabilities and agility by using firm archived website data from the Wayback Machine. This work relies on publicly-posted web information that is not only essential for many firms' marketing and sales efforts but also may reflect important strategic content, particularly for small and medium sized enterprises (SMEs) (Li et al. 2016). We also contribute to the innovation literature by (a) using a computational, machine learning approach to quantify firm change over time, and (b) employing distinct, non-overlapping sources to measure dynamic capabilities and performance (c.f., Laaksonen and Peltoniemi 2016).

The case context for this work is the green goods manufacturing industry, which from a qualitative standpoint shows both underlying environmental dynamism and associated needs for changes in internal capabilities. The time period under study follows the 2008 economic crisis and US government incentives to encourage the manufacturing and adoption of sustainability related goods and services, indicating a high-velocity environment requiring market and policy learning as well as incremental strategic adjustments (c.f., Eisenhardt and Martin 2000).

In sum, we ask two related research questions: "Are green goods firms that sense and seize on market opportunities more likely to experience positive performance outcomes?" and "Can website data help us answer this first question?" While only addressing change over one-time period, we find that there is an inverse U-shaped relationship between seizing, operationalized as topical website change, and firm performance. Conversely, changes in sensing, as measured through R&D keyword mentions, are not associated with sales growth. Our exploratory study has implications for the way in which future research on small firm growth and technology commercialization could be undertaken to address market dynamism and firms' propensity to respond with increasing agility. Although more work is needed to refine the method, we contend that measuring firm change through

websites is an appropriate empirical approach for studying firm agility in the dynamic capability framework.

The paper is organized as follows: first, we position our review of the literature within the dynamic capabilities framework. Second, we introduce the case study context in more detail. Third, we submit an extensive overview of the research design to introduce the reader to our website measures and supporting methods. Fourth, we present the findings of our exploratory econometric analysis, which uses an instrumental variable approach to account for endogeneity of website topical change in the sales growth model. Finally, the discussion and conclusion summarizes the results and assess implications for research, policy, and practice.

2 Literature review

Organizational agility and performance can be explained by the dynamic capabilities framework, i.e., sensing, seizing, and managing threats (Roberts and Grover 2012; Teece et al. 2016). Sensing, according to Teece (2007), is the process of ascertaining environmental changes—for example, in science and technology, customer, and supplier landscapes—and directing R&D efforts accordingly. Sensing is related to exploration, which involves the outward examination of new information, applications of existing know-how, and dynamism of markets and science and technology (March 1991; Teece 2017). Seizing, conversely, includes business model definition and operationalization, securing complementary assets, implementing routines for decision making, and instituting leadership, communication, and cultural supports (Teece 2007). Seizing is related to exploitation, which refers to the operational aspects of following through on innovations and realizing efficiencies (Lavie et al. 2010; March 1991).

Research on ambidexterity acknowledges the motivations for, opportunities associated with, and challenges related to engaging in both exploration and exploitation (Raisch and Birkinshaw 2008; Rothaermel and Alexandre 2009; Tushman et al. 2010). A broad body of research shows ambidexterity results in positive performance outcomes, including sales growth, particularly in environments characterized by high market and technological uncertainty (O'Reilly and Tushman 2013). For example, using a longitudinal sample of 605 public technology firms, Caspin-Wagner et al. (2012) find an inverse-U shaped relationships between exploration, as measured by a firm R&D expenditures, and performance. They also find a similar relationship between exploitation activities and performance, suggesting that optimal levels of exploration and exploitation exist simultaneously.

Sensing and seizing, as put forth here, are examples of “first-order” dynamic capabilities (Winter 2003), which allow firms to reallocate resources and pursue market opportunities that may not exist in their current operational purview. First-order capabilities can be contrasted with ordinary capabilities, which enable operations and recurring (present) revenue streams, and second-order dynamic capabilities, which help firms “learn how to learn” (Schilke 2014b).

For a new small firm, dominant or ordinary routines are not likely to be well formed, but this is precisely why smaller firms are often perceived as being more nimble than their large firm counterparts (Arora and Gambardella 1994; Rothwell 1989). While all firms may become more steadfast in response to competitive threats (Killaly 1998; Toh and Kim 2013), young small firms may be more likely to benefit from a proactive approach to growth than from competitive aggressiveness, compared to larger, well-established

firms (Lumpkin and Dess 2001), suggesting that they may be more likely to change when encountering environmental uncertainty. In the absence of higher order routines that can alter an SME's lower-level capabilities, entrepreneurial managers may instead redirect resources and pursue strategic shifts directly (Teece 2012). Kickul and Gundry (2002) find that prospector (i.e., scanning) strategies mediate the relationship between proactive entrepreneur personalities and desirable innovation outcomes.

Ambidexterity in small firms appears more often sequentially rather than simultaneously due to limited resources and a lack of organizational slack (O'Reilly and Tushman 2013). That is, small firms are unlikely to sense and seize at the same time. In a related study, for example, Ebben and Johnson (2005) study efficiency and flexibility strategies and find that small firms that mix both strategies perform less well than those that focus primarily on one approach or another. If ambidexterity is likely to appear sequentially, what might cause a small firm to sense and seize insufficiently—or too much? We consider several possible explanations.

First, entrepreneurs often face resource constraints that limit their ability to act on new information and fully exploit new opportunities (Elston and Audretsch 2007; Newman 2010). Consequently, entrepreneurs restrict the types of goals and growth opportunities they seek based on available resources (Sarasvathy 2001). In some cases, entrepreneurs may not have the means to sense widely, though in other cases, these constraints may catalyze unconventional modes of experimentation, setting the foundation for radical innovations (Keupp and Gassmann 2013).

Second, in the mind of the entrepreneur, ingrained schemas, or ways of organizing and classifying observations about the world (Thagard 2005), may limit the breadth and depth of perceptions, a phenomena likely exacerbated by entrepreneurs who fail to respond appropriately to key developments or who do not wish to grow in the first place (Soly-mossy 2009; Wiklund et al. 2003). In this case, certain phenomena of critical importance for the firm's operations and long-term survival may systemically "fail to register", and as a result the entrepreneur has fewer alternatives to consider (March 1994; Simon 1997).

Too much seizing, on the other hand, may suggest deviations from firm-specific capabilities, i.e., that the firm is spreading itself too thin in the short-run in the pursuit of positive growth outcomes. Some scholars acknowledge that an unfettered propensity to change may result in declining performance outcomes (Schilke 2014b), and this is a particularly salient consideration for small firms which have limited resource to begin with. In addition, as argued by Spender (1989), industries often share a common set of knowledge, norms, and ways of doing things—or an "industry recipe". The recipe is abstract enough to be translated from one firm to another in the same industry, yet detailed enough to offer opportunities for vicarious learning.

Recent conceptual work by Teece et al. (2016) argues that agility is so in vogue as a management philosophy (e.g., by way of pivoting) that organizations operating in low to moderate uncertainty environments may be likely to over-emphasize the importance of agility and thereby stress resources that could otherwise be used for efficiency building and risk-reduction purposes. Even within highly competitive environments, Yuan et al. (2010) find that flexibility in the coordination of resources is a more important moderator than the development of new resources alone in the (positive) relationship between product innovation and firm performance.

One way to engage in proactive sensing activities is through R&D investments. Cohen and Levinthal (1990) remark that absorptive capacity helps a firm gauge the merit of information and knowledge that it may in turn exploit. Research outputs and intellectual property may be secured from alliance partners, accessed via informal contacts, or developed

in-house, with the intention of informing commercialization efforts. By investing in R&D, the small firm is not only able to stay abreast of exogenous scientific and technological changes but also promote the development of important inventions (Rosenberg 1990). Higher levels of R&D have been associated with small firms' abilities to invest early in the product development lifecycle, which is itself a type of dynamic capability (Stam and Wennberg 2009). Saemundsson and Candi (2017) suggest that identifying opportunities in new high-technology firms depends more on the adaptation of existing technical knowledge to known problem areas than on the identification of new customer wants and needs.

In sum, as we attempt to assess components of agility, we posit that greater activity in sensing through investment and engagement in R&D is likely to lead to improved small firm performance. For seizing, we advance a non-linear expectation. Too much seizing could be one way of deviating from common practices in an industry; that is, whenever firms look to each other to observe the competitive landscape, drastic change is more the exception than the rule. Thus, not enough change or too much change suggests an under- or over-extending of firm resources that negatively impacts firm performance.

3 Case study context: the green goods manufacturing industry

We use the green goods manufacturing industry as a case to study small firm agility because of (1) contemporaneous and rapid changes in policy and incentives (e.g., subsidies and other public investments) impacting the industry in the time period under investigation, i.e., 2008–2011; (2) the underlying dynamism of the technology and market; and (3) the role of local exploration and exploitation processes for small green goods firms.

Green goods firms can be defined as firms that produce manufactured items with environmental or natural resource benefits when used by other businesses, organizations or households (Shapira et al. 2013). Small firms in green goods manufacturing represent an important case, where the survival of small firms is closely tied to their abilities to adapt to rapid technological and market changes as well as government policy shifts (Patzelt and Shepherd 2011). In the policy domain, US energy policy during the George W. Bush administration saw an emphasis on hydrogen fuel and energy security and efficiency through the Energy Independence and Security Act of 2007. The Obama administration continued to support green goods development through a range of policies including R&D investments from the American Recovery and Reinvestment Act (ARRA) of 2009, of which 12% was allocated to the energy area under Title IV of ARRA to build infrastructure and conduct R&D in energy efficiency and renewables. In addition, the Obama administration oversaw the startup of the Advanced Research Projects Agency-Energy (ARPA-E), Energy Frontier Research Centers, Energy Innovation Hubs, and greenhouse gas emission reporting. A range of renewable and nuclear energy options were promoted, although issues such as the Solyndra bankruptcy filing and the Fukushima Daiichi nuclear accident were highlighted in public discussions. In recent years, targets of interest for investment in clean-tech have shifted from biomass to solar to wind to smart grid to water and waste management (PriceWaterhouseCoopers 2014). Since 2017, notwithstanding the Trump administration's efforts to backtrack from renewable energy and to revitalize coal production, US investment in renewable energy, clean transportation, and other green sectors has continued to increase (Bloomberg New Energy Finance 2019). These fluctuations form the backdrop highlighting the necessity of sensing and seizing by green goods firms if they wish not only to survive but also to grow.

The geographic focus of a firm's marketing and operations may play an important part in the influence of agility on growth. On one hand, after the economic downturn in 2008, the US government (as well as other governments) expressed interest in the growth of the green goods industry as a means to create new jobs in the economy. Because of the regulated nature of the broader energy sector, and due to the policy agenda of the US government to increase sustainable and green jobs, firms with increases in a local or domestic focus may be more likely to grow. That is, strong connections to local partners, suppliers, customers, and other stakeholders, who encourage information sharing, opportunity seeking, and geographic-based embeddedness, may positively enhance sensing and seizing activities and contribute to better performance (Becheikh et al. 2006; MacKinnon et al. 2002; Moulaert and Sekia 2003; Simmie 2005). On the other hand, the Brookings Institute found that most of these (green) jobs are in mature industries in manufacturing and public services rather than in new industries that could propel economic growth (Muro et al. 2011), and thus an increase in local ties would have limited effect on performance because presumably those local ties already existed and were being exploited to a sufficient extent.

4 Data and methods

Measuring deviations from a firm's existing course in real-time is challenging, though the need to do so is especially critical in high-velocity markets, where the fast-changing environment requires adaptation (Eisenhardt and Martin 2000). Feedback provided from managers and other respondents via surveys and interviews can be expensive and suffer from low response rates and panel attrition (Bartholomew and Smith 2006; Baruch 1999). Additionally, manager responses about dynamic capabilities and firm performance cannot be evaluated independently of one another because perceptions of performance and actions tend to reinforce each other (Laaksonen and Peltoniemi 2016). Such recall and response bias problems introduce measurement error when key explanatory variables are not observed directly and respondents selectively rationalize behaviors and decisions to conform to known performance outcomes and "conventional storylines" (March and Sutton 1997, p. 702). Finally, the phenomena under study—i.e., how firms position themselves for competitive advantage and what that looks like at any given time—often moves faster than traditional research designs involving primary data collections can track.

Computational social science approaches, including the use of website data and other forms of new media, present new opportunities for measuring firm change and attempts at shaping and responding to the environment. Firms use the web to give away or pilot digital products to gauge market demand; they develop online communities to promote a brand; and they create websites to share information and collect feedback (Choo 2006). Ashurst et al. (2012) show that ICT platforms are increasingly used to solicit feedback from stakeholders both within and external to the organization, while Fischer and Reuber (2011) find that social media furthers entrepreneurial sense-making processes because of its ability to facilitate a wide range of social interactions.

Websites, in particular, offer a view of a firm's seizing capabilities because websites may show changes in market orientation, product portfolio, and alliance networks (Arora et al. 2013, 2015; Li et al. 2016). Websites as information and marketing vehicles, may also show how small firms sense the market and portray their internal capabilities. For example, Gök et al. (2015) examine the use of research and development related terms on firm websites and determine that derivative website variables, while subject to some measurement

error, appear to convey R&D activities not captured in traditional sources (e.g., databases on firm R&D expenditure and grants and patenting and publishing outputs). Ebben and Johnson (2005) use website data to code flexibility and efficiency strategies for small Finnish firms, and when comparing these measures to survey-derived responses, they find convergent validity between measures derived from the two sources.

Using website data, however, is not without limitations, which include sampling biases and endogeneity between website change (as a proxy for evolution in firm capabilities) and growth. These limitations inform the research design and are further discussed in the concluding section of this paper.

4.1 Sampling strategy

The empirical analysis in this paper focuses on US small and medium-sized enterprises (SMEs) that entered the green goods sector with a year of establishment in the 2003–2007 time period. Concerning the definition of green goods, the formal US Bureau of Labor Statistical definitions of green industries were judged to be too broad, and green patents were judged to be too narrow. Instead, we used a multi-phase approach to identify green goods sector companies. The first phase began with a keyword based approach that was applied to the “line of business” field in the Dun & Bradstreet (D&B) Million Dollar database. [For a discussion on the use of D&B Million Dollar database in small firm research, see Carlton (1983), Brown and Medoff (2003), Neumark et al. (2011) and Reynolds (2017)]. More than 100 keywords were used to identify relevant companies in three green goods sub-segments: *environmental* (e.g., biological treatment, air pollution, environmental monitoring, land remediation, waste management, water treatment, recovery and recycling), *renewable energy* (e.g., wave and tidal, biomass, wind, geothermal, photovoltaic/solar), and *emerging low carbon* (e.g., alternative fuel vehicles, alternative fuels, electrochemical processes, batteries, carbon capture and storage, building technologies) (Shapira et al. 2013).

We applied additional filters for manufacturers with US headquarters, year of establishment in the 2003–2007 timeframe, and with 250 or fewer employees. Despite the year-of-establishment criterion, some older firms which re-incorporated between 2003 and 2007 appeared in the results. On further inspection of some of these re-incorporations, the re-incorporated firm exhibited a new orientation towards green goods. In total, this process yielded a frame of 2505 US enterprises. Manufacturers comprised 80% of these enterprises, but enterprises outside out of the primary North American Industry Classification System (NAICS) codes for manufacturing were considered because some of their offerings suggested green goods production.

Further work was undertaken to verify whether enterprises within this broad frame were valid green goods manufacturing firms. This second phase of population refinement involved the application of a four-point coding scheme by two different coders and reviewed by a senior researcher to each of these enterprises based on the text in the Dun and Bradstreet “line of business” field to determine the enterprise’s relevance to green manufacturing. Enterprises not involved in manufacturing or lacking a green product offering were removed from the list. This coding scheme reduced the number of enterprises to 62% of the original list. This coding scheme disproportionately affected manufacturers with primary NAICS classifications in the wood products, printing and furniture industries; these enterprises were more apt to be considered less relevant and removed from the list because although they produced resource-based goods, the goods were not relevant to green goods manufacturing. In addition, a group of nonmanufacturing enterprises in the

agriculture, mining, and construction industries were found to not manufacture goods and were removed from the list.

In a third phase, each of these enterprises was examined to determine whether or not it had a website. This criterion reduced the number of enterprises to 700. The addition of a website provided for more information about the enterprises, so another round of green goods relevance scoring was performed. Again, enterprises found not to be doing manufacturing or not having a green product offering were taken out of the list. In addition, those websites comprised of non-HTML pages (e.g., developed in Adobe Flash) were not included in the sample as our analytical software requires textual input. As a result, the sample was reduced to 298.

The proportions of enterprises across the primary NAICS codes were slightly different between the first phase and the second and third phases. As previously indicated, the second and third phases had proportionately fewer enterprises in the wood products/printing/furniture industries and in agriculture/mining/construction (Appendix Table 5). Likewise, the final phase of relevance coding emphasized more durables industries. Nevertheless, the total percentage of enterprises in manufacturing industry codes rose only slightly between the first phase list and the final phase: from 79% in the first phase/list to 86% in the final phase/list.

Next, firms with missing website data, e.g., such that a seizing (i.e., firm change) score could not be computed, were removed from the dataset. A total of 223 observations were available for analysis. Of these 223 firms, 78 were in the emerging low carbon subarea, 124 were in the environmental area, with the remaining 21 being classified as renewables. A weighting approach described below was used to account for over- and under-representation of certain firms in the 2SLS model.

As a final step to examine the representativeness of our sample, we compared the employment growth rate of our final sample of 223 firms with the population of private green goods firms, as defined by the US Bureau of Labor and Statistics (BLS). The rate of growth of private US green goods and services employment was 4.9% between 2010 and 2011. (Manufacturing employment growth during this same period was 2.9%.) Our sample shows an employment growth rate of 12.0%. Due to the smaller database that is used in this analysis compared to the BLS benchmarks, the mean employment growth for this database is skewed by a few outliers such as Tesla Motors, Inc (which started during our study period). The median employment growth rate for individual firms in our database is zero, and 73.5% of the firms have zero employment growth, with another 10.3% having negative employment growth over this period. Overall, the majority of the sample firms are at or below national green goods employment growth benchmarks.

4.2 Dependent variable

The dependent variable, derived from D&B, is *logged sales growth*. We computed growth, a continuous interval variable, by subtracting logged revenues in 2012 from logged revenues in 2010. Because of data quality and availability issues, in 30 select cases, sales growth was computed by subtracting 2014 revenue figures from 2012 revenue. Fifteen firms lacked sufficient sales data from D&B (in two distinct time periods) to determine sales growth and thus were dropped from the dataset. Four of these fifteen firms were acquired.

4.3 Website method and variables

We gathered data from company websites using a web crawling method adapted for social sciences (Arora et al. 2015). Websites have advantages (including accessibility, timeliness, and scope of data) and limitations (as unstructured company self-reports) for understanding innovation behavior (see consideration of limitations towards the end of this paper, and for further discussion of website mining advantages and limitations, see Gök et al. 2015). Company websites for each firm were extracted from 2008 to 2011 using the Wayback Machine. The Internet Archive, available at <http://archive.org>, is a repository of historical web pages saved periodically over time. By 2016 over its 20 year history, the Wayback machine held 273 billion webpages consisting of 361 petabytes (Goel 2016).

Leetaru (2015) documents two important considerations for researchers to be aware of in their use of the Wayback Machine. First, the Wayback machine adheres to “robots.txt” directives, which signal crawlers to exclude sites from automated web crawling. Consequently, firms with “robots.txt” exclusion directives are missing from the archive. Second, the Wayback machine is a decentralized architecture of crawlers, and there is no single policy as to how often or how comprehensively websites are crawled. Leetaru quotes Mark Graham, Director of the Wayback Machine, who comments, “I would expect any researcher would be remiss to not take the fluid nature of the web, and the crawls of the [Internet Archive], into consideration.”

Despite its limitations, the Wayback Machine has been used in other research contexts and, despite issues of lack of coverage of some websites in some years, has been found to produce information with content validity over time (Murphy et al. 2007). To ensure sufficient coverage and capture any updates, we recrawled (i.e., scraped again) any website with what appeared to be erroneous data, and as a result, many of these company websites in the Wayback Machine were completely crawled in 2012 and again in 2014. We also account for bias of missing data in our analysis, as described further below.

IBM Content Analytics (ICA)¹ was employed as a software tool for web extraction because of its ability to efficiently scrape websites, to allow the user to define keyword definition sets via parsers, and to deploy such parsers to collections of documents. After scraping was conducted, and parsing rules had been deployed on the dataset, ICA delivered data regarding the amount of times the words within the model appeared. XML files of these data were exported from IBM Content Analytics to a proprietary Java-based program, and the numbers of terms were summarized on a per-variable basis. Keyword-derived website variables include *R&D difference* and *local difference*. *R&D difference* captures the degree of research and development intensity found on firm websites in 2008–2009, while *local difference* reflects the change in local geographic focus between 2008–2009 and 2010–2011, relative to the firm’s headquarters location.

Each variable was normalized based on the number of pages collected during a given year (with duplicates excluded). Additionally, because Wayback coverage of firm websites can be sparse from year-to-year, despite our care in recrawling websites, we grouped these variables into two time periods, 2008–2009 and 2010–2011, by summing each normalized-year variable in the period and dividing by two.

The validity of these keyword-based variables is studied elsewhere. For example, the R&D keywords include “lab” OR “laboratory” OR “research” OR “development” OR “R&D” OR “researcher” OR “scientist”. Ongoing work suggests that this set of keywords is statistically

¹ Now IBM Watson Explorer Content Analytics.

albeit weakly correlated with patenting intensity using a non-parametric test of significance ($\tau=0.16$, $z=7.64$) (Arora 2019). Prior work by Gök et al. (2015) finds that various instantiations of R&D keyword combinations deployed on website data, while highly correlated with one another due to the overlapping key terms, are not statistically correlated with other non-website related measures, such as total number of publications, R&D expenditures and number of grants. That said, not all SMEs patent and even fewer publish, so measures of correlation cannot tell a complete story pertaining to the value of R&D keyword-based variables.

For *local difference*, our method successfully captures different intensities of geographic foci for a given SME; we assure validity for this measure by manually assessing selected firm websites from the distribution of scores in the two time periods, and from the distribution of *local difference* scores across both periods. For instance, several firms that show higher levels of local geographical presence in the second time period opened local offices in their respective metropolitan statistical area. In contrast, one firm, Tesla Motors, Inc., retained a negative local difference score; the firm developed a strong national and international focus on its website in the second time period, even as it advertised two locations in California in the second time period, but only one in the first.

The key variables of interest in this study are proxies for strategic sensing (*R&D difference*, as noted above) and seizing (*dpq*, as explained here). We operationalize seizing through estimating the topical change in website content from one period (2008–2009) to the next (2010–2011). Topical change on websites is an appropriate proxy measure for firm seizing, we argue, because of two reasons: (1) the extent to which a firm changes its topical focus reflects adjustments in the way it organizes, accesses, and markets its capabilities and/or pursues customers (we provide an example of this after explaining our method), and (2) marketing efforts, including those put forth on websites, are unlikely to deviate too much from a firm's actual capabilities to deliver to its prospective customer base, lest the firm suffer reputational harm. We review the endogenous relationship between topical change and firm growth below, and potential limitations of this approach in the discussion session.

We used latent Dirichlet allocation (LDA), a popular topic modeling algorithm, to identify latent topics in the website data. Topic modeling is based on probabilistic models of the generation of texts. Based on observed words in documents (in this case web pages), topic modeling uncovers a “hidden structure” to identify the probability distribution of words to K topics and the probability distribution of K topics to M documents in a corpus (Blei 2012). By estimating these distributions, it is possible to establish which words belong together to form topics, what mixtures of topics are found in documents, and which topics belong together in a broader corpus of firm-specific website data. The distribution of words and documents can be expressed as:

$$p(w|\alpha, \beta) = \int p(\theta|\alpha) \left(\prod_{n=1}^N \sum_{z_n} p(z_n|\theta) p(w_n|z_n, \beta) \right) d\theta \quad (1)$$

$$p(D|\alpha, \beta) = \prod_{d=1}^M \int p(\theta_d|\alpha) \left(\prod_{n=1}^{N_d} \sum_{z_{dn}} p(z_{dn}|\theta_d) p(w_{dn}|z_{dn}, \beta) \right) d\theta_d \quad (2)$$

where w and D represent words and documents respectively, θ represents a mixture of topics z_n , N the number of words in a document, M the number of documents in a corpus, and β is a matrix of probabilities indicating the probability that the j th word belongs in the i th topic. Finally α is a parameter defining the distribution of the mixture of topics. In LDA,

Fig. 1 Technology architecture for operationalizing the firm change variable

the distribution $p(\theta|\alpha)$ is chosen to be a Dirichlet distribution with shape given by the vector parameter α (Blei et al. 2003).

To prepare the website data for topical analysis, we followed a number of cleaning and preparation steps (Fig. 1). First, as with the keyword variables, duplicates were removed, along with any non-English textual content. We also removed punctuation characters, company names and company urls, the latter of which were found to bias the topics around company-specific content. At this point, LDA was trained on 2008–2009 website data to develop baseline probability distributions of (a) latent topics-to-web page documents and (b) observed words to web page documents. We employed $K=100$ topics over 1000 iterations with smoothing parameters $\alpha=50/K$ and $\beta=0.01$. We refer the reader to Steyvers and Griffiths (2007), who discuss this parameter setting in producing robust LDA results.

Our overarching goal is not to isolate and present specific topical content; rather, we are interested in a broad measure of topical change over time as a proxy for seizing. For example, small and medium-sized firms report the breadth of their product portfolio, the geographic scope of their operations, and their approach to investor relations, etc. through their websites, and this content changes over time, thus conveying some degree of strategic change. We measured change by creating firm-level documents for the time periods 2008–2009 and 2010–2011, and then using the trained LDA model, we inferred the latent

probability distribution of topics to those single firm-level documents. A firm-level document contains all the firm web-pages for the specific time period (excluding duplicates), aggregated in one record.

An asymmetric firm-level measure captures the extent to which the probability distribution of inferred topics in 2008–2009 differs from the inferred probability distribution of topics in 2010–2011 (Steyvers and Griffiths 2007). It is calculated as follows, where D measures the distance between the topical distribution vector p for a given firm i in 2008–2009 and the topical distribution q for the same firm in 2010–2011:

$$D(p, q) = \sum_i^{K=100} p_i * \log_2 \left(\frac{p_i}{q_i} \right) \quad (3)$$

Identifying topics is a generative (i.e., bottom-up) process based on observed words. Hence, there is no guarantee that topical order and underlying content will remain consistent between one run and the next. To safeguard against this potential bias in measurement error, we ran the training and inference process (bottom right portion of Fig. 1) 30 times and used the average of the information distance measure as the sample observation for firm sensing and seizing (Canini et al. 2009). While topical content and order will change from one run to another, we contend that this approach deemphasizes outlier observations by incorporating a greater number of distance estimates. The resulting variable representing the extent of firm change is signified by dpq . The squared value is denoted as dpq squared (distance between p and q squared) in the ensuing models.

To better illustrate the intuition behind our measure of firm change via website topical content, we present descriptive results of one LDA run. In this run, 99 of 100 topics have face validity in that a human can easily review the top most frequently occurring words belonging to a topic and offer a label or two to summarize that topic. We exemplify the labeling and its relationship with the underlying words through presenting four illustrative topics: “wastewater treatment”, “engineering systems”, “technology solutions”, and “toxicity and hazardous effects”, chosen to show the breadth of topics captured by LDA (Table 1).

To exemplify how topic modeling and our information distance measures capture firm change, we describe the evolution of two firms’ websites with respect to the four sample topics in Table 1. Upon founding, one Colorado firm specializing in water treatment solutions began servicing a nearby industrial company in natural gas. As the Colorado firm grew, it expanded its product and service offerings to the food and beverage industry, which also has a need for water purification. As a result, the firm’s website as a whole reduced emphasis on industrial water treatment (topic 0) in favor of a broader portfolio of technical and service solutions (topics 9 and 49). The Colorado firm’s dpq score (3.071) for this LDA run was well above the sample average (1.303).

In contrast, a Pennsylvania firm offering fifteen non-toxic cleaners, degreasers, and washers had a dpq score (0.994) below the sample mean, indicating that the firm’s website content changed very little between the two time periods. However, the LDA inference process was still able to isolate small changes in topical content. For example, in 2008 this firm produced nine material safety data sheets for its fifteen products. In 2010, the company added material safety data sheets for all fifteen products. Because each data sheet stresses the non-toxic and non-hazardous profile of each product, the probability of the toxicity and hazard related topic (64) occurring increased in the second time period. The Pennsylvania

Table 1 Sample topics (4/100) with top ten words trained on all firm website data, 2008–2009, from one run of LDA

Topic 0 “wastewater treatment”	Topic 9 “engineering services”
Treatment	Experience
Wastewater	Engineering
Water	Design
System	Work
Bacteria	Skills
Sludge	Management
Systems	Project
Waste	Team
Solids	Generator
Grease	Engineer
Topic 49 “technology solutions”	Topic 64 “toxicity and hazardous effects”
Products	Section
Solutions	Material
Customers	Water
Quality	Product
Product	Information
Technology	Skin
Systems	Contact
Applications	Hazardous
Design	Safety
Services	Data

The LDA model training process was repeated to train 30 distinct LDA models, which were run in inference mode to produce topical distributions on 2010–2011 data

firm did not seize any net-new business opportunity, and thus its *dpq* score was low, less than 1.0. Moreover, whereas the Colorado firm experienced sales growth between the first and second time periods, the Pennsylvania firm did not.

4.4 Instruments

Topical website change is likely endogenous in the sales growth model.² Profitable firms are not only more likely to have the capital on hand to explore and exploit new business opportunities (Fauchart and Keilbach 2009) but also may have more resources to invest in website development, which may introduce new topics in self-disclosed narratives. To address this endogeneity, we instrument on topical change (both *dpq* and *dpqsq*) using website variables that are correlated with website topical change over time but uncorrelated (or weakly correlated) with sales growth. These instruments include: (1) *year site appears*,

² We also consider whether the other website variables, *R&D difference* and *local difference*, are endogenous in the sales growth model. Our tests for endogeneity show this is not the case.

or the year the site first appears on the Wayback Machine; (2) *logged words difference*, or the difference between the number of words observed in each time period, 2008–2009 and 2010–2011, logged; and (3) *redesigns*, the number of site redesigns since 2004. A *site redesign* is defined as a significant change in the color, imagery, and style presentation of a firm website from 1 year to the next. A site redesign may also incorporate functional enhancements, e.g., adding a substantive capability such as an e-commerce feature. Our review of website redesigns show primarily incremental, aesthetic changes in the final sample of websites (Appendix Table 6). This variable is coded manually.

The rationale behind these instruments is as follows: While firms may change the presentation of their website content over time (e.g., to look more contemporary), this should not significantly add to the long-term operating expense of running such sites; therefore, the number of site redesigns, controlling for the year the site appears, should not correlate with the structural error term. At the same time, a redesign *likely* introduces some new topics as the firm rebrands itself (e.g., as it seeks to find new markets for its existing capabilities and complementary assets). We interact, *redesigns* with *year site appears* to capture the possibility that sites with longer histories could be more likely to experience redesigns, and together this interactive relationship could instrument for topical change. Similarly, adding new words to a website is a near costless activity; topics may be added or removed as a result, but such an endeavor is unlikely to correlate with the error term in such a way as to explain sales growth. Refer to the correlation table in the results section for confirmation that the instruments appear to pass face validity. Results of more robust statistical checks are also reported below.

4.5 Controls

We collect from D&B data on employee growth from 2010 to 2012 (*logged employee growth*). A second control variable captures environmental dynamism: the change in industry-level gross output in the second time period (i.e., from 2011 to 2010) (*logged industry growth*). Put another way, this variable controls for broader economic growth in these industries. To calculate this value, we map a firm's primary North American Industry Classification System (NAICS) code, as obtained from D&B, to US Bureau of Economic Analysis yearly gross output reports. We include this variable because the literature on dynamic capabilities emphasizes differences in rate of industry growth with consequential implications for firm-level growth (Schilke 2014a). Finally, we include two dichotomous industry controls for *environmental* and *renewable energy*, with *low carbon* as the reference group.

4.6 Model identification

We use a 2SLS regression model to account for endogeneity between topical change and sales growth. Formally, we estimate the reduced form as follows:

$$\begin{aligned} \log(\text{sales growth}) = & B_0 + B_1 dpq + B_2 dpq^2 + B_3 \text{local difference} \\ & + B_4 R \& D \text{ difference} + B_5 \log(\text{industry growth}) \\ & + B_6 \log(\text{employees growth}) \\ & + B_7 \text{environmental} + B_8 \text{renewable energy} + u \end{aligned} \quad (4)$$

where u is the error term. In the fully specified 2SLS model, the first stage regresses topical change (both dpq and $dpqsq$) on all exogenous variables, including the three instruments

noted above. The second stage regresses *logged sales difference* on the estimated parameters of *dpq* and *dpqsq*, as well as *R&D difference*, *local-diff*, and the control variables.

5 Results

Descriptive statistics and pairwise correlations are shown in Table 2 for all 2SLS variables. The average number of employees in 2010 was 104.63, with the largest firm employing 3061 people. (Note that all firms were classified as being small in 2003–2007 but may have grown, sometimes by a substantial number in subsequent years.) The mean number of pages per firm site in 2008–2009 was 40.87 (min. = 1 and max = 830), while in 2010–2011, the mean was 47.39 (min. = 1 and max = 1467). The seizing proxy website variable, *dpq* is always positive, with some firms exhibiting large changes in website content and others showing relatively little variation. As shown in Table 2, there are no strong pairwise correlations variables outside of the relationship between *dpq* and *dpqsq*, although some correlations are significant at $p < 0.05$, e.g., as is the case with the positive relationship between *logged employee growth* and *logged sales growth*, as well as the positive relationships between the instruments, *redesigns* and *logged word difference*, and the topical change variables, *dpq* and *dpq squared*.

Before examining the results for substantive findings, we evaluated the merits of our instrumental variable approach using traditional tests of partial correlation, endogeneity and over-identification. For model 4 presented below, the first stage in a 2SLS model with multiple instruments can be used to assess $\text{Cov}(z, x) \neq 0$. An F-statistic in the first stage of a 2SLS regression of greater than seven indicates strong instruments. In models A1 and A2 (Appendix Table 7), we reject the hypothesis H_0 : all instruments = 0 at the 95% confidence level, though the F-statistics (3.461 for *dpq* and 4.565 for *dpqsq*) suggest moderate, not strong, instrumentation.

The coefficient on *redesigns* is positive and highly significant for both models A1 and A2 with *dpq* and *dpqsq* as the dependent variables, respectively, suggesting that each additional redesign introduces some amount of topical change to a website's content, holding all other variables constant. Contrary to our intuition, however, the coefficient on *year site appears* is negative but only marginally significant in A1. (This variable is not significant in A2.) The interaction term between *year site appears* and *redesigns* is highly significant in both models, and the coefficient is negative, implying that the effect of each additional redesign decreases with firms whose sites appear relatively later vis-à-vis firms whose sites are older. The effect of *logged words difference* is positive but insignificant in Model A1 and positive and highly significant in A2.

Tests of endogeneity reject H_0 : instruments are exogenous in the Durbin test ($p = 0.012$) and in the Wu–Hausman test ($p = 0.014$) at $\alpha = 0.05$. Lastly, testing for overidentification shows large p values ($p > 0.10$) for the Sargan and Basman tests; consequently, we cannot reject the null hypothesis that the instruments are valid and the model is correctly specified. Overall, we have sufficient evidence that *dpq* and *dpq squared* are in fact endogenous and that the instruments, although not as strong as we would like, serve appropriately for estimating *logged sales growth*.

We now turn to the substantive modeling results. 2SLS model specifications 1–4 in Table 3 reflect an increasing number of variables in their respective specifications. Model 5, as discussed below, introduces a weighting scheme as a robustness check on top of the model 4 specification.

Table 2 Pairwise correlations and descriptive statistics, $n = 223$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	Mean	SD
Logged sales growth (1)	1.00												0.11	0.98
dpq (2)	-0.02	1.00											1.58	2.28
dpq squared (3)	-0.05	0.89*	1.00										7.68	31.97
<i>Redesigns (4)</i>	0.15*	0.16 [†]	0.13 [†]	1.00									0.67	0.71
<i>Site year appears (5)</i>	0.20*	0.04	-0.03	0.02	1.00								2003.12	3.55
<i>Logged word diff (6)</i>	-0.01	0.14 [†]	0.17 [†]	0.08	0.17*	1.00							0.22	1.84
Logged industry growth (7)	0.02	0.11	0.08	-0.02	0.06	0.12 [‡]	1.00						7.40	2.16
Logged employee growth (8)	0.35*	0.03	0.03	0.11	0.05	0.02	-0.04	1.00					0.10	0.61
Environmental (9)	-0.10	-0.13 [‡]	-0.12 [‡]	-0.16*	-0.07	-0.11	-0.07	-0.05	1.00				0.56	0.50
Renewable energy (10)	0.13 [‡]	0.10	0.14*	0.17*	0.18*	0.06	-0.15	0.03	-0.36*	1.00			0.09	0.29
R&D difference (11)	-0.12 [‡]	0.08	0.09	-0.01	0.05	0.12 [‡]	0.03	-0.08	0.07	-0.03	1.00		0.06	1.09
Local difference (12)	0.18*	-0.05	-0.05	0.04	0.10	0.13 [†]	0.03	0.06	-0.12 [‡]	0.18*	0.08	1.00	0.13	1.08

Instruments shown in italics

*Significant at $\alpha = 0.01$; [†]Significant at $\alpha = 0.05$; [‡]Significant at $\alpha = 0.10$

Table 3 2SLS results on logged sales growth, n=223

	(1)	(2)	(3)	(4)	(5)
<i>dpq</i>	1.03** (0.49)	1.13** (0.53)	1.06** (0.50)	0.94** (0.45)	0.72* (0.43)
<i>dpq squared</i>	-0.06* (0.03)	-0.07* (0.03)	-0.06* (0.03)	-0.06** (0.03)	-0.04 (0.03)
<i>Logged industry growth</i>		-0.04 (0.05)	-0.03 (0.05)	-0.02 (0.05)	-0.00 (0.04)
<i>Logged employee growth</i>		0.54*** (0.16)	0.50*** (0.16)	0.51*** (0.15)	0.42* (0.24)
<i>Local difference</i>			0.18** (0.09)	0.16* (0.09)	0.15* (0.08)
<i>R&D difference</i>			-0.10 (0.09)	-0.10 (0.09)	-0.11 (0.07)
<i>Environmental</i>				0.08 (0.21)	0.09 (0.19)
<i>Renewable energy</i>				0.43 (0.37)	0.46 (0.33)
Constant	-1.05* (0.55)	-0.95* (0.55)	-0.90* (0.52)	-0.96* (0.54)	-0.86* (0.47)
Observations	232	223	223	223	223
Weights applied	No	No	No	No	Yes
<i>df</i>	2	4	6	8	8
Wald χ^2 (<i>df</i>)	4.83	16.48	21.51	25.28	21.66
Prob > χ^2	0.089	0.002	0.002	0.001	0.006

Standard errors in parentheses

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

In Model 1, *dpq* is significant and positive ($p < 0.05$) with a coefficient of 1.03, and *dpq squared* is marginally significant and negative ($p < 0.10$) with a coefficient of -0.06, suggesting that topical change as an endogenous proxy to firm seizing is associated with the outcome variable, *logged sales growth*, in an inverse U-shaped fashion. In other words, a one-unit increase of topical change on websites is associated with a sales growth increase, but this increase declines with increasing levels of topical change.

Model 2 adopts the time varying control variables. *Logged employee growth* is highly significant at $\alpha = 0.01$: a 10% increase in a firm's number of employees results in a 2.1% increase in firm sales growth. *Logged industry growth* does not impact firm sales growth. In Model 2, the coefficient on *dpq* becomes significant ($p < 0.01$) and goes up to 1.13, while *dpqsq* remains marginally significant and negative. The whole model becomes significant at Wald $\chi^2(4) = 16.48$, $p = 0.002$.

Model 3 includes the *R&D difference* variable, but it is not significant. Model 3 also introduces the *local difference variable*, which shows a positive and significant coefficient (0.18, $p < 0.05$): A one-unit increase in the difference in local mentions by a firm vis-à-vis its headquarters is associated with an 18% increase in firm sales growth. Changes in localization intensity appears to capture some of the explanatory power of the firm seizing proxy variable (i.e., *dpq*); removing *R&D difference* from Model 3 does not impact the magnitude of

Fig. 2 Using *dpq* to predict logged sales growth while holding other variables at their means

coefficients *dpq* and *dpqsq* in Model 2. While these results are not conclusive, it appears that *dpq* and *local difference* measure some aspects of firm seizing. Model 3 as a whole remains highly significant.

Model 4 is the fully specified, unweighted 2SLS model, and it is significant at $\alpha=0.01$. All explanatory variables of interest are (marginally or better) significant in the same direction as mentioned above. For example, *dpq* and *dpqsq* are significant at $p < 0.05$, though *local difference* is only marginally significant at $p < 0.10$. In Fig. 2, we show a graphical depiction of the curvilinear relationship between *dpq* and *dpq squared* while holding all other variables in M4 at their means. This curve shows that most firms in our dataset are likely to experience increasing levels of firm sales growth with increasing amounts of website topical change until a *dpq* score of between 8 and 9; thereafter a select few with very high website change scores are predicted to experience lower relative sales growth. Removing firms with very high-levels of website change produces model results where the (endogenous) relationship between firm seizing and firm growth appears linear in nature.

To account for observations with missing values, we introduce an inverse probability weighting scheme to estimate the likelihood of being included in the final sample of 223 firms (c.f., Solon et al. 2015). Recall that some of the 298 firms did not have sufficient website data over the two-time period span to be included in the sample, while others were missing other data derived from D&B. We include this analysis as a post hoc robustness check addressing potential sample bias and to illustrate the representativeness of website data more generally for small firm innovation studies. We estimate the log-odds of being in the final sample of 223 firms vis-à-vis the initial set of 298 firms as:

$$\ln \frac{\Pr(\text{in sample})}{1 - \Pr(\text{in sample})} = B_0 + B_1 \log(\text{employees}) + B_2 \text{year firm founded} \\ + B_3 \text{out of business} + B_4 \text{northeast US} \\ + B_5 \text{midwest US} + B_6 \text{southern US} \quad (5)$$

Table 4 Logit results for probability of being included in the final sample, a dichotomous variable

Variable	Odds ratio	SE	z	p > z
<i>Logged employees</i>	1.12	0.11	1.19	0.23
<i>Year firm founded</i>	0.95	0.03	(1.49)	0.14
<i>Out of business</i>	0.03	0.03	(3.32)	0.00
<i>Located in northeast</i>	0.55	0.26	(1.28)	0.20
<i>Located in Midwest</i>	0.31	0.14	(2.63)	0.01
<i>Located in South</i>	0.23	0.09	(3.73)	0.00
Constant	4.47E+47	3.23E+49	1.52	0.129

75 out of 298 firms have insufficient data to be used in the 2SLS model, n = 298

Logged employees is the number of employees in 2010; *year founded* is the year the firm was established; and *out of business* is coded 1 if the firm experienced a market exit in the time period under study, 0 otherwise. Finally, a firm may be found in any one of the four main US Census Bureau defined regions: *Northeast*, *South*, *Midwest*, and *West*. The reference group for region is *West*.

Results of the logit model (Eq. 5, Table 4) suggest that weighting is necessary to account for bias in the sample. The odds ratios suggest that firms that have gone out of business are less likely to be represented ($p < 0.01$). Additionally, firms in the Midwest and South are much less likely to be in the final sample than firms in the West ($p < 0.01$). These findings suggest systemic differences in the likelihood of the existence of firm websites across the sample. Further, this result points to a clear need to better understand differences in firm website publishing behavior, a topic to which we return in the discussion below.

Comparing Model 4 with Model 5 in Table 3, which includes probability weights of being included in the final sample, we see that the linear firm change measure of dpq remain positive and significant, but only marginally so. The coefficient on $dpqsq$ is negative and no longer significant, while the coefficients for *local difference* and *logged employee growth* remain positive but turn marginally significant.

6 Discussion and conclusions

This paper has put forward a methodological approach for operationalizing the concept of firm agility and examining the relationship between agility and growth outcomes. We build a measure of strategic change through unstructured text mining of firm websites using topic modeling based on LDA. To construct this measure, we collect archived website data for US green goods small manufacturers from 2008 to 2011 using the Wayback Machine and using web scraping methods (Arora et al. 2015). With this relatively novel source of data, we overcome an important barrier in studying small firms: a dearth of data. Among the small green goods manufacturers in our sample, only 20% had any scientific publications or patents, preventing us from constructing any indicators based on this type of public record (Li et al. 2016). Business databases have limited, if not inconsistent, records for these small firms as well. On the contrary, as shown through this work, we are able to learn a great amount about firm activities and strategies from their websites.

The exploratory findings suggest that strategic change (or seizing), as measured by firm website topical change over time, has a curvilinear, inverse U-shaped relationship with

sales growth. This is consistent with the expectation indicated following our literature review. We take into account the problem of endogeneity between topical change on firm websites and sales growth by instrumenting topical change with several variables characterizing the firm websites, which are, by definition, correlated with topical changes but not with sales growth. Additionally, increasing levels of local presence as emphasized on firm websites is associated positively with sales growth. In contrast, changes in sensing, as measured by differences in R&D levels between the first time period and the second, do not appear to have an impact on sales growth outcomes. Although, as discussed below, there are methodological limitations related to this finding, it could also be that a longer period of elapsed time (than measured in our study) is needed for firms to see business results from prior R&D activities.

Our use of website topical change as a proxy for firm change captures aspects of both ordinary and dynamic capabilities (Laaksonen and Peltoniemi 2016). The single time-period view of topics comprising a firm's website conveys substantive, self-reported capabilities ranging from the firm's product and service portfolio, and marketing approach (e.g., use of science- and technology-based language vis-à-vis verbiage focused on applications and industries), often grounded in a firm's history and personnel. The longitudinal view of these firm websites and how they change over time may reveal higher-order dynamic capabilities at play, i.e., the extent to which firms are able to modify the substantive routines that may result in enhanced performance outcomes.

However, not all firms are willing and able to invest in dynamic capabilities. Winter (2003) argues that firms may strategize on an ad hoc basis if the cost–benefit analysis does not show a defensible reason for building and maintaining dynamic capabilities. In this sense, high topical change scores could be the result of either substantive changes as a result of seizing or ad hoc strategizing. In any case, larger scores indicate substantial deviations from the firm's previous capabilities (or at the very least, the way the firm marketed itself), while smaller scores suggest less (or even a lack of) change.³ Future work is needed to refine variables that the current topical change measures capture but may not sufficiently isolate. With our topical change measure, our approach has the advantage of allowing for inter-firm comparison across a variety of firm behaviors and activities that comprise what we put forth as strategic movements. Yet, this probabilistic model-based measurement conceals the details of the actual seizing activities. Whether the changes reflect product/process iterations, realignment of market orientations, or changes in product portfolios, these specific activities are not reflected in the final measures, and thus we cannot identify them in our regression models. Admittedly, this limitation has constrained our ability to link the results to a specific context and interpret its policy implications.

Even so, we contend that the need for multiple observations per firm over a longer period of time may enhance the study of strategic and organizational change, and we argue that website and other new media data are well-suited for this purpose. Because agility is an increasingly important driver of competitive advantage (Teece et al. 2016), and because of the aforementioned dearth of data on small innovative firms, researchers need more cost-effective, richer, and unobtrusive sources from which to study the evolving nature of innovation and decision-making. Findings from this line of work, including future studies, will better inform practitioners and policymakers about the context and boundary conditions

³ A lack of observed change does not necessarily indicate the absence of dynamic capabilities; managerial intention in the presence of no or little change is needed to definitively comment on dynamic capabilities (Arend and Bromiley 2009).

that make such agility valuable and necessary, given specific industry and macro-economic contexts.

We illustrate our position through the prism of pivoting (O'Connor and Klebahn 2011). This is a popular and influential idea in business schools and among practitioners who claim that the “secret sauce” for successful startups such as Groupon, Pay Pal, and Twitter is decisive, rapid and early shifts in business strategies based on frequent engagement with customers. Through these pivots, startups reach out to customers in early stages of product development and make deliberate as well as often dramatic changes to the business model in response to customer feedback received. To some extent, the idea of pivoting is an antidote to technophile-dominated startups where the teams spend much of their time “in the laboratory” instead of interacting with real users. Thus, a crucial task of startups is to learn “what customers really want” through pivots, which is echoed by Ries in his popular *The Lean Startup* (2011). However, the concept of pivoting, while appealing to entrepreneurs, is not without criticism. McGinn (2012) argues that an overemphasis on pivoting might encourage a culture of opportunism, in which people starting new companies engage in “hopscotching” between widely different ideas. As McGinn points out, the latest trend of pivoting might overlook passion, commitment and sustained efforts, which are well recognized as key qualities of successful technology entrepreneurship (Stevenson 1983).

The pivoting concept has clear ties to organizational theory: in dynamic capabilities parlance, the pivot is essentially an iterative process of sensing and seizing to manage threats proactively and capitalize on new business opportunities (Teece et al. 2016). Zahra et al. (2006) propose that over time new ventures “calcify” meta routines (i.e., dynamic capabilities facilitating change) through a progression from ad hoc improvisation to trial-and-error to eventual resource-intensive experimentation.

Experimentation is increasingly seen as a viable option as new ventures, especially those that are software based, adopt lean principles emphasizing hypothesis testing and customer-driven approaches to product development (Humble et al. 2015). The green goods manufacturing industry is notably different than software based industries that are amenable to extreme rapid cycle iterations and frequent pivoting, however (Humble et al. 2015). We wish to highlight three distinctions here: First and foremost, we posit that the timeframe between noticeable changes in any manufacturing sector, at least as observed on firm websites, is not on the order of weeks (or sprint cycles in software) but rather months or years: tangible goods, especially those of a green goods nature, necessitate relatively longer lead times for changes in technology, supply chain coordination, and environmental practices (Chung and Wee 2011). Secondly, the consequences of any new pivot are likely more important in green goods manufacturing than in software or Internet related industries, which experience lower fixed costs, potentially less liability for faulty products, and less costly channels for repair.

Our study design addresses change in two periods and does not explicitly reflect the role of customer input. Yet, in line with these cautionary tales, our primary results suggest that there are two kinds of hazards to SMEs operating in uncertain and fluctuating technology-intensive markets such as the green goods manufacturing sector. The first hazard is the lack of ability, willingness, or resources to invest in seizing activities. If these entrants lack the flexibility to experiment with and make changes in response to technological or market fluctuations, they will experience limits in their ability to grow. The second hazard stems from an overemphasis on strategic changes. The risk here is that growing firms become locked-into the notion of change for change's sake.

Can websites be used to study agility and rapid cycle pivoting? Despite its popularity, the pivoting concept has not been deeply analyzed in business and management literature.⁴ We suggest the answer is yes, but with important caveats. In this work, we used a sample of 223 US-based green goods SMEs and their archived website information to explore the relationship between strategic change and the growth of innovative small manufacturing firms. Taking advantage of recent developments in information technologies including web scraping and topic modeling, our study develops a measure of strategic change by quantifying the evolution in firm website topics. We showed that this measure is reasonably robust in detecting changes in firm behavior.

However, the pivot is not all about seizing; it is sensing, too. We used changes in R&D keyword mentions as a proxy of sensing, but the pivoting concept focuses more on customer feedback processes than absorptive capacity. As a result, it is then necessary to consider how a firm's customer feedback channels can be detected in the website data collection and variable operationalization process. While there may be some pointers to such channels on the website itself (e.g., in the topics that emerge), we posit that websites themselves may already be fine-tuned to elicit customer feedback and interaction as marketing vehicles. The point of hypothesis-driven testing in the pivoting framework is to solicit feedback and improve value capture of products and services (Humble et al. 2015), and we posit that every time a website substantially changes, it could be a way of improving the customer experience. Alternatively, it could be the case that websites are not able to determine sensing activities through customer channels and thus additional data sources (e.g., such as survey instruments) will be needed.

There are other limitations associated with using company websites as an information source. Websites are essentially self-reported by companies and have not been validated in a consistent way by third party data providers. One issue is that firms differ in what they publish on their sites. To some extent, variations may represent sub-sectoral differences. For example, biotechnology firms may be more apt to report information on patents, publications, and clinical trials while firms in the energy domain may be more apt to report information about customers and partnerships (Youtie et al. 2012). Variations may also exist over time; for example, at one point a firm may desire high visibility to attract visits from potential funders or customers while at another time, the firm may seek to "go dark" and reveal very little information while working on a new application or trade secret (Arora et al. 2012). The variation in information on a website over time is one reason that we re-scraped websites of the green good SMEs in our dataset and normalized results over multiple years. A second issue is that information embedded in images, PDFs, or Flash displays are not captured by our approach, which requires HTML text-based data in order to process. We manually reviewed the websites we scraped and found that non-textual information was not extensive for the SMEs in our dataset, but this type of checking should be performed whenever website data are used. In sum, we regard web scraping (subject to manual verification of websites, c.f. Arora et al. 2015) as a value-adding source particularly when data from official statistics, scholarly publications, or other patent and business information sources are lacking.

Within the confines of these limitations, we position our contributions through an original, systematic, and quantitative analysis of a conceptualization of strategic change and

⁴ An exception is Marx et al. (2014) on new entrants in the voice recognition market, where successful startups pivot their strategies from competition in standalone applications to cooperation with innovation partners (such as Smartphone makers) in bundled offerings.

associated small firm behaviors using website data. Our results have research, managerial, and public policy implications for promoting entrepreneurship. The narratives that emerge on websites are self-determined and updated on firm-specific terms and schedules. As a result, we are able to ask questions about the entrepreneurial narratives that emerge (Autio et al. 2014; Zott and Huy 2007). This is an approach that could be used to probe other relevant entrepreneurial research topics, for example, how the adoption of new media channels, including websites and social media, influences a firm's ability to access complimentary resources, attract and retain talent, and survive and grow. A common feature of this line of research is the use of data-intensive methods to analyze enterprise behaviors exhibited online, with the mapping of these behaviors to existing or new types of entrepreneurial business models and innovation patterns. In addition to web mining, other accessible platforms and sources (such as LinkedIn or Twitter) present themselves as sources of big data for business and innovation analyses. As our study illustrates, careful attention must be given to the reliability of such sources. Consideration of data protection and privacy is a further prerequisite. Nonetheless, as sources of big data grow both in scope as well as scale, we anticipate ongoing value for entrepreneurship research in finding ways to use these sources to understand increasingly dynamic business environments and firms.

Acknowledgements This work was supported by the UK Economic and Social Research Council (Grant No. ES/J008303) and the US National Science Foundation (Grant No. 1646773). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors, and do not necessarily reflect the views of the Economic and Social Research Council, the National Science Foundation, or Ernst & Young, LLP.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix

See Tables 5, 6 and 7.

Table 5 Industry distribution of enterprises in selection phases

Industries	NAICS code(s)	First phase: original list (%)	Second phase: first round of relevance coding (%)	Third phase: has a website (%)	Fourth phase: final round of relevance coding (%)
Food and beverage manufacturing	311, 312	0.7	0.7	0.4	0.0
Textile, leather manufacturing	313, 314, 315, 316	0.6	0.7	1.0	0.0
Wood, paper, printing, furniture manufacturing	321, 322, 323, 337	11.7	6.3	5.9	0.7
Chemicals, petroleum manufacturing	324, 325	22.0	33.3	30.3	18.0
Rubber, plastics, nonmetallic minerals manufacturing	326, 327	1.8	1.3	1.3	0.7
Metals manufacturing	331, 332	3.6	2.1	1.6	1.3
Machinery manufacturing	333	15.1	16.8	18.3	30.7
Computer manufacturing	334	7.8	4.5	6.6	11.0
Electrical appliance manufacturing	335	12.4	15.0	15.9	21.3
Transportation manufacturing	336	2.4	1.2	0.9	1.3
Other manufacturing	339	1.3	1.2	1.3	1.0
Wholesale and retail	400–488	4.4	5.2	5.9	3.7
Services	511–924	4.8	4.7	6.0	8.7
Other nonmanufacturing	112–238	7.6	2.2	1.7	1.7
No NAICS code in list		3.9	4.7	3.1	0.0
		100.0	100.0	100.0	100.0

Table 6 Types of websites redesigns for a selection of firms in the samples

Firm name (anonymized)	Total redesigns	Incremental redesigns	Qualitative description of the redesign(s), including any functional and incremental changes
AB01	3	3	Incremental changes only, limited to new banners, color schemes, and menu
CD02	3	2	One functional change in 2009 adding a distributors log-in feature plus a toll free number for sales; in 2010 and 2011 two additional incremental changes with removal of distributors log-in feature
EF03	3	2	Incremental redesigns in 2009 in 2011. Addition of web ordering feature in late 2011 (otherwise website remains the same)
GH04	2	2	In 2008, firm already maintained an e-commerce capability; incremental redesigns in 2010 and 2011
IJ05	2	2	In 2008, firm already maintained an e-commerce capability; incremental redesigns in 2009 and 2011
KL06	2	2	Incremental redesigns in 2009 in 2011 with new color schemes and layout (2011 only)
MN07	1	1	Barebones site enhanced in 2010 with new layout, logo, and color scheme
OP08	1	1	Barebones site enhanced in 2010 with new layout, logo, and color scheme
QR09	1	1	Incremental design in 2011 with new color scheme, layout, and logo
ST10	0	0	No substantive change to home page
UV11	0	0	No substantive change to home page
WX12	0	0	No substantive change to home page

Firms were chosen randomly within each number of total redesigns

Table 7 First stage results on endogenous variables *dpq* and *dpqsq* for model 3 presented in Table 3, $n = 223$

	A1 <i>dpq</i>	A2 <i>dpqsq</i>
<i>Logged industry growth</i>	0.12* (0.07)	1.42 (0.98)
<i>Logged employees difference</i>	0.09 (0.24)	1.76 (3.39)
<i>Local difference</i>	−0.20 (0.14)	−2.86 (1.96)
<i>R&D difference</i>	0.18 (0.14)	2.87 (1.92)
<i>Environmental</i>	−0.38 (0.33)	−4.42 (4.52)
<i>Renewable energy</i>	0.61 (0.57)	15.25* (7.90)
<i>Redesigns</i>	350.92*** (128.98)	5722.60*** (1783.26)
<i>Year site appears</i>	0.10* (0.06)	0.86 (0.79)
<i>Redesigns * year site appears</i>	−0.17*** (0.06)	−2.85*** (0.89)
<i>Logged word difference</i>	0.12 (0.08)	2.33** (1.16)
<i>Constant</i>	−198.43* (114.07)	−1730.09 (1577.07)
R-squared	0.107	0.128
F(4, 212), H_0 : all instruments = 0	3.461	4.565
Prob > F	0.009	0.002

F statistic below tests whether all instruments = 0

Standard errors in parentheses

*** $p < 0.01$; ** $p < 0.05$; * $p < 0.1$

References

- Arend, R. J., & Bromiley, P. (2009). Assessing the dynamic capabilities view: Spare change, everyone? *Strategic Organization*, 7(1), 75–90. <https://doi.org/10.1177/1476127008100132>.
- Arora, A., & Gambardella, A. (1994). The changing technology of technological change: General and abstract knowledge and the division of innovative labour. *Research Policy*, 23(5), 523–532. [https://doi.org/10.1016/0048-7333\(94\)01003-X](https://doi.org/10.1016/0048-7333(94)01003-X).
- Arora, S. K. (2019). *Comparing website measures on R&D with patenting indicators*. Retrieved July 10, 2019 from <https://tinyurl.com/pat-web-compare>.
- Arora, S. K., Li, Y., Youtie, J., & Shapira, P. (2015). Using the wayback machine to mine websites in the social sciences: A methodological resource. *Journal of the Association for Information Science and Technology*. <https://doi.org/10.1002/asi.23503>.
- Arora, S. K., Porter, A. L., Youtie, J., & Shapira, P. (2012). Capturing new developments in an emerging technology: An updated search strategy for identifying nanotechnology research outputs. *Scientometrics*, 95(1), 351–370. <https://doi.org/10.1007/s11192-012-0903-6>.

- Arora, S. K., Youtie, J., Shapira, P., Gao, L., & Ma, T. (2013). Entry strategies in an emerging technology: A pilot web-based study of graphene firms. *Scientometrics*, 95(3), 1189–1207. <https://doi.org/10.1007/s11192-013-0950-7>.
- Ashurst, C., Freer, A., Ekdahl, J., & Gibbons, C. (2012). Exploring IT-enabled innovation: A new paradigm? *International Journal of Information Management*, 32(4), 326–336. <https://doi.org/10.1016/j.ijinfomgt.2012.05.006>.
- Atkinson, R. D., & Lind, M. (2018). *Big is beautiful: Debunking the myth of small business*. Cambridge, MA: MIT Press.
- Autio, E., Kenney, M., Mustar, P., Siegel, D., & Wright, M. (2014). Entrepreneurial innovation: The importance of context. *Research Policy*, 43, 1097–1108. <https://doi.org/10.1016/j.respol.2014.01.015>.
- Bartholomew, S., & Smith, A. D. (2006). Improving survey response rates from chief executive officers in small firms: The importance of social networks. *Entrepreneurship Theory and Practice*, 30(1), 83–96. <https://doi.org/10.1111/j.1540-6520.2006.00111.x>.
- Baruch, Y. (1999). Response rate in academic studies—A comparative analysis. *Human Relations*, 52(4), 421–438. <https://doi.org/10.1177/001872679905200401>.
- Becheikh, N., Landry, R., & Amara, N. (2006). Lessons from innovation empirical studies in the manufacturing sector: A systematic review of the literature from 1993–2003. *Technovation*, 26(5–6), 644–664. <https://doi.org/10.1016/j.technovation.2005.06.016>.
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4), 77. <https://doi.org/10.1145/2133806.2133826>.
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3, 993–1022.
- Bloomberg New Energy Finance. (2019). *Clean energy investment trends, 2018*. Retrieved July 10, 2019 from <https://data.bloomberglp.com/professional/sites/24/BNEF-Clean-Energy-Investment-Trends-2018.pdf>.
- Brown, C., & Medoff, J. L. (2003). Firm age and wages. *Journal of Labor Economics*, 21(3), 677–697. <https://doi.org/10.1086/374963>.
- Canini, K., Shi, L., & Griffiths, T. (2009). Online inference of topics with latent Dirichlet allocation. In *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics, PMLR* (Vol. 5, pp. 65–72). <http://proceedings.mlr.press/v5/canini09a.html>.
- Carlton, D. W. (1983). The location and employment choices of new firms: An econometric model with discrete and continuous endogenous variables. *The Review of Economics and Statistics*, 65(3), 440. <https://doi.org/10.2307/1924189>.
- Caspin-Wagner, K., Ellis, S., & Tishler, A. (2012). Balancing exploration and exploitation for firm's superior performance: The role of the environment. *Academy of Management Proceedings*, 2012(1), 17177. <https://doi.org/10.5465/AMBPP.2012.17177abstract>.
- Choo, C. W. (2006). *The knowing organization*. New York: Oxford University Press.
- Chung, C.-J., & Wee, H.-M. (2011). Short life-cycle deteriorating product remanufacturing in a green supply chain inventory control system. *International Journal of Production Economics*, 129(1), 195–203. <https://doi.org/10.1016/j.ijpe.2010.09.033>.
- Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 35(1), 128. <https://doi.org/10.2307/2393553>.
- Covin, J. G., & Slevin, D. P. (1988). The influence of organization structure on the utility of an entrepreneurial top management style. *Journal of Management Studies*, 25(3), 217–234. <https://doi.org/10.1111/j.1467-6486.1988.tb00033.x>.
- Ebben, J. J., & Johnson, A. C. (2005). Efficiency, flexibility, or both? Evidence linking strategy to performance in small firms. *Strategic Management Journal*, 26(13), 1249–1259. <https://doi.org/10.1002/smj.503>.
- Eisenhardt, K. M., & Martin, J. A. (2000). Dynamic capabilities: What are they? *Strategic Management Journal*, 21(10–11), 1105–1121. [https://doi.org/10.1002/1097-0266\(200010/11\)21:10<113c::AID-SMJ133%3e3.0.CO;2-E](https://doi.org/10.1002/1097-0266(200010/11)21:10<113c::AID-SMJ133%3e3.0.CO;2-E).
- Elston, J. A., & Audretsch, D. B. (2007). The role of risk in entrepreneurial behavior. In M. Minniti (Ed.), *Entrepreneurship: The engine of growth* (Vol. 12, pp. 101–118). Westport, CT: Praeger.
- Fauchart, E., & Keilbach, M. (2009). Testing a model of exploration and exploitation as innovation strategies. *Small Business Economics*, 33(3), 257–272. <https://doi.org/10.1007/s11187-008-9101-6>.
- Fischer, E., & Reuber, A. R. (2011). Social interaction via new social media: (How) can interactions on Twitter affect effectual thinking and behavior? *Journal of Business Venturing*, 26(1), 1–18. <https://doi.org/10.1016/j.jbusvent.2010.09.002>.
- Goel, V. (2016). *Defining web pages, web sites and web captures*. Retrieved July 10, 2019 from <https://blog.archive.org/2016/10/23/defining-web-pages-web-sites-and-web-captures/>.

- Gök, A., Waterworth, A., & Shapira, P. (2015). Use of web mining in studying innovation. *Scientometrics*, 102(1), 653–671. <https://doi.org/10.1007/s11192-014-1434-0>.
- Humble, J., Molesky, J., & O'Reilly, B. (2015). *Lean enterprise: How high performance organizations innovate at scale*. Sebastopol, CA: O'Reilly Media Inc.
- Kassicieh, S. K., Kirchhoff, B. A., Walsh, S. T., & McWhorter, P. J. (2002). The role of small firms in the transfer of disruptive technologies. *Technovation*, 22(11), 667–674. [https://doi.org/10.1016/S0166-4972\(01\)00064-5](https://doi.org/10.1016/S0166-4972(01)00064-5).
- Keupp, M. M., & Gassmann, O. (2013). Resource constraints as triggers of radical innovation: Longitudinal evidence from the manufacturing sector. *Research Policy*, 42(8), 1457–1468. <https://doi.org/10.1016/j.respol.2013.04.006>.
- Kickul, J., & Gundry, L. (2002). Prospecting for strategic advantage: The proactive entrepreneurial personality and small firm innovation. *Journal of Small Business Management*, 40(2), 85–97. <https://doi.org/10.1111/1540-627X.00042>.
- Killaly, B. (1998). Gains from change: The impact of market and mode of service changes on performance in the US international telecommunication services industry. *Industrial and Corporate Change*, 7(4), 759–767. <https://doi.org/10.1093/icc/7.4.759>.
- Laaksonen, O., & Peltoniemi, M. (2016). The essence of dynamic capabilities and their measurement. *International Journal of Management Reviews*. <https://doi.org/10.1111/ijmr.12122>.
- Lavie, D., Stettner, U., & Tushman, M. L. (2010). Exploration and exploitation within and across organizations. *The Academy of Management Annals*, 4(1), 109–155. <https://doi.org/10.1080/19416521003691287>.
- Leetaru, K. (2015). How much of the Internet does the Wayback Machine really archive? *Forbes*. Retrieved December 2, 2018 from <https://www.forbes.com/sites/kalevleetaru/2015/11/16/how-much-of-the-internet-does-the-wayback-machine-really-archive>.
- Li, Y., Arora, S. K., Youtie, J., & Shapira, P. (2016). Using web mining to explore Triple Helix influences on growth in small and mid-size firms. *Technovation*. <https://doi.org/10.1016/j.technovation.2016.01.002>.
- Lumpkin, G., & Dess, G. G. (2001). Linking two dimensions of entrepreneurial orientation to firm performance. *Journal of Business Venturing*, 16(5), 429–451. [https://doi.org/10.1016/S0883-9026\(00\)00048-3](https://doi.org/10.1016/S0883-9026(00)00048-3).
- MacKinnon, D., Cumbers, A., & Chapman, K. (2002). Learning, innovation and regional development: A critical appraisal of recent debates. *Progress in Human Geography*, 26(3), 293–311. <https://doi.org/10.1191/0309132502ph371ra>.
- March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science*, 2(1), 71–87. <https://doi.org/10.1287/orsc.2.1.71>.
- March, J. G. (1994). *Primer on decision making: How decisions happen*. New York: The Free Press.
- March, J. G., & Sutton, R. I. (1997). Organizational performance as a dependent variable. *Organization Science*, 8(6), 698–706.
- Marx, M., Gans, J., & Hsu, D. (2014). Dynamic commercialization strategies for disruptive technologies: Evidence from the speech recognition industry. *Management Science*, 60(12), 3103–3123. <https://doi.org/10.1287/mnsc.2014.2035>.
- McGinn, D. (2012). *Too many pivots, too little passion what's wrong with today's entrepreneurship*. Retrieved July 10, 2019 from <https://hbr.org/2012/09/too-many-pivots-too-little-passion>.
- Moulaert, F., & Sekia, F. (2003). Territorial innovation models: A critical survey. *Regional Studies*, 37(3), 289–302. <https://doi.org/10.1080/0034340032000065442>.
- Muro, M., Rothwell, J., & Saha, D. (2011). *Sizing the clean economy: A National and Regional Green Jobs Assessment*. Retrieved July 10, 2019 from <https://www.brookings.edu/research/sizing-the-clean-economy-a-national-and-regional-green-jobs-assessment/>.
- Murphy, J., Hashim, N. H., & O'Connor, P. (2007). Take me back: Validating the Wayback machine. *Journal of Computer-Mediated Communication*, 13(1), 60–75. <https://doi.org/10.1111/j.1083-6101.2007.00386.x>.
- Neumark, D., Wall, B., & Zhang, J. (2011). Do small businesses create more jobs? New evidence for the United States from the National Establishment Time Series. *Review of Economics and Statistics*, 93(1), 16–29. https://doi.org/10.1162/REST_a_00060.
- Newman, A. L. (2010). What You want depends on what you know: Firm preferences in an information age. *Comparative Political Studies*, 43(10), 1286–1312. <https://doi.org/10.1177/0010414010369068>.
- O'Connor, C., & Klebahn, P. (2011). The strategic pivot: Rules for entrepreneurs and other innovators. *Harvard Business Review*. Retrieved July 10, 2019 from <https://hbr.org/2011/02/how-and-when-to-pivot-rules-fo>.

- O'Reilly, C. A., & Tushman, M. L. (2013). Organizational ambidexterity: Past, present, and future. *Academy of Management Perspectives*, 27(4), 324–338. <https://doi.org/10.5465/amp.2013.0025>.
- Patzelt, H., & Shepherd, D. A. (2011). Recognizing opportunities for sustainable development. *Entrepreneurship Theory and Practice*, 35(4), 631–652. <https://doi.org/10.1111/j.1540-6520.2010.00386.x>.
- PriceWaterhouseCoopers. (2014). *Cleantech MoneyTree™ Report: Q4 2013*. Retrieved April 21, 2019 from <http://www.pwc.com/us/en/technology/publications/moneytree-cleantech-venture-funding-report.jhtml>.
- Raisch, S., & Birkinshaw, J. (2008). Organizational ambidexterity: Antecedents, outcomes, and moderators. *Journal of Management*, 34(3), 375–409. <https://doi.org/10.1177/0149206308316058>.
- Reynolds, P. D. (2017). Erratum to: When is a firm born? Alternative criteria and consequences. *Business Economics*, 52(1), 81–83. <https://doi.org/10.1057/s11369-017-0032-6>.
- Ries, E. (2011). *The lean startup: How today's entrepreneurs use continuous innovation to create radically successful businesses*. New York: Random House LLC.
- Roberts, N., & Grover, V. (2012). Investigating firm's customer agility and firm performance: The importance of aligning sense and respond capabilities. *Journal of Business Research*, 65(5), 579–585. <https://doi.org/10.1016/j.jbusres.2011.02.009>.
- Rosenberg, N. (1990). Why do firms do basic research (with their own money)? *Research Policy*, 19(2), 165–174. [https://doi.org/10.1016/0048-7333\(90\)90046-9](https://doi.org/10.1016/0048-7333(90)90046-9).
- Rothaermel, F. T., & Alexandre, M. T. (2009). Ambidexterity in technology sourcing: The moderating role of absorptive capacity. *Organization Science*, 20(4), 759–780. <https://doi.org/10.1287/orsc.1080.0404>.
- Rothwell, R. (1989). Small firms, innovation and industrial change. *Small Business Economics*, 1(1), 51–64. <https://doi.org/10.1007/BF00389916>.
- Saemundsson, R. J., & Candi, M. (2017). Absorptive capacity and the identification of opportunities in new technology-based firms. *Technovation*, 64–65, 43–49. <https://doi.org/10.1016/j.technovation.2017.06.001>.
- Sarasvathy, S. D. (2001). Causation and effectuation: Toward a theoretical shift from economic inevitability to entrepreneurial contingency. *Academy of Management Review*, 26(2), 243–263. <https://doi.org/10.5465/AMR.2001.4378020>.
- Schilke, O. (2014a). On the contingent value of dynamic capabilities for competitive advantage: The non-linear moderating effect of environmental dynamism. *Strategic Management Journal*, 35(2), 179–203. <https://doi.org/10.1002/smj.2099>.
- Schilke, O. (2014b). Second-order dynamic capabilities: How do they matter? *Academy of Management Perspectives*, 28(4), 368–380. <https://doi.org/10.5465/amp.2013.0093>.
- Shapira, P., Gök, A., Klochikhin, E., & Sensier, M. (2013). Probing “green” industry enterprises in the UK: A new identification approach. *Technological Forecasting and Social Change*. <https://doi.org/10.1016/j.techfore.2013.10.023>.
- Simmie, J. (2005). Critical surveys edited by Stephen Roper innovation and space: A critical review of the literature. *Regional Studies*, 39(6), 789–804. <https://doi.org/10.1080/00343400500213671>.
- Simon, H. A. (1997). *Administrative behavior* (4th ed.). New York: Free Press.
- Solon, G., Haider, S. J., & Wooldridge, J. M. (2015). What are we weighting for? *Journal of Human Resources*, 50(2), 301–316.
- Solymossy, E. (2009). Entrepreneur and organization: Symbiotic change and transition. In M. A. Galindo, J. Guzman, & D. Riberio (Eds.), *Entrepreneurship and business* (pp. 269–290). Berlin, Heidelberg: Springer.
- Spender, J.-C. (1989). *Industry recipes. An inquiry into the nature and sources of managerial judgement*. Oxford: Blackwell.
- Stam, E., & Wennberg, K. (2009). The roles of R&D in new firm growth. *Small Business Economics*, 33(1), 77–89. <https://doi.org/10.1007/s11187-009-9183-9>.
- Stevenson, H. H. (1983). *A perspective on entrepreneurship*. Cambridge, MA: Harvard Business School Press.
- Steyvers, M., & Griffiths, T. (2007). Probabilistic topic models. In T. Landauer, D. McNamara, S. Dennis, & W. Kintsch (Eds.), *Latent semantic analysis: A road to meaning*. Hillsdale, NJ: Laurence Erlbaum.
- Teece, D. J. (2007). Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal*, 28(13), 1319–1350. <https://doi.org/10.1002/smj.640>.
- Teece, D. J. (2012). Dynamic capabilities: Routines versus entrepreneurial action. *Journal of Management Studies*, 49(8), 1395–1401. <https://doi.org/10.1111/j.1467-6486.2012.01080.x>.
- Teece, D. J. (2017). Towards a capability theory of (innovating) firms: Implications for management and policy. *Cambridge Journal of Economics*, 41(3), 693–720. <https://doi.org/10.1093/cje/bew063>.

- Teece, D. J., Peteraf, M., & Leih, S. (2016). Dynamic capabilities and organizational agility: Risk, uncertainty, and strategy in the innovation economy. *California Management Review*, 58(4), 13–35. <https://doi.org/10.1525/cmr.2016.58.4.13>.
- Thagard, P. (2005). *Mind: Introduction to cognitive science* (2nd ed.). Cambridge: A Bradford Book.
- Toh, P. K., & Kim, T. (2013). Why put all your eggs in one basket? A competition-based view of how technological uncertainty affects a firm's technological specialization. *Organization Science*, 24(4), 1214–1236. <https://doi.org/10.1287/orsc.1120.0782>.
- Tushman, M., Smith, W. K., Wood, R. C., Westerman, G., & O'Reilly, C. (2010). Organizational designs and innovation streams. *Industrial and Corporate Change*, 19(5), 1331–1366. <https://doi.org/10.1093/icc/dtq040>.
- Wiklund, J., Davidsson, P., & Delmar, F. (2003). What do they think and feel about growth? An expectancy-value approach to small business managers' attitudes toward growth. *Entrepreneurship Theory and Practice*, 27(3), 247–270. <https://doi.org/10.1111/1540-8520.t01-1-00003>.
- Winter, S. G. (2003). Understanding dynamic capabilities. *Strategic Management Journal*, 24(10), 991–995. <https://doi.org/10.1002/smj.318>.
- Youtie, J., Hicks, D., Shapira, P., & Horsley, T. (2012). Pathways from discovery to commercialisation: Using web sources to track small and medium-sized enterprise strategies in emerging nanotechnologies. *Technology Analysis & Strategic Management*, 24(10), 981–995. <https://doi.org/10.1080/09537325.2012.724163>.
- Yuan, L., Zhongfeng, S., & Yi, L. (2010). Can strategic flexibility help firms profit from product innovation? *Technovation*, 30(5–6), 300–309. <https://doi.org/10.1016/j.technovation.2009.07.007>.
- Zahra, S. A., Sapienza, H. J., & Davidsson, P. (2006). Entrepreneurship and dynamic capabilities: A review, model and research agenda*. *Journal of Management Studies*, 43(4), 917–955. <https://doi.org/10.1111/j.1467-6486.2006.00616.x>.
- Zott, C., & Huy, Q. N. (2007). How entrepreneurs use symbolic management to acquire resources. *Administrative Science Quarterly*, 52(1), 70–105. <https://doi.org/10.2189/asqu.52.1.70>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.