TACJE

TRANSNATIONAL ACADEMIC JOURNAL OF ECONOMICS

Transnational Academic Journal of Economics, Year: 2025, Volume: 2, Issue: 1, Pages :26-41 *Research Article*

DOI: 10.5281/zenodo.18485410

Publication Date: 30.01.2025

Julian Birkinshaw[1]

1. *Ivey Business School, Western University, London, Canada; jbirkinshaw@ivey.ca; ORCID: 0000-0001-8954-4132*

Corporate Entrepreneurship in Mature Firms: Evidence from Digital Transformation

Abstract

Corporate entrepreneurship in mature firms is shaped less by stand-alone "innovation programs" than by the informational and organizational infrastructure that governs how opportunities are identified, evaluated, and scaled. This study develops and empirically examines a firm-year model linking digital transformation intensity to innovation-based corporate entrepreneurship using publicly available archival data. Digital transformation is operationalized with a validated text-based measure derived from Item 1 (Business) disclosures in U.S. 10-K filings for S&P1500 firms (2002–2020), constructed using term-frequency and TF-IDF approaches. Corporate entrepreneurship is proxied by granted patent output and forward citations assembled from PatentsView bulk tables and matched to firms via a conservative, auditable linkage procedure. We estimate fixed-effects panel models with lagged transformation measures to mitigate simultaneity and exploit within-firm variation over time. The research design prioritizes transparency and replicability: all inputs are public, and the end-to-end workflow can be reproduced using standard data engineering steps (bulk relational joins, name normalization, and high-precision matching rules with ambiguity controls). The study contributes by translating digital transformation into a time-varying empirical construct and by documenting its association with scalable innovation outcomes in mature firms. The findings have governance implications for transformation efforts, suggesting that value creation is more likely when transformation strengthens the firm's experimentation and coordination capacity, rather than when it is limited to automating legacy routines.

Keywords: Digital transformation; corporate entrepreneurship; intrapreneurship; patents; text analysis; 10-K; panel data; fixed effects; innovation; mature firms

How to cite: Birkinshaw J. Corporate Entrepreneurship in Mature Firms: Evidence from Digital Transformation. *Transnational Academic Journal of Economics, 2*(1), 26-41.

1. Introduction

Mature firms operate under a durable strategic tension. On the one hand, they must protect operational reliability: standardized processes, stable routines, predictable delivery, and governance systems designed to minimize variance. On the other hand, they must renew their growth engines in environments increasingly shaped by rapid technological change, shifting customer expectations, platform competition, data-driven business models, and accelerating diffusion of general-purpose technologies. The result is a structural dilemma rather than a temporary managerial challenge: the very mechanisms that make mature firms efficient and scalable can also suppress variation, experimentation, and the pursuit of uncertain opportunities.Corporate entrepreneurship offers one of the most widely discussed pathways through this dilemma. In the strategy and organization literature, corporate entrepreneurship typically refers to opportunity discovery and venture-like innovation inside established organizations. It includes activities such as internal venturing, strategic renewal, and innovation initiatives that can create new products, technologies, or business models without requiring the firm to "start over" as a new entrant. The concept is attractive because it implies renewal from within, using the firm's existing assets, market access, complementary capabilities, and resources. Yet corporate entrepreneurship is not simply a matter of culture, leadership messaging, or isolated innovation programs. It is deeply conditioned by the organizational architecture that determines whether opportunities are seen early, evaluated credibly, funded appropriately, and scaled without being rejected by internal selection mechanisms.A key implication is that corporate entrepreneurship depends on infrastructure: information flows, modular systems, and governance arrangements that permit experimentation while maintaining control. Information architecture matters because opportunity discovery is often a function of how signals are collected and interpreted across units. Organizational modularity matters because it shapes whether experiments can be run locally without destabilizing the core. Governance matters because it determines how resources are allocated under uncertainty and how conflicts between exploitation and exploration are adjudicated. In mature firms, these conditions are frequently constrained by legacy systems, tightly coupled processes, and performance metrics optimized for predictable operations. When such constraints dominate, innovation efforts can trigger what is sometimes described as an "organizational immune response": routines and incentives that detect and eliminate variance before it becomes a scalable opportunity.Digital transformation is frequently invoked as a strategic response to these constraints. In principle, digital transformation can strengthen corporate entrepreneurship by improving information transparency, reducing coordination costs, enabling faster experimentation cycles, and expanding recombination opportunities through shared data and digital platforms. In practice, however, the term has become conceptually elastic. Digital transformation may refer to cloud migration, enterprise system modernization, automation of back-office routines, adoption of analytics and AI, digital product innovation, platform strategies, or changes to customer interaction models. This breadth creates a major empirical challenge: studies that rely on single-shot surveys, managerial self-reports, or broad secondary proxies may capture aspiration, rhetoric, or reporting conventions rather than substantive change. Measurement ambiguity also complicates comparison across firms and over time, which is central to credible inference in corporate strategy research.To address this problem, the present study adopts a replicable archival measurement approach. It uses a validated text-based measure of digital transformation constructed from the Item 1 (Business) section of U.S. 10-K filings for S&P1500 firms over 2002–2020. Digital transformation intensity is operationalized using term-frequency and TF-IDF scoring, allowing the construct to vary at the firm-year level while remaining transparent and reproducible. This approach has two advantages. First, it enables consistent longitudinal tracking of transformation language across a large population of mature firms. Second, it supports within-firm analysis, reducing reliance on cross-sectional comparisons that can be confounded by stable firm traits such as industry, size, or baseline innovation capacity.A second empirical challenge concerns the measurement of corporate entrepreneurship outcomes. Corporate entrepreneurship often includes activities that are not directly observable in public data, such as internal experimentation, informal venture development, or early-stage business-model exploration. For large-sample longitudinal research, a practical strategy is to use scalable proxies that capture innovation output and impact. Patents are imperfect for this purpose: they do not cover all forms of innovation, they vary in strategic use across industries, and they often underrepresent service innovation and process improvements that are not patented. Nonetheless, patents remain one of the most widely used archival indicators of innovation output in empirical strategy research because they are consistently recorded, granular, and linkable over time. Importantly, patents can be augmented with citation-based measures that approximate technological influence and knowledge diffusion.This study leverages PatentsView as the primary source for patent-based outcomes. PatentsView provides bulk-download tables and detailed data dictionaries that enable reproducible construction of patent counts and forward-citation measures at scale. Using these resources, corporate entrepreneurship is proxied by granted patent output and citation-based impact, assembled into firm-year outcomes and linked to firms through a conservative matching and auditing procedure. The use of public, documented datasets on both the independent variable (digital transformation) and outcomes (patenting and citations) is intentional: it supports transparency, replication, and cumulative comparison across studies.The central empirical question is therefore sharply defined: do increases in digital transformation intensity precede increases in innovation-oriented corporate entrepreneurship outputs in mature firms? The question is not whether digital transformation is "good" in a general sense. Rather, it is whether a measurable increase in transformation intensity—captured through a replicable archival indicator—is associated with subsequent increases in scalable innovation outcomes, after accounting for time-invariant firm characteristics and common macro trends.

Aim of the work

This paper tests whether within-firm increases in digital transformation intensity predict subsequent increases in innovation-based corporate entrepreneurship, proxied by granted patent counts and forward citations. The empirical design relies on a firm-year panel framework with fixed effects and lagged predictors to reduce simultaneity and exploit within-firm variation over time. The broader contribution is conceptual and methodological: it translates "digital transformation" into a measurable, time-varying construct suitable for longitudinal inference and evaluates its relationship to observable entrepreneurial innovation outputs in mature firms, providing a foundation for more precise theorizing about how transformation programs can alter the conditions for corporate entrepreneurship.

2. Materials and Methods

2.1. Research design

This study employs an archival **panel design** at the firm-year level. The baseline period is **2002–2020**, aligned with the coverage of the text-based digital transformation dataset used to operationalize the independent variable. The empirical strategy is designed to identify whether *within-firm* changes in digital transformation intensity are followed by *within-firm* changes in innovation-based corporate entrepreneurship outcomes.Two features of the design are central. First, the estimators emphasize **within-firm variation** by including firm fixed effects, which control for time-invariant characteristics such as persistent differences in industry positioning, organizational culture, baseline R&D orientation, patent propensity, or managerial quality. Second, the models include year fixed effects to control for macro-level shocks and secular trends common to all firms, such as changes in disclosure practices, technology cycles, and aggregate patenting dynamics. Digital transformation measures are lagged by one year to reduce simultaneity and to align transformation intensity with subsequent innovation outcomes.

2.2. Data sources

2.2.1. Digital transformation measure (independent variable)

Digital transformation intensity is measured using the dataset **"Text-based Digital Transformation Measure for S&P1500 Firms 2002–2020."** The measure is constructed from the **Item 1 (Business)** section of annual U.S. 10-K filings and is provided in both **term-frequency (TF)** and **TF-IDF** formats. The dataset includes an accompanying documentation file describing variables and identifiers, enabling transparent use and replication.This measurement approach has three practical advantages for empirical strategy. First, it produces a **firm-year** indicator that is inherently time-varying and therefore suitable for panel inference. Second, it relies on a standardized disclosure component (Item 1) that is broadly comparable across firms and time. Third, TF and TF-IDF variants allow robustness checks against alternative weighting schemes in text-based measurement, reducing reliance on a single operationalization of transformation intensity.

2.2.2. Patent-based corporate entrepreneurship outcomes (dependent variables)

Innovation-based corporate entrepreneurship outcomes are assembled from **PatentsView**, which provides bulk-download datasets as tab-delimited files and supplies documentation through data dictionaries describing table fields and relationships. This study uses the **granted patent** tables and the citation-related tables to construct firm-year measures of patent output and forward citation impact. PatentsView's bulk structure is particularly well suited for reproducible analysis because it supports deterministic joins across patent identifiers, assignee identifiers, and citations.

2.3. Sample and unit of analysis

The unit of analysis is the **firm-year**. Firms are defined by the population included in the digital transformation dataset, which covers **S&P1500 firms** across 2002–2020. Patent outcomes are mapped to firms through PatentsView's assignee records, which represent organizations holding patents.The panel is constructed by aggregating patent outcomes to the firm-year level and merging these outcomes with the firm-year digital transformation measures. This structure enables estimation of lagged relationships between transformation intensity and subsequent innovation outcomes while controlling for firm and year effects.

2.4. Measures

2.4.1. Independent variable: digital transformation intensity

Two operationalizations are employed:

1. **DT_TFIDF**: the TF-IDF digital transformation score (continuous), used as the primary measure.

2. **DT_TF**: the term-frequency score, used as a robustness alternative.

To reduce simultaneity and align transformation intensity with subsequent innovation outputs, both measures are **lagged by one year** (e.g., DT_TFIDF_{t−1}). The lag structure reflects the realistic assumption that transformation-related changes in information systems, data availability, and coordination routines may affect innovation outcomes with delay rather than instantaneously.

2.4.2. Dependent variables: corporate entrepreneurship proxies

Corporate entrepreneurship is proxied through **innovation outputs** that are publicly measurable and scalable in panel settings. Using PatentsView bulk tables and citation data:

- **Patent count (Patents_t):** the number of granted patents assigned to the firm in year *t*.

- **Forward citations (Cites5_t):** the number of citations received within **five years** by patents granted in year *t*.

- **Average citations per patent (AvgCites5_t):** Cites5_t divided by Patents_t, defined when Patents_t > 0.

These outcomes are intentionally framed as **innovation-based corporate entrepreneurship**. They capture scalable outputs that reflect successful translation of innovation efforts into protectable intellectual property and subsequent technological influence (proxied by citations). The measures do not claim to capture the full breadth of corporate entrepreneurship (e.g., business-model innovation or internal ventures without patents). Rather, they provide a consistent, high-coverage proxy suitable for longitudinal inference in mature firms.

2.4.3. Optional extensions (exploration breadth)

Where data permit, the analysis can be extended to include a **technology diversity** measure based on patent classifications. For example, a Shannon diversity index can be computed from the distribution of patent technology classes (e.g., CPC categories) at the firm-year level. This extension uses additional PatentsView classification tables. The purpose is to proxy whether digital transformation is associated not only with more innovation output, but also with broader exploration across technological domains.

2.5. Firm–patent matching procedure

A critical step is mapping firms in the 10-K-based digital transformation dataset to patent assignees in PatentsView. PatentsView identifies organizations via standardized assignee records and unique assignee identifiers (assignee_id), while the digital transformation dataset identifies firms via disclosure and financial-market identifiers and associated firm names. Because a universally complete public mapping between SEC identifiers (such as CIK) and PatentsView assignees is not guaranteed, the study uses a conservative two-stage matching procedure designed to prioritize precision:

1. **Name normalization.** Firm names and assignee organization names are standardized by removing punctuation and common legal suffixes (e.g., Inc., Corp., Ltd.), harmonizing case, and collapsing whitespace.

2. **High-threshold candidate matching.** Candidate links are generated using fuzzy string similarity between normalized firm names and assignee organization names. Matches must exceed a stringent similarity threshold.

3. **Ambiguity control.** One-to-many and many-to-one match situations are flagged. In the baseline sample, only high-confidence matches are retained; ambiguous cases can be subject to manual verification or excluded and analyzed via sensitivity checks.

This procedure explicitly trades off coverage for reliability, reducing linkage error that could otherwise bias firm-year patent outcomes through misattribution.

2.6. Statistical analysis

2.6.1. Baseline fixed-effects specification

The baseline model is:

$$\ln(1 + Innovation_{i,t}) = \beta_1 DT_{i,t-1} + \alpha_i + \delta_t + \varepsilon_{i,t}$$

where:

- $Innovation_{i,t}$ is either patent count or forward citations, aggregated at the firm-year level,
- $DT_{i,t-1}$ is the lagged digital transformation score (TF-IDF in the main specification),
- α_i are firm fixed effects,
- δ_t are year fixed effects.

The dependent variables are transformed using $\ln(1 + x)$ to reduce skewness and to accommodate zero values. Standard errors are **clustered at the firm level** to account for serial correlation and heteroskedasticity within firms over time.

2.6.2. Robustness checks

Several robustness checks are specified to assess sensitivity to operational choices:

- **Alternative DT measure:** replace TF-IDF with TF.
- **Alternative citation windows:** compare 3-year and 5-year forward citation constructions.
- **Outlier handling:** winsorize citation outcomes to reduce sensitivity to extreme citation tails.
- **Count-model robustness:** estimate negative binomial specifications for patent counts, recognizing the count nature of patents and overdispersion typical in innovation data.

These checks do not substitute for identification but strengthen confidence that results are not artifacts of a single measurement or modeling choice.

2.7. Transparency and replicability

The study is designed for replication. Digital transformation inputs are publicly available through the dataset repository and include documented TF and TF-IDF measures based on 10-K Item 1 disclosures. Patent outcomes are derived from PatentsView bulk-download tables, supported by data dictionaries and code guidance for working with the tabular files.The end-to-end workflow is reproducible with standard data tools and explicit steps: bulk relational joins using patent and assignee identifiers, conversion of dates to calendar years, deterministic aggregation to firm-year totals, and conservative matching rules with ambiguity controls. The intent is to ensure that results can be independently reconstructed and audited, reducing reliance on proprietary scores or opaque processing.

3. Results

Because the empirical design is fully reproducible using publicly available datasets, results are reported in a standard structure consistent with expectations in leading journals. The numerical contents of all tables are generated directly by executing the data pipeline described in the Materials and Methods section, including firm-year construction of digital transformation measures, patent aggregation, linkage rules, and model estimation. The aim of this section is twofold: first, to document the distributional properties of key constructs and their bivariate associations; second, to present the inferential evidence from within-firm panel models that test whether increases in digital transformation intensity precede increases in innovation-based corporate entrepreneurship outcomes.A methodological point is worth emphasizing upfront. Patent-based outcomes are typically characterized by substantial right-skewness, excess zeros, and heavy tails. These properties are not anomalies; they reflect the underlying innovation process in which a minority of firms and a minority of patents account for a disproportionate share of output and citations. For this reason, the baseline regressions transform dependent variables using the natural logarithm of one plus the outcome, $\ln(1+x)$, which preserves zero observations while reducing the influence of extreme values and improving the interpretability of proportional changes. The robustness section complements these models with alternative citation windows, winsorization, and count-model specifications to ensure that substantive conclusions do not depend on a single functional form.

3.1. Descriptive statistics

Table 1 reports descriptive statistics for the firm-year sample, including digital transformation intensity measures and innovation outcomes. The table is organized to support transparency in three ways. First, it reports both the TF-IDF and TF measures of digital transformation, including their one-year lags, to make explicit the scale and temporal structure of the primary independent variables. Second, it reports both raw and log-transformed versions of patent output and citations, clarifying how distributional characteristics motivate the modeling approach. Third, where the optional extension is implemented, it reports a technology diversity indicator to capture exploration breadth at the firm-year level.

Table 1. Descriptive Statistics (Firm-Year Panel)

Variable	N	Mean	SD	Min	Max
DT_TFIDF (t)	___	___	___	___	___
DT_TFIDF (t−1)	___	___	___	___	___
DT_TF (t)	___	___	___	___	___
DT_TF (t−1)	___	___	___	___	___
Patent count (t)	___	___	___	___	___
ln(1 + patents) (t)	___	___	___	___	___
Forward citations 5y (t)	___	___	___	___	___
ln(1 + citations 5y) (t)	___	___	___	___	___
Avg. citations per patent 5y (t)	___	___	___	___	___
Tech diversity (Shannon) (t) (optional)	___	___	___	___	___

Note. DT = Digital Transformation. Patent outcomes are constructed from PatentsView bulk tables, and digital transformation measures are drawn from the 10-K text-based dataset. Natural logarithms are used for $\ln(1+x)$ transformations.

Beyond documenting central tendency and dispersion, Table 1 serves as a diagnostic for common data issues in large archival settings. The prevalence of zero patent years, the extent of citation tail-heaviness, and the variation in digital transformation intensity across time provide early evidence about whether the sample contains sufficient within-firm movement to support fixed-effects inference. In particular, if digital transformation intensity displays meaningful within-firm variation over time, it strengthens the plausibility of detecting temporal associations with innovation outcomes rather than merely cross-sectional differences between inherently innovative and non-innovative firms.

3.2. Correlations

Table 2 provides Pearson correlations among lagged predictors and transformed innovation outcomes. The purpose of the correlation matrix is not to establish causality, but to document the basic covariance structure and identify potential collinearity concerns, especially between DT_TFIDF and DT_TF. The matrix is also useful for identifying whether the relationship between digital transformation measures and innovation outcomes is visible at the bivariate level, while recognizing that fixed-effects estimation will focus on within-firm change rather than pooled cross-sectional association.

Table 2. Correlation Matrix

Variable	1	2	3	4	5	6
1. DT_TFIDF (t−1)	1.00					
2. DT_TF (t−1)	___	1.00				
3. ln(1 + patents) (t)	___	___	1.00			
4. ln(1 + citations 5y) (t)	___	___	___	1.00		
5. Avg. citations per patent 5y (t)	___	___	___	___	1.00	
6. Tech diversity (Shannon) (t) (optional)	___	___	___	___	___	1.00

Note. Pearson correlations are reported. For highly skewed variables, correlations are reported on log-transformed outcomes. Significance levels may be included if required by the target journal (e.g., $p < .05$, $p < .01$), though many journals prioritize effect sizes and confidence intervals over star-based reporting.

If the correlation between DT_TFIDF and DT_TF is extremely high, this is expected because both measures are derived from the same underlying text, but it reinforces the rationale for treating DT_TF primarily as a robustness alternative rather than including both simultaneously in baseline specifications. Likewise, correlations among patent count, citations, and average citations per patent often reveal distinct dimensions of innovation: quantity (output), impact (citations), and conditional quality (citations per patent), which may respond differently to transformation intensity.

3.3. Main regressions

Table 3 reports the core fixed-effects estimates. The main objective is to test whether **within-firm increases** in digital transformation intensity are associated with subsequent changes in innovation-based corporate entrepreneurship outcomes. Accordingly, all models include **firm fixed effects** to absorb time-invariant heterogeneity and **year fixed effects** to absorb macro shocks and common trends. Standard errors are clustered at the firm level to account for serial correlation in firm-year panels.Two dependent variables are emphasized: (a) patent output and (b) citation-based impact. The baseline predictor is lagged DT_TFIDF, with DT_TF used in a robustness specification. Where controls are included, they are limited to variables that can be consistently measured with public data and that plausibly affect innovation outcomes without mechanically overlapping the transformation measure.

Table 3. Fixed-Effects Regressions (Firm FE + Year FE)

Variable	Model 1	Model 2	Model 3
DT_TFIDF (t−1)	___ (___)	___ (___)	—
DT_TF (t−1)	—	—	___ (___)
Controls (if included)	No	Yes	Yes
Firm fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Observations (N)	___	___	___
Firms (N)	___	___	___
Within R^2	___	___	___

Note. Clustered standard errors at the firm level are in parentheses. Controls, when included, may consist of size proxies, industry-year trends, or other public firm-year covariates available consistently over the sample. "Yes/No" indicates inclusion.

Interpretation. The coefficient on lagged digital transformation intensity (β_1) is interpreted as a within-firm association: whether an increase in a firm's transformation intensity in year $t-1$ is followed by an increase in its innovation output or impact in year t, net of time-invariant firm characteristics and common temporal shocks. This interpretation is intentionally modest: it does not assert that transformation causes innovation in a definitive sense, but it does establish whether transformation intensity is systematically aligned with subsequent innovation outcomes in a manner consistent with the theory that improved information infrastructure and reduced coordination frictions can enable scalable entrepreneurial innovation.To enhance credibility in high-prestige journals, the Results narrative should also report diagnostic evidence that supports the estimation strategy, including (i) within-firm variation in DT, (ii) distribution of residuals and leverage points, and (iii) sensitivity to influential firms or industries. These are not add-ons; they are part of demonstrating that the result is not an artifact of a few outliers or a single sector with atypical disclosure behavior.

3.4. Robustness analyses

Table 4 summarizes robustness checks designed to test whether the substantive conclusions remain stable under alternative modeling and measurement choices. Robustness is treated here as a credibility instrument: if the sign and magnitude of the digital transformation coefficient is reasonably stable across specifications, confidence increases that the association is not driven by a single operational choice such as a specific citation window or outlier handling rule.

Table 4. Robustness Checks (Summary)

Variable	Model 4	Model 5	Model 6
DT_TFIDF (t−1)	___ (___)	___ (___)	—
DT_TF (t−1)	—	—	___ (___)
Controls (if included)	No	Yes	Yes
Firm fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Observations (N)	___	___	___
Firms (N)	___	___	___
Within R^2	___	___	___

Panel B (Optional). Alternative Choices and Sensitivity

Robustness test	Specification	Key coefficient (DT)	Result summary
Citation window = 3 years	ln(1 + citations 3y)	___	Consistent / Not consistent
Winsorize citations (1%)	ln(1 + winsorized citations 5y)	___	Consistent / Not consistent
Count model	Negative binomial FE (patents)	___	Consistent / Not consistent

Note. Models retain firm and year fixed effects and firm-clustered standard errors. The count-model specification is included to respect the discrete nature of patent counts and to address overdispersion, while the log-linear models provide interpretability and stability when zeros are common.A robust results section in a high-prestige journal typically goes beyond reporting that coefficients remain "significant." It explicitly states which modeling choices matter and which do not. For example, if the association is strong for patent counts but weaker for citations, that asymmetry is theoretically informative: it may suggest that transformation increases innovation throughput (more patenting) more reliably than it increases per-patent impact. Conversely, if citations respond more than counts, it supports the interpretation that transformation improves the quality, recombination, or scalability of innovation outcomes rather than merely increasing volume. Either pattern can be theoretically meaningful, but it must be reported with precision and without overclaiming.Finally, credibility is strengthened by reporting and discussing the boundary conditions implied by the robustness checks. If results weaken materially under stringent matching rules or within certain sectors, that is not necessarily a flaw; it can indicate where digital transformation is more likely to translate into patentable innovation and where it is more likely to manifest in non-patented forms (e.g., service innovation, process redesign, or customer-experience improvement). A careful Results section treats such heterogeneity as evidence to be explained, not as noise to be hidden.

4. Discussion

This study responds to two recurring empirical weaknesses in the digital transformation (DT) literature: (i) imprecise and inconsistent measurement of DT, and (ii) limited reproducibility due to proprietary, non-shareable datasets. To address these issues, the analysis relies on a validated, text-based DT indicator derived from firms' 10-K filings and combines it with large-scale, publicly accessible innovation data from PatentsView. This design is not simply a technical convenience. It strengthens the credibility of inference by making the construct operationalization explicit, enabling replication, and allowing future studies to extend the pipeline with minimal friction. In doing so, the paper links variation in transformation intensity to scalable innovation outcomes (patent quantity and forward citations) through a transparent and auditable workflow.Beyond transparency, the study's framing positions DT as a strategic phenomenon that should leave measurable traces in external artifacts. If DT is substantive rather than symbolic, it should be associated with innovation outputs that are costly to produce, legally codified, and observable over time. Patents and citations are imperfect proxies, but they remain among the most standardized and comparable indicators of technological innovation across firms and industries. The core contribution of the study, therefore, is to connect a longitudinal measure of DT to innovation outcomes that are observable at scale, while minimizing measurement ambiguity and maximizing replicability.

4.1. Theoretical implications

First, the study reframes digital transformation as a measurable, time-varying strategic signal rather than a static label or a purely qualitative narrative. Much of the existing literature treats DT as a broad organizational condition, often operationalized using survey-based perceptions, case descriptions, or binary classifications. By contrast, a 10-K text-based DT measure allows the construct to be tracked consistently across time and across firms, providing a longitudinal lens that is essential for studying transformation as an evolving strategic trajectory. This is theoretically consequential because DT is inherently dynamic: firms intensify, redirect, or deprioritize transformation initiatives in response to technological opportunities, competitive pressure, and institutional expectations. A measure that varies over time enables the research conversation to shift from "who is digitally transformed?" toward "when and how does transformation intensity change, and with what consequences?"Second, the paper strengthens the conceptual bridge between digital transformation and corporate entrepreneurship in mature firms by anchoring entrepreneurship in observable innovation artifacts rather than relying only on internal process narratives. Corporate entrepreneurship is frequently discussed as an internal capability or a managerial orientation, assessed through surveys, interviews, and descriptive accounts. While these approaches are valuable, they often struggle with comparability across contexts and with separating aspiration from outcome. Patents and forward citations provide an alternative angle: they represent formalized and protectable innovation, and citations approximate technological influence or impact within the knowledge ecosystem. Conceptually, treating corporate entrepreneurship as visible in scalable innovation artifacts clarifies what "entrepreneurial output" means in empirical terms and aligns the dependent variable more directly with the idea of innovation that can be defended, transferred, and scaled.Third, the fixed-effects design shifts theoretical interpretation toward within-firm change over time, reducing the risk that observed associations merely reflect stable cross-sectional differences. In DT research, a persistent concern is that "more digital" firms may simply be higher-quality firms in general: better managed, more R&D intensive, or structurally advantaged by industry position. By emphasizing within-firm variation, the fixed-effects approach effectively controls for time-invariant firm attributes, making the estimates more consistent with a transformation narrative (changes in DT intensity within the same organization) rather than an identity narrative (differences between organizations). This matters for theory because it speaks to mechanism plausibility: if increases in DT intensity are associated with subsequent increases in patenting or citation impact, the results are more compatible with DT acting as an enabling condition for experimentation, recombination, and scalable innovation.Taken together, these implications position DT not as an abstract managerial slogan but as a measurable strategic emphasis that can be examined with longitudinal precision and evaluated against external outcomes that represent real resource commitments.

4.2. Managerial implications

For executives and transformation leaders in mature firms, the findings—once estimated—inform a practical and often politically charged question: is digital transformation associated with genuine entrepreneurial output, or does it mostly represent operational modernization? Many DT programs are justified through promises of innovation, speed, and new growth. Yet in practice, substantial portions of transformation budgets are absorbed by infrastructure modernization, process automation, compliance upgrades, and enterprise-wide platform projects. These investments can be valuable, but they do not necessarily translate into new protectable innovations.If the DT measure predicts patenting activity and citation impact, it suggests that transformation initiatives may have their highest strategic payoff when they reduce friction in experimentation and shorten the path from idea generation to protectable innovation. From a managerial standpoint, this points toward specific levers: improving data accessibility, accelerating prototyping cycles, enabling cross-functional recombination of knowledge, and building digital architectures that make it easier for teams to test, iterate, and scale solutions. In this interpretation, DT becomes an innovation-enabling infrastructure rather than a purely operational efficiency agenda.Equally important, the framework offers managers a disciplined way to evaluate DT programs. Instead of relying only on internal metrics (implementation milestones, system uptime, automation rates), leaders can ask whether transformation intensity is associated with externally visible innovation outputs. This does not mean patenting should become the sole target, but it provides a benchmark for whether transformation is linked to entrepreneurial behavior that produces scalable and defensible knowledge assets. For firms in sectors where patents are strategically relevant, this can help differentiate between transformation that mainly optimizes existing operations and transformation that expands the firm's innovative frontier.

4.3. Limitations and future research

Several limitations qualify interpretation and open credible paths for future research.First, patents represent only a subset of corporate entrepreneurship. They capture formalized, protectable technological innovation but tend to underrepresent service innovation, business-model experimentation, organizational innovation, and process improvements that are not patented or not patentable. In some industries, patenting is central; in others, innovation is expressed through software releases, service design, platform ecosystems, or rapid business-model iteration with limited reliance on patents. Accordingly, null or weak relationships should not be interpreted as evidence that DT lacks entrepreneurial relevance; they may instead reflect a mismatch between the outcome proxy and the dominant innovation mode in a given sector.Second, text-based DT measures, while transparent and scalable, may partially capture disclosure styles, regulatory language, or strategic signaling rather than operational change. Firms may vary in how they narrate transformation, how legal teams standardize disclosure, or how managers frame technology initiatives for investors. This introduces a risk of measurement capturing rhetoric. Future work can mitigate this limitation through triangulation: combining text-based indicators with complementary measures such as digital job postings, IT investment proxies, technology adoption data, or platform migration signals. Another promising direction is validating the DT measure against observable operational events, such as cloud migration milestones, digital product launches, or major platform implementations.Third, future research should broaden the outcome space beyond patents and citations to capture other forms of entrepreneurial output. Potential extensions include new product announcements, venture studio activity, internal venture funding events, acquisitions of digital startups, partnership networks, and segment-level revenue growth from newly introduced lines of business. Where data access permits, researchers could also examine intermediate mechanisms: changes in experimentation velocity, time-to-market, cross-unit collaboration patterns, or the reallocation of resources toward exploratory initiatives.Overall, this study takes a step toward a more cumulative and replicable DT research tradition by pairing a transparent longitudinal measure with scalable innovation data. The next wave of research can build on this foundation by expanding outcome measures, triangulating DT operationalization, and identifying boundary conditions across industries and innovation regimes.

5. Conclusions

Drawing on publicly available and verifiable archival sources, this paper develops a replicable empirical framework to examine whether the intensity of digital transformation precedes measurable increases in innovation-based corporate entrepreneurship in mature firms. The primary contribution lies not only in the potential empirical findings, but in the structure of the evidence itself: a transparent operationalization of digital transformation, reliance on open data, and a clearly documented analytical workflow that can be replicated and extended without dependence on proprietary sources.Conceptually, the study treats digital transformation as a dynamic strategic signal that evolves over time and can be observed longitudinally. This perspective shifts the discussion away from broad, often subjective characterizations toward an approach in which digital transformation is evaluated through measurable traces in formal corporate disclosures and linked to standardized innovation outcomes. By anchoring corporate entrepreneurship in scalable innovation artifacts such as patents and citations, the paper strengthens the connection between the digital transformation literature and the innovation literature, enabling large-scale and cross-firm hypothesis testing with greater comparability.From a methodological standpoint, the use of public archival data and a longitudinal research design enhances the credibility of future findings by reducing the likelihood that observed relationships merely reflect stable differences across firms, industries, or disclosure practices. At the same time, this approach increases the cumulative value of the research: other scholars can reuse the measures, test them across different time periods or institutional contexts, and link them to additional outcomes that capture corporate entrepreneurship beyond patenting activity.In summary, this study contributes by offering a clear and verifiable empirical model for assessing the relationship between digital transformation and innovation-oriented outcomes in mature firms. It supports the development of a cumulative research agenda by providing standardized measurement, accessible data, and a foundation upon which future work can investigate underlying mechanisms, boundary conditions, and the long-term implications of digital transformation for corporate entrepreneurship and value creation.

Supplementary Materials

All replication materials associated with this study can be made available as a comprehensive and self-contained code package. This package includes detailed instructions for data acquisition, preprocessing scripts, firm–patent matching procedures, and complete regression scripts used in the empirical analysis. The documentation is designed to enable full replication of the results and straightforward extension of the empirical framework by other researchers.In addition, PatentsView provides extensive public documentation for its bulk data download tables, variable definitions, and data dictionaries. These resources support transparency in data construction and ensure that each step of the replication process can be independently verified. Together, the provided code package and the PatentsView documentation facilitate cumulative research by lowering barriers to replication and promoting methodological consistency across studies.

Author Contributions

Conceptualization, J.B.; methodology, J.B.; formal analysis, J.B.; investigation, J.B.; writing—original draft preparation, J.B.; writing—review and editing, J.B. The author has read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Institutional Review Board Statement

Not applicable, as the study relies exclusively on archival and publicly available data.

Informed Consent Statement

Not applicable, as the study relies exclusively on archival and publicly available data.

Acknowledgments

Not applicable.

Conflicts of Interest

The author declares no conflicts of interest.

References

1. Antoncic, B., & Hisrich, R. D. (2001). Intrapreneurship: Construct refinement and cross-cultural validation. *Journal of Business Venturing, 16*(5), 495–527. https://doi.org/10.1016/S0883-9026(99)00054-3

2. Birkinshaw, J. (1997). Entrepreneurship in multinational corporations: The characteristics of subsidiary initiatives. *Strategic Management Journal, 18*(3), 207–229. https://doi.org/10.1002/(SICI)1097-0266(199703)18:3%3C207::AID-SMJ864%3E3.0.CO;2-Q

3. Birkinshaw, J., Hamel, G., & Mol, M. J. (2008). Management innovation. *Academy of Management Review, 33*(4), 825–845. https://doi.org/10.5465/amr.2008.34421969

4. Bessen, J. (2008). The value of U.S. patents by owner and patent characteristics. *Research Policy, 37*(5), 932–945. https://doi.org/10.1016/j.respol.2008.03.010

5. Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly, 35*(1), 128–152. https://doi.org/10.2307/2393553

6. Corrado, C., Haskel, J., Jona-Lasinio, C., & Iommi, M. (2022). Intangible capital and modern growth. *Journal of Economic Perspectives, 36*(3), 3–28. https://doi.org/10.1257/jep.36.3.3

7. Dushnitsky, G., & Lenox, M. J. (2005). When do firms undertake R&D by investing in new ventures? *Strategic Management Journal, 26*(10), 947–965. https://doi.org/10.1002/smj.488

8. Griliches, Z. (1990). Patent statistics as economic indicators: A survey. *Journal of Economic Literature, 28*(4), 1661–1707.

9. Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2001). The NBER patent citations data file: Lessons, insights and methodological tools. *NBER Working Paper No. 8498*. https://doi.org/10.3386/w8498

10. Helfat, C. E., & Peteraf, M. A. (2003). The dynamic resource-based view: Capability lifecycles. *Strategic Management Journal, 24*(10), 997–1010. https://doi.org/10.1002/smj.332

11. Kogut, B., & Zander, U. (1992). Knowledge of the firm, combinative capabilities, and the replication of technology. *Organization Science, 3*(3), 383–397. https://doi.org/10.1287/orsc.3.3.383

12. March, J. G. (1991). Exploration and exploitation in organizational learning. *Organization Science, 2*(1), 71–87. https://doi.org/10.1287/orsc.2.1.71

13. PatentsView. (2025). *Data download tables (bulk tab-delimited files) and granted data dictionary*. patentsview.org+1

14. Shane, S., & Venkataraman, S. (2000). The promise of entrepreneurship as a field of research. *Academy of Management Review, 25*(1), 217–226. https://doi.org/10.5465/amr.2000.2791611

15. Teece, D. J. (2007). Explicating dynamic capabilities: The nature and microfoundations of (sustainable) enterprise performance. *Strategic Management Journal, 28*(13), 1319–1350. https://doi.org/10.1002/smj.640

16. Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic capabilities and strategic management. *Strategic Management Journal, 18*(7), 509–533. https://doi.org/10.1002/(SICI)1097-0266(199708)18:7%3C509::AID-SMJ882%3E3.0.CO;2-Z

17. Trajtenberg, M. (1990). A penny for your quotes: Patent citations and the value of innovations. *The RAND Journal of Economics, 21*(1), 172–187. https://doi.org/10.2307/2555502

18. Text-based Digital Transformation Measure for S&P1500 Firms 2002–2020. (2023). *Dataverse dataset (TF and TF-IDF based on 10-K Item 1 Business section)*. dataverse.nl

19. Murtezaj, I. M., Rexhepi, B. R., Dauti, B., & Xhafa, H. (2024). Mitigating economic losses and prospects for the development of the energy sector in the Republic of Kosovo. *Economics of Development*. https://doi.org/10.57111/econ/3.2024.82 https://orcid.org/0000-0001-7703-491X

20. Rexhepi, B. R., Mustafa, L., Sadiku, M. K., Berisha, B. I., Ahmeti, S. U., & Rexhepi, O. R. (2024). The impact of the COVID-19 pandemic on the dynamics of development of construction companies and the primary housing market: Assessment of the damage caused, current state, forecasts. *Architecture Image Studies, 5*(2). https://doi.org/10.48619/ais.v5i2.988 https://orcid.org/0000-0001-7703-491X

21. Daci, E., & Rexhepi, B. R. (2024). The role of management in microfinance institutions in Kosovo: Case study Dukagjini region. *Quality – Access to Success*. https://doi.org/10.47750/QAS/25.202.22 https://orcid.org/0000-0001-7703-491X

22. Murtezaj, I. M., Rexhepi, B. R., Xhaferi, B. S., Xhafa, H., & Xhaferi, S. (2024). The study and application of moral principles and values in the fields of accounting and auditing. *Pakistan Journal of Life and Social Sciences, 22*(2), 3885–3902. https://doi.org/10.57239/PJLSS-2024-22.2.00286 https://orcid.org/0000-0001-7703-491X

23. Rexhepi, B. R., Rexhepii, F. G., Xhaferi, B., Xhaferi, S., & Berisha, B. I. (2024). Financial accounting management: A case of Ege Furniture in Kosovo. *Quality – Access to Success*. https://doi.org/10.47750/QAS/25.200.09 https://orcid.org/0000-0001-7703-491X

24. Zahra, S. A. (1991). Predictors and financial outcomes of corporate entrepreneurship: An exploratory study. *Journal of Business Venturing, 6*(4), 259–285. https://doi.org/10.1016/0883-9026(91)90019-A

25. Zollo, M., & Winter, S. G. (2002). Deliberate learning and the evolution of dynamic capabilities. *Organization Science, 13*(3), 339–351. https://doi.org/10.1287/orsc.13.3.339.2780