How Valuable Is FinTech Innovation?

Mark A. Chen
J. Mack Robinson College of Business, Georgia State University

Qinxi Wu
Hankamer School of Business, Baylor University

Baozhong Yang
J. Mack Robinson College of Business, Georgia State University

We provide large-scale evidence on the occurrence and value of FinTech innovation. Using data on patent filings from 2003 to 2017, we apply machine learning to identify and classify innovations by their underlying technologies. We find that most FinTech innovations yield substantial value to innovators, with blockchain being particularly valuable. For the overall financial sector, internet of things (IoT), robo-advising, and blockchain are the most valuable innovation types. Innovations affect financial industries more negatively when they involve disruptive technologies from nonfinancial startups, but market leaders that invest heavily in their own innovation can avoid much of the negative value effect. (*JEL* G14, G20, G29, G39)

Received May 31, 2017; editorial decision September 30, 2018 by Editor Andrew Karolyi.

In recent years, the rapid advance of FinTech, or financial technology, has attracted considerable attention within the finance industry. Many observers have welcomed the rise of FinTech, claiming that newly emerging technologies have the potential to radically transform financial services by making transactions less expensive, more convenient, and more secure.[1] Worldwide, external funding for FinTech development has been rising quickly. During the first half of 2018, global investment in FinTech companies totaled \$57 billion, a striking increase from \$38.1 billion for all of 2017 (KPMG 2018b).

We are grateful to Itay Goldstein, Wei Jiang, and Andrew Karolyi (the Sponsoring Editors for the RFS FinTech Initiative); Paul Tetlock (the RFS FinTech Workshop discussant); and two anonymous referees for valuable guidance and suggestions. We also appreciate helpful feedback from Vikas Agarwal, Jerry Hoberg, James Hodson, Andreas Hoepner, Lixin Huang, Dalida Kadyrzhanova, Omesh Kini, Laura Xiaolei Liu, and Xiaoyan Zhang and participants at the 2017 RFS FinTech Workshop, the 2018 RFS FinTech Conference, and the 2018 Western Economic Association International Conference, and seminar participants at Georgia State University, Peking University, and Tsinghua University. Yuan Gao, Justin Lee, Han Ma, An Nguyen, Honglin Ren, and Liang Zhang provided capable research assistance. Send correspondence to Mark A. Chen, J. Mack Robinson College of Business, Georgia State University, 35 Broad St. NW, Atlanta, GA 30303; telephone: 404-413-7339. E-mail: machen@gsu.edu.

[1] See, for example, *The Economist* (2015), McKinsey & Co. (2015), Harrist (2017), and Carney (2017).

doi:10.1093/rfs/hhy130

Furthermore, large financial institutions and technology firms are increasingly investing in FinTech innovation.[2]

Yet, despite the widespread interest in FinTech, little is currently known about exactly how it will affect existing financial firms and their business models. Which specific types of new FinTech will be most valuable to their innovators? Will FinTech discoveries help incumbent financial institutions reduce costs and better engage customers, resulting in higher future profits? Or will new FinTech enable upstart firms to erode established firms' competitive advantage, leading to lower profits and value throughout the industry? Such questions are difficult to answer in the absence of systematic evidence on FinTech innovation.

In this paper, we provide the first large-scale evidence on the occurrence and value of FinTech innovation. For this purpose, we construct a novel data set of published FinTech patent applications over the 2003–2017 period. Our data set draws on the Bulk Data Storage System (BDSS) of the United States Patent and Trademark Office (USPTO) and includes both the full text of each patent filing and identifying information on original patent assignees.[3] By using the identifying information to ascertain whether assignees are public firms, private firms, or individual inventors, we can obtain a much more complete picture of innovation activity than would be possible with public-company filings alone.

A fundamental challenge in studying FinTech innovation is that there currently exists no standard definition of what "FinTech" is and what specific technologies the term encompasses. One of the key goals of our study is thus to develop an objective, data-based definition and classification of FinTech innovation. To this end, we exploit the rich textual data in our sample of patent filing documents. We first assemble a new lexicon of finance-related terms and use it to narrow down the set of patent filings to those relating to financial services. We then apply several families of machine-learning algorithms to the textual data to identify FinTech innovations and classify them into seven key technology categories: cybersecurity, mobile transactions, data analytics, blockchain, peer-to-peer, robo-advising, and internet of things (IoT).[4]

Using our sample of classified patent filings, we document a number of basic facts about FinTech innovation. For example, we find that publicly traded companies as a group have driven only a minority of FinTech innovations to date. Indeed, private companies and nonfirm individuals account for about 62.7% of FinTech filings in the sample. Of the FinTech filings from companies, about 57.8% are in fact from technology companies outside of the financial services industries. Among the seven FinTech categories, cybersecurity and

2 See Nash (2016) and Russo (2017).

3 The BDSS platform, first released to the public in October 2015, succeeds the discontinued Google USPTO Bulk Downloads service. Although BDSS does not appear to have been used before by finance researchers, a number of papers have used Google USPTO Bulk Downloads to obtain information on assignee names, grant dates, and citations (see, e.g., He and Tian 2016; Kogan et al. 2017; Mann 2018).

4 In Section 1, we define these seven technology categories and provide illustrative examples of each.

mobile transactions have experienced the most total innovation over the sample period. Blockchain is currently the smallest but fastest-growing category of FinTech innovation.

To explore the implications of FinTech innovation, we develop a new methodology for estimating the value of a patent filing to one or more publicly traded firms. Our valuation approach is based on observed stock market responses to USPTO disclosures of patent filings. Importantly, our approach accounts for the market's anticipation of different types of patent filings made by different filers.[5] We start by estimating innovation arrival intensities via Poisson regression models that account for factors such as technology type, time effects, a patent filer's prior innovation experience, and filer fixed effects. Then, for each patent filing, we combine its predicted count intensity with a firm's stock price movements to infer the underlying value of the innovation to the firm.

Using our valuation approach, we examine how much firms in the financial services sector stand to gain from their own FinTech innovations. The calculations show that a FinTech innovation's private value (i.e., the value accruing to the innovator) is typically large and positive. For instance, in 2017 dollars, the median private value of a FinTech innovation is about $46.7 million, which is much higher than the median private value of $3.1 million for other financial innovations. Overall, the FinTech innovation types that are most valuable to innovators are blockchain, cybersecurity, and robo-advising. When we control in multivariate regressions for time-varying firm characteristics, firm fixed effects, and patent quality measures, we find that blockchain and robo-advising emerge as the most valuable types, thus underscoring the economic importance of these new segments of the FinTech space.

We extend our valuation method to study how FinTech innovations affect the financial services sector and its key component industries: banking, payment processing, brokerage, asset management, and insurance.[6] Our calculations show that, for the financial sector as a whole, the typical FinTech innovation brings positive value. The most valuable innovation types are IoT, robo-advising, and blockchain, with median value impacts of $24.5 billion, $15.5 billion, and $8.1 billion, respectively (2017 dollars). Nonetheless, we find substantial variation in value across different technology-industry pairings. Among mobile transaction innovations, for example, the median value impact is negative for the banking industry but positive for the payments industry.

What explains the wide cross-sectional variation in the value effects of FinTech innovation? We argue that the value effects of an innovation are driven by two key factors: (1) how inherently disruptive is the underlying technology;

---

5 Throughout, we use the term "filer" to refer to the original assignee of a patent application filing.

6 Because our method requires the use of stock-price data, any inferences or conclusions we draw about an innovation's value impact to an industry must be tempered by the fact that we cannot directly measure value effects for privately held firms within the industry.

and (2) whether the innovator poses a competitive entry threat to the industry. To study this issue, we use technological spillovers emanating from individual, nonfirm inventors to construct a data-based measure of "disruptiveness" across technology-industry pairs. Consistent with theories of disruptive innovation (e.g., Christensen 1997; Christensen and Raynor 2003; Downes and Nunes 2013), we find that a FinTech innovation tends to destroy significantly more industry value when its underlying technology is disruptive and when it originates from a young, nonfinancial firm ("FinTech startup").

Next, we examine how FinTech innovation affects value from the viewpoint of individual incumbent firms, that is, market-share leaders and their rivals. Theoretical considerations suggest that disruptive innovation by potential entrants can be especially harmful to an industry's market leaders, which are sluggish in adapting to change and focused on existing customers (Tang 1988; Christensen and Raynor 2003). On the other hand, industry-wide disruption might be advantageous to market leaders because, compared to rivals, they have larger scale economies and more financial resources with which to innovate new lines of business (Dasgupta and Stiglitz 1980; Scherer 1980; Blundell, Griffith, and van Reenen 1999; Czarnitzki, Etro, and Kraft 2014). Our empirical tests support the latter prediction and also suggest that market leaders' ability to avoid harm from disruptive outside innovation is strongly linked to the amount of resources that they devote to their own research and development (R&D).

Our paper is connected to a number of different literatures. First, our work complements the sizable body of research that utilizes patent data to study innovation activity by firms (e.g., Griliches, Hall, and Pakes 1991; Hall, Jaffe, and Trajtenberg 2005; Kortum and Lerner 2000; Lerner 2009; Brav et al. 2018). While this literature has delivered valuable insights about corporate patenting and innovation in general, much of the research relies on patent grant data and thus cannot fully capture the FinTech innovation activity that has occurred over just the past few years.[7] By focusing on patent applications and utilizing the BDSS data, we can mitigate the data truncation problems inherent in relying on patent grants and thus provide a more complete picture of very recent trends and patterns in FinTech innovation.

Second, our work builds upon a stream of research that uses stock price data to study the value of innovation (see, e.g., Pakes 1985; Austin 1993; Hall, Jaffe, and Trajtenberg 2005; Nicholas 2008; Kogan et al. 2017). The methods that we develop extend this literature by recognizing the count-based nature of innovation events over time, thus permitting more precise estimates of the true value impact of such events. More generally, our approach of combining stock price reactions with predicted Poisson arrival intensities could be useful

[7] Empirical studies document that, for observed patent grants, the average time between application and granting is about 2 years (e.g., Hall, Jaffe, and Trajtenberg 2005; Seru 2014; Cornaggia et al. 2015). Some applications remain under review for much longer than 2 years and thus may not be captured by currently available grants data.

for studying other types of recurring, partially anticipated phenomena, such as revisions to analyst estimates, sequences of corporate news releases, or waves of mergers or bankruptcies.

Third, our findings contribute to the finance, strategy, and economics literatures that explore the role of innovation in shaping industry competition. Theoretical research has modeled how innovation from outside of an industry can harm or benefit incumbent firms (see, e.g., Lieberman and Montgomery 1988; Henderson and Cockburn 1996; Christensen 1997; Adner 2012) and how incumbents can use their own innovation to protect themselves from outside threats (see Dasgupta and Stiglitz 1980; Gilbert and Newbery 1982; Aghion et al. 2001; Aghion and Griffith 2005). Testing such theories is challenging because of the difficulty of obtaining a large data sample of competitive threats from innovation. Our work employs a new data set and provides systematic evidence of how innovation by potential entrants can affect individual firms within an industry.

Finally, our approach to identifying and classifying FinTech patent filings contributes to the literature that applies textual analysis and machine learning to finance and economics. Researchers have used text-based methods to study news articles, online forum postings, corporate filings, and analyst reports (e.g., Antweiler and Frank 2004; Tetlock, Saar-Tsechansky, and Macskassy 2008; Hanley and Hoberg 2010; Loughran and McDonald 2011; Jegadeesh and Wu 2013; Hoberg and Phillips 2016; Bellstam, Bhagat, and Cookson 2017; Manela and Moreira 2017). Other work studies the application of machine-learning methods to economics (e.g., Kleinberg et al. 2015; Glaeser et al. 2016; Naik, Raskar, and Hidalgo 2016; Athey and Wager 2018; Athey, Tibshirani, and Wager 2018). A number of the machine-learning algorithms that we use for text classification appear to be new to the finance domain and can be applied to study a broad set of questions relating to patent filings, legal documents, media stories, and other textual data.

1. Categories of FinTech

What is FinTech? Although FinTech can be broadly defined as any technology that enables or enhances the provision of financial services, such a definition is of limited use for empirically identifying and classifying real-world FinTech. To proceed with our analysis, we therefore require a typology that (1) distinguishes innovation within the FinTech space from other types of financial or scientific innovation; and (2) articulates the key technological differences among different instances of FinTech innovation.

We begin with the premise that FinTech ultimately consists of the set of recently developed *digital computing technologies* that have been applied—or that will likely be applied in the future—to financial services. Then, based on a

Table 1
Categories of FinTech

Category definition	Key technologies	Real-world examples
Cybersecurity: Hardware or software used to protect financial privacy or safeguard against electronic theft or fraud	Encryption, tokenization, authentication, biometrics	*Diebold iris-scanning ATM, Mastercard Biometric Card, USAA face recognition login, Experian CreditLock*
Mobile transactions: Technologies that facilitate payments via mobile wireless devices, such as smartphones, tablets, and wearables	Smartphone wallets, digital wallets, near-field communication	*Apple Pay, Android Pay, PayPal Mobile Express Checkout, Venmo, Square*
Data analytics: Technologies and algorithms that facilitate the analysis of transactions data or consumer financial data	Big data, cloud computing, artificial intelligence, machine learning	*Equifax NeuroDecision credit scoring, JPMorgan Contract Intelligence (COiN), Bloomberg Social Sentiment Analytics*
Blockchain: Distributed ledger technologies with a primary application to financial services	Cryptocurrency, proof-of-work, smart contracts, directed acyclic graphs	*Bitcoin, Ripple Payment Network, Visa B2B Connect, Nasdaq Linq asset trading platform*
Peer-to-peer (P2P): Software, systems, or platforms that facilitate consumer-to-consumer financial transactions	Crowdfunding, P2P lending, customer-to-customer payments	*GoFundMe, Kickstarter, Lending Club, Prosper Marketplace, Zelle*
Robo-advising: Computer systems or programs that provide automated investment advice to customers or portfolio managers	Artificial intelligence, big data, machine learning	*Betterment, E-Trade Core Portfolios, Schwab Intelligent Portfolios, Vanguard Personal Advisor Services*
Internet of things (IoT): Technologies relating to smart devices that gather data in real time and communicate via the internet	Smart devices, near-field communication, wireless sensor networks, actuators	*UnitedHealthcare Motion F.I.T. tracker, Nationwide SmartRide telematics, Travelers Insurance smart home sensors*

This table shows a proposed typology of FinTech. The definitions, technologies, and examples listed are based on the authors' reading of news articles, industry reports, and surveys.

general reading of various articles and reports,[8] we formulate a broad typology of FinTech comprising seven categories: cybersecurity, mobile transactions, data analytics, blockchain, peer-to-peer (P2P), robo-advising, and IoT. Table 1 provides brief definitions of these FinTech categories and lists key technologies and real-world examples associated with each category.

It is apparent from Table 1 that some categories (e.g., data analytics) are quite broad, and their constituent technologies are already in widespread use across many financial industries. Other categories (e.g., P2P or robo-advising) are associated with a narrower set of industries. In some cases, key technologies underlying a FinTech category also belong to another, broader category. For example, robo-advising shares big data, artificial intelligence, and machine-learning technologies with the broader category of data analytics. In situations in which an innovation involves a technology that spans multiple categories, the intended finance application should dictate which category the innovation belongs to.

[8] See in particular *The Economist* (2015), Orton-Jones (2017), Robinson (2017a, 2017b), Reklaitis (2018), and KPMG (2018a).

Finally, it is important to note that not all the key technologies listed in Table 1 automatically qualify as being FinTech. Indeed, technologies are partly defined by their intended use. Our typology requires that the actual (or intended) main use case of a technology lie within the field of financial services in order for the technology to be considered FinTech. Thus, a new blockchain designed for supply-chain management or a new machine-learning algorithm for predicting weather patterns would not be considered FinTech since the primary intended applications of these innovations do not fall within the domain of financial services.

2. Data

Our source of patent filings data is the Bulk Data Storage System (BDSS) provided by the U.S. Patent and Trademark Office (USPTO). First released to the public in October 2015, BDSS is updated weekly and provides comprehensive coverage of patent filings from 1976 to the present. The BDSS platform offers two key advantages relative to other data sets, such as the commonly used National Bureau of Economic Research (NBER) patent database. First, BDSS permits bulk downloads of the full text of patent applications.[9] Second, for applications after March 2001, BDSS reports the dates on which patent applications were first disclosed to the public.[10] Having exact patent application disclosure dates is critical for our purposes because the valuation methodology we develop in Section 3 makes use of stock price data around public news of innovations.

To construct a sample for our study, we first use BDSS to obtain information on the 4,680,587 total patent applications published by the USPTO between January 1, 2003 and September 7, 2017. Of these, 2,243,484 patent applications are identified in BDSS as having been filed by public firms, private firms, and individuals located in the United States. We identify each patent filer's type by the assignee type and the assignee/applicant addresses as published in BDSS.

We gather the International Patent Classification (IPC) codes associated with each patent application and then restrict the sample to applications belonging to either IPC Class G or IPC Class H. The union of these very broad patent classes ("Class G&H") covers the areas that are potentially related to digital computing, which is a technology that underlies all FinTech categories as discussed in

[9] Our study uses patent applications rather than patent grants because many FinTech patents applied for during the past few years have not yet been granted. Thus, focusing only on news of FinTech patent grants could result in a severe truncation bias.

[10] Under the American Inventors Protection Act of 1999, patent applications filed on or after November 29, 2000 must be publicly disclosed by the USPTO no later than 18 months after the initial filing date. Disclosure can occur sooner than 18 months after filing if an inventor requests early publication.

Section 1.[11] With this restriction, we obtain a sample of 1,181,162 Class G&H patents filed by U.S. companies or individuals.

2.1 Identifying and classifying FinTech innovations

To identify and classify FinTech patent filings in accordance with our typology outlined in Section 1, we proceed in two main steps. First, we use text-based filtering to narrow down the large set of Class G&H patent filings to those that are plausibly related to financial services. Second, we apply several families of machine-learning algorithms to automatically generate classifications of the filings. In the following three subsections, we describe the text-based filtering process, the training and application of machine-learning classifiers, and the results of the classification.

2.1.1 Text-based filtering to exclude nonfinancial innovations. We construct a new lexicon of financial terms that can be used to exclude nonfinancial patent filings from the overall sample. This approach is related to prior research in finance that has used word lists or textual analysis to measure sentiment, detect bias, or classify subjects in news media or financial filings.[12] To build our list of filtering terms, we start with two publicly available glossaries. First, we obtain all terms from Campbell R. Harvey's *Hypertextual Finance Glossary*[13] (the November 8, 2016, version), a widely used online compendium of finance terms that serves as the basis for the glossaries of numerous media companies, such as *New York Times*, *Forbes*, and *CNN Money*. Second, we gather all terms from the online *Oxford Dictionary of Finance and Banking*, 5th Edition, published by Oxford University Press. Combining these two lists of terms and excluding acronyms, we obtain a total of 11,196 unique single-word and multi-word finance terms. From the combined glossary, we create a list of all terms that can be unambiguously associated with financial services, including single-word terms (e.g., "bourse," "futures," or "chargeback") and any 2- or 3-grams contained within glossary terms (e.g., "credit card," "bond indenture," or "mutual fund"). We add to our filtering list a small handful of additional words that have recently gained recognition as financial terms, such as "bitcoin," "cryptocurrency," and "crowdfunding." Our final list has a total of 487 unique finance-related terms.[14]

---

[11] Class G corresponds to Physics and includes computing, calculating, counting, information and communication technology, and other categories. Class H corresponds to Electricity and covers basic electric elements, generation of electricity, applied electricity, basic electronic circuits, and their control; radio or electric communication techniques; and other areas. Full descriptions of IPC classes are available at http://www.wipo.int/classifications/ipc/en.

[12] Tetlock (2007) applies large-scale, quantitative content analysis to examine of the effects of media sentiment on stock returns and trading volume. Tetlock, Saar-Tsechansky, and Macskassy (2008) and Kelly and Tetlock (2013) study how text-based measures of negative sentiment relate to firm fundamentals, retail trades, and stock returns. Loughran and McDonald (2011) develop a lexicon of finance words and use it to study how negative sentiment in 10-K filings relates to outcomes such as stock returns, trading volume, fraud, and earnings surprises.

[13] Available at https://people.duke.edu/~charvey/Classes/wpg/glossary.htm.

[14] Our list of finance-related filtering terms is available on request.

We apply our filtering list to the sample of all Class G&H patents filed by U.S. companies and individuals. Specifically, we retain patent filings that meet two requirements: (1) at least one filtering term appears in the title, abstract, summary, or claims sections of the filing; (2) a second, different filtering term appears anywhere in the filing document. Using this filtering strategy, we obtain a total of 67,948 patent filings that are potentially related to financial services.

2.1.2 Using machine learning to identify and classify FinTech innovations. Next, we apply supervised machine-learning methods to classify the filtered sample of patent filings based on their textual data. The application of these methods requires three basic steps: text preprocessing, creation of a training sample, and training one or more algorithms to produce a classification. For ease of exposition, we briefly summarize each step here and provide more detailed descriptions in Appendix A.

Step 1: Preprocessing text contained in filings

We preprocess the text of each patent application document using approaches that are standard practice in text-based analysis, including tokenization, stemming, removal of stopwords, and removal of very common terms (see, e.g., Gentzkow, Kelly, and Taddy forthcoming). Using a "bag of words" approach, we map each filing document into a numerical vector of "term frequency-inverse document frequency" scores, where a word's score reflects how important the word is to a document within the broader collection of documents.

Step 2: Constructing a training sample

To construct a training sample for the machine-learning algorithms, we first compile a list of firms featured in the annual FinTech surveys of six different magazines. We add to this list the firms in Compustat that are among the top-10 most-prolific patent filers within each of five financial industries: commercial banking, payment processing, brokerage, asset management, and insurance.[15] Out of all Class G&H patent applications filed during 2003–2017 by firms on our list, we select a random subsample of 1,000. We review and manually classify the 1,000 filings into nine groups (seven for the FinTech categories, one for other financial filings, and one for nonfinancial filings). Using these groups as the basis for a simple nearest-centroid classifier, we classify the entire sample of 67,948 text-filtered filings into the nine groups and then select 200 from each group. The resultant collection of 1,800 filings, which we manually reclassify as needed, constitutes our training sample.

[15] We identify these five industry groups using six-digit NAICS codes as follows: asset management: 523920, 523930; banking: 522298, 522120, 522291, 522220, 522110, 522210; brokerage: 523110, 523120; insurance: 524210, 524127, 524113, 524130, 524126, 524114; and payments: 522320.

Step 3: Applying machine-learning classifiers

The key step in classifying the broader set of 67,948 text-filtered filings is to train and use one or more supervised machine-learning algorithms. Rather than simply rely on a single approach, we use several different families of algorithms, each of which has been well-studied and successfully applied to classification problems in other domains.[16] The two types of classifiers that are most central to our work are *support vector machines (SVM)* and *neural networks*. Appendix A briefly explains these two approaches and reports the specific hyperparameters and design choices that we use with each approach. For comparison purposes, we also train and apply several other well-known classifiers: *Naïve Bayes (NB)*, *k-nearest neighbor (kNN)*, *random forest*, and *gradient boosting*. Since these additional methods do not play a direct role in our final classification, we do not discuss them in detail and instead refer the reader to Hastie, Tibshirani, and Friedman (2009).

2.1.3 Performance of the machine-learning classifiers. Table 2 reports the performance of the various machine-learning classifiers we apply to our sample.[17] We consider a total of seven individual classifier models: naïve Bayes, kNN, random forest, gradient boost, linear multiclass SVM, nonlinear multiclass SVM (Gaussian kernel), and neural network. For each classifier, we report performance according to four commonly used measures: accuracy, precision, recall, and F1 score.[18] Each performance measure for a given classifier is calculated as an average performance score across the seven FinTech categories. In practice, the F1 metric is generally considered to be superior to the others if the data potentially contain many true negatives. Thus, when fitting each individual classifier model, we select parameters to maximize the F1 score.

Panel A of Table 2 shows that, among the individual classifiers tested, the neural network classifier and the linear SVM classifier are the top two models in terms of F1 score and accuracy. To achieve stronger classification performance, we use a simple majority-rule "voting" classifier that aggregates predictions of the linear SVM, Gaussian SVM, and neural network models.[19] (In the small

[16] An early version of our paper relied on just one algorithm, support vector machines, to classify patent texts. We thank several audience members at workshops and seminars for suggesting that we also use other well-known algorithms to improve the quality of the classification and the overall reliability of our empirical results.

[17] We evaluate classifier performance using a 10-fold cross-validation approach. This involves first splitting the training sample into 10 random subsets. For each subset $j = 1, 2, \ldots, 10$, we use the union of the remaining nine subsets as a training sample to fit the model and use subset j as a test sample to assess out-of-sample performance. We then average the model's performance across the 10 sample splits. This cross-validation method has the advantage of being robust and utilizing the entire training sample (see, e.g., chapter 5 of James et al. 2013).

[18] *Accuracy* is one minus the ratio of the number of incorrect category predictions to the total number of observations. *Precision* is the ratio of true positives to the sum of true positives and true negatives. *Recall* is the ratio of true positives to the sum of true positives and false positives. $F1$ is the harmonic mean of precision and recall.

[19] The voting classifier is a popular and effective method of aggregating the results from different algorithms to achieve better performance (see, e.g., Schapire et al. 1998).

Table 2
Performance of machine-learning methods in classifying patent filings

A. Out-of-sample performance of different machine-learning methods

	NB (%)	kNN (%)	Random forest (%)	Gradient boost (%)	Linear SVM (%)	Gaussian SVM (%)	Neural network (%)	Voting classifier (%)
Accuracy	73.2	74.0	78.5	79.6	82.2	81.7	81.9	82.6
Precision	80.2	67.5	83.4	75.2	80.4	82.1	77.9	80.8
Recall	50.2	69.4	60.1	69.0	72.0	69.8	74.3	72.7
F1 score	58.6	68.1	67.5	71.6	75.7	74.9	76.0	76.3

B. In-sample performance of the voting classifier for the training sample

Category	# of filings in true category	# of filings in predicted category	Precision (%)	Recall (%)	Classification of the filtered sample
0 (nonfinancial)	573	567	99.3	98.3	45,011
1 (cybersecurity)	188	187	98.9	98.4	3,607
2 (mobile transactions)	175	176	97.7	98.3	1,659
3 (data analytics)	110	119	91.6	99.1	602
4 (blockchain)	69	68	100.0	98.6	136
5 (P2P)	86	88	96.6	98.8	488
6 (robo-advising)	73	73	98.6	98.6	377
7 (IoT)	59	58	98.3	96.6	270
8 (non-FinTech financial)	467	464	98.3	97.6	15,798
Total	1,800	1,800	97.7	98.3	67,948

This table summarizes the performance of text-based machine-learning methods used to identify and classify FinTech patent filings. Texts of patent filings are obtained from the USPTO's Bulk Data Storage System (BDSS). Panel A shows the out-of-sample performance of each machine-learning method. Performance is measured by ten-fold cross-validation using the training sample constructed in Section 2.1. *Accuracy* is one minus the ratio of incorrect category predictions to total observations. *Precision* is the ratio of true positives to the sum of true positives and true negatives. *Recall* is the ratio of true positives to the sum of true positives and false positives. *F1 score* is the harmonic mean of Precision and Recall. Model parameters are selected to maximize the F1 score of models in cross-validation. The parameters for the models are as follows: $k=5$ for kNN, $C=0.6$ for Linear SVM, and $\gamma=1.0$, $C=2.0$ for Gaussian SVM. The neural network classifier has three hidden layers with 1,792, 307, and 53 neurons in the first, second, and third hidden layer, respectively. The voting classifier aggregates the results of linear SVM, Gaussian SVM, and neural network under a majority voting rule. Panel B shows the in-sample performance of the voting classifier on the training sample and shows the number of filings predicted to be in each category according to the final classification of the text-filtered sample.

handful of cases in which all three classifiers predict different categories, the tie is broken by using the neural network classifier prediction.) The voting classifier outperforms all other methods with an accuracy of 82.6% and F1 score of 76.3%. This is therefore the classifier we use throughout the rest of the paper.

After fitting the voting classifier to the training sample, we use the fitted model to obtain a final classification for the entire text-filtered sample of 67,948 patent filings. Panel B of Table 2 reports the in-sample classification performance. As seen in the table, the voting algorithm has in-sample precision of 97.7% and recall of 98.3%. The last column of panel B shows the number of cases in the text-filtered sample that the voting classifier assigns to each category. Based on the final classification, the sample consists of about 66.2% nonfinancial filings, 23.3% non-FinTech financial filings, and 10.5% FinTech filings.

2.2 Matching filings with data on public and private firms

For the subset of finance-related filings made by public or private firms, we assemble data on key filer characteristics. To this end, we start with assignees' names as reported in BDSS and then conduct a name match using CRSP/Compustat, company websites, Google, and various other public online sources. Because the name matching sometimes involves ambiguity (e.g., due to acquisitions of assignees or the presence of name variants in BDSS), we employ several consistency checks and filtering strategies to ensure that assignee firms are properly identified. Also, we exclude from the sample a handful of filings for which the name matching reveals that the filer is in fact a university, a foreign company, or a U.S. subsidiary of a foreign company. Appendix B provides a detailed description of the steps involved in the name matching.

Next, for each firm in the name-matched sample, we gather data (where available) on financials, stock prices, SIC and NAICS industry codes, and year of founding. We obtain data on industry codes and year of founding from D&B Hoover's, Standard & Poor's NetAdvantage, LexisNexis company profiles, Bloomberg, Bizapedia, and Google. For data on stock prices and financials, we use CRSP and Compustat.

Table 3 shows the number of patent filings removed by the filtering and name matching steps described above. Out of our set of 67,948 text-filtered filings classified by machine learning, 22,937 are finance-related. From these 22,937 filings, we remove 1,792 problematic cases that are missing requisite data or were filed by universities or non-U.S. entities. This leaves a sample of 21,145 filings by U.S.-based firms or individuals, of which 14,634 are non-FinTech financial filings and 6,511 are FinTech filings.

2.3 Descriptive statistics

Table 4 reports the frequencies of FinTech patent applications filed by various groups of innovators. As seen in the table, public firms, private firms, and individuals filed, respectively, 37.3%, 23.0%, and 39.7% of all FinTech applications. Nonfinancial firms are an important group of innovators, accounting for 34.8% of FinTech filings. FinTech startups—which we define as nonfinancial firms founded no more than 8 years prior—account for nearly one-fourth of all filings made by nonfinancial companies. Among public firms in the financial services industries, banks are by far the most active innovators, followed by payment processing companies.

Table 5 considers whether certain types of innovators exhibit more activity in certain FinTech categories. Panel A shows that public firms dominate most of the seven categories, but private firms substantially contribute to total firm-based innovation activity in robo-advising (59.4%), mobile transactions (42.0%), data analytics (37.6%), and cybersecurity (36.0%). Nonfinancial firms innovate heavily in cybersecurity, mobile transactions, and P2P, whereas financial companies account for more than half of all blockchain and IoT filings made by firms. Individual (nonfirm) inventors are also key participants in

Table 3
Construction of the patent filings sample

Filtering step	Observations removed	Remaining observations
Extract all patent filings disclosed from January 2003 to September 2017		4,680,587
Remove patent filings with non-U.S. BDSS identifying information	2,437,103	2,243,484
Remove patent filings not belonging to IPC Classes G or H	1,062,322	1,181,162
Filter out nonfinancial patent filings using list of financial terms	1,113,214	67,948
Remove filings classified as nonfinancial by machine-learning algorithms	45,011	22,937
Remove filings with missing address and name information	173	22,764
Remove filings by universities, non-U.S. firms, and subsidiaries of non-U.S. firms	56	22,708
Remove filings by firms with missing industry code or missing founding date	1,563	21,145
Remaining sample of financial patent filings:	21,145	
Non-FinTech filings		
by firms	8,056	
by individuals	6,578	
FinTech filings		
by firms	3,923	
by individuals	2,588	

This table shows the steps involved in construction of the sample of patent application filings. Data on patent filings are drawn from the Bulk Data Storage System (BDSS) provided by the U.S. Patent and Trademark Office. Text filtering and machine learning methods are used as described in Section 2.1 to classify filings. Data on company founding dates and NAICS industry classifications are gathered from Compustat, LexisNexis, Hoover's, S&P NetAdvantage, and public websites.

Table 4
FinTech innovation activity by innovator type

	Number of FinTech patent filings
U.S. filers	
Individuals	2,588
Public firms	2,429
Private firms	1,494
U.S. financial services firms	
Public financial firms	1,425
Asset management	5
Banking	806
Brokerage	22
Insurance	56
Payments	536
Private financial firms	229
U.S. nonfinancial firms	
FinTech startups	548
Other nonfinancial firms	1,721

This table shows the frequencies of FinTech patent filings by various types of innovators. The sample of FinTech filings is constructed from patent applications drawn from the USPTO's Bulk Data Storage System (BDSS). FinTech startups are nonfinancial firms with a founding date no more than 8 years prior. Financial industry groupings are based on primary NAICS codes as detailed in Footnote 15. Data on company founding dates and NAICS industry classifications are obtained from Compustat, LexisNexis, Hoover's, S&P NetAdvantage, and public websites.

Table 5
FinTech innovation activity by technology category and innovator type

A. FinTech innovation activity by firms and individuals

Technology Category	Innovator type: Public firm	Private firm	Financial firm	Non-financial firm	FinTech startup	Individual inventor
Cybersecurity	1,179	664	749	1,094	258	1,510
Mobile transaction	569	412	311	670	171	468
Data analytics	234	141	211	164	32	202
Blockchain	60	27	47	40	21	19
P2P	207	101	131	177	33	160
Robo-advising	71	104	104	71	26	147
IoT	109	45	101	53	7	82

B. Industry-level FinTech innovation activity

	Most-active industry		2^{nd}-most-active industry	
Technology category	Industry	# of filings	Industry	# of filings
Cybersecurity	Banking	380	Payments	358
Mobile transactions	Payments	175	Banking	129
Data analytics	Banking	143	Payments	33
Blockchain	Banking	34	Payments	11
P2P	Payments	68	Banking	59
Robo-advising	Asset mgmt.	33	Banking	30
IoT	Banking	60	Insurance	41

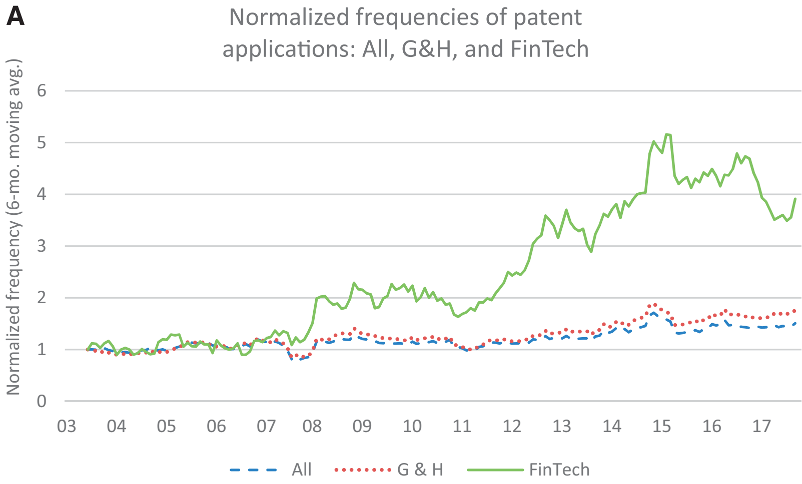
This table shows, by technology category, the frequencies of FinTech patent filings made by various innovator types. Technology categories are determined via text filtering and machine-learning methods as described in Section 2.1. Panel A reports frequencies for public firms, private firms, financial firms, nonfinancial firms, FinTech startups (nonfinancial firms with a founding date no more than 8 years prior), and individuals. Panel B reports frequencies for the two industries that have the highest numbers of innovations within each technology category. Financial industry groupings are based on primary NAICS codes as detailed in Footnote 15. Data on company founding dates and NAICS industry classifications are obtained from Compustat, LexisNexis, D&B Hoover's, S&P NetAdvantage, Bloomberg, Bizapedia, and other public web-based sources.

FinTech innovation, especially in cybersecurity and robo-advising. As seen in panel B, banking and payments firms dominate nearly all categories of innovation, but some categories show narrowly focused activity by firms in other industries. For example, asset management is the most active industry in terms of robo-advising, and insurance is the second-most active industry with regards to IoT.

Next, we present evidence on broad trends in FinTech innovation activity over time. Given the very rapid and volatile rate of development of some FinTech technologies (e.g., blockchain), we calculate innovation at a monthly frequency and focus mainly on 6-month moving averages of patent filing counts. We count a patent filing based on its USPTO publication date, which is typically 18 months (and occasionally less) after the initial filing.

Figure 1 displays time-series trends in filing activity from June 2003 to August 2017 for three classes of patent applications: FinTech, Class G&H, and All. Panel A shows FinTech patent applications increase approximately fourfold over 14 years.[20] The increase is particularly rapid from 2010 to 2014,

20 To facilitate comparisons, we normalize each group's 6-month moving average series in panel A by dividing by the respective group's moving average in June 2003.



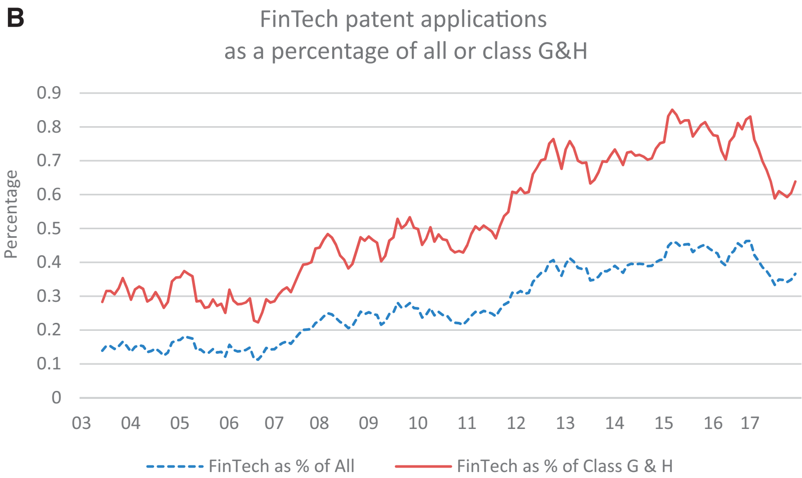


Figure 1
Innovation over time: FinTech, Class G&H, and all
This figure shows, from June 2003 to August 2017, monthly frequencies of different classes of patent applications. Frequencies are calculated as 6-month moving averages of patent filing counts. Panel A shows frequencies for three groups of filings: FinTech filings, Class G&H filings, and all filings. (To facilitate comparisons, we normalized the frequency series for each group by dividing by the respective group's frequency in June 2003.) Panel B shows the frequency of FinTech filings as a percentage of all filings and as a percentage of Class G&H filings. FinTech applications are identified by applying text-based filtering and machine learning to patent filings drawn from the USPTO's Bulk Data Storage System (BDSS).

both relative to IPC Class G&H patent applications and relative to total patent applications. Indeed, panel B shows that, over the sample period, FinTech filings more than double as a percentage of Class G&H filings and nearly triple as a percentage of all filings. The strong relative growth in FinTech filings

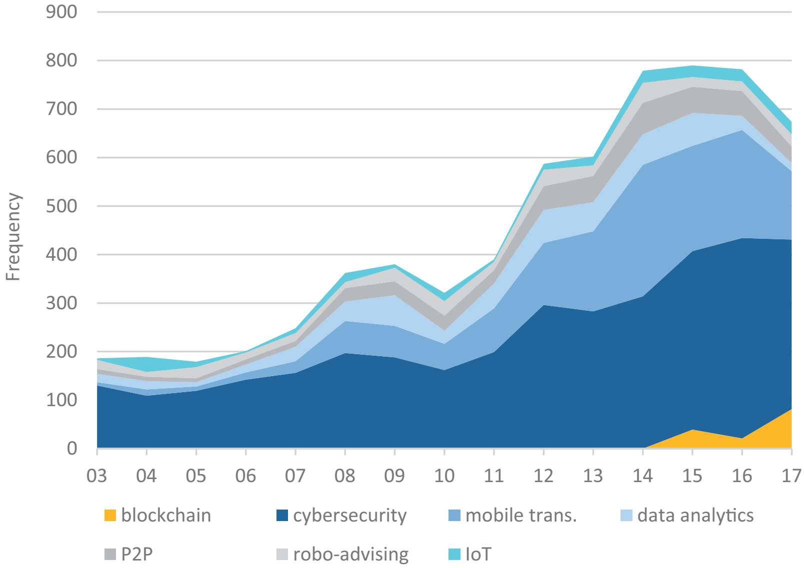


Figure 2
FinTech patent filings over time, by technology category
The stacked area chart in this figure shows, for each year from 2003 to 2017, the number of newly published FinTech patent applications within each of seven technology categories: cybersecurity, mobile transactions, data analytics, blockchain, P2P, robo-advising, and internet of things. For 2017, the number of filing events in each of the last 4 months is estimated as the monthly average of filings in June, July, and August. FinTech applications are identified by applying text-based filtering and machine learning to patent filings drawn from the USPTO's Bulk Data Storage System (BDSS).

corroborates anecdotal evidence that FinTech innovation has been rapidly accelerating over the past decade.

The stacked area chart in Figure 2 shows the annual number of newly published filings within each FinTech category listed in Table 1.[21] It is apparent that innovation in most FinTech categories has accelerated over the sample period. The time-series changes are particularly apparent for blockchain, cybersecurity, mobile transactions, and P2P. Of all the seven patent filing types, blockchain has experienced the highest rate of growth in the last few years. Indeed, when blockchain filings first appear in the sample in 2015, they account for only 5% of all FinTech filings. By 2017, blockchain emerged as the third-largest category of FinTech innovation. Figure 2 also reveals dramatic changes between 2003 and 2017 in the distribution of filings across categories. For example, the percentage of all filings in the cybersecurity category sharply declines from 70% in 2003 to 52% in 2017. By contrast, the share of filings related to mobile transactions increases from 4% to 22%.

[21] Because our data do not cover the last 4 months of 2017, we extrapolate for these months from the average monthly filing count for June–August 2017.

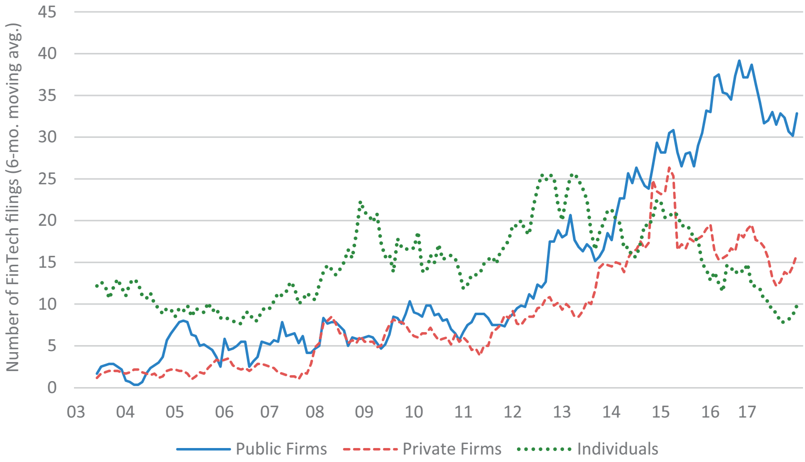


Figure 3
FinTech innovation over time by public firms, private firms, and individuals
This figure shows 6-month moving averages, from June 2003 to August 2017, of the numbers of published FinTech patent applications filed by public firms, private firms, and individuals. FinTech applications are identified by applying text-based filtering and machine learning to patent filings drawn from the USPTO's Bulk Data Storage System (BDSS).

Which types of innovators are most responsible for the dramatic growth in FinTech innovation? Figure 3 plots 6-month moving averages of the numbers of FinTech patent applications filed by each of three innovator groups: public companies, private companies, and individuals. Notably, individuals contribute the largest portion of FinTech innovation in earlier years. Company innovators lag in FinTech patent filings until 2014 but thereafter gain a leading position. Given that financial resources as well as economies of scale and scope can be important for innovation productivity, it is unsurprising that public companies are most responsible for the recent spike in FinTech applications. Nevertheless, private firms also play an important role and contribute about one quarter of FinTech filings after 2013.

Figure 4 considers whether FinTech innovation is primarily driven by financial companies or by firms operating outside of the financial sector. Panel A reveals an acceleration in FinTech patent applications by both groups, especially in recent years. It is noteworthy that nonfinancial firms dominate the filing activity in most time periods. This fact suggests that FinTech innovation often relies critically on basic advances in nonfinancial areas, such as computer science, IT, and software technology. It is also of interest to compare the filing activity of well-established financial firms versus young outsider firms. Panel B of Figure 4 plots innovation frequencies of publicly traded financial firms and FinTech startups. The plot shows that FinTech startups have filed a large number of patent applications over the past several years. Concurrently, the

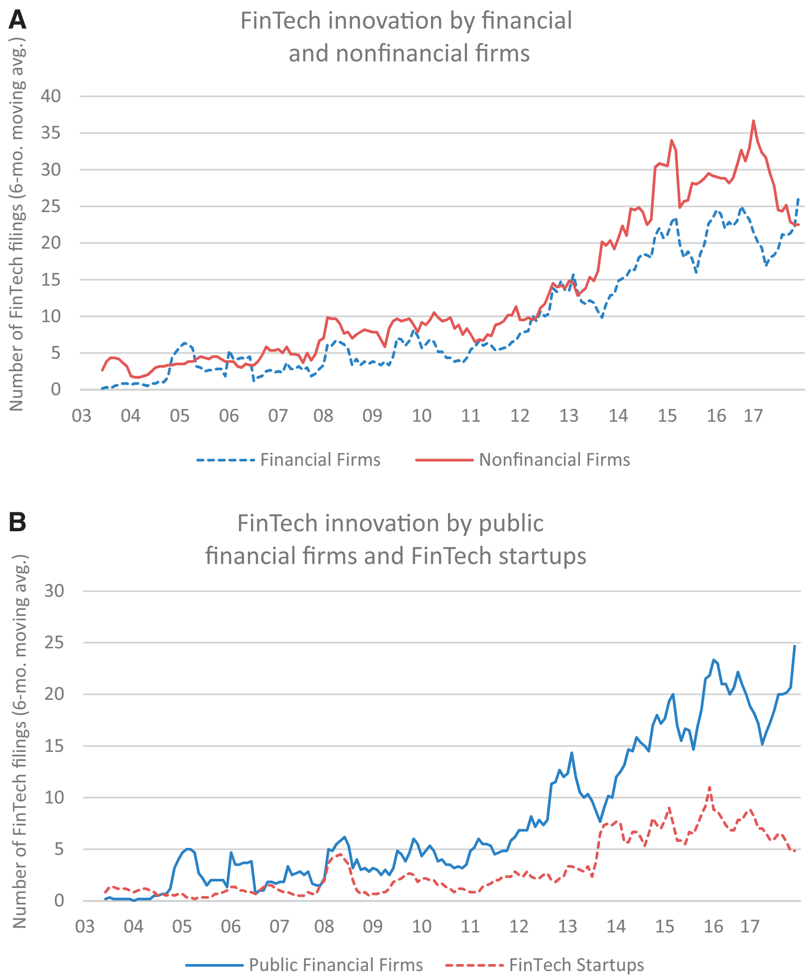


Figure 4
FinTech innovation over time by financial and nonfinancial firms
This figure shows 6-month moving averages, from June 2003 to August 2017, of the numbers of published FinTech patent applications filed by different groups of firms. Panel A shows moving averages for financial firms and nonfinancial firms. Panel B shows moving averages for public financial firms and FinTech startups. FinTech startups are defined as firms operating outside of the traditional financial services sector that were founded 8 years or less prior to the filing news date. FinTech applications are identified by applying text-based filtering and machine learning to patent filings drawn from the USPTO's Bulk Data Storage System (BDSS).

number of applications by public financial firms has strongly increased. This pattern is consistent with the idea that incumbents may often pursue FinTech innovation for defensive or preemptive purposes.

3. The Value of FinTech Innovation

3.1 Estimating values with stock market reactions

Our empirical analysis requires having reliable estimates of the values of individual FinTech innovations. Since the work of Pakes (1985) and Austin (1993), the literature on corporate innovation has recognized that stock price reactions can be used to study the value of new patents. What is less well appreciated, however, is that the price response to a patent event reflects a surprise component: market investors may anticipate an event's future arrival and partially incorporate this anticipation into a firm's stock price today. Thus, without correcting for rational anticipation, the abnormal stock-price reaction to a patent event will give a biased estimate of the intrinsic value of the innovation. Most studies of price reactions to patents do not explicitly account for partial anticipation by investors. A notable exception, however, is Kogan et al. (2017), who estimate patent values using anticipation-adjusted price reactions to patent grants.

In reality, the market may anticipate not just one future innovation during a given time period, but possibly two, three, four, or more. We proceed to outline a method for recovering the underlying value of a FinTech innovation in the presence of anticipation of multiple innovation events. Our method is sufficiently general that it can be used with different models of patent count data (Poisson, negative binomial, zero-inflated Poisson, etc). However, for simplicity, we focus here on the well-known Poisson count distribution used by Hausman, Hall, and Griliches (1984) and others in studies of patenting activity.

Let V_0 be the intrinsic value of a firm without a patent event and let V^* be the incremental value of one patent event to the firm. Assume the number of patents, N, that will occur during the time interval $(t, t+T]$ follows a Poisson count distribution:

$$\Pr(N=m|I_t)=\frac{\lambda^m e^{-\lambda}}{m!}, \quad m=0,1,\ldots \tag{1}$$

where I_t is the information set of market participants at time t. Let the incremental time $t+T$ value to the firm be mV^* if exactly m patent events occur. Then the ex ante market value of the firm before any patent event occurs is

$$\bar{V}_0=V_0+\sum_{m=1}^{\infty}\frac{\lambda^m e^{-\lambda}}{m!}\left(mV^*\right)=V_0+\lambda V^* \tag{2}$$

Assuming that the patent events are independent, then the occurrence of one patent event yields a conditional distribution over total end-of-period patents that is effectively a Zero-Truncated Poisson distribution:

$$\Pr(N=m|N\geq 1, I_t)=\frac{\lambda^m e^{-\lambda}}{(1-e^{-\lambda})m!}, \quad m=1,2,\ldots \tag{3}$$

Therefore, the *ex post* market value of the firm after one patent event occurs is equal to

$$\begin{aligned}\bar{V}_1 &= V_0 + \sum_{m=1}^{\infty} \Pr(N=m|N \geq 1, I_t) m V^* \\ &= V_0 + \sum_{m=1}^{\infty} \frac{\lambda^m e^{-\lambda}}{(1-e^{-\lambda})m!} m V^* \\ &= V_0 + \frac{\lambda}{1-e^{-\lambda}} V^*\end{aligned} \tag{4}$$

From Equations (2) and (4), it follows that the incremental value of a patent is given by

$$V^* = \frac{\Delta \overline{V}}{\frac{\lambda}{1-e^{-\lambda}} - \lambda} = \frac{e^{\lambda}-1}{\lambda} \Delta \overline{V}, \tag{5}$$

where $\Delta \overline{V} \equiv \overline{V}_1 - \overline{V}_0$ is the observed change to the market value of the firm upon occurrence of the patent event. Equation (5) gives a straightforward method of calculating the incremental value of a patent, V^*, from observational data. In particular, the observed market value change $\Delta \overline{V}$ can be computed from abnormal stock price reactions, and the Poisson intensity parameter λ can be estimated from an empirical model of patent counts like in Hausman, Hall, and Griliches (1984).

3.2 Estimating innovation intensities

To construct time-varying estimates of the intensity parameter λ in the above model, we fit a series of Poisson regressions using innovator-year panel data on patent filing counts. Since innovation intensities could depend systematically on both the nature of the underlying technology and on innovator characteristics, we fit separate models for different pairwise combinations of technology type and innovator type (public firms, private firms, and individuals). In total, we estimate $8 \times 3 = 24$ different regression models, including 21 models for the seven FinTech categories and three "benchmark" models for other (non-FinTech) financial innovations.

In the case of public firms, for a given technology category k we estimate the following regression using maximum likelihood estimation (MLE):

$$\begin{aligned}\log(\lambda_{i,k,t}) = \alpha &+ \beta_1 Size_{i,t} + \beta_2 RD_{i,t} + \beta_3 RD_{i,t-1} + \beta_4 RD_{i,t-2} + \beta_5 RD_{i,t-3} \\ &+ \beta_6 Age_{i,t} + \beta_7 PriorFinTech_{i,t} + \beta_8 PriorOtherFinancial_{i,t} \\ &+ \beta_8 PriorNonFinancial_{i,t} + \gamma_i + \delta_t + \varepsilon_{i,k,t}\end{aligned}, \tag{6}$$

where i and t are indices for the innovating firm and year, respectively. In the regression, $Size_{i,t}$ is total assets (in 2003 dollars); $RD_{i,t-n}$ is R&D expenditures

$n+1$ years prior to the current year (in 2003 dollars); *Age* is the number of years since founding of the company; $PriorFinTech_{it}$, is the company's stock of FinTech applications before year t; $PriorOtherFinancial_{i,t}$ is the company's stock of non-FinTech financial applications before year t; $PriorNonFinancial_{i,t}$ is the company's stock of nonfinancial filings in Class G&H before year t; and γ_i and δ_t capture innovator and year fixed effects, respectively. All nonindicator variables are in a natural log form.

In the case of private firms, we estimate the following regression:

$$\begin{aligned}\log(\lambda_{i,k,t}) = \alpha + \beta_1 Age_{i,t} + \beta_2 PriorFinTech_{i,t} + \beta_3 PriorOtherFinancial_{i,t} \\ + \beta_4 PriorNonFinancial_{i,t} + \gamma_i + \delta_t + \varepsilon_{i,k,t}\end{aligned}. \quad (7)$$

Likewise, we estimate the following regression for individual innovators:

$$\begin{aligned}\log(\lambda_{i,k,t}) = \alpha + \beta_1 PriorFinTech_{i,t} + \beta_2 PriorOtherFinancial_{i,t} \\ + \beta_3 PriorNonFinancial_{i,t} + \gamma_i + \delta_t + \varepsilon_{i,k,t}\end{aligned}. \quad (8)$$

Table 6 reports the results of the Poisson regressions. Panels A, B, and C are based on filer-years corresponding to public firms, private firms, and nonfirm individuals, respectively.[22] As seen in the table, for most categories, public firms that are larger tend to file more FinTech patent applications. Among private firms, firm age and the extent of prior non-FinTech filings are strong positive predictors of FinTech innovation. Finally, for individuals, the most consistent predictor of FinTech filing activity appears to be prior innovation experience in non-FinTech financial areas.

3.3 The private value of FinTech innovation

In this section, we examine how much value publicly traded financial companies obtain from their own FinTech patent filings. To infer these "private values," we combine the Poisson intensities estimated above with cumulative market-adjusted abnormal returns (CARs) around news of patent filings. Specifically, for an innovation of technology type k filed by company i and published on date t, the empirical analogue of Equation (5) yields an estimate of the innovation's value to the company:

$$V_{i,k,t}^{*,Own} = \frac{e^{\hat{\lambda}_{i,k,t}} - 1}{\hat{\lambda}_{i,k,t} \times n_{i,t}} CAR_{i,t} M_{i,t}, \quad (9)$$

where $\hat{\lambda}_{i,k,t}$ is the predicted firm-level innovation intensity from the Poisson regressions estimated in Section 3.2 (winsorized at the 1st and 99th percentiles);

[22] Note that, in each panel, the sample sizes in Columns 1 through 7 are smaller than the sample size in Column 8. The difference is because the Column 8 regressions include filer-years where the filer made at least one financial patent application during the sample period, whereas the other regressions only include filer-years where the filer made at least one FinTech application during the sample period.

Table 6
Poisson count models of FinTech and financial innovation by public firms, private firms, and individuals

A. Public firms

	Cybersecurity (1)	Mobile trans. (2)	Data analytics (3)	Blockchain (4)	P2P (5)	Robo-advising (6)	IoT (7)	Other financial (8)
Total assets	0.943***	0.907***	−1.277**	−35.903*	−0.355	2.017**	1.410	0.776***
	(0.193)	(0.349)	(0.541)	(20.399)	(0.486)	(0.827)	(0.887)	(0.066)
R&D	0.073	1.753***	−0.172	59.590**	1.988	4.745	−2.000	−0.042
	(0.304)	(0.557)	(1.720)	(29.945)	(1.362)	(3.061)	(1.724)	(0.140)
R&D_1	0.048	−0.673	1.908	34.657**	3.738	−3.871	2.863	−0.094
	(0.300)	(0.660)	(2.245)	(16.662)	(2.336)	(3.456)	(2.490)	(0.204)
R&D_2	−0.236	0.364	0.287	−22.221**	−7.361***	−0.350	−5.094	−0.091
	(0.288)	(0.662)	(2.050)	(10.288)	(2.344)	(2.692)	(6.064)	(0.228)
R&D_3	−0.273	−0.935*	−1.241	29.204**	2.111*	0.407	6.846	0.047
	(0.248)	(0.514)	(1.443)	(14.246)	(1.129)	(1.565)	(5.348)	(0.168)
Age	−0.201	−0.093	−3.693**	37.648	−0.295	−5.210*	−12.298***	−1.194***
	(0.591)	(1.329)	(1.514)	(139.437)	(2.107)	(2.658)	(2.827)	(0.250)
Prior applications (FinTech)	0.111	−0.092*	0.244	−5.566	−1.126***	−0.447	0.541*	0.300***
	(0.110)	(0.174)	(0.309)	(3.711)	(0.339)	(0.430)	(0.279)	(0.040)
Prior applications (other financial)	0.253**	0.197	0.358	27.971**	1.243***	0.405	0.378	0.018
	(0.101)	(0.176)	(0.278)	(12.487)	(0.305)	(0.513)	(0.278)	(0.038)
Prior applications (nonfinancial)	0.123*	0.035	−0.332*	−22.240*	0.010	0.010	−0.972***	0.172***
	(0.069)	(0.135)	(0.201)	(12.153)	(0.200)	(0.333)	(0.242)	(0.034)
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	1,757	1,757	1,757	1,757	1,757	1,757	1,757	2,852

B. Private firms

	Cybersecurity (1)	Mobile trans. (2)	Data analytics (3)	Blockchain (4)	P2P (5)	Robo-advising (6)	IoT (7)	Other financial (8)
Age	1.145***	1.730***	3.399***	11.987**	3.547***	0.541	0.190	0.905***
	(0.166)	(0.306)	(0.628)	(5.550)	(0.725)	(0.447)	(1.583)	(0.079)
Prior applications (FinTech)	−0.555***	−1.103***	−1.643***	−2.790**	−1.105***	−3.663***	−1.967***	1.115***
	(0.090)	(0.137)	(0.267)	(1.304)	(0.270)	(0.512)	(0.698)	(0.041)
Prior applications (other financial)	0.619***	0.818**	0.966***	8.111***	0.867**	0.408	−0.459	−1.265***
	(0.101)	(0.174)	(0.281)	(2.982)	(0.369)	(0.389)	(0.612)	(0.044)
Prior applications (nonfinancial)	0.576***	1.120***	1.745***	0.997	1.372***	2.900***	1.141*	0.723***
	(0.082)	(0.125)	(0.277)	(1.295)	(0.271)	(0.524)	(0.656)	(0.038)
Firm fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	7,365	7,365	7,365	7,365	7,365	7,365	7,365	23,745

Table 6
(Continued)

C. Individuals

	Cybersecurity (1)	Mobile trans. (2)	Data analytics (3)	Blockchain (4)	P2P (5)	Robo-advising (6)	IoT (7)	Other financial (8)
Prior applications (FinTech)	−1.684***	−1.525***	−0.965***	−74.508	−1.792***	−6.367***	−2.503***	2.205***
	(0.083)	(0.140)	(0.154)	(225.781)	(0.252)	(0.598)	(0.416)	(0.112)
Prior applications (other financial)	1.598***	1.247***	0.094	61.036	1.233***	0.857	0.586	−3.448***
	(0.120)	(0.219)	(0.470)	(406.907)	(0.345)	(0.866)	(0.674)	(0.058)
Prior applications (nonfinancial)	−0.035	−0.252	−0.443***	−7.663	0.654	−0.019	0.712*	0.691***
	(0.081)	(0.171)	(0.215)	(191.575)	(0.467)	(0.475)	(0.398)	(0.066)
Individual fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	22,275	22,275	22,275	22,275	22,275	22,275	22,275	85,875

This table reports the results of Poisson regressions estimating the count intensities of FinTech and other financial patent applications from 2003 to 2017. Panels A, B, and C are based on filer-years corresponding to public firms, private firms, and individual inventors, respectively. Each regression in each panel is estimated for a specific technology type. Regressions in Columns 1 through 7 include only filer-years where the filer made at least one FinTech patent application during the sample period. Regressions in Column 8 include only filer-years where the filer made at least one financial patent application during the sample period. *Total Assets* is measured as of the fiscal year immediately preceding the year of a filing news event. *R&D*, *R&D_1*, *R&D_2*, and *R&D_3* are research and development expenditures 1, 2, 3, and 4 years, respectively, before the year of a filing news event (missing values of R&D are treated as zero). *Prior applications (FinTech)* is a filer's number of prior FinTech applications. *Prior applications (other financial)* is a filer's number of prior non-FinTech financial applications. *Prior applications (nonfinancial)* is a filer's number of prior nonfinancial applications in Class G&H. *Age* is the number of years since a firm's founding. All independent variables shown are in natural logarithms. Each regression also includes filer fixed effects and year fixed effects. Robust standard errors are reported in parentheses. *$p<.1$; **$p<.05$; ***$p<.01$.

$n_{i,t}$ is the number of filings by company i that are published on date t; $CAR_{i,t}$ is calculated over a 4-day window starting 2 trading days before the publication date t; and $M_{i,t}$ is the firm's market capitalization 5 trading days prior to date t. Innovation values estimated from (9) are converted into 2003 U.S. dollar values.

Table 7 reports summary statistics for private values of innovation within nine different groups: the seven distinct FinTech categories, the set of all FinTech innovations, and the set of non-FinTech financial innovations. The table also reports mean CARs for each of the nine groups. As seen in the table, FinTech innovations create economically sizeable private value: the average value to the innovator is \$19.7 million, and the median value is \$35 million. By comparison, non-FinTech financial innovations yield a much lower median value of \$2.3 million, although the average value is close to that of FinTech innovations. These magnitudes are consistent with the average and median U.S. patent values of \$19.1 million and \$6.0 million, respectively, that Kogan et al. (2017) estimate using data on patent grants.[23] Across innovation categories, the mean CARs are mostly positive. The mean and median private values for some categories have opposite signs, suggesting a high degree of skewness in the distributions. Nevertheless, the median values are almost all positive, with the sole exception being data analytics. The largest median values are in blockchain (\$98.1 million), cybersecurity (\$52.9 million), and robo-advising (\$49.1 million).

To assess statistical significance of the mean and median private values, we use a two-stage bootstrapping procedure, which is described in detail in Appendix C. We do not use standard tests (e.g., t-tests or nonparametric sign tests) here because they could lead to biased inferences.[24] Table 7 reports, in parentheses, the bootstrapped (two-tailed) p-values for means and medians. The results show that the positive medians for cybersecurity, mobile transactions, blockchain, and IoT are all statistically different from zero. Medians for each of the other three categories within FinTech are not significantly different from zero, but the positive median for all FinTech innovations is highly statistically significant. Thus, taken together, the p-values in Table 7 support the view that FinTech innovations tend to bring substantial private value to innovators.[25]

[23] See Table 1 in Kogan et al. (2017). To facilitate comparisons, we report the values in Kogan et al. (2017) in terms of their equivalents in 2003 U.S. dollars.

[24] Using standard tests in the current setting gives rise to three difficulties. First, the Poisson regressions in Table 6 introduce estimation error that can differ systematically by technology category and/or innovator type, and such differences can propagate via Equation (5) to innovation values. Second, the innovation values themselves may be heteroskedastic. Third, it is evident that the value distributions are skewed and, as such, depart substantially from normality.

[25] Although we use per-patent values in our main analysis, it is of interest to gauge the aggregate importance of each FinTech category. To do so, we divide each patent filing's private value by the ratio of 1-year lagged nominal U.S. GDP to 2002 nominal U.S. GDP and then sum the resulting scaled values within each category. We find that cybersecurity is the most valuable FinTech category, with an aggregate private value of \$28.01 billion. This is

Table 7
The private value of FinTech innovation

Innovation type	N	Mean CAR (%)	Value Mean	Value Median	Value SD	Value p10	Value p90
Cybersecurity	643	0.26	57.7	52.9	1,658.4	−819.8	902.6
			(0.410)	(0.004)			
Mobile transactions	271	0.42	43.1	18.9	1,013.8	−607.7	707.9
			(0.456)	(0.092)			
Data analytics	181	−0.31	−98.1	−45.3	1,803.8	−715.2	663.4
			(0.492)	(0.166)			
Blockchain	42	0.24	−105.9	98.1	975.2	−653.8	264.5
			(0.544)	(<0.001)			
P2P	95	−0.33	−30.1	1.2	884.0	−668.6	744.6
			(0.772)	(0.874)			
Robo-advising	54	0.04	93.5	49.1	791.9	−1,142.3	1,096.6
			(0.322)	(0.278)			
IoT	86	0.30	−20.8	32.2	973.4	−397.1	540.9
			(0.916)	(0.096)			
All FinTech	1,372	0.17	19.7	35.0	1,439.9	−668.6	734.1
			(0.592)	(<0.001)			
Other financial	2,719	0.20	20.7	2.3	3,141.0	−1,081.0	1,206.5
			(0.676)	(0.588)			

This table reports summary statistics for the private (own-firm) value effect of different categories of FinTech innovation. Values, measured in millions of 2003 U.S. dollars, are calculated like in Equation (9) from public financial firms' abnormal stock returns around news of their own patent filings. Cumulative abnormal returns (CARs) are calculated over the 4-day event window beginning 2 days before the date of patent filing news. Two-tailed p-values for means and medians, reported in parentheses, are calculated using a two-stage bootstrapping method as described in Appendix C.

Next, we use multivariate regressions to investigate how private values of FinTech innovations depend on the underlying technologies. First, to mitigate the effects of skewness and outliers in the value distribution, we apply a logarithmic transformation to the estimated private values using the following formula:

$$V=\begin{cases} \log(1+V^*), & \text{if } V^*>0 \\ -\log(1-V^*), & \text{if } V^*<0 \end{cases}, \tag{10}$$

where V^* is the estimated value from Equation (9) and V is the transformed value to be used as the dependent variable in regressions. We then estimate regressions of the following form:

$$V_{i,k,t}^{OWN}=\alpha_i+\beta' TechnologyDummies_k+\Gamma' X_{i,k,,t}+\varepsilon_{i,k,t}, \tag{11}$$

where $V_{i,k,t}^{OWN}$ is the log-transformed private value to firm i of its own patent filing of technology type k on date t. In these regressions, *TechnologyDummies* are binary variables that capture the different FinTech types (mobile transactions is the omitted category). The term X includes controls for firm size, firm age, prior FinTech filings, prior filings in other (non-FinTech)

followed by mobile transactions and robo-advising, which account for aggregate private values of $10.44 billion and $3.96 billion, respectively.

financial areas, and prior nonfinancial filings in Class G&H. Also included are controls for patent quality, patent breadth, and firm and year fixed effects. All nonindicator controls are in a natural log form.

Table 8 reports the regression results. Column 1 shows that blockchain is the most valuable category of innovation, followed by robo-advising. These categories are associated with significantly more valuable innovation compared to the baseline category of mobile transactions. The other four categories (cybersecurity, data analytics, P2P, and IoT), however, do not significantly differ from mobile transactions. Columns 2 and 3 show that, after controlling for firm size, firm age, prior innovation activity, patent breadth, and patent quality, blockchain continues to be the most valuable category. This finding is consistent with the suggestion in industry reports and the popular press that blockchain technology can offer large potential benefits in terms of future cost savings in financial services.

3.4 Industry-level value of FinTech innovation

As a next step, we examine the value impact that FinTech innovations have on financial industries. To measure industry-level value effects, we start by calculating, for each FinTech patent filing, the value-weighted 4-day CARs across firms in a given financial industry or in the financial sector as a whole. Then, for an industry i and a patent filing of technology type k filed by innovator j and published on date t, we estimate the industry-value impact using the following empirical analogue to Equation (5):

$$V^{*,IND}_{i,j,k,t} = \frac{e^{\hat{\lambda}_{j,k,t}} - 1}{\hat{\lambda}_{j,k,t} \times n_{k,t}} CAR_{i,t} M_{i,t}, \tag{12}$$

where $\hat{\lambda}_{j,k,t}$ is the predicted firm-level innovation intensity from the Poisson regressions estimated in Section 3.2; $n_{k,t}$ is the total number of filings in category k that are published on date t; $CAR_{i,t}$ is the 4-day value-weighted CAR for industry i beginning 2 trading days before the patent publication date t; and $M_{i,t}$ is the total market capitalization of the industry 5 trading days prior to date t. Equation (12) applies to patent filings by all of the major innovator types: public firms, private firms, and nonfirm individuals.

We summarize the industry value impact of FinTech innovations in Table 9. Each cell in the table shows, for a given technology-industry pair, the median value impact from innovations within that technology category. Also reported in each cell are the mean value impact (in brackets) and a two-tailed p-value (in parentheses) for a test of zero median. Medians and means are expressed in millions of 2003 dollars. To calculate p-values, we use a similar two-stage bootstrapping procedure as for Table 7, except that here we focus on a patent's industry-value impact rather than its private-value impact.

Column 1 of Table 9 shows that, for the financial sector as a whole, IoT, robo-advising, and blockchain are the most valuable types of innovation, translating

Table 8
Technology categories and the private value of FinTech innovation

	(1)	(2)	(3)
Cybersecurity	0.245	0.256	0.260
	(0.533)	(0.525)	(0.564)
Data analytics	−0.260	−0.280	−0.221
	(0.623)	(0.569)	(0.659)
Blockchain	2.022***	2.060***	2.037***
	(0.652)	(0.627)	(0.667)
P2P	−0.259	−0.245	−0.289
	(1.029)	(0.961)	(1.010)
Robo-advising	1.341*	1.287*	1.289*
	(0.731)	(0.659)	(0.689)
IoT	0.704	0.645	0.791
	(0.947)	(0.896)	(0.819)
Total assets		1.171	1.257
		(1.711)	(1.705)
Age		−2.183	−2.17
		(5.706)	(5.838)
Prior applications (fintech)		0.313	0.267
		(0.665)	(0.669)
Prior applications (other financial)		1.439	1.404
		(1.517)	(1.516)
Prior applications (nonfinancial)		−0.877	−0.826
		(1.289)	(1.291)
Patent claims			−0.407
			(0.239)
Patent classes			0.084
			(0.425)
Firm fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Observations	1,367	1,367	1,367
R^2	.058	.064	.065

This table reports regressions relating the private value of FinTech innovation to the category of the underlying technology. Innovation values, measured in millions of 2003 U.S. dollars, are calculated like in Equation (9) from public financial firms' abnormal stock returns around news of their own patent filings. The dependent variable in the regressions is a log transformation of the innovation value (see Equation (10)). *Cybersecurity, Data analytics, Blockchain*, *P2P*, *Robo-advising, IoT*, and *Mobile transactions* are indicators for technology categories. *Total assets* is in 2003 U.S. dollars and is measured as of the fiscal year immediately prior to the year of a news event. *Age* is the number of years since founding. *Prior applications (FinTech)* is the count of a firm's prior FinTech applications. *Prior applications (other financial)* is the count of a firm's prior non-FinTech financial applications. *Prior applications (nonfinancial)* is the count of a firm's prior nonfinancial applications in Class G&H. *Patent claims* is the number of independent claims in a filing. *Patent classes* is the number of IPC patent codes spanned by a filing. All nonindicator controls are in natural logarithm form. Standard errors, reported in parentheses, are clustered at the firm level. *$p<.1$; **$p<.05$; ***$p<.01$.

into median values of \$18,348 million, \$11,625 million, and \$6,053 million, respectively, in 2003 dollars. However, not all FinTech categories bring positive value to the financial sector. The data analytics category, for example, is associated with a median value impact of −\$5,862 million. Thus, some types of FinTech innovation appear likely to erode future profits throughout the financial sector by opening the door to new business models, new entrants, and increased competition.

Columns 2 to 6 report the technology-specific value effects for each of the five financial industries. Some FinTech categories have innocuous or beneficial effects across all five industries, consistent with the notion that these types of innovations will help firms lower costs across a wide range of financial

Table 9
Industry value effects of FinTech innovation

Fintech category	Number of obs.	Financial services (1)	Banking (2)	Payment processing (3)	Brokerage (4)	Asset management (5)	Insurance (6)
Cybersecurity	3,222	296.9	−121.2	21.7	32.5	37.3	102.1
		[−4,347.3]	[−3,953.2]	[2.2]	[9.0]	[49.6]	[−454.8]
		(0.760)	(0.988)	(<0.001)	(<0.001)	(0.006)	(0.184)
Mobile transactions	1,430	−2,575.4	−3,811.5	49.3	25.6	13.2	−96.1
		[−9,570.5]	[−8,868.1]	[−13.0]	[−39.3]	[−35.0]	[−615.2]
		(0.106)	(0.050)	(<0.001)	(0.146)	(0.596)	(0.410)
Data analytics	562	−5,862.1	−6,350.7	28.7	38.1	46.7	−211.1
		[−21,584.7]	[−21,361.9]	[25.0]	[−137.3]	[77.0]	[−187.5]
		(0.034)	(0.202)	(0.010)	(0.176)	(0.700)	(0.516)
Blockchain	102	6,053.4	5,762.8	−13.5	81.0	−43.8	245.1
		[−1,793.2]	[−3,567.9]	[287.9]	[93.9]	[25.2]	[1,367.8]
		(0.004)	(0.016)	(0.074)	(0.486)	(0.946)	(0.004)
P2P	438	−674.1	−3,808.2	80.2	−170.1	83.2	−335.2
		[−28,702.3]	[−25,253.5]	[26.9]	[−59.6]	[66.9]	[−3,483.0]
		(0.986)	(0.686)	(0.076)	(0.140)	(0.370)	(0.520)
Robo-advising	305	11,624.8	10,494.4	31.2	299.2	162.8	346.3
		[6,266.3]	[6,588.3]	[196.9]	[447.4]	[270.9]	[−1,237.2]
		(0.290)	(0.388)	(0.050)	(0.076)	(0.136)	(0.338)
IoT	218	18,347.7	18,551.9	12.3	124.7	−179.4	−599.4
		[18,039.4]	[20,610.3]	[−178.0]	[−89.7]	[−186.0]	[−2,117.2]
		(0.062)	(0.034)	(0.558)	(0.544)	(0.556)	(0.470)

This table reports median and mean industry value effects of different categories of FinTech innovation. The sample of innovations consists of FinTech patent applications filed by public firms, private firms, and nonfirm inventors. Mean value effects appear in brackets below median value effects. Value effects are in millions of 2003 U.S. dollars and are calculated from value-weighted portfolios of public firms' abnormal stock returns around patent filing news (see Equation (12)). Two-tailed p-values for medians, reported in parentheses, are calculated using a two-stage bootstrapping method as described in Appendix C.

services and products. Other FinTech categories have more heterogeneous effects. For example, the median value impact of blockchain innovation is significantly positive for banking and insurance but significantly negative for payment processing. This suggests that, while blockchain technology can help banks and insurers better serve the needs of customers, it will also eventually disrupt traditional payment systems. Overall, the univariate evidence in Table 9 shows that different categories of FinTech innovation can have very different aggregate value effects, both within and across industries.

4. Disruption, Competition, and the Value of FinTech Innovation

In this section, we empirically examine what determines the wide cross-sectional variation in the value impact of FinTech innovation on industries and firms. We posit that the value effects of an innovation crucially depend on two considerations: (1) how inherently disruptive the underlying technology is to existing lines of business in an industry and (2) whether the innovation is likely to be used by the innovator to enter an industry and compete with incumbent firms. In what follows, we develop and test several hypotheses based on these two considerations.

4.1 Hypothesis development

While innovation can help existing firms reduce costs and improve product quality, it also can yield destructive effects that threaten the very existence of entire industries (Schumpeter, 1942; Aghion and Howitt, 1992; Klette and Kortum, 2004). An innovation could be harmful to an industry because the technology underlying the innovation is *disruptive*, that is, at odds with existing business models and processes. In the case of FinTech, anecdotal evidence suggests that some innovations can be disruptive, some nondisruptive, and others both. For example, P2P platforms that disintermediate lending may be disruptive to traditional banking. Cybersecurity and mobile innovations may be complementary to the traditional payments industry. Blockchain could pose a substantial future threat to the payments industry, yet at the same time numerous banks have embraced it to help lower the costs of interbank settlement.[26]

The disruptiveness of the technology behind an innovation, however, is not the only factor that determines the innovation's value impact. Another key factor is how the innovator will use the innovation. Prior research shows that both industry incumbents and potential entrants have incentives to seek disruptive innovations (see, e.g., Arrow 1962; Christensen 1997; Christensen and Raynor 2003; Etro 2004; Nerkar and Roberts 2004; Cockburn and MacGarvie 2011). But, whereas potential entrants leverage disruptive innovations to successfully launch novel products and gain footholds in new markets, incumbents use disruptive innovations to preempt other innovators[27] or to secure first-mover advantages against future threats. This distinction leads to two basic predictions. First, a disruptive innovation will more adversely affect industry value if the innovation comes from a potential entrant rather than an incumbent. Second, among the innovations of potential entrants, those based on disruptive technologies will be more detrimental to industry value than those based on nondisruptive technologies.

How will different types of FinTech innovation affect the key players within an industry? For instance, will market-share leaders be more affected by a potential entrant's innovation if the innovation is built on disruptive technology? Market leaders tend to focus on their existing customers rather than changing their business lines (Christensen and Raynor 2003), and they are often sluggish in adapting to change because of large sunk costs in fixed assets and marketing (Tang 1988). Moreover, leaders may suffer from "incumbent inertia" that makes them unwilling to alter existing routines, adjust pricing, or cannibalize product lines (see, e.g., Bresnahan 1985; Lieberman and Montgomery 1988;

[26] See, for example, Arnold (2016) and Vigna (2017).

[27] For example, an incumbent firm could preemptively file patent applications to establish intellectual property rights over new technologies without intending to use the technologies in commerce (see Aghion and Griffith 2005; Acemoglu and Akcigit 2012).

Gilbert 2005). This suggests a straightforward empirical prediction: market-share leaders within an industry will experience a greater value loss when a potential entrant's innovation is disruptive than when it is nondisruptive.

A competing prediction is that the magnitude of value loss to market leaders will be smaller when an entrant's innovation is disruptive. That is, market leaders may be able to gain additional advantage over industry rivals in a disruptive environment. For example, market leaders enjoy technical economies of scale and ample financial resources with which they can develop their own new technologies for reducing costs and retaining customers (Blundell, Griffith, and van Reenen 1999; Czarnitzki, Etro, and Kraft 2014). Also, compared to other firms that are mainly focused on imitation, market leaders have greater capacity to develop entirely new product lines that are insulated from industry-wide disruption (Lieberman and Montgomery 1988).

4.2 Results

4.2.1 Measuring disruptiveness. To test our predictions about disruption and competition, we first construct an empirical, data-based measure of disruptiveness.[28] Here, the challenge is that we do not want our measure to directly account for strategic or competitive effects, for example, the possibility that an innovator will use its new technology to enter an industry and acquire rents from incumbents. Rather, we want our measure to capture the fundamental industry effects of the technology behind an innovation. To proceed, we rely on the idea that, even when an entity does not enter competitively into an industry, its innovations can still affect the industry via technological spillovers. This idea is related to the argument of Aghion and Howitt (1992) that a private research sector can be destructive to an industry because private research firms do not internalize the disruptive costs arising from their innovations.

Our empirical definition of technology disruptiveness is based on how negatively the FinTech patent filings of *nonfirm innovators* affect an industry's value. Specifically, for a given industry, we define a FinTech category to be disruptive if it is one of the two most negative categories in terms of median value impact from nonfirm filings. It is important to emphasize that the median calculations exclude public-firm and private-firm filings. The reason for this is that company filers are not "neutral" innovators—they represent a competitive threat of entry into the industry.[29] By focusing on nonfirm patent filers, we can largely abstract away from the confounding effects of strategic entry and capture disruptiveness in a clean manner.

Table 10 lists each FinTech innovation category that our definition identifies as being disruptive, along with the associated median industry-value impact.

[28] Specifying on theoretical grounds which technologies are disruptive to which industries would be impractical given the large number of different technology-industry pairings in our sample.

[29] For the same reason, we do not simply rely on the median industry-value effects reported in Table 9 to measure disruptiveness.

Table 10
Disruptive FinTech categories by financial industry

	Industry				
	Banking	Payments	Brokerage	Asset management	Insurance
Most-disruptive technology category	Blockchain [−9,117.5]	IoT [−209.1]	P2P [−591.5]	Blockchain [−535.4]	Robo-advising [−526.7]
2nd-most-disruptive technology category	Cybersecurity [−368.2]	Blockchain [−74.9]	Robo-advising [−334.6]	IoT [−253.2]	P2P [−358.6]
Diff. in medians (top-two disruptive categories vs. others)	−4,744.0	−224.5	−427.7	−270.6	−418.4
p-value, permutation test for diff. in medians	.021	<.001	<.001	<.001	.028

This table shows disruptive FinTech categories for each financial industry. Disruptive FinTech categories are those that have either the most negative or second-most negative median industry value impact (calculated using only the patent filings of nonfirm inventors) among all categories. The median industry value impact for each technology-industry pair is reported in brackets. Industry value impacts are in millions of 2003 U.S. dollars and are calculated from value-weighted portfolios of public firms' abnormal stock returns around patent filing news (see Equation (12)). The third row of the table shows, for each industry, the difference between the median value impact from disruptive innovations and the median value impact from nondisruptive innovations. The last row in the table reports one-tailed p-values for tests that the median disruptive value impact is less than the median nondisruptive value impact. Tests of differences are based on permutation tests (see Appendix D).

For each industry, we use a permutation test (details provided in Appendix D) to test whether the median value impact of disruptive innovations is more negative than the corresponding median from all other innovations. The p-values from these permutation tests confirm that our definition of disruptiveness is well-formulated with respect to industry-value impact. It is also noteworthy that some of the results in Table 10 stand in stark contrast to our earlier findings about the median industry-value impact across all patent filings. For example, recall from Table 9 that blockchain filings by all innovators (individuals and firms together) have a positive median value of $5.8 billion for the banking industry. But Table 10 shows that blockchain-related filings by nonfirm innovators have a *negative* median value impact, -$9.1 billion, for banking. Such differences underscore the importance of restricting attention to nonfirm innovators when gauging the inherent disruptiveness of FinTech.

4.2.2 Industry-level value impact: Disruption and competition. To test our predictions about how disruption and competition relate to the industry-value effects of FinTech innovation, we first identify innovators that pose an elevated threat of competitive entry. Intuitively, it is the young, nonfinancial, innovating firms that should present the largest competitive threat to an industry because they do not have established business lines that would suffer should their innovation cause widespread disruption throughout the industry. We thus define a *FinTech Startup* to be a patent-filing firm in our sample that has a nonfinancial industry code and is no more than 8 years old as measured from its date of founding.[30]

[30] Our main qualitative results are similar if we define FinTech startups according to a 6-year age cutoff or a 10-year age cutoff.

We estimate panel regressions that explain the value impact of innovations on the five financial industries. The sample includes all patent filing news events associated with public-firm and private-firm innovators. Specifically, we estimate variants of the following:

$$\begin{aligned} V_{i,j,k,t}^{IND} = \beta_0 + \beta_1 Disruptive_{i,k} + \beta_2 FinTechStartup_j \times Disruptive_{i,k} \\ + \beta_3 FinTechStartup_j \times NonDisruptive_{i,k} + \Gamma' X_{i,j,k,t} + \varepsilon_{i,j,k,t} \end{aligned}, \quad (13)$$

where $V_{i,j,k,t}^{IND}$ is the log-transformed (see Equation (10)) value effect on industry i of the filing news event on date t associated with innovator j and technology type k. In these regressions, $Disruptive_{i,k}$ and $Nondisruptive_{i,k}$ are indicator variables equal to 1 and 0, respectively, for disruptive innovation events; $FinTechStartup_j$ is an indicator equal to 1 if the innovator j is a FinTech startup; and $X_{i,j,k,t}$ represents a set of control variables. We estimate regressions with ordinary least squares and cluster standard errors at the industry-technology level.

The coefficients β_2 and β_3 allow us to separately test both of our predictions about the effects of disruptive FinTech and strategic entry. Specifically, if disruptive innovations are more negative when brought by potential entrants rather than incumbents, then we expect β_2 to be negative. Likewise, if potential entrants' innovations are more detrimental when they are disruptive in nature, then we expect β_2 to be strictly less than β_3. We do not have strong *a priori* expectations on the coefficient β_1 since disruptive innovations by incumbent firms could have ambiguous effects on industry value.

The control variables in our regressions include the following innovator characteristics: age since founding, prior stock of FinTech applications, prior stock of non-FinTech financial applications, and prior stock of nonfinancial filings (all in a natural log form). These variables proxy for a filer's general ability to drive FinTech innovation. To capture the potential quality and importance of patent filings, we also control for the number of independent claims in a filing and the number of IPC codes covered by a filing (both in a natural log form). Finally, to account for time-invariant industry characteristics as well as potential innovation cycles, we control for industry fixed effects, year fixed effects, and, in some specifications, industry-year and technology-year fixed effects.

Table 11 reports the results. In Column 1, we first estimate a basic specification that includes *Disruptive* and *FinTechStartup* but excludes any interaction terms, filer-level controls, or patent-level controls. The coefficient estimates show that the extent of disruptiveness does not seem to explain the value impact of FinTech innovations. However, the coefficient on *FinTechStartup* is negative and significant (p-value =.041), which implies that innovations coming from FinTech startups are generally more harmful to industry value than are innovations from other types of firms.

In Column 2, we test our industry-level predictions from Section 4.1 by replacing *FinTechStartup* with its pairwise interactions with *Disruptive* and

Table 11
Disruptive innovation, FinTech startups, and industry value

	(1)	(2)	(3)	(4)
Disruptive	0.153	0.281	0.267	−0.058
	(0.216)	(0.230)	(0.228)	(0.199)
FinTech startup	−0.315**			
	(0.148)			
FinTech startup × Disruptive		−1.139**	−1.188***	−1.102**
		(0.503)	(0.420)	(0.460)
FinTech startup × Nondisruptive		−0.154	−0.207	−0.422
		(0.150)	(0.270)	(0.264)
Filer's age			0.001	−0.112
			(0.093)	(0.090)
Filer's prior applications (FinTech)			−0.018	−0.004
			(0.105)	(0.111)
Filer's prior applications (other financial)			0.040	0.065
			(0.082)	(0.087)
Filer's prior applications (nonfinancial)				−0.043
				(0.034)
Patent claims				−0.075
				(0.114)
Patent classes				−0.210
				(0.163)
Industry fixed effects	Yes	Yes	Yes	No
Year fixed effects	Yes	Yes	Yes	No
Industry × Year fixed effects	No	No	No	Yes
Technology × Year fixed effects	No	No	No	Yes
Observations	18,090	18,090	18,090	18,090
R^2	.007	.007	.007	.029

This table reports regressions explaining the value impact on financial industries from FinTech innovations by public and private firms. The industry value impact of an innovation is in millions of 2003 U.S. dollars and is calculated from a value-weighted portfolio of public firms' abnormal stock returns around patent filing news (see Equation (12)). The dependent variable in the regressions is a log transformation of the industry value impact (see Equation (10)). *Disruptive* is an indicator equal to 1 if a filing's FinTech type is disruptive to the target industry according to the definition in Table 10. *FinTech startup* is an indicator equal to 1 if the patent filer is a firm existing outside of the financial services sector with an age of 8 years or less (measured from the founding date). *Filer's Age* is the number of years since the filer's founding. *Filer's prior applications (FinTech)* is the number of prior FinTech applications. *Filer's prior applications (other financial)* is the number of prior non-FinTech financial applications. *Filer's prior applications (nonfinancial)* is the number of prior nonfinancial applications in Class G&H. *Patent claims* is the number of independent claims in the patent filing. *Patent classes* is the number of IPC patent codes spanned by a filing. All nonindicator controls are in natural logarithm form. Standard errors, reported in parentheses, are clustered at the industry-technology level. $^{*}p<.1$; $^{**}p<.05$; $^{***}p<.01$.

Nondisruptive. The regression shows that *FinTechStartup*×*Disruptive* has a negative and significant coefficient. In other words, consistent with our first prediction, disruptive innovations are incrementally more negative to an industry when they originate from potential entrants rather than from other firm types. Also, a t-test reveals that the coefficient on *FinTechStartup*×*Disruptive* is significantly less than the coefficient on *FinTechStartup*×*Nondisruptive* (two-sided p-value = .066). This lends support to our second prediction, namely, that an innovation from a FinTech startup should be more harmful to an industry's value when the technology behind the innovation is inherently disruptive.

When controls are added in Column 3 for a filer's capacity to drive and leverage FinTech innovation, the coefficient on *FinTechStartup*×*Disruptive* becomes slightly more negative and remains statistically significant. Moreover, the two interaction coefficients continue to differ according to a two-sided test

(p-value = .063). Finally, in Column 4 we introduce controls for patent quality and also include a full set of industry-year and technology-year indicators that can account for innovation cycles within industries or technologies. In this complete specification, the coefficient on *FinTechStartup*×*Disruptive* is once again negative and significantly less than the coefficient on *FinTechStartup*×*Nondisruptive*. Based on these results, we conclude that disruptiveness and the threat of competitive entry have an interactive effect: when both factors are present in an innovation, the destructive value effects on an industry can be particularly large.

4.2.3 Firm-level value impact: Market-share leaders and disruption. In this section, we test our competing predictions about how the value loss to a market-share leader will vary with the disruptiveness of a FinTech startup's innovation. We define a time-varying indicator variable, *Leader*, equal to one if an incumbent firm is in the top quartile within its industry in terms of revenue-based market share. We then focus on the subsample of innovations by FinTech startups and estimate a series of regressions explaining the firm-level value impact. Specifically, we estimate regressions of the following form:

$$V^{FIRM}_{i,j,k,t} = \beta_0 + \beta_1 Disruptive_{i,k} + \beta_2 Leader_{i,t} + \beta_3 Leader_{i,t} \times Disruptive_{i,k} + \Gamma' X_{i,j,k,t} + \varepsilon_{i,j,k,t}, \tag{14}$$

where $V^{FIRM}_{i,j,k,t}$ is the log-transformed value effect[31] on firm i resulting from the filing news event on date t associated with innovator j and technology type k. The set of controls, $X_{i,j,k,t}$, includes filer and patent characteristics as well as firm, year, industry, and technology indicators or their interactions. We estimate the regressions with OLS and firm-level clustering of standard errors.

Table 12 reports the regression results. In Columns 1 through 3, the estimated coefficient on *Leader*×*Disruptive* is positive and highly statistically significant. Also, the sum of coefficients for *Disruptive* and *Leader*×*Disruptive* is positive and significant across all three regressions. These facts confirm the prediction that market-share leaders find FinTech startup innovations to be less harmful when the underlying technology is disruptive rather than nondisruptive. It is also worth noting that, in all three regressions, the coefficient on *Leader* is negative and highly significant. Although not one of our main predictions, this finding suggests that *nondisruptive* innovation from startups is more harmful to leaders than to other firms in the industry.

How are market leaders able to avoid much of the harm from FinTech startups' disruptive innovation? One plausible explanation is that leaders typically enjoy large financial resources and technical economies of scale, and

[31] The calculation of these value effects proceeds in a manner similar to what was done for private values in Section 3.3 (see Equations (9) and (10)).

Table 12
Incumbent firms and the value effects of innovation by FinTech startups

	(1)	(2)	(3)
Leader × Disruptive	0.283***	0.283***	0.281***
	(0.037)	(0.037)	(0.036)
Disruptive	0.015	0.007	−0.020
	(0.012)	(0.013)	(0.022)
Leader	−0.281***	−0.280***	−0.260***
	(0.054)	(0.054)	(0.053)
Total assets	−0.071***	−0.071***	−0.077***
	(0.025)	(0.025)	(0.025)
Filer's age		−0.064***	−0.081***
		(0.014)	(0.015)
Filer's prior applications (FinTech)		0.048***	0.025***
		(0.006)	(0.006)
Filer's prior applications (other financial)		−0.121***	−0.069***
		(0.011)	(0.011)
Filer's prior applications (nonfinancial)		0.008	0.003
		(0.006)	(0.006)
Patent claims			0.016
			(0.011)
Patent classes			−0.014
			(0.016)
Firm fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	No
Industry × Year fixed effects	No	No	Yes
Technology × Year fixed effects	No	No	Yes
Observations	310,416	310,416	310,416
R^2	.004	.005	.012

This table reports regressions explaining the value impact on financial industry market leaders and followers from innovations by FinTech startups. An innovation's value impact on a firm is in millions of 2003 U.S. dollars and is calculated like in Equation (9) from the firm's abnormal stock returns around news of the patent filing. FinTech startups are defined as firms operating outside of the traditional financial services sector that are no more than 8 years from their founding. The dependent variable in the regressions is a log transformation of the value impact on a firm (see Equation (10)). *Leader* is an indicator equal to 1 if a firm's revenue market share is in its industry's top quartile. *Disruptive* is an indicator equal to 1 if a filing's FinTech type is disruptive as defined in Table 10. *Total assets* is measured in 2003 dollars and is in a natural log form. Table 11 describes *Filer's age, Filer's prior applications (FinTech), Filer's prior applications (other financial), Filer's prior applications (nonfinancial), Patent claims*, and *Patent classes*. Standard errors, reported in parentheses, are clustered at the firm level. $^*p<.1$; $^{**}p<.05$; $^{***}p<.01$.

they can invest heavily in their own innovation to protect themselves from the adverse effects of disruption. To investigate this explanation, we restrict our sample to the set of disruptive innovations from FinTech startups and consider three continuous measures of investment in innovation over the most recent fiscal year: *R&D*, which is simply the log-transformed amount of R&D spending in the year prior to the filing news event; *R&D Intensity*, which is equal to R&D spending divided by total revenues; and *R&D/Assets*. All three measures are computed with Compustat data and winsorized at the 0.1% and 99.9% levels to account for outliers. We then estimate regression models similar to those in Equation (14), except that *Disruptive* and *Leader×Disruptive* are replaced, respectively, with one of the continuous measures of R&D investment and its interaction with *Leader*.

Table 13 reports the results of these regressions. As seen in Column 1, *R&D* has an insignificant coefficient, but the coefficient on *Leader×R&D* is

Table 13
Incumbent R&D spending and disruptive innovation by FinTech startups

	(1)	(2)	(3)
Leader × R&D	0.445*		
	(0.260)		
R&D	−0.127		
	(0.258)		
Leader × R&D intensity		17.329***	
		(5.257)	
R&D intensity		0.314	
		(5.102)	
Leader × R&D/assets			68.371***
			(7.275)
R&D/assets			−2.084
			(6.001)
Leader	0.015	0.018	0.016
	(0.070)	(0.070)	(0.070)
Total assets	−0.164***	−0.168***	−0.164***
	(0.041)	(0.042)	(0.041)
Control variables	Yes	Yes	Yes
Firm fixed effects	Yes	Yes	Yes
Industry × Year fixed effects	Yes	Yes	Yes
Technology × Year fixed effects	Yes	Yes	Yes
Observations	115,195	115,195	115,195
R^2	.005	.005	.005

This table reports regressions explaining the value impact on market leaders and market followers from disruptive innovations by FinTech startups. An innovation's value impact on a firm is in millions of 2003 U.S. dollars and is calculated like in Equation (9) from the firm's abnormal stock returns around news of the patent filing. FinTech startups are defined as firms operating outside of the traditional financial services sector that are no more than 8 years from their founding. The dependent variable in the regressions is a log transformation of the value impact (see Equation (10)). *Leader* is an indicator equal to 1 if a firm's revenue market share is in the industry's top quartile, and 0 otherwise. *R&D* is log-transformed R&D spending in 2003 dollars. *R&D intensity* is R&D spending divided by total revenues. *R&D/assets* is R&D spending divided by the book value of total assets. R&D spending, sales, and total assets are from the most recent fiscal year preceding the year of a filing news event. *Total assets* is measured in 2003 dollars and is in a natural log form. Each regression includes all filer-level and patent-level control variables appearing in Table 12. Standard errors, reported in parentheses, are clustered at the firm level. *$p<.1$; **$p<.05$; ***$p<.01$.

positive and significant at 10%. When we use *R&D Intensity* and *R&D/Assets* as alternative measures of investment in Columns 2 and 3, the interaction coefficients increase substantially in both magnitude and statistical significance. The economic magnitudes of the interaction effects are sizable. Consider, for instance, the coefficient estimate for *Leader×R&D Intensity* in Column 2. Supposing that the value impact from a startup's disruptive innovation is negative, the estimate implies, ceteris paribus, that a 1-percentage-point increase in R&D intensity from its mean translates into roughly a 16.2 percent reduction in the negative value impact for leader firms.[32] We conclude, therefore, that the ability to invest heavily in innovation does play an important

[32] We estimate this reduction in value impact by calculating an approximate average partial effect of R&D intensity in the regression, conditional on the predicted value impact being negative. Specifically, for each observation, we first calculate the predicted (log-transformed) value from the regression and infer the predicted value impact from the log-transformation in Equation (10). We then calculate the predicted value impact for a changed R&D intensity, keeping other variables fixed, and thus obtain the change in value impact for each observation. The mean level of R&D intensity in the sample is about 1.03%.

role in helping market leaders avoid harm from FinTech startups' disruptive innovation.

5. Conclusion

The dramatic surge of interest in FinTech over the past few years has highlighted the need for a better understanding of the value of technological innovations in this space. Our paper provides large-scale evidence to help fill this gap. Using a new data set constructed from the full document texts of patent applications, we study the occurrence of FinTech innovation and the value that it brings to innovators, industries, and incumbent firms.

Our data set enables the application of text-based machine learning to classify FinTech innovations according to their underlying technologies. To obtain value estimates of such innovations, we use a new method that combines observed stock price reactions with estimated Poisson arrival intensities. We find that FinTech innovations are generally valuable to innovators and to the financial sector as a whole. However, for some financial industries, certain types of FinTech innovation can have an adverse value impact. We find that value effects on an industry are more negative when an innovation comes from a young, nonfinancial firm and brings forth disruptive technology. Also, market-share leaders tend to suffer less harm from outside disruptive innovation if they have invested heavily in their own R&D.

We note limitations to our approach of using market reactions to study the value of FinTech innovation. First, despite being an important outcome of the innovation process, patent applications reflect only part of firms' innovation activities. Some firms may have unsuccessful innovation attempts; others may choose to forego the patenting process altogether and rely on trade secrets to protect their discoveries. Second, we cannot accurately measure the direct costs of FinTech innovation due to the aggregated nature of R&D spending. Third, stock-price data are not well suited to studying the impact that FinTech is likely to have on non-U.S. firms, privately held firms, customers, and employees. For many firms and individuals throughout the financial services sector, FinTech innovation will undoubtedly lower costs. At the same time, by paving the way for more automation, FinTech also has the potential to reduce employment and welfare. Future research can study the very important issue of what broader societal impacts FinTech innovation will have going forward.

Appendix A. Using Machine Learning to Classify Patent Filings

A.1 Text Preprocessing

We preprocess the text of each filing using standard methods in applied text analysis (see, e.g., the survey article by Gentzkow, Kelly, and Taddy forthcoming). First, for each financial patent application, we create a document that contains the text of its title, abstract, summary, and claims sections and then "tokenize" the document into a sequence of words. Because English words have different inflections (e.g., singular and plural noun forms or different verb tenses), we then use a stemmer to obtain the unique stems of words. For example, “transforms,” “transformed,” and “transforming” are all converted into the word stem “transform.” We also exclude very common words (so-called “stopwords”), such as “the,” “and,” and “a,” because these words do not contain information useful for classification of patent documents. For this purpose, we use the stopword list of Jegadeesh and Wu (2017) but exclude from this list any words that appear in one or more terms in our financial lexicon constructed in Section 2.1.1. We also rank words by the number of documents they appear in and remove the top-50 words (except for those that appear in at least one of our financial lexicon terms). These top-50 words are generic terms, such as “method” or “background,” that appear in a vast number of patent filings and thus are not helpful for implementing text-based classification methods.

Next, we convert documents into numeric vectors using a “bag of words” approach. Each document is mapped into a numerical vector that contains a frequency score for each individual word in the document. Frequency scores are calculated using the popular “term frequency-inverse document frequency” (*tf-idf*) method (Jones 1972). *Tf-idf* provides a statistic that represents the importance of a word in a document contained in a corpus, or collection of documents. The *tf-idf* statistic increases with the number of times a word appears in a document, while it decreases in the frequency of the word in the corpus. Therefore, *tf-idf* adjusts for both varying lengths of documents and varying commonality of words in the collection of documents.

A.2 Creating a Training Sample

We start by compiling a group of likely FinTech innovators from six annual lists: the *Forbes*’ Fintech 50, the Fintech 100 list of H2 Ventures/KPMG, *American Banker’s* Fintech 100, and the “FinTech companies to watch” surveys in *Inc, Entrepreneur*, and *Fastcompany*. These six sources differ in their focus and selection methods. For instance, the *Forbes* and H2 Ventures/KPMG lists focus on new startup companies, while the *American Banker* list features technology companies that derive a large share of their revenues from financial services. In the case of *Inc, Entrepreneur*, and *Fastcompany*, we use the 2016 editions of the lists. For the other three periodicals, we use both the 2015 and 2016 editions. After combining the various lists and keeping only U.S. firms, we obtain an overall list of 72 distinct firms.

We then augment our list with a set of publicly traded, “traditional” financial firms that are observed to file patent applications. Specifically, we identify the top 10 most prolific Compustat companies—in terms of total patent applications filed and published during the sample period—within each of several industry subgroups.[33] The industry subgroups, based on six-digit NAICS codes as described in the text (Footnote 15), include the following: commercial banking, payment processing, brokerage, asset management, and insurance. In total, we add 44 distinct publicly traded companies from these financial industries to our list.

Next, for each firm in our combined list of likely FinTech innovators, we extract all of the firm’s associated Class G&H patent applications during the sample period. This provides us with 11,431 filings, from which we randomly draw 1,000 cases. We read through and manually classify these 1,000 filings into nine different categories: the seven FinTech types from Table 1, other (non-FinTech) financial filings, and filings unrelated to financial services.

Note that the random sample of 1,000 documents includes only patent filings by companies in the list of likely FinTech innovators we constructed from magazines and Compustat. Therefore, the random sample might not be representative of all FinTech filings by U.S. companies and individuals. To address this issue, we use the 1,000 filings to bootstrap the creation of a more representative training sample. This is done by applying a simple nearest-centroid classifier to the 67,948 text-filtered patent filings identified in Section 2.1.1. Specifically, for each text-filtered patent filing,

[33] Identifying the top patent filers among public financial services companies requires matching assignee names in the BDSS bulk data with the names of Compustat firms. For this purpose, we use exact name matches combined with manual matching.

we calculate its distance to a patent category $i=0,\ldots,8$ by averaging its distance to all category-i patents in the random sample, where distance between two patent filings is defined as one minus the cosine similarity between the filings.

The result of the centroid classifier is a nine-way classification of all 67,948 patent filings in the text-filtered sample. From the filings in each of the resulting nine categories, we then select the 100 patents that are closest to the centroid plus an additional 100 random filings. Thus, in total we have $9\times 200=1,800$ filings. We read through these filings and manually reclassify them as needed. The resulting classified set of 1,800 filings serves as the training sample for all of our machine-learning algorithms.

A.3 Main Classification Algorithms: Support Vector Machines and Neural Networks

One classification method we use is support vector machines (SVM). SVM is a class of supervised machine-learning algorithms that has been applied to a wide range of problems, such as text categorization, image classification, handwriting analysis, and classification of biological proteins. Given a training sample for which each data point is already assigned to one of two classes, Linear SVM searches for an $(n-1)$-dimensional hyperplane that separates the two groups in n-dimensional space and maximizes the margin, that is, the distance from the hyperplane to the closest data point. A key advantage of SVM is that it can deliver good classification performance even for data with a large number of features compared to the sample size. This makes SVM especially well-suited for classifying texts since document collections typically have large numbers of distinct words (see, e.g., Joachims, 1998).

We also use a *Neural Networks* approach to classification. Originally motivated by the brain structure of humans and animals, research on artificial neural networks has recently led to new, deep-learning models that perform well in complex classification tasks involving large data sets. Neural networks have been widely applied in numerous fields, such as natural language processing, speech recognition, computer vision, autonomous driving, and medical sciences. A typical neural network has three parts: (1) an input layer of neurons that are used to receive the input data; (2) an output layer of neurons that produce predictions and results; and (3) one or more hidden layers of neurons in the middle that connect the input and output layers. The number and configuration of hidden layers turns out to be crucial for efficient implementation of the network training process. Goodfellow et al. (2016) provide a detailed treatment of modern machine-learning techniques based on neural networks.

A.4 Hyperparameters and Design Choices for Classifiers

The hyperparameters used in the classifiers are as follows: $k=5$ for kNN, $C=0.6$ for Linear SVM, and $\gamma=1.0$, $C=2.0$ for Gaussian SVM. For the neural network classifier, we follow the suggestion of Masters (1993) and Shibata and Ikeda (2009) and use a geometric pyramidal layout to pin down the numbers of neurons in the various hidden layers. We fit neural network classifiers with one, two, three, four, and five hidden layers and select the configuration with three hidden layers since this yields the best F1 score. In the first, second, and third hidden layers that follow after the input layer, there are 1792, 307, and 53 neurons, respectively.

Appendix B. Matching Patent Filings to Data on Public and Private Firms

We attempt to gather data on key characteristics of the public and private firms that filed patent applications in Categories 1 through 8 (finance-related innovations). Starting with all of the assignees' names reported in BDSS, we conduct a name match against CRSP/Compustat, D&B Hoover's, Standard & Poor's NetAdvantage, LexisNexis, Bloomberg, company websites, Google, and other publicly available websites.

The name matching process for public and private firms often involves some ambiguity. In some cases, an assignee firm appears in the USPTO database with multiple names. In other cases, different assignee firms may have similar names. To ensure that name matches are valid, we extract from USPTO filings the city and state of each assignee firm and look for an exact name-city-state match. If either the assignee's name or address does not match, we use additional information from online Google searches, Bloomberg company profiles, or company websites to determine whether the match is valid.[34] Another source of ambiguity arises from the fact that assignee firms are occasionally acquired by other firms during the sample period, and both acquirers and targets appear in the databases we use. Given the large number of private assignee firms, tracking the merger and acquisition histories for all firms is difficult. Therefore, we screen for cases in which the assignee's name contains keywords in its current parent company's name, and in such cases we use the parent's information for matching.[35] Also, we note that some filers are coded in BDSS as being U.S. firms, but they are in fact universities, foreign companies, or subsidiaries of foreign companies.[36] We remove from the sample a small number of filings that correspond to such cases.

For each public and private firm in the name-matched sample, we gather available data on financials, stock prices, SIC and NAICS industry codes, and the firm's year of founding. The data on stock prices and financials are from CRSP and Compustat. Data on industry codes and founding year are drawn from D&B Hoover's, Standard and Poor's NetAdvantage, LexisNexis, Bloomberg company profiles, Google, Bizapedia, and other web-based sources.

Finally, we observe that, due to IPOs and exchange listings, a small number of filers are publicly traded at the disclosure of an application but were privately held when the application was originally filed. Thus, for each patent filing by a public firm, we compare the filing date to the time window of the firm's available stock prices in CRSP/Compustat. If the filing date does not locate in the time window, we treat the filer as a private firm in the Poisson regression analysis of Table 6.

Appendix C. Using Two-Stage Bootstrapping to Calculate *p*-Values (Tables 7 and 9)

As described in Footnote 24, standard tests for significance of means and medians are not appropriate in the context of Table 7 due to measurement error, heteroskedasticity, and nonnormality. Therefore, we develop a two-stage bootstrapping approach for testing whether the means and medians reported in Table 7 are significantly different from zero (see, e.g., Efron (1979), Efron and Tibshirani (1986), and Mooney and Duval (1993) for general treatments of bootstrapping methods).

In the first stage of our procedure, we bootstrap each of the 24 Poisson regressions in Table 6 to obtain distributions of the estimated innovation intensities. Specifically, we resample each regression sample 200 times with replacement to obtain bootstrapped samples, each of which has the same number of (filer-year) observations as the original regression sample. We then conduct the Poisson regressions on each bootstrapped sample and obtain predicted innovation intensities for each filer-year observation. Pooling the results together, we obtain a distribution of 200 fitted innovation intensities (lambdas) for each filer-year observation in each regression.

The second stage involves building upon the fitted lambdas from the first stage to bootstrap innovation values at the filing level. Specifically, within each innovation category in Table 7, we

[34] For instance, Mass Catalyst Corporation, an assignee located in Arlington, Texas, according to BDSS data, has one possible match in Hoover's with the same corporation name but a different city: Dallas, Texas. A Google search in this case verifies that the match is correct.

[35] An example of such a case is the patent filer Visa International Service Association. Since this filer operates as a subsidiary of Visa Inc. and since its name contains the word "Visa," we use Visa Inc.'s company data for matching purposes.

[36] For example, we exclude filings from Samsung Pay, Inc. which is a U.S. subsidiary of the South Korean conglomerate Samsung. We also exclude filings, such as USPTO document number 20070040019, which was filed by University of Nevada-Las Vegas but incorrectly classified in BDSS as having been filed by a U.S. company.

resample the set of patent filings 1,000 times with replacement to create bootstrapped samples. Each bootstrapped sample of filings is then merged with the appropriate set of 200 possible fitted innovation intensities from the first stage. Simulated private values are calculated like in Equation (9) from a filing's CAR and from each of the 200 fitted lambdas. Note that each of the 1,000 two-stage bootstrapped samples of values consists of $200 \times N$ observations, where N is the number of actual filings in the innovation category. We then calculate the mean and median values from each two-stage bootstrapped sample to obtain distributions of the two statistics. Without assuming normality of the distributions, we can use the nonparametric percentile method (see, e.g., Mooney and Duval 1993), which entails sorting each statistic and using the 2.5th and 97.5th percentiles as the lower and upper bounds of the 95% confidence interval. Similarly, we can obtain specific p-values of the statistics by comparing the bootstrapped distributions of the statistics against zero.

For Table 9, we can also apply two-stage bootstrapping to compute p-values for tests of whether median industry-level values differ from zero. The bootstrapping procedure used here resembles the one used for Table 7 but differs in two respects: (1) the filing-level bootstrap resampling is conducted within each technology type-industry pair; and (2) the bootstrapped industry-level value impacts are computed using Equation (12) rather than Equation (9).

Appendix D. Permutation Tests (Table 10)

Estimation error from the Poisson regressions in Table 6 could systematically affect the calculation of industry-level values. Therefore, for Table 10 we rely on nonparametric permutation tests (see, e.g., Good 1994) to assess whether the median value impact of disruptive technologies on an industry is significantly more negative than that of nondisruptive technologies. The permutation test for a given industry is implemented as follows. First, we first divide the original sample of all patent filings by individual inventors into those that are disruptive and those that are nondisruptive and then calculate the median values for each group. The difference between the two (i.e., disruptive median minus nondisruptive median) is taken to be the original statistic.

We then compute, for each filing in the original sample, all of the 200 possible industry values using the fitted innovation intensities from the first stage of our two-stage bootstrapping procedure (see Appendix C). We then resample this bootstrapped sample *without* replacement, randomly assigning $200 \times D$ observations to an artificial disruptive group and the rest to an artificial nondisruptive group, where D is the number of disruptive filings in the original sample. For a given resampling iteration, we calculate the difference in median values of the two artificial groups (i.e., disruptive median minus nondisruptive median). The resampling is repeated 1,000 times to obtain 1,000 permutation statistics. Finally, we locate the original statistic on the permutation distribution. The p-value for the permutation test is the fraction of the 1,000 permutation statistics that are no larger than the original statistic.

References

Acemoglu, D., and U. Akcigit. 2012. Intellectual property rights policy, competition and innovation. *Journal of the European Economic Association* 10:1–42.

Adner, R. 2012. *The wide lens: A new strategy for innovation*. London: Penguin.

Aghion, P., and R. Griffith. 2005. *Competition and growth: Reconciling theory and evidence*. Cambridge: MIT Press.

Aghion, P., C. Harris, P. Howitt, and J. Vickers. 2001. Competition, imitation and growth with step-by innovation. *Review of Economic Studie*s 68:467–92.

Aghion, P., and P. Howitt. 1992. A model of growth through creative destruction. *Econometrica* 60:323–51.

Antweiler, W., and M. Frank. 2004. Is all that talk just noise? The information content of internet stock message boards. *Journal of Finance* 59:1259–94.

Arnold, M. 2016. Financial industry faces extreme disruption in payments. *Financial Times*, September 26.

Arrow, K. 1962. Economic welfare and the allocation of resources for invention. In *The rate and direction of inventive activity*, ed. R. Nelson, 609–25. Princeton, NJ: Princeton University Press.

Athey, S., J. Tibshirani, and S. Wager. 2018. Generalized random forests. arXiv, arXiv:1610.01271v4, https://arxiv.org/abs/1610.01271.

Athey, S., and S. Wager. 2018. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association* 113:1128–42

Austin, D. 1993. An event-study approach to measuring innovative output: The case of biotechnology. *American Economic Review* 83:253–58.

Bellstam, G., S. Bhagat, and A. Cookson. 2017. A text-based analysis of corporate innovation. Working Paper, University of Colorado.

Blundell, R., R. Griffith, and J. van Reenen. 1999. Market share, market value and innovation in a panel of British manufacturing firms. *Review of Economic Studie*s 66:529–54.

Brav, A., W. Jiang, S. Ma, and X. Tian. 2018. How does hedge fund activism reshape corporate innovation? *Journal of Financial Economic*s 130:237–64

Bresnahan, T. 1985. Post-entry competition in the plain paper copier market. *American Economic Association* 75:15–19.

Carney, M. 2017. The promise of Fintech—Something new under the sun? Speech, Wiesbaden, Germany, January 25.

Christensen, C. 1997. *The innovator's dilemma*. Cambridge, MA: Harvard University Press.

Christensen, C., and M. Raynor. 2003. *The innovator's solution*. Cambridge, MA: Harvard University Press.

Cockburn, I., and M. MacGarvie. 2011. Entry and patenting in the software industry. *Management Science* 57:915–33

Cornaggia, J., Y. Mao, X. Tian, and B. Wolfe. 2015. Does banking competition affect innovation? *Journal of Financial Economics* 115:189–209.

Czarnitzki, D., F. Etro, and K. Kraft. 2014. Endogenous market structures and innovation by leaders: An empirical test. *Economic*a 81:117–39.

Dasgupta, P., and J. Stiglitz. 1980. Industrial structure and the nature of innovative activity. *Economic Journal* 90:266–93.

Downes, L., and P. Nunes. 2013. Big bang disruption. *Harvard Business Review*, March.

The Economist. 2015. The Fintech revolution. May 9.

Efron, B. 1979. Bootstrap methods: Another look at the jackknife. *Annals of Statistics* 7:1–26.

Efron, B., and R. J. Tibshirani. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* 1:54–75.

Etro, F. 2004. Innovation by leaders. *Economic Journal* 114:281–303.

Gentzkow, M., B. Kelly, and M. Taddy. Forthcoming. Text as data. *Journal of Economic Literature*.

Gilbert, C. 2005. Unbundling the structure of inertia: Resource versus routine rigidity. *Academy of Management Journal* 48:741–63

Gilbert, R., and D. Newbery. 1982. Preemptive patenting and the persistence of monopoly. *American Economic Revie*w 72:514–26.

Glaeser, E., A. Hillis, S. Kominers, and M. Luca. 2016. Predictive cities crowdsourcing city government: Using tournaments to improve inspection accuracy. *American Economic Revie*w 106:114–18

Good, P. 1994. *Permutation tests: A practical guide to resampling methods for testing hypotheses*. New York: Springer.

Goodfellow, I., Y. Bengio, and A. Courville. 2016. *Deep learning*. Cambridge, MA: MIT Press.

Griliches, Z., B. Hall, and A. Pakes, 1991. R&D, patents, and market value revisited: Is there a second (technological opportunity) factor? *Economics of Innovation and New Technology* 1:183202.

Hall, B., A. Jaffe, and M. Trajtenberg. 2005. The NBER Patent Citation Data File: Lessons, insights and methodological tools. NBER Working Paper.

———. 2005. Market value and patent citations. *RAND Journal of Economics* 36:16–38

Hanley, K., and G. Hoberg, 2010. The information content of IPO prospectuses. *Review of Financial Studies* 23:2821–64.

Harrist, M. 2017. How Fintech is powering the global economy. *Forbes*, March 14.

Hastie, T., R. Tibshirani, and J. Friedman. 2009. *Elements of Statistical Learning*. New York: Springer.

Hausman, J., B. Hall, and Z. Griliches. 1984. Econometric models for count data with an application to the patents-R&D relationship. *Econometrica* 52:909–38.

He, J., and X. Tian. 2016. Do short sellers exacerbate or mitigate managerial myopia? Evidence from patenting activities. Working Paper, University of Georgia and Tsinghua University.

Henderson, R., and I. Cockburn. 1996. Scale, scope, and spillovers: The determinants of research productivity in drug discovery. *RAND Journal of Economics* 27:32–59.

Hoberg, G., and G. Phillips. 2016. Text-based network industries and endogenous product differentiation. *Journal of Political Economy* 124:1423–65.

James, G., D. Witten, T. Hastie, and R. Tibshirani. 2013. *An introduction to statistical learning: With applications in R*. New York: Springer.

Jegadeesh, N., and D. Wu, 2013. Word power: A new approach for content analysis. *Journal of Financial Economics* 110:712–29.

———. 2017. Deciphering Fedspeak: The information content of FOMC meetings. Working Paper, Emory University and University of Michigan.

Joachims, T. 1998. Text categorization with Support Vector Machines: Learning with many relevant features. In *Proceedings of the 10th European Conference on Machine Learning*, eds. C. Nédellec and C. Rouveirol, 137–42. New York: Springer.

Jones, K. S. 1972. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation* 28:11–21.

Kelly, E., and P. Tetlock, 2013. How wise are crowds? Insights from retail orders and stock returns. *Journal of Finance* 68:1229–65.

Kleinberg, J., J. Ludwig, S. Mullainathan, and Z. Obermeyer. 2015. Prediction policy problems. *American Economic Review* 105:491–95.

Klette, T., and S. Kortum, 2004. Innovating firms and aggregate innovation. *Journal of Political Economy* 112:986–1018.

Kogan, L., D. Papanikolaou, A. Seru, and Stoffman. 2017. Technological innovation, resource allocation, and growth. *Quarterly Journal of Economics* 132:665–712.

Kortum, S., and J. Lerner. 2000. Assessing the contribution of venture capital to innovation. *RAND Journal of Economics* 31:674–92.

KPMG. 2018a. Global report on fintech investment trends. *The Pulse of FinTech Q4 2017*, February 13.

KPMG. 2018b. Global Fintech investment soars to record US$57B in first half of 2018. *The Pulse of Fintech 2018*, July 30.

Lerner, J. 2009. The empirical impact of intellectual property rights on innovation: Puzzles and clues. *American Economic Review* 99:343–48.

Lieberman, M., and D. Montgomery. 1988. First-mover advantages. *Strategic Management Journal* 9:41–58.

Loughran, T., and B. McDonald, 2011. When is a liability not a liability? Textual analysis, dictionaries, and 10-Ks. *Journal of Finance* 66:35–65.

Manela, A., and A. Moreira, 2017. News implied volatility and disaster concerns. *Journal of Financial Economics* 123:137–62.

Mann, W. 2018. Creditor rights and innovation: Evidence from patent collateral. *Journal of Financial Economics* 130:25–47.

Masters, T. 1993. *Practical neural network recipes in C++*. New York: Academic Press.

McKinsey & Co. 2015. Beyond the hype: Blockchains in capital markets. Working Paper, McKinsey Working Papers on Corporate & Investment Banking.

Mooney, C. and R. Duval. 1993. *Bootstrapping: A nonparametric approach to statistical inference*. Sage Publications.

Naik, N., R. Raskar, and C. Hidalgo. 2016. Cities are physical too: Using computer vision to measure the quality and impact of urban appearance. *American Economic Review* 106:128–32.

Nash, K. S. 2016. Big banks stake fintech claims with patent application surge. *Wall Street Journal*, May 10.

Nerkar, A., and P. Roberts. 2004. Technological and product-market experience and the success of new product introductions in the pharmaceutical industry. *Strategic Managemen*t 25:779–99.

Nicholas, T. 2008. Does innovation cause stock market runups? Evidence from the great crash. *American Economic Revie*w 98:1370–96.

Orton-Jones, C. 2017. The second wave of FinTech. *The Economist*, February 9.

Pakes, A. 1985. On patents, R&D, and the stock market rate of return. *Journal of Political Economy* 93:390–409.

PricewaterhouseCoopers. 2017. Redrawing the lines: Fintech's growing influence on financial services. *PwC Global FinTech Report*.

Reklaitis, V. 2018. How much do you know about FinTech? Time to find out. *Wall Street Journal*, January 7.

Robinson, E. 2017a. FinTech, the buzzword finance loves and hates: QuickTake Q&A, *Bloomberg Businessweek*, February 24.

Robinson, E., 2017b. Revolution or hype? *Bloomberg Businessweek*, April 2.

Russo, C. 2017. Goldman and Google are among the most active blockchain investors. *Bloomberg*, October 17.

Scherer, F. 1980. *Industrial market structure and economic performance*, 2nd ed. Chicago, IL: Rand McNally.

Schumpeter, J. 1942. *Capitalism, socialism and democracy*. London: Allan and Unwin.

Seru, A. 2014. Firm boundaries matter: Evidence from conglomerates and R&D activity. *Journal of Financial Economics* 111:381–405.

Schapire, R., Y. Freund, P. Bartlett, and W. Lee. 1998. Boosting the margin: A new explanation for the effectiveness of voting methods. *Annals of Statistics* 26:1651–86.

Shibata, K., and Y. Ikeda, 2009. Effect of number of hidden neurons on learning in large-scale layered neural networks. In *ICROS-SICE International Joint Conference*, eds. S. Duan, Y. Wu, Z. Ma, and G. Lu. New York: IEEE Publishing.

Tang, M. 1988. An economic perspective on escalating commitment. *Strategic Managemen*t 9:79–92.

Tetlock, P. 2007. Giving content to investor sentiment: The role of media in the stock market. *Journal of Finance* 62:1139–68.

Tetlock, P., M. Saar-Tsechansky, and S. Macskassy. 2008. More than words: Quantifying language to measure firms' fundamentals. *Journal of Finance* 23:1437–67.

Vigna, P. 2017. The newest bank blockchain: Will this be the breakthrough? *Wall Street Journal*, February 28.