Limitation of Firm Fixed Effects Models and the Missing R&D-Patent Relation: New Methods and Evidence

Hui-Ching Chuang[a], Po-Hsuan Hsu[b,*], Chung-Ming Kuan[c], Jui-Chung Yang[d]

[a] *Department of Statistics, National Taipei University, Taiwan*
[b] *Department of Quantitative Finance, National Tsing Hua University, Taiwan*
[c] *Department of Finance, National Taiwan University, Taiwan*
[d] *Department of Economics, National Taiwan University, Taiwan*

Abstract

The common practice to include firm fixed effects in empirical research may eliminate the explanatory power of important economic factors that are persistent. We use the intuitive R&D-patent relation to illustrate this point. Our review of recent studies suggests a surprising pattern that R&D input only positively explains patent output in half of prior regression estimations. This "missing link" can be attributed to the persistence of R&D and patents that causes the between-firm variation to be absorbed by firm fixed effects. We consider adjusted Hausman-Taylor estimators and advanced machine learning methods, and find that both methods deliver a clear positive R&D-patent relation in data and outperform fixed effects estimators in simulations. In particular, advanced machine learning methods suggest that only 10-20% of firm dummies are informative in regressions for the R&D-patent relation and that including other non-informative firm dummies may bias the identification. This paper thus offers two ready-to-use econometric methods to serve as a "second opinion" for empirical researchers working with explanatory variables that strongly correlate with between-individual unobservables.

JEL Codes: C01, C13, G30, O30
Keywords: LASSO, double machine learning, sparse modelling, within estimator, unobserved heterogeneity

*Corresponding author: Po-Hsuan Hsu. Address: Department of Quantitative Finance, College of Technology Management, National Tsing Hua University, Taiwan. We are grateful for helpful comments from Brad Barber, Dong Beom Choi (discussant), Jillian Grennan, Kristine Hankins, Daehyun Kim, Kiridaran Kanagaretnam (discussant), Kai Li, Tavy Ronen, Ayako Yasuda, and seminar and conference participants at Max Planck I&E Seminar, SFS Cavalcade Asia-Pacific 2024, 2024 UC Davis-FMA Napa Finance Conference, 2024 FMA Asia Pacific Conference, the 23rd Taiwan Symposium on Innovation Economics and Entrepreneurship, and the 16th NYCU Finance Conference. We provide our codes and example data to demonstrate our procedures and to facilitate their applications via the following webpage: `https://github.com/hcchuang/Limitation-of-Firm-Fixed-Effects-Models-and-the-Missing-R-D-Patent-Relation`. We acknowledge the financial support from the National Science and Technology Council, Taiwan.

Email addresses: `hcchuang@gm.ntpu.edu.tw` (Hui-Ching Chuang), `pohsuanhsu@mx.nthu.edu.tw` (Po-Hsuan Hsu), `ckuan@ntu.edu.tw` (Chung-Ming Kuan), `jcyang1225@ntu.edu.tw` (Jui-Chung Yang)

1. Introduction

A common practice in the economics and finance literature is to include firm fixed effects in regression models to control or absorb all firm-specific, time-invariant factors. This approach is widely adopted because it is challenging for researchers to identify the "correct" regression models (and to address potential endogeneity); thus, to be conservative, they may choose to estimate firm fixed effects regressions and claim that their test results are not subject to the influence of firm-specific unobservables.[1] However, without theoretical modeling or appropriate econometric designs, such prevailing (almost by default) use of firm fixed effects may "throw the baby out with the bathwater" by eliminating the explanatory power of important economic factors that are persistent in time and mainly vary across firms. In this paper, we aim to propose appropriate and feasible econometric methodologies to tackle this issue.

Our first challenge is the fact that no one knows the underlying data generating process of financial data, which makes it difficult for us to justify whether the explanatory power of variables of interest is indeed biased by the inclusion of firm fixed effects. We clear this concern by focusing on the positive relation between R&D input and patent output that is intuitive and has economic foundation.

While the R&D-patent relation has been examined several decades ago (e.g., Hausman et al., 1984; Pakes and Griliches, 1984; Griliches, 1990; Kortum, 1993), it has attracted more attention lately possibly due to two reasons. First, accelerating technological advancement and intensifying business competition make it necessary for corporations to heavily invest in R&D.[2] Firms' intention to create patents to protect their intellectual property is certainly the most important incentive for their substantial R&D investments.[3] Second, there is a surge of studies on "new factors" that are able to explain corporate innovation performance (see the reviews of Ederer and Manso, 2011; He and Tian, 2018; Lerner and Seru, 2022). Given the important role of R&D input in the production of patents, these studies often include R&D-related proxies in their regressions as a control variable.

Our literature review suggests a surprising and puzzling pattern: the role of R&D input is often missing in prior studies because 40% to 50% of their estimates show insignificant or even

[1] This practice, as a solution for endogeneity concerns, has been discussed in Angrist and Pischke (2009), Imbens and Wooldridge (2009), and Roberts and Whited (2013).

[2] Amazon leads all U.S. public firms by spending $73 billion in 2022, followed by Alphabet and Meta which spent $40 and $35 billion, respectively, in the same year. Source: https://www.fdiintelligence.com/content/feature/global-innovation-leaders-2022-edition-82527

[3] Other reasons include the investment in the ability to absorb external knowledge and technology (Cohen and Levinthal, 1989; Arora, Belenzon and Sheer, 2017).

negative coefficients on R&D input (more details are provided in Section 2). These findings of an insignificant R&D-patent relation can be directly attributed to the prevailing use of firm fixed effects (Aghion et al., 2013; Balsmeier et al., 2017).[4] A more thorough discussion of this issue is found in Hall et al. (2005, p.26): "*We have chosen not to include fixed firm effects for a number of reasons... [b]ecause R&D tends to change slowly over time, a firm's R&D intensity is highly correlated with its individual effect; in fact, it is an important component of what creates differences across firms, so removing these effects would entail an overcorrection.*"[5]

In fact, the missing role of R&D input leads to two concerns about recent studies and the new factors they propose to explain patent output. First, if R&D input cannot explain patent output in a regression model, then to what extent can we accept the explanatory power of those new factors? This issue can be easily fixed if prior studies also consider regression models without firm fixed effects and, in turn, recover the positive coefficient on R&D input. Unfortunately, most papers we review only report results from firm fixed effects models. This leads to our second concern: if the role of a new factor is only examined using firm fixed effects regressions, its explanatory power may not be substantial for patent output; instead, it may only capture the time-series variation of patent output, which is relatively small when compared with the cross-sectional variation (Hausman et al., 1984; Hall et al., 2005). In addition, statistical inferences based on firm fixed effects regressions may be dominated by firms that are featured with larger within-firm variation (i.e., time-series variation) (Breuer and DeHaan, 2024). To the best of our knowledge, there is a lack of appropriate econometric tools to address the aforementioned issue for more reliable statistical inferences. Thus, a common practice in the literature is not to include firm fixed effects (Hall et al., 2005; Noel and Schankerman, 2013), which may also introduce alternative biases.

An even more thorny issue is the prominent possibility that unobserved heterogeneity exists in some firms but not others. For example, some managers are aggressive in investing in R&D and pursuing patent output but others are not, or some firms have a strong, innovation-oriented culture, while others do not. To appropriately model the aforementioned plausible scenarios, we cannot resort to models that only allow (or disallow) the use of dummies for all individual firms. (Throughout this paper, we use the term "firm fixed effects" to denote the inclusion of dummies

[4]Aghion et al. (2013) wrote: *"In the specifications where we include fixed effects, the coefficient on the R&D stock falls significantly."* Balsmeier et al. (2017) wrote *"Alternative regressions with R&D investments scaled by total assets reveal a significant positive effect only in specifications without firm fixed effects. Inclusion of controls for time-invariant firm heterogeneity leads to statistically insignificant results."*

[5]This issue is also pointed out in Hausman et al. (1984), Pakes and Griliches (1984), Blundell et al. (1995), Zhou (2001), Hall et al. (2007), and Noel and Schankerman (2013). Small variation in within-groups amplifies the measurement error issue (Griliches and Hausman, 1986; Jennings et al., 2024).

for *all* firms. On the other hand, we often use the term "firm dummies" when we discuss some dummies for *specific* firms.) Thus, an econometric methodology that can *select* which individual firm dummies are informative to the R&D-patent relation is called for.

This paper proposes two approaches to mitigate potential bias due to firm fixed effects. First, we propose the adjusted Hausman and Taylor ("adjusted HT" henceforth) approach. Hausman and Taylor (1981) approach is well-known to address the time-invariant variable of interest to be correlated to firm heterogeneity. As shown in Wooldridge (2002, Ch. 11) and Hansen (2022, Ch. 17), the Hausman-Taylor estimator can be obtained using Generalized Method of Moments (GMM) in which "demeaned" independent variables form moment conditions. We implement the GMM using the moment conditions for demeaned independent variables including R&D that is rarely time-varying.

Second, we use advanced machine learning based on the Least Absolute Shrinkage and Selection Operator (LASSO, Tibshirani, 1996), which has been used to select predictors and control variables (Chinco et al., 2019; Feng et al., 2020; Gu et al., 2020; Erel et al., 2021) and individual dummies (Kock and Tang, 2019; Chan and Mátyás, 2022; Semenova et al., 2023).[6] In particular, we use the post-regularization LASSO ("PRL" henceforth, Chernozhukov et al., 2015) estimation and its extension based on double machine learning ("DML" hereafter, Chernozhukov et al., 2018) to select individual firm dummies in explaining firm-level patent outputs.[7]

We then compare our estimation and simulation outcomes of adjusted HT and PRL estimators to those from simple OLS or Poisson regressions without and with firm fixed effects to present a comprehensive picture for the R&D-patent relation, as well as to offer new insights concerning regression model choices.

Our comparison starts with empirical estimation. We construct panel data of firm-year observations by collecting patent data from the PatentsView database and by collecting accounting and financing data of publicly-listed firms from the CRSP and Compustat database. We then estimate the extent to which firm-level patent output (measured by the number of new patents filed or the number of forward citations of these patents) can be explained by an extensive list of economic variables (e.g., R&D input, firm size, cash holding). Similar to the majority of prior studies in the literature, we implement OLS and Poisson regression estimation in which dependent variables are

[6] Our paper differs from theirs because we utilize LASSO as the causal inference instrument to select appropriate individual firm dummies in linear and Poisson regression models.

[7] While Semenova et al. (2023) also employ LASSO for selecting individual dummies within the DML framework, our objectives differ remarkably. Semenova et al. (2023) aim to debias the Nickell (1981) bias in dynamic panel models, whereas our focus is on identifying the coefficients of time-persistent variables.

the logarithmic and raw values of patent output, respectively.

Our OLS regression results suggest a robust pattern that R&D input positively explains patent output when regression models include an extensive list of control variables but do *not* include firm fixed effects (FE). The economic magnitude of this positive relation is eliminated or severely weakened when firm fixed effects are included in regressions, which is consistent with the concerns we discussed earlier. In addition, as suggested in Breuer and DeHaan (2024), we examine the within- and between-firm variations (i.e., time-series and cross-sectional variations, respectively) of our R&D and patent variables. We find that these variables' between-firm variation is much larger than their within-firm variation, which further verifies the influence of firm fixed effects on the missing R&D role.

We find that, contrary to FE regressions, the adjusted Hausman-Taylor method and the OLS without firm FE deliver a positive R&D and patents relation. We then use the PRL estimator and its extension based on the DML method, allowing us to select an appropriate model containing only important covariates, including individual firm dummies. We find that these two advanced methods select about 10% to 20% of firm dummies that have distinct explanatory power other than conventional control variables and should be included in regression models.[8] More importantly, the coefficients on R&D input based on the three estimators we propose are statistically significant, and their economic magnitude is much closer to those from OLS models without firm FE than those with firm FE. These results and prior analyses suggest that most firm dummies do not play a crucial role in identifying the R&D-patent relation (or the determinants of innovation output in general). These findings also highlight the possibility that the bias from adding all firm dummies may be costlier than that from not adding any dummies.

We then implement simulations based on the data generation process (DGP) of Cohn et al. (2022) to model between- and within-firm variations in panels and to examine the efficacy and efficiency of different estimators. As between-firm variation increases relative to within-firm variation, fixed effects estimators tend to have lower rejection rates and higher error rates than other estimators including OLS models without firm fixed effects, adjusted Hausman-Taylor estimators, and PRL estimators. This finding is consistent with prior studies' argument: the high correlation between persistent R&D and (unobservable) firm features hinders the ability of fixed effects estimators to capture the genuine R&D-patent relation. Moreover, we show that the proposed PRL methods outperform adjusted Hausman-Taylor estimators and OLS models without firm fixed ef-

[8]For the oracle property of LASSO in the selection of FE, see Kock and Tang (2019).

fects, highlighting the advantage of machine learning in handling persistent explanatory variables.

We also consider Poisson regressions that are common for count-based dependent variables in the economics and finance literature (Cohn et al., 2022). Our empirical evidence shows that, similar to our results in OLS regressions, the existence of firm FE also prevents Poisson models from delivering a positive R&D-patent relation. When we implement PRL and DML methods for Poisson regressions, we again find that the estimated coefficients on R&D input are much closer to those from Poisson regressions without firm FE than those with. Simulations suggest that Poisson regressions with firm FE underperform those without in efficacy and efficiency, and that the PRL method performs the best among all estimators. Thus, the bias that we attempt to highlight and solve in this paper is a general issue across different econometric models.

This paper provides feasible solutions for empirical researchers who are concerned with persistent explanatory variables that likely correlate with important between-firm unobservables. The first solution includes the PRL and DML methods, which are easy to implement (as illustrated in the link on our title page) and can handle omitted variable issues without strong model assumptions. More importantly, these methods not only resolve the debate on whether firm fixed effects should be included or not, but also enable researchers to decide exactly which firm dummies to be added in regressions. In addition, our empirical evidence that machine learning methods select only a small portion of individual firm dummies and that their estimates for R&D input are much closer to OLS or Poisson models without firm fixed effects suggests that the bias from including firm fixed effects may overpower the bias from not adding any at all. Thus, a claim that using firm fixed effects is a "conservative" choice can be challenged.

At the very least, researchers should consider checking between- and within-firm variations as well as estimating regression models without firm fixed effects to justify their regression model choices. When the economic magnitude and statistical significance of the coefficients of interest vary widely across regressions with and without firm fixed effects, researchers should discuss the variations (cross section versus time series) of their variables of interest.

This paper adds to the innovation economics, corporate finance, and applied econometrics literature as follows. First, noting that empirical studies tend to solve firm-specific, time-invariant unobservables issues by using fixed effects models (e.g., Angrist and Pischke, 2009; Imbens and Wooldridge, 2009; Roberts and Whited, 2013), we illustrate the potential biases of such a practice by using the intuitive R&D-patent relation. More importantly, we offer several feasible and easy-to-implement methodologies to enable economists to analyze the effects of economic variables that are persistent in time, such as ownership structure and managerial capability. In particular, we

provide explanations that they may use to justify their choice of regression specifications without firm fixed effects (or with only a limited set of firm dummies).

From a broader perspective, we demonstrate how to apply modern machine learning techniques to empirical research, in particular for the selection of relevant covariates (e.g., Chinco et al., 2019; Feng et al., 2020; Gu et al., 2020; Erel et al., 2021) and individual dummies (Kock and Tang, 2019; Chan and Mátyás, 2022; Semenova et al., 2023). Without theoretical/structural models, it is often challenging for researchers to determine input variables to be included in regressions and to address endogeneity (e.g., Coles et al., 2012; Roberts and Whited, 2013). The methods we propose offer researchers a relatively more objective approach for covariate selection; moreover, this approach is robust to the existence of between and within heterogeneity.

This study also adds to the literature by empirically justifying prior studies' choice of not including firm fixed effects to estimate knowledge production functions (Pakes and Griliches, 1984; Blundell et al., 1995; Hall et al., 2007; Noel and Schankerman, 2013). The advanced machine learning techniques we use to restore the positive R&D-patent relation facilitate economists' exploration of new factors in knowledge production. The fact that only a small portion of individual firm dummies are selected by the PRL and DML methods indicates that placing too much emphasis on firm fixed effects may restrict researchers' attention with respect to within-firm heterogeneity, which would cause them to miss important explanatory variables for between-firm heterogeneity.

The rest of our paper proceeds as follows. In Section 2, we review the literature on knowledge production functions and recent studies on new factors of firms' patent output. In Section 3, we present our data sources and empirical designs. In Section 4, we examine the R&D-patent relation in OLS regressions without and with firm fixed effects. In Section 5, we implement the adj-HT, PRL, and DML methods to reexamine the R&D-patent relation. In Section 6, we examine the R&D-patent relation in Poisson regressions. We conclude this paper with Section 7. Technical details of PRL and DML methods and supplementary tables and figures are provided in the Internet Appendix.

2. The R&D-Patent Relation in the Recent Literature

As an essential class of intangible capital and intellectual property, patents have been widely regarded as an important indicator of firms' innovation output (Kamien and Schwartz, 1975; Griliches, 1990).[9] With the surge of technology competition and patent lawsuits, researchers have

[9]Griliches (1998, chap. 13) noted *"[n]othing else even comes close in the quantity of available data, accessibility, and the potential industrial, organizational, and technological detail."*

gained more interest in identifying and analyzing the driving forces for firms' patent output. Their accumulated efforts in this direction have led to a burgeoning literature on the new factors of corporate innovation.

The foundation for this literature can be traced back to a Cobb-Douglas production function for the creation of knowledge capital (e.g., Pakes and Griliches, 1984; Hall et al., 1986; Kortum, 1993). Such a knowledge production function can be tested using reduced-form OLS regressions with a patent-based dependent variable, and many explanatory variables that reflect different input factors (including R&D) that may contribute to knowledge accumulation. With some modifications, the function can also be tested using Poisson or negative binomial regressions.

After a thorough review of the literature,[10] we focus on 36 papers that are most relevant to the R&D-patent relation: they examine how patent output is explained by variables of interest and include R&D input as an independent variable. Table 1 provides an overview of these studies, which were published from 2007 to 2021. In these papers, patent output measures include the number of patents a firm produce in a period (patent counts), the number of forward citations received by these patents (sometimes referred to as "citation-weighted patent counts"), and their modifications (e.g., truncation-adjusted patent counts, forward citations per patent).[11] Among these papers, 30 include both patents and forward citations in their studies; of the rest, four studies use only patents, and two use only forward citations. Their R&D input measures can be classified as R&D intensity and R&D stock. R&D intensity includes R&D expenses (or their multi-year sum) scaled by total assets, sales, or market equity. R&D stock includes R&D expenses (or their multi-year sum) in logarithm. Among the 36 papers, 27 use R&D intensity and 10 use R&D stock (note: one paper uses both of them).

We find that 21 papers include at least one regression with firm fixed effects (FEs), and 15 studies do not consider firm FEs models. In addition, consistent with the findings in Cohn et al. (2022), OLS regressions are more popular in the literature: 30 studies use OLS regressions as the major estimation method. Only six (four) studies use the Poisson (negative binomial) model.[12]

[10]We start our search with papers covered by the review papers of Ederer and Manso (2011), He and Tian (2018), and Lerner and Seru (2022) as well as our own reading lists. We then restrict the publications to 197 journal articles published in the following journals: Econometrica, Journal of Accounting and Economics, Journal of Financial Economics, Journal of Financial and Quantitative Analysis, Journal of Law and Economics, Management Science, Review of Financial Studies, The American Economic Review, The Journal of Finance, and The Review of Economics and Statistics. Among these journal articles, 1.52% were published in the 1980s, 4.06% in the 1990s, 15.74% in the 2000s, and 62.44% in the 2010s.

[11]We do not consider other patent-based variables such as basicness, exploration, exploitation, originality, and generality.

[12]For those studies that passed our screening conditions on R&D, patent, and citation definitions, we find only one

We analyze the statistical significance of the R&D-patent relation (i.e., whether the R&D coefficients are significantly positive at the 5% level or lower) in these selected studies that differ in estimation methods, dependent variables (patents or citations), and R&D input (intensity or stock). We first focus on studies using OLS estimation. Panels (a) and (b) of Figure 1 present an overview of studies using the number of patents as the dependent variable, and Panels (c) and (d) of Figure 1 present an overview of studies using the number of forward citations as the dependent variable.

As every article presents multiple estimates on the R&D-patent/citation relation, we design Figure 1 in two ways: an article-based counting method in Panels (a) and (c); and a regression-based counting method in Panels (b) and (d). For article-based counting, we present the number of articles in which at least one of the estimates satisfies the marked condition in each step (i.e., R&D intensity or stock, with or without firm FEs, and R&D coefficients being significantly positive or not). Thus, the numbers in the parentheses of the child nodes do not exactly add up to the numbers in the parenthesis of the parent node. For our regression-based counting method, we present the number of all regression estimates of all articles that satisfy the marked condition.

Panel (a) of Figure 1 starts with 28 papers that consider patent counts as the dependent variable. Among 24 studies that use R&D intensity as the input, 11 control for firm FEs. In this strand, 7 out of 11 papers (63.6%) include any significantly positive R&D estimates. On the other hand, 6 of 11 papers (54.5%) include any negative or insignificantly positive R&D estimates. Among the other 14 studies that do not control for firm FEs, all include at least one significantly positive R&D estimate (and only 2 out of 14 (14.3%) report any negative or insignificantly positive R&D estimates). As for the studies based on R&D stock tht appear in the lower part of this figure, we do not discuss them due to their small sample size (only five).

Panel (b) starts with 134 regression estimates and presents largely similar results. Among 122 regressions that use R&D intensity as the input, 66 control for firm FEs. Next, 28 out of 66 regressions (42.4%) post negative or insignificantly positive R&D coefficients. Among the other 56 regressions without firm FEs, only five regressions fail to discover a significant R&D-patent relation.

Panels (c) and (d) of Figure 1 analyze the R&D-citation relation in the same framework, and we find similar patterns as with the R&D-patent relation. As a result, we argue that the R&D-patent

paper that uses a random effect model (Gomes-Casseres et al., 2006), one that uses a Fama Macbeth procedure (Cohen et al., 2013), one that uses a hierarchical linear model (Griffin et al., 2021), and two that use Tobit models (Faleye et al., 2011; Celikyurt et al., 2014). We exclude those studies from our discussion.

relation is missing in a substantial number (roughly half) of prior studies based on OLS regressions.

Figure IA.1 in the Internet Appendix offers an overview of prior studies based on count models (i.e., Poisson and negative binomial regressions).[13] Similar to the prior figure, Panels (a) and (c) of this figure are based on the article-based counting method; and Panels (b) and (d) are based on the regression-based counting method.

Due to the small number of papers based on count models, we focus on the regression-based counting method. Panel (b) of Figure IA.1 starts with 50 regressions that use patent counts as the dependent variable. Among 48 regressions that use R&D stock as the input, 45 control for firm FEs or a pre-sample mean. In this strand, 18 of 45 regressions (40.0%) present negative or insignificantly positive R&D estimates. Among the other 3 regressions that do not control for firm FEs, all present significantly positive R&D estimates. Finally, we identify seven studies using citation counts as the dependent variable in count-based models.

Our literature review connects the missing R&D effect to firm FEs. Figure 1 and Figure IA.1 collectively present a pattern that the important role of R&D input is not supported by a substantial part (40% to 50%) of prior studies' regressions with firm FEs. On the other hand, a positive R&D-patent relation is much more likely to hold in regressions without firm FEs. This pattern is consistent with the findings of Aghion et al. (2013) and Balsmeier et al. (2017), and supports our concern for "throwing the baby out with the bathwater." As discussed in the economics literature, controlling for firm FEs is problematic in the estimation of the R&D-patent relation because R&D input changes slowly over time and is highly correlated with firms' individual effects (Hausman et al., 1984; Pakes and Griliches, 1984; Blundell et al., 1995; Zhou, 2001; Hall et al., 2007; Noel and Schankerman, 2013). Moreover, small variation in within-firm may result in large measurement errors (Griliches and Hausman, 1986; Jennings et al., 2024).

Nevertheless, as we are well aware of heterogeneity in firm-level unobservables and associated endogeneity concerns, it is unrealistic to simply ignore firm FEs in empirical analyses. Thus, a more appropriate econometric methodology for identifying the explanatory power for R&D input (as well as other important factors that are persistent in time) is needed.

[13] As the high-dimensional FEs Poisson regression has not been available until recently, we also include the following studies that use the pre-sample mean of patent output as a control variable to capture between-firm heterogeneity: Aghion et al. (2013); Bena et al. (2017); Blanco and Wehrheim (2017); Bloom et al. (2013); Custódio and Metzger (2014); Galasso and Simcoe (2011).

3. Data and Empirical Methods

In this section, we discuss how we construct our sample for our empirical analysis, how we measure patent output, how we measure R&D input, and how we examine the R&D-patent relation with and without firm fixed effects.

3.1. Sample formation

To measure firms' patent output, we collect patent and citation data from the PatentsView database maintained by the US Patent and Trademark Office (USPTO). The PatentsView database provides detailed information on all patents granted by the USPTO since 1976. We then link the PatentsView data to publicly listed firms in the Center for Research in Security Prices ("CRSP" henceforth) database and the Compustat Industrial Annual files ("Compustat" hereafter) that provide firms' accounting and financial information using the patent assignee information of Kogan et al. (2017). We exclude financial and utility firms with four-digit standard industrial classification (SIC) codes in 6000-6999 and 4900-4999. We only consider the ordinary shares traded on NYSE, AMEX, and NASDAQ. Firms with negative and missing total assets are excluded. In addition, we collect institutional ownership data from the Thomson Reuters Institutional Holdings (13F) database.

We consider the following three common measures to capture firm-level innovation output: Patent denotes the number of new patents filed by a firm to the USPTO per year (and is granted later), Citation denotes the number of forward citations received by these patents, and Adjcitation denotes the number of adjusted forward citations received by these patents. For each patent, we calculate its number of adjusted forward citations as its number of forward citations scaled by the average number of forward citations of all patents that are assigned in the same subsection and are granted in the same year (Seru, 2014). The first measure Patent is simple yet widely used in the economics and finance literature. The two citation measures offer quality-adjusted proxies for firms' innovation output because patents that are more technologically important are likely to receive more forward citations (Trajtenberg, 1990; Hall et al., 2005; Aghion et al., 2013).

Two issues are worth discussion before we proceed. First, we use a patent's application year rather than its grant year to construct the aforementioned measures because the former time placer is closer to the occurrence of firms' innovation output (Hall et al., 2005) and facilitates our analysis for the input-output relation between R&D and patents. Second, for a firm that does not have any patent filing record in a given year, we enter zeros for its Patent, Citation, and Adjcitation.

Our first sample includes 86,341 firm-year observations of 11,544 unique firms from 1975 to 2000. We intentionally end our sample in 2000 to be consistent with most prior studies' samples. In an unreported robustness check, we consider an extended sample period from 1976 to 2010 and find consistent results. Given that not all firms file for patents, we also construct our second sample, which includes 45,913 firm-year observations of 4,312 unique firms that have at least one patent filled during the same period, because some prior studies only consider patenting firms in their samples (e.g., Bloom et al., 2013; Aghion et al., 2013).

3.2. Patent production function estimation

As discussed earlier, the economics literature has suggested a Cobb-Douglas production function for the relation between R&D input and patent output (e.g., Pakes and Griliches, 1984; Hall et al., 1986; Kortum, 1993) To empirically examine the R&D-patent relation, we use the OLS regression to estimate the following model:

$$Innov_{i,t+1} = \gamma\, R\&D_{i,t} + \mathbf{x}_{i,t}^{\top}\boldsymbol{\beta} + \alpha_i + \epsilon_{i,t}, \tag{1}$$

in which the dependent variable $Innov_{i,t+1}$ is ln(1+Patent), ln(1+Citation), or ln(1+Adjcitation) of firm i in year $t+1$ for OLS regressions. We also consider the Poisson regression, in which $Innov_{i,t+1}$ is specified to be Poisson-distributed, with:

$$\mathbb{E}\left(Innov_{i,t+1}\middle|\, R\&D_{i,t}, \mathbf{x}_{i,t}\right) = \exp\left(\gamma\, R\&D_{i,t} + \mathbf{x}_{i,t}^{\top}\boldsymbol{\beta} + \alpha_i\right), \tag{2}$$

in which the dependent variable $Innov_{i,t+1}$ is Patent or Citation of firm i in year $t+1$.

The primary variable of interest, $R\&D$, denotes firm i's R&D input in year t. Following the literature, we consider both R&D intensity and R&D stock. Since firms' R&D input highly correlates with firm size, we also include the logarithmic value of a firm's market equity, ln(ME), at calendar year-end in our main results.

Our primary proxy for R&D input in Equations (1) and (2) is R&D/Asset (R&D intensity), which is defined as firm i's five-year accumulated R&D expenses divided by total assets in the fiscal year-end of year t, as total assets appear to be the most common scaling variable for R&D input in prior studies. Nevertheless, we also use market equity to scale five-year accumulated R&D expenses in our robustness checks (i.e., R&D/ME). Since $R\&D$ defined in this way has been scaled, we include firm size in our regressions. Our second proxy for R&D input is ln(1+R&D) (R&D stock), which is defined as the logarithmic value of firm i's five-year accumulated R&D expenses

in the fiscal year-end of year t plus one. When we use this R&D input variable, we do not include variables related to firm size in regressions to avoid estimation biases due to collinearity.

We acknowledge that some firms may choose not to report their R&D expenses (Koh and Reeb, 2015); thus, we also include an indicator variable (R&D missing) that equals one for firms without R&D expenses and zero otherwise. In addition, $\mathbf{x}$ denotes a set of firm characteristics that previous studies have shown as explanatory variables of firms' innovation output.[14] All explanatory variables are lagged for one year and winsorized at the 1st and 99th percentiles.

Table 2 reports the summary statistics and correlation coefficients of the innovation measures and explanatory variables. Panel A lists the pooled mean, median, standard deviation (Stddev), minimum (Min), 25th percentile (Q1), 75th percentile (Q3), and maximum (Max) for the entire sample. The distribution of patent measures is similar to that of prior studies' summary statistics. We find that 43% of firm-year observations missing R&D reporting. Five-year R&D input amounts to 13% and 15% of total assets and market equity, respectively. Panel B reports the pairwise correlation coefficients of innovation measures and explanatory variables. We find that patent output is highly correlated with R&D input. For example, ln(1+Patent)'s correlation coefficients with ln(1+R&D), R&D/Asset, and R&D/ME are 0.55, 0.17, and 0.08, respectively. All of these correlation coefficients are statistically significant.

3.3. Within and between variations

As we explained earlier, the insignificant and negligible explanatory power of R&D input for patent output can be attributed to the fact that both R&D and patents are persistent across years (i.e., the within variation is small) and mainly vary across firms (i.e., the between variation is

[14]This set includes the following: ln(1+Age) is the natural logarithmic value of one plus the number of years since a firm's first appearance in Compustat. ln(K/L) is the natural logarithmic value of net property, plant, and equipment (PPENT) divided by the number of employees in Compustat. Tobin Q proxies a firm's value and is defined as the market value of equity plus the book value of assets minus the book value of equity minus balance sheet deferred taxes, divided by the book value of assets. ROA, the return on asset, is the profitability measure, which is calculated as the operating income before depreciation divided by the book value of total assets. Leverage controls for the financial leverage behavior in the innovation process, defined as the long-term debt plus current debt, divided by total assets. Cash/Asset is the liquidity of firms' cash flow measures and is the ratio of cash to total asset. KZ index, following Kaplan and Zingales (1997), is the financial constraint index, which is calculated as -1.002×(Cash flow)+0.283×(Tobin Q)+3.139×(Leverage)-39.368×(Dividends)-1.315×(Cash), in which Cash flow is the gross income before extraordinary items, and Dividends is calculated as the sum of common and preferred stocks dividends. Cash flow, Dividend, and Cash are all scaled by the lag net PPENT. InstOwn represents institutional ownership and is defined as the total institutional ownership from 13F filings divided by the total number of outstanding shares. (1-HHI) is one minus the Herfindahl-Hirschman index, and (1-HHI) square captures the nonlinear relation of market competition and innovation activities. Herfindahl-Hirschman index is the sum of squared sales-based market shares of all firms in the same industry. The industry is defined using the 3-digit SIC code.

large). To further verify this finding, we examine the between and within variations of proxies for patent output and R&D input for our baseline model.

Table 3 reports the variations of the patent and R&D variables for our baseline model. The "Overall Stddev" is the standard deviation of the overall pooled observations. "Between Stddev" first calculates the firm-level averages for each variable of interest, and then calculates the standard deviation for these averages. "Within Stddev" measures the within firm variation of each variable. It calculates the standard deviations of a variable of interest with firm clusters, and then averages those standard deviations. We calculate the ratio of between-to-within variation and within-to-overall in the last two columns.[15]

We present our estimation results based on the full sample in Panel A of Table 3. As expected, the within variation is rather small compared to the between variation. For example, the within variation of ln(1+Patent) is 0.430, which is only 38.9% of its total variation of 1.105. A similar ratio is found in ln(1+Citation) and ln(1+Adjcitation). These patterns suggest that the cross-sectional variation of firms' innovation output is much more than the time-series variation. In addition, we also find that R&D input's within variation is smaller than its between variation. For instance, the within and between variations of R&D/Asset are 0.117 and 0.265, respectively. The contrast is even greater for ln(1+R&D): its within and between variations are 1.329 and 4.507, respectively. These findings indicate that the cross-sectional variation dominates the time-series variation for R&D input.

In Panel B of Table 3, we present the same statistics for the sample of patenting firms and find largely consistent results. All results presented in Table 3 support our argument that the missing R&D-patent relation can be attributed to the fact that the between variation dominates the within variation of R&D and patent measures. In addition, the neglected dominance of the between variation may result in biases in statistical inferences as pointed out in Griliches and Hausman (1986); Jennings et al. (2024); Breuer and DeHaan (2024).

[15]We report the within and between standard deviation using the `xtsum` command in `STATA`. The reported number is bias-adjusted and may not sum to the overall deviation for unbalanced panels. See `ttps://www.stata.com/support/faqs/statistics/decomposed-variances-in-xtsum`

4. OLS Estimation without and with Firm Fixed Effects

4.1. Full sample

We first implement the reduced-form OLS regression estimation as specified in Equation (1) and Section 3.2.[16] We use the full sample with 86,341 firm-year observations of 11,544 unique firms from 1975 to 2000.

In our first set of analyses, we use R&D intensity (R&D/Assets) to capture R&D input and include ln(ME) in Equation (1). Panel A of Table 4 reports the estimation results when we measure a firm's innovation output using the logarithm of one plus the number of patents filed each year (ln(1+Patent)). For brevity, we omit the estimates of control variables in the main text but provide them in Internet Appendix Table IA.1.

Columns (1) present estimation results using firm and year fixed effects (FEs). The coefficients on R&D/Assets are insignificant, with values of 0.041. In contrast, when we exclude firm FEs but retain year FEs (Columns 2), the coefficients increase to 0.593 and become statistically significant. This stark contrast indicates that firm fixed effects absorb the impact of R&D on innovation output, rather than year specific factors.

Panels B and C of Table 4 extend this analysis to alternative measure of innovation output: the logarithm of one plus the number of forward citations (ln(1+Citation)) and adjusted forward citations (ln(1+Adjcitation)), respectively. Consistent with the patent count results, the coefficients on R&D/Assets are insignificant when firm and year FEs are included, with values of –0.051 and 0.033 for ln(1+Citation) and ln(1+Adjcitation), respectively. In sharp contrast, omitting firm FEs (Columns 2) leads to a significant increase in the coefficients to 1.396 for ln(1+Citation), and to 0.590 for ln(1+Adjcitation), all significant at the 1% level.

We also consider market equity (ME) as an alternative scaling factor for R&D intensity. Panel A.1, B.1, and C.1 of the Table 5 presents our results when R&D intensity is defined as the five-year accumulated R&D expense scaled by market equity (R&D/ME). Panels A, B, and C present the results when the dependent variables are the logarithmic value of the number of patents, forward citations, and adjusted-forward citations plus one, respectively. We arrange column layout similar to the Table 4. For brevity's sake, we omit the coefficients of other control variables.

In Column (1) of Panels A.1, B.1, and C.1—which include firm and year fixed effects—the coefficients on R&D/ME are insignificant and even negative: 0.026, –0.024, and 0.021, respectively.

[16] We implement the fixed effects estimation using the `reghdfe` package of Correia (2016) in `STATA` that is designed to deal with OLS regression with high dimensional fixed effects.

In contrast, when we omit firm fixed effects in Columns(2), the coefficients on R&D/ME increase significantly and become statistically significant, restoring the positive relationship between R&D and innovation output. Column (2) shows that the coefficients rise to 0.502, 1.040, and 0.506. This suggests that firm fixed effects absorb the impact of R&D on innovation, rather than time -specific factors.

In our second set of analyses, we consider the R&D stock (ln(1+R&D)) as R&D input and do not include ln(ME) in Equation (1) to avoid collinearity due to firm size. We report our results in Table 5. In Panel A.2 for patent, the coefficient on ln(1+R&D) is 0.023 in in Column (1). Without firm FEs (column 2), the coefficients on ln(1+R&D) increase sharply to 0.140. Although these coefficients are significant at the 1% level, they are small in economic magnitude and suggest that the R&D-elasticity of patent output drops by 85% when firm fixed effects included. A similar pattern is found in Panels B.2 and C.2 for citation and adjusted citation, respectively. The coefficients on ln(1+R&D) drop from 0.286 (without firm fixed effects) to 0.036, and from 0.143 to 0.022, respectively—corresponding to reductions of 87% and 85%.

As some prior studies focus on patenting firms only, we also consider a sample of firms that have at least one patent filled during our sample period. This reduced yet more focused sample includes 45,913 firm-year observations of 4,312 unique firms. Internet Appendix Table IA.2 lists the results. Similar to prior results based on our full sample, we find that the explanatory power of R&D input is substantial and statistically significant when we do not consider firm fixed effects. However, once these firm fixed effects are added into our regressions, the coefficients on R&D input become either statistically insignificant or much smaller in economic magnitude.

Our results thus far suggest that, in the absence of firm fixed effects, R&D input measured by R&D intensity or stock has significant explanatory power for firms' innovation output measured by patent and (adjusted-)citation counts. However, such an intuitive input-output relation becomes insignificant or is severely weakened by more than 85% once we add firm fixed effects into our regressions. This pattern is consistent with concerns raised by Hall et al. (2005) and many other economic studies: as R&D investment is a persistent variable across time, including firm fixed effects may underestimate the true effect of R&D input.

4.2. Adjusted Hausman-Taylor estimation

To quantitatively evaluate the effect of R&D input on the innovation output, in this paper, we consider a novel identification strategy inspired by the Hausman-Taylor estimator (Hausman and Taylor, 1981) for the coefficients of time-invariant regressors. We specify an empirical model

outlined as follows:

$$Innov_{i,t+1} = \gamma\, R\&D_{i,t} + \mathbf{x}_{i,t}^{\top}\boldsymbol{\beta} + \alpha_i + \epsilon_{i,t}, \tag{3}$$

where $Innov_{i,t+1}$ denotes the innovation output of firm i at year $t+1$, $R\&D_{i,t}$ denotes the R&D input of firm i at year t, and $\mathbf{x}_{it}$ is a k-dimensional vector of time-varying exogenous control variables, with $t = 1, 2, \ldots, T_i$ for $i = 1, 2, \ldots, N$. α_i denotes the firm fixed effects. Because of the persistent nature of R&D input, accurately identifying γ using the fixed effects estimator is challenging.

Hausman and Taylor (1981) studied the identification of the coefficient of the time-invariant variable. To fix the idea, consider the following model:

$$y_{i,t+1} = z_i\gamma + \mathbf{x}_{it}^{\top}\boldsymbol{\beta} + \alpha_i + \epsilon_{it}. \tag{4}$$

In equation (4), $y_{i,t+1}$ and z_i respectively correspond to $Innov_{i,t+1}$ and $R\&D_{i,t}$ in Equation (3). However, Equation (4) imposes an even stricter assumption by treating z_i as completely time-invariant.[17] Assume that both $\mathbf{x}_{it}$ and z_i are uncorrelated with the idiosyncratic error term $\{\epsilon_{it}\}$ across all time horizons. Additionally, following Hausman and Taylor (1981), we assume that $\mathbf{x}_{it}$ is exogenous to firm fixed effects (α_i); that is, for all $s, t = 1, 2, \ldots, T_i$ and $i = 1, 2, \ldots, N$, $\mathbb{E}(\mathbf{x}_{it}\epsilon_{is}) = \mathbf{0}$, $\mathbb{E}(\mathbf{x}_{it}\alpha_i) = \mathbf{0}$, and $\mathbb{E}(z_i\epsilon_{it}) = 0$.[18] Hausman and Taylor (1981) propose to identify γ and $\boldsymbol{\beta}$ and using the following moment conditions without specifically estimating α_i:

$$\mathbb{E}\left[\tilde{\mathbf{x}}_{it}^{\top}\left(y_{i,t+1} - z_i\gamma - \mathbf{x}_{it}^{\top}\boldsymbol{\beta}\right)\right] = \mathbf{0}, \tag{5}$$

$$\mathbb{E}\left[\mathbf{x}_{it}^{\top}\left(y_{i,t+1} - z_i\gamma - \mathbf{x}_{it}^{\top}\boldsymbol{\beta}\right)\right] = \mathbf{0}, \tag{6}$$

where $\tilde{\mathbf{x}}_{it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_i$ is the de-meaned version of $\{\mathbf{x}_{it}\}$, in which $\bar{\mathbf{x}}_i = T_i^{-1}\sum_{t=1}^{T_i}\mathbf{x}_{it}$ is the time-series mean of $\{\mathbf{x}_{it}\}$. Given the moment conditions, γ and $\boldsymbol{\beta}$ can be consistently estimated by the 2SLS (Wooldridge, 2002; Hansen, 2022) using $\tilde{\mathbf{x}}_{it}$ as internal instrumental variables.[19] Our

[17] Therefore, after applying the within transformation, z_i entirely vanishes, making it impossible to identify γ with the fixed effects estimator.

[18] To maintain relevance, this paper focuses on a simplified version of the original model, considering only the case where $\mathbf{x}_{it}$ is exogenous. The original model also allows at least a part of $\mathbf{x}_{it}$ is endogenous. For a comprehensive discussion, see Hansen (2022).

[19] $\mathbf{x}_{it}$ will be included in the 2SLS estimation by default. Hausman and Taylor (1981) also propose a feasible generalized least squares (FGLS) type procedure. Under the assumption that the idiosyncratic errors ϵ_{it} and the fixed effects α_i are strictly mean independent and homoskedastic, one can consistently estimate $\sigma_\epsilon^2 = \mathrm{var}(\epsilon_{it})$ and $\sigma_\alpha^2 = \mathrm{var}(\alpha_i)$. These estimates can then be utilized to estimate γ and $\boldsymbol{\beta}$.

generalized method of moments (GMM) estimators for γ and $\boldsymbol{\beta}$ can be consistently estimated as follows: we start with the two-stage least squares (2SLS) estimators, use the residual ($y_{i,t+1} - z_i\hat{\gamma}^{2SLS} - \mathbf{x}_{it}^{\top}\hat{\boldsymbol{\beta}}^{2SLS}$) for the covariance of the moments, and then use its inverse as the weighting matrix to re-estimate Model (4) without α_i to obtain GMM estimators $\hat{\gamma}^{GMM}$ and $\hat{\boldsymbol{\beta}}^{GMM}$.

Back to our research question, the population mean of R&D (for which we use $\overline{R\&D}_i$ as an estimate) may correlate with firm fixed effects (α_i) by nature. Therefore, although $R\&D_{i,t}$ itself cannot be used for moment conditions, we can use the de-meaned version of R&D, $\widetilde{R\&D}_{i,t}$=$R\&D_{i,t} - \overline{R\&D}_i$, as moment conditions because α_i will be differenced away and, like $\tilde{\mathbf{x}}_{i,t}$, $\widetilde{R\&D}_{i,t}$ are uncorrelated with the idiosyncratic error term $\{\epsilon_{i,t}\}$. Thus, we have the following moment conditions for Equation (3):

$$\mathbb{E}\left[\widetilde{R\&D}_{it}\left(Innov_{i,t+1} - \hat{\gamma}R\&D_{i,t} - \mathbf{x}_{it}^{\top}\hat{\boldsymbol{\beta}}\right)\right] = 0; \tag{7}$$

$$\mathbb{E}\left[\tilde{\mathbf{x}}_{it}^{\top}\left(Innov_{i,t+1} - \hat{\gamma}R\&D_{i,t} - \mathbf{x}_{it}^{\top}\hat{\boldsymbol{\beta}}\right)\right] = \mathbf{0}; \tag{8}$$

$$\mathbb{E}\left[\mathbf{x}_{it}^{\top}\left(Innov_{i,t+1} - \hat{\gamma}R\&D_{i,t} - \mathbf{x}_{it}^{\top}\hat{\boldsymbol{\beta}}\right)\right] = \mathbf{0}. \tag{9}$$

As a result, the adjusted Hausman-Taylor estimates for γ (the loading on R&D input) in Equation (3) can be estimated as follows: in the first stage, we regress $R\&D_{i,t}$ on $\tilde{\mathbf{x}}_{it}$ and $\widetilde{R\&D}_{i,t}$ together with control variables $\mathbf{x}_{it}$ and year (or industry-year) fixed effects. In the second stage, we regress $Innov_{i,t+1}$ on predicted $R\&D_{i,t}$, $\mathbf{x}_{it}$, and year (or industry-year) fixed effects for $\hat{\gamma}^{2SLS}$.

For our GMM estimation for γ, we start with the 2SLS estimators, use the residuals ($Innov_{i,t+1} - \hat{\gamma}^{2SLS}R\&D_{i,t} - \mathbf{x}_{it}^{\top}\hat{\boldsymbol{\beta}}^{2SLS}$) for the covariance of the moments, and then use its inverse as the weighting matrix to re-estimate Equation (3) without α_i to obtain GMM estimators $\hat{\gamma}^{GMM}$.[20] These estimated γs are our adjusted Hausman-Taylor estimates. The efficacy and efficiency of the Hausman-Taylor estimates will be discussed in Section 4.4.

4.3. Results from the adjusted Hausman-Taylor method

We apply the adjusted Hausman-Taylor method (adjHT) to our full sample using R&D/Asset as our R&D input and present our main results in Column (3) of the Table 4 (as well as the full table in Internet Appendix Table IA.1). Panels A, B, and C are estimation results for dependent variables of ln(1+Patent), ln(1+Citation), and ln(1+Adjcitation), respectively. Column (3) presents adjusted HT estimates with year fixed effects.

[20]It is easy to implement in statistical package. For instance, the STATA code will be: `ivregress gmm y control (rd = demean_rd demean_control), wmatrix(cluster firmID)`

First, the coefficients on R&D input are consistently positive and significant across all specifications and dependent variables. The coefficient estimates are 0.220, 0.551, and 0.231 for patents, citations, and adjusted citations with year fixed effects, respectively. The magnitude of the R&D/Asset coefficients using adjusted HT is lower than those obtained via OLS without firm fixed effects in Column (2).

More importantly, the R&D coefficients under the adjusted HT method are closer to those from OLS without firm FEs rather than those with firm fixed effects (Column (1)). For example, in patent regression, the estimate from OLS with firm fixed effects is 0.041, compared to 0.220 from the adjusted HT method and 0.593 from OLS without firm fixed effects. This suggests that including firm fixed effects may significantly underestimate the impact of R&D input on innovation output. Conversely, OLS without firm effects fails to account for firm-specific unobserved factors, potentially biasing the R&D effect by ignoring persistency in R&D activity within each firms. Overall, the adjusted HT approach offers a more balanced estimate of the R&D impact, capturing both firm-level variation and the effects of R&D input on innovation.

In line with our previous analysis, we use alternative R&D measures to assess the robustness of the R&D effect in the Table 5. Consistent with the results using R&D/Assets, all coefficients in Column (3) for R&D/ME and ln(1+R&D) are positive and highly significant. The adjusted HT estimates generally exhibit smaller magnitudes compared to the OLS estimates without firm fixed effects. However, when compared to the OLS estimates with firm fixed effects, the adjusted HT estimates show greater magnitudes. For instance, consider the regression of patents on R&D/ME (Panel A.1). The coefficients in Columns (1), (2), and (3) are 0.026, 0.2502 and 0.139 for OLS with firm fixed effects, OLS without firm fixed effects, and adjusted HT, respectively.

Lastly, we apply the adjusted Hausman-Taylor method to a subsample of firms that have filed at least one patent during the sample period. The results are consistent with our full-sample analysis, reinforcing the robustness of our findings that the adjusted HT method yields estimates closer to OLS without firm fixed effects, balancing biases from firm heterogeneity and omitted factors. Detailed estimates of the patenting firms sample can be found in the Internet Appendix Table IA.2.

4.4. Simulation study on the within- and between-variation

In this section, we conduct simulation experiments to investigate how varying levels of between- and within-firm R&D effects affect the efficacy and efficiency of different estimators. Specifically, we examine both discrete and continuous outcomes, embedding a within-variation design into the

data generation process (DGP) outlined by Cohn et al. (2022).

We simulate panel data observations $\{y_{i,t}, z_{i,t}, x_{i,t}\}$, where i indexes the firm and t indexes the year. $y_{i,t}$ is the discrete or the continuous dependent variable, and $z_{i,t}$ and $x_{i,t}$ proxy for R&D and control variables, respectively. Our primary interest is understanding how the between-firm and within-firm variations in $z_{i,t}$ affect the estimators. For this purpose, we define $z_{i,t} = 0.5\eta_i + 0.5\nu_{i,t}$ where η_i and $\nu_{i,t}$ are independently drawn from normal distributions truncated at the 1st and 99th percentiles, with standard deviations σ_η and σ_ν, respectively. η_i represents time-invariant firm fixed effects, and $\nu_{i,t}$ is the firm-and-time varying component. Consequently, the ratio of between variation to the total variation is given by:

$$\text{Between variation ratio} = \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_\nu^2},$$

We let the between variation ratio run from 0.71 to 0.91 to analyze its impact on estimation results. A higher value between variation ratio reflects an increase in the proportion of between variation relative to total variation, with a corresponding decrease in within variation. We hypothesize that the fixed effects estimators will deteriorate as the proportion of between variation increases.

To generate the discrete or continuous $y_{i,t}$, we adopt the DGP described by Cohn et al. (2022). For discrete outcomes, $y_{i,t}$ is drawn from a negative binomial distribution with conditional mean: $E(y_{i,t}|z_{i,t}, x_{i,t}) = \exp(\gamma z_{i,t} + x_{i,t})$, where we set the true value of γ as 0.2. The overdispersion parameter of negative binominal distribution is set as constant α, which controls the degree of heterogeneity relative to the standard Poisson distribution. A larger value of α corresponds to a greater dispersion of $y_{i,t}$. For continuous outcome, $y_{i,t}$ is generated using a mixture model. Each mixture component follows a $\chi^2(1)$ distribution. The number of mixture components is determined by the product of the conditional mean (as above) and a log-normal random variable with mean one and variance controlled by a parameter α.

For both count and continuous outcomes, α is set to 0.001, 0.5, and 2, representing cases of limited overdispersion, moderate overdispersion, and extreme overdispersion, respectively. In addition, $x_{i,t}$ is sampled from a normal distribution with zero mean and standard deviation of 2. The data is generated as a balanced panel, with 500 firms observed over 10 years each, and the Monte Carlo replicate number is 10,000.[21]

We first calculate the rejection rate for each simulated panel, defined as the percentage of

[21] Our simulation design is a modification of the code provided in Cohn et al. (2022), available at `https://data.mendeley.com/datasets/nc37vsf2bp/1`.

the 10,000 simulated panels in which the hypothesis of $\gamma = 0$ is rejected at the 5% significant level. The higher rejection rate implies that we have more confidence in identifying the significant R&D-Patent relation. The second and the third metrics concern the precision and variability of γ estimates: The mean square error (MSE) is the mean of the sum of squares of the estimated deviation from the actual value, and the mean absolute error (MAE) uses the absolute distance instead of the square distance.

Figure 2 compares performance of OLS with firm fixed effects ("FE", black triangle line), OLS without firm fixed effects ("OLS", blue circle line), and adjusted Hausman-Taylor ("adjHT", brown diamond line) estimators for discrete outcomes. FE consistently shows the lowest rejection rate and the highest MSE and MAE across all between-variation ratios and over levels. In low-overdispersion cases (subplot a-(i)), the FE rejection rate declines from 13.1% to 6.6% as the between-variation ratio increases. The drop is even steeper in high-overdispersion cases (subplot a-(iii)), reaching the lowest rejection rate as 4.7%. FE also demonstrates the highest MSE, which worsens with increasing between-variation ratios: in low-overdispersion cases (subplot b-(i)), the MSE is modest but peaks at 53.8 in high-overdispersion settings (subplot b-(iii)). MAE results follow the same trend as MSE, reinforcing the inefficiency of FE estimator.

The adjHT estimator's rejection rate outperforms that of FE estimators, with average increases ranging from 0.4% to 1.5%. The adjHT estimators also have lower MSE and MAE compared to FE, with MSE values ranging from 3.1 to 20.9, approximately 52% to 64% of the MSE for FE estimators. A similar reduction is observed in MAE results. The MAE values of adjHT account only for 81% to 85% of the MAE of the FE estimator. OLS estimators demonstrate robust performance across all settings, consistently achieving the highest rejection rates, lowest MSE, and lowest MAE. This underscores its reliability in this simulation design.

We find similar results for continuous outcomes in Internet Appendix Figure IA.2, mirroring the patterns in Figure 2. Both discrete and continuous designs underscore the inefficiency of FE estimators, consistent with empirical R&D studies where limited time variation and explanatory variables highly correlated with firm effects hinder the ability of fixed effects estimators to capture the genuine R&D-patent relation.

5. Machine Learning Methods

Some prior literature argues that not controlling for firm fixed effects may result in some non-negligible, omitted-variable bias for some firms or some industries. However, if we control for *all*

firm dummies, we may underestimate the effect of R&D input on patent output as shown in the prior section. Given that machine learning methods are able to facilitate the selection of predictors and control variables (Chinco et al., 2019; Feng et al., 2020; Gu et al., 2020; Erel et al., 2021; Reeb and Zhao, 2021) and individual dummies (Kock and Tang, 2019; Chan and Mátyás, 2022; Semenova et al., 2023), we use modern machine learning techniques to "select" only some firm dummies, so we may draw a more appropriate statistical inference for the R&D-patent relation. In particular, we propose the following two methods: the post-regularization LASSO regression (PRL) and double machine learning (DML).

5.1. Post-regularization LASSO regression (PRL)

The first technique is the post-regularization LASSO (PRL) estimator (Chernozhukov et al., 2015). Similar to the Frisch-Waugh-Lovell estimator (FWL, Frisch and Waugh, 1933; Lovell, 1963), the PRL estimator partials out nuisance variables before the estimation and inference of the parameters of interest. However, different from FWL, the PRL estimator selects predictors and control variables using LASSO, one of the state-of-the-art machine learning algorithms.

Based on the observation that *Innov* has a rather small within variation compared to the between variation, we consider a variation of Equation (1), which only controls for a few important firm dummies:

$$Innov_{i,t+1} = \beta_0 + \beta_1 R\&D_{i,t} + \mathbf{x}_{i,t}^\top \boldsymbol{\beta}_2 + \sum_{s\in\mathcal{S}} \alpha_s \text{dummy}_{s,i} + u_{i,t}, \tag{10}$$

in which $\mathcal{S}$ is a nonempty subset of $\{1, 2, \dots N\}$ that contains a few important firm dummies, and $\text{dummy}_{s,i}$ is the dummy variable such that $\text{dummy}_{s,i} = 1$ if $s = i$, and 0 otherwise.

Without loss of generality, $R\&D_{i,t}$ and $\mathbf{x}_{i,t}$ can also be written as:

$$R\&D_{i,t} = \gamma_{1,0} + \sum_{s\in\mathcal{S}} \gamma_{1,s}\text{dummy}_{s,i} + v_{1,i,t}, \text{ and } \mathbf{x}_{i,t} = \boldsymbol{\gamma}_{2,0} + \sum_{s\in\mathcal{S}} \boldsymbol{\gamma}_{2,s}\text{dummy}_{s,i} + \mathbf{v}_{2,i,t}, \tag{11}$$

Equations (10) and (11) lead to the following moment conditions:

$$\mathbb{E}\left[\left(\mathbf{x}_{i,t} - \boldsymbol{\gamma}_{2,0} - \sum_{s\in\mathcal{S}} \boldsymbol{\gamma}_{2,s}\text{dummy}_{s,i}\right)\left(Innov_{i,t+1} - \beta_0 - \beta_1 R\&D_{i,t} - \mathbf{x}_{i,t}^\top\boldsymbol{\beta}_2 - \sum_{s\in\mathcal{S}} \alpha_s\text{dummy}_{s,i}\right)\right] = \mathbf{0},$$

$$\mathbb{E}\left[\left(R\&D_{i,t} - \gamma_{1,0} - \sum_{s\in\mathcal{S}} \gamma_{1,s}\text{dummy}_{s,i}\right)\left(Innov_{i,t+1} - \beta_0 - \beta_1 R\&D_{i,t} - \mathbf{x}_{i,t}^\top\boldsymbol{\beta}_2 - \sum_{s\in\mathcal{S}} \alpha_s\text{dummy}_{s,i}\right)\right] = 0.$$

We note that when $\mathcal{S} = \{1, 2, \ldots, N\}$ (i.e., when all firm dummies are non-zero), the moment conditions are identical to the fixed effects model estimation. In this sense, fixed effects models can be viewed as a special case.

While $\mathcal{S}$ is known, the parameters of interest can be estimated by GMM. However, in practice, $\mathcal{S}$ is usually unknown and must be selected by researchers. In this paper, we follow Chernozhukov et al. (2015) and use LASSO.

The PRL method is a three-step procedure. First, we regress all of the variables of interest, including the patent output $Innov_{i,t+1}$, the R&D input $R\&D_{i,t}$, and other controls $\mathbf{x}_{i,t}$ on all N dummies using LASSO:

$$\min_{\{\alpha_j\}_{j=0}^{N}} \quad \sum_{i=1}^{N}\sum_{t=1}^{T}\left(z_{it} - \alpha_0 - \sum_{j=1}^{N}\alpha_j \text{dummy}_{j,i}\right)^2 + \lambda\sum_{j=1}^{N}|\alpha_j|. \tag{12}$$

In Equation (12), z_{it} denotes the variables of interest, and λ is the tuning parameter that controls the extent of shrinkage and selection. Given a non-zero tuning parameter λ, the shrinkage and selection operator would *shrink* the estimated coefficients toward zero. Furthermore, among all coefficients, the ones of dummies that are *not important* in predicting the values of z_{it} will be shrunk to zero, and those dummies with non-zero estimated coefficients are thus *selected* by LASSO. The value of λ is determined using the Bayesian information criterion (BIC). Second, we regress the variables of interest on the LASSO-selected dummies using the post-LASSO ordinary least squares to obtain the residuals. Third, we regress the residuals from the equation of $Innov_{i,t+1}$ (Equation (10)) on the residuals from the equations of $R\&D_{i,t}$ and $\mathbf{x}_{i,t}$ (Equation (11)).[22]

In the Internet Appendix A, we show that, under regularity conditions and the observation that the variables of interest depend only on a nonempty subset $\mathcal{S}$, the PRL estimator is asymptotically normally distributed. Technically, as long as the *sparsity* condition holds (i.e., the number of strong dummies is bounded from above by an order of $\sqrt{NT}/\ln N$), the PRL estimator has a conventional asymptotic distribution, which allows a conventional inference with respect to the coefficients of interest. Intuitively, since important dummies have been controlled for at the first stage, we are able to prevent the omitted-variable bias. On the other hand, since unimportant dummies are not selected, we are able to identify the coefficient of the persistent R&D input without potential underestimation.

[22]This method is straightforward to implement in statistical software. For example, the STATA code is: `poregress y rd control, controls(i.firmID) vce(cluster firmID)`.

5.2. Double Machine Learning

In addition to the PRL method, we also consider the double machine learning (DML) approach (DML, Chernozhukov et al., 2018), which is an extension of the PRL method. Before the estimation, the DML method partitions the sample into two independent folds: the main fold and the auxiliary fold. As in the PRL method, first we regress all of the variables of interest, including the patent output $Innov_{i,t+1}$, the R&D input $R\&D_{i,t}$, and other controls $\mathbf{x}_{i,t}$ on all N dummies using LASSO as in Equation (12), and then we regress the variables of interest on the LASSO-selected dummies using the post-LASSO ordinary least squares. However, instead of using the full sample as in the PRL method, we only use the auxiliary fold for the partialing-out regression. Finally, we plug in the main fold of the sample into the estimated equations, obtain the residuals, and regress the residuals from the equation of $Innov_{i,t+1}$ (Equation (10)) on the residuals from the equations of $R\&D_{i,t}$ and $\mathbf{x}_{i,t}$ (Equation (11)).

To further improve efficiency, Chernozhukov et al. (2018) propose the use of *cross-fitting*. Thus, after we obtaining our estimates, we reverse the procedure, using the main fold for the first stage of variable-selection and the auxiliary fold for the estimation of parameters of interest, and take the average of the two estimates.[23] In general, the sample-splitting step in DML may split the sample into K folds with a finite $K \geq 2$. In our empirical studies, we split the sample into 5 folds.

Compared to the PRL method, Chernozhukov et al. (2018) argue that sample splitting can improve the performance in inference. Intuitively, because of the independence of the main fold and the auxiliary fold, the dependence between two estimation stages is reduced. In the Internet Appendix A, we show that the DML estimator has a smaller deviation from the asymptotic approximation than the PRL method.

Notably, Semenova et al. (2023) also employ LASSO for selecting individual dummies within the DML framework. However, our objectives differ markedly. While Semenova et al. (2023) focus on debiasing the Nickell (1981) bias in dynamic panel models, our primary goal is to identify the coefficients of time-persistent variables.

5.3. Results from the PRL and DML methods

We apply the aforementioned estimation methods to our full sample using R&D/Asset as our R&D input and present our main results also in Table 4 (as well as the full table in the Internet Appendix Table IA.1). Panels A, B, and C are estimation results for dependent variables of

[23]This method is straightforward to implement in statistical software. For example, the STATA code is: `xporegress y rd control, controls(i.firmID) vce(cluster firmID) xfolds(#fold)`.

ln(1+Patent), ln(1+Citation), and ln(1+Adjcitation), respectively. Columns (4) to (5) are PRL estimates with selected firm and year dummies, and DML estimates with the same set of dummies.

In Panel A of Table 4 for patent output, we find that the explanatory power of R&D input is highly significantly positive. More importantly, the coefficients on R&D input are closer to the OLS estimates without firm FEs (Column (2)) than to those with firm FEs (Column (1)). For example, we find that the PRL estimates are 0.199, while the DML estimates are 0.213. These values are smaller than the estimates of 0.593 reported in Columns (2), which do not account for firm FEs. However, they are significantly larger than the FE estimates of 0.041 in Columns (1). The same patterns are found in Panels B and C for citations and adjusted citations, respectively. Results presented in Table 4 suggest that selecting important firm dummies restores the positive R&D-patent relation. Furthermore, the observation that both PRL and DML estimates deliver R&D-elasticity that is closer to an OLS model without firm FEs suggests that simply including dummies for all firms may not be the best practice in empirical analysis.

To further understand the size of "optimal firm fixed effects" based on our two methods, we also report the numbers of dummies for firms, years, and industry-years selected by the PRL and DML methods in Table 4. Before the selection, there exist 11,570 firms and year dummies (11,544+26) in Columns (4) and (5). Columns (4) of Panel A in Table 4 show that the PRL method selects 1,241 dummies out of 11,570 firm-year dummies, corresponding to a ratio of 10.73% dummies selected in the patent regression. In Columns (5), DML selects 15.01% of the dummies. Internet Appendix Table IA.3 presents the selected top 100 firms. For the citation (adjusted citation) regression, the selected ratios are 4.54% for PRL and 8.18% for DML (10.32% for PRL and 16.27% for DML).

We also perform several robustness checks and find consistent results. First, we use alternative R&D measures to examine the robustness of the R&D effect. Table 5 provides the coefficient estimates and the corresponding selected ratio for alternative R&D intensity, R&D/ME. We also use ln(1+R&D) as the proxy for R&D input and exclude ln(ME) from regression models. In the Internet Appendix Tables IA.2, we implement the same estimation as in Table 4 and 5 using a sample of firms with at least one patent in our sample period.

We find that about 10% to 20% dummies are selected by two methods, supporting our proposal that it is inappropriate to simply include dummies for all firms in the sample. Our evidence also offers an important insight: most firm dummies are not crucial for identifying the effect of R&D input on patent output. In addition, the R&D coefficient estimates based on these two methods are closer to our OLS estimates without firm FEs, which suggests that the bias from adding firm FEs overpowers the bias from not adding any of them. Thus, it could be inappropriate for researchers

to claim that the inclusion of firm FEs is a "conservative" choice in their study of the determinants of innovation performance.

In this section, we present two feasible solutions: post-regularization LASSO regression and double machine learning, both of which are easy for empirical researchers to implement (our codes and example data are provided in footnotes and the link on the title page). These methods not only resolve the debate for adding firm FEs or not but also enable researchers to decide exactly which firm dummies should be added into models for better identification.

5.4. Simulation results

This section presents a simulation study to evaluate the performance of machine-learning-based estimators under the DGP in Section 4.4. Both discrete and continuous outcomes are examined, with variations in the ratio of between-firm variation and overdispersion levels to assess their effects on rejection rates, mean squared error, and mean absolute error. Since the key distinction between PRL and DML lies only in sample splitting, we only present the simulations for PRL estimates.

Figure 2 shows that PRL (PRL, red square line) achieves the highest rejection rate for discrete outcomes. In the low between-variation ratio and low-overdispersion scenario (subplot a-(i)), PRL achieves the highest rejection rate at 43.3%, outperforming OLS (39.0%), adjHT (14.6%), and FE (13.1%). PRL consistently demonstrates superior rejection rates which help to identify the non-zero link between $z_{i,t}$ variables and outcomes across all levels of between-variation ratios and overdispersion.

In terms of parameter efficiency, PRL also achieves the lowest MSE and MAE among all estimators. For example, in the low-overdispersion scenario ($\alpha = 0.001$, subplot b-(i)), PRL achieves an MSE of 2.04 to 2.16, remaining low in the high-overdispersion scenario ($\alpha = 2$, subplot b-(iii)) at 2.44 to 2.74. By comparison, FE estimators exhibit substantially higher MSE values, ranging from 8.06 to 32.69 in subplot b-(i) and 14.81 to 53.86 in subplot b-(iii). Unlike FE, PRL maintains stable MSE and MAE values as between-group variation increases. Additionally, the distinction between PRL and OLS becomes more pronounced in MAE, where the MAE of OLS averages 1.28 times larger than that of PRL.

The high rejection rate, low error levels, and stability of PRL are not limited to discrete outcomes but extend to continuous outcomes, as shown in Figure IA.2 of the Internet Appendix. Our simulation exercise thus highlights PRL's robustness and efficiency in parameter inference and estimation across various outcome types, levels of outcome overdispersion, and degrees of between-group variation in the variable of interest.

6. Poisson Estimation

6.1. Poisson regressions without and with firm fixed effects

To account for the discrete nature of patent and citation counts, we also examine Poisson regression models in this section. Similar to our earlier design for OLS regressions, we estimate Poisson regressions in Equation (2) using the Quasi-Maximum Likelihood of (Silva and Tenreyro, 2006).[24] All explanatory variables have been defined earlier and the standard errors are also clustered by firm. Table 6 presents our results when we use R&D/Asset as the proxy for R&D input and the full sample. Panel A (B) presents main results when the dependent variable is patent counts (citation counts). Column (1) is for Poisson regressions including firm and year FEs, and Column (2) is Poisson regressions without firm FEs. (Internet Appendix Table IA.4 provides a full set of estimates.)

Panel A and B of Table 6 indicate that the R&D coefficients are negative when firm FEs are included in Columns (1). Specifically, the estimates for patent counts are –0.248, while for citation counts they are –0.125, respectively. However, once we remove firm FEs in Columns (2), the R&D input regains its significance and positive association with patent counts. The coefficients increase substantially to 2.305, for patent counts, and to 2.094 for citation counts, all statistically significant. Overall, Poisson regression results are consistent with our earlier OLS regression results in the following way: a strong positive R&D-patent relation exists without firm FEs; however, such a relation is seriously reduced or eliminated once we start to include firm FEs in regressions.

We also consider the following robustness checks in Table 7. First, we use alternative R&D measures to examine the robustness of our results in the count models. Panels A.1 and B.1 present results of R&D/ME in explaining patent counts and citation counts, respectively. We find that the same patterns as R&D/Assets in the Table 6 above. In addition, we also use ln(1+R&D) as the proxy for R&D input and drop ln(ME) in Poisson regressions. Panel A.2 and B.2 shows that the coefficients on R&D input are significantly positive in models without firm FEs. When we introduce firm FEs to regressions, the coefficients on ln(1+R&D) drop by more than 90%. For example, in the patent count model, the coefficient is 0.054 in Columns (1) of Panel A.2 when firm FEs are included, but it increases to 0.790 when firm FEs are excluded.

Third, we focus on a sample of firms that have at least one patent granted during the sample period in Internet Appendix Table IA.5. In Column (2), the coefficient on R&D/Asset is 1.663

[24] We use `ppmlhdfe` package of Correia et al. (2020) in `STATA` that is designed to deal with Poisson regressions with high dimensional fixed effects.

(1.480) with statistical significance for patent (citation) output when we do not add any firm FEs. However, Columns (1) show (significantly) negative coefficients on R&D input when we account for firm FEs. Panels A.2 and B.2 present results of R&D/ME in explaining patent counts and citation counts, respectively. We find that the same patterns as R&D/Assets. Finally, when we measure R&D input using ln(1+R&D) and exclude ln(ME) in the sample of patenting firms in Panel A.3 and B.3 the same table, we find that the existence of firm FEs greatly reduces the economic magnitude of the R&D effect.

All results presented in this section suggest that the biases driven by firm FEs are not specific to OLS regressions. In fact, such biases appear to be a prevailing phenomenon that deserves researchers' attention and treatment.

6.2. Post-regularization LASSO and double machine learning in Poisson regressions

To select important firm dummies, we use the post-regularization LASSO Poisson regression (PRL Poisson) (Belloni et al., 2016). Specifically, consider the variation of Equation (2):

$$\mu_{i,t} := \mathbb{E}\left(Innov_{i,t+1} \middle| R\&D_{i,t}, \mathbf{x}_{i,t}\right) = \exp\left(\alpha_0 + \sum_{s\in\mathcal{S}} \alpha_s \text{dummy}_{s,i} + \beta_1 R\&D_{i,t} + \mathbf{x}_{i,t}^\top \boldsymbol{\beta}_2\right), \quad (13)$$

in which $\mathcal{S}$ is a nonempty subset of $\{1, 2, \dots N\}$ and contains a few important firm dummies, and $\text{dummy}_{s,i}$ is the dummy variable such that $\text{dummy}_{s,i} = 1$ if $s = i$, and 0 otherwise. Again, $R\&D_{i,t}$ and $\mathbf{x}_{i,t}$ can be written as:

$$R\&D_{i,t} = \gamma_{1,0} + \sum_{s\in\mathcal{S}} \gamma_{1,s}\text{dummy}_{s,i} + v_{1,i,t}, \text{ and } \mathbf{x}_{i,t} = \boldsymbol{\gamma}_{2,0} + \sum_{s\in\mathcal{S}} \boldsymbol{\gamma}_{2,s}\text{dummy}_{s,i} + \mathbf{v}_{2,i,t}, \quad (14)$$

Equations (13) and (14) lead to the following moment conditions:

$$\mathbb{E}\left[\left(\mathbf{x}_{i,t} - \boldsymbol{\gamma}_{2,0} - \sum_{s\in\mathcal{S}} \boldsymbol{\gamma}_{2,s}\text{dummy}_{s,i}\right)\left(Innov_{i,t+1} - \mu_{i,t}\right)\right] = \mathbf{0},$$

$$\mathbb{E}\left[\left(R\&D_{i,t} - \boldsymbol{\gamma}_{1,0} - \sum_{s\in\mathcal{S}} \boldsymbol{\gamma}_{1,s}\text{dummy}_{s,i}\right)\left(Innov_{i,t+1} - \mu_{i,t}\right)\right] = 0.$$

Again, when $\mathcal{S}$ is known, the parameters can be estimated using the moment conditions above by GMM. When $\mathcal{S}$ is unknown, we use Poisson LASSO and LASSO to select the dummy variables. Specifically, we perform Poisson LASSO of $Innov_{i,t+1}$ on $R\&D_{i,t}$, $\mathbf{x}_{i,t}$, and all dummies to select

the dummies, and fit a Poisson regression of $Innov_{i,t+1}$ on $R\&D_{i,t}$, $\mathbf{x}_{i,t}$, and the selected dummies. We denote the coefficients of the selected dummies as $\tilde{\alpha}_s$. We then perform LASSO of $R\&D_{i,t}$ and $\mathbf{x}_{i,t}$ on the dummies, and fit linear regressions of $R\&D_{i,t}$ and $\mathbf{x}_{i,t}$ on the selected dummies again. Internet Appendix B provides more detailed on the PRL Poisson estimations.[25] In the Internet Appendix C, we show that under regularity conditions and the observation that the variables of interest depend only on a nonempty subset $\mathcal{S}$, the PRL Poisson estimator is asymptotically normally distributed.

6.3. Results form PRL and DML methods in Poisson regressions

Table 6 also displays our results based on R&D/Asset for patent counts (Panel A) and citation counts (Panel B). Columns (3) and (4) are estimates of the PRL Poisson and DML Poisson, respectively. (The full tables of all estimates are provided the Table IA.4 in the Internet Appendix.)

First, we find that the PRL and DML estimates of the R&D coefficients in the Poisson model are significantly positive. The PRL and DML estimates are closer to the Poisson estimates without any firm fixed effects. For example, in Panel A of Table 6, the PRL Poisson estimate is 2.407, which is close to 2.305–the estimate from Poisson regressions without firm fixed effects in Column (2). Additionally, we define the selected ratio as the number of selected dummies divided by the total number of dummies. For the PRL Poisson and DML methods, these selected ratios are again 11% to 21% for models with firm and year fixed effects. Panel B for citation counts shows similar results.

We also use alternative R&D measures to examine the robustness of the R&D effect in the Poisson count models. Table 7 provides the coefficient estimates and the corresponding selection ratio for alternative R&D intensity, R&D/ME, and R&D stock, ln(1+R&D). We implement the same estimation as in Table 6 using a sample of firms with at least one patent in our sample period in Internet Appendix Table IA.5. All these findings suggest that while Poisson regressions with firm fixed effects may underestimate the impact of R&D on innovation, PRL and DML methods in the Poisson framework consistently deliver a positive relation between R&D and innovation outputs by effectively accounting for firm-specific effects without entirely absorbing the variables of interest.

[25]The STATA codes are easy to implement as: `popoisson y rd control, controls(i.firmID) vce(cluster firmID) coef` for PRL Poisson; and `xpoisson y rd control, controls(i.firmID) vce(cluster firmID) coef xfolds(#fold)` for DML Poisson.

6.4. Simulations in Poisson regressions

This section compares the rejection rate, MSE, and MAE of Poisson with firm fixed effects, Poisson without firm fixed effects, and PRL Poisson methods. The data-generating process follows the design described in Section 4.4. Figures 3 and the Internet Appendix Figure IA.3 illustrate the performance of Poisson with firm fixed effects ("FE Poisson", black triangle line), Poisson without firm fixed effecs ("Poisson", blue circle line), and PRL Poisson ("PRL Poisson", red square line) for discrete and continuous outcomes, respectively.

Subplot (a) of Figures 3 and Figure IA.3 displays the rejection rates for discrete and continuous outcomes, respectively. Across the three estimators, PRL Poisson demonstrates the highest rejection rates in medium and high-overdispersion scenarios, and matching the 100% rejection rate achieved by Poisson for both outcome types.

In contrast, FE Poisson consistently delivers the lowest rejection rates, particularly as the between-group variation increases. In low-overdispersion scenarios (subplot a-(i)), its rejection rate declines to 95.6% for discrete outcomes and drops sharply to 58.6% for continuous outcomes. This performance gap becomes even more pronounced in medium-overdispersion scenarios (subplot a-(ii)) and high-overdispersion scenarios (subplot a-(iii)), where the rejection rate of FE Poisson deteriorates further. The disparity is especially evident for continuous outcomes, highlighting the limited ability of FE Poisson to maintain robust rejection rates under higher overdispersion or between-group variation.

The MSE and MAE of FE Poisson increase when the between variation increases, and deteriorate more sharply in the high overdispersion cases. For example, subplot b-(iii) of Figures 3 shows that the MSE of FE Poisson increases from 3.95% to 15.99%, compared to the low overdispersion subplot b-(i) in the range between 0.07% and 0.30%. The same pattern applies to MAE, as subplot c-(iii) shows that the MAE of FE Poisson increases from 15.66% to 31.49%, compared to the low overdispersion subplot c-(i) in the range between 2.14% and 4.35%.

The MSE of Poisson and PRL Poisson depends on the dispersion levels of the outcomes. In low dispersion scenario for discrete outcome case (subplot b-(i) and c-(i) in Figure 3), both approaches maintain low and stable error rates, with MSEs ranging from 0.01% to 0.02% and MAEs between 0.83 and 1.02%. In median overdispersion case (subplot b-(ii) and c-(ii)), both Poisson and PRL Poisson continue to outperform FE Poisson. Among the three methods, PRL Poisson method achieves the smallest MAE.

In extreme-overdispersion scenarios (subplot b-(iii) and c-(iii)), trade-offs emerge between FE Poisson and PRL Poisson. At lower levels of between-group variation, FE Poisson performs bet-

ter, but as between-group variation increases, PRL Poisson becomes the preferred method due to its ability to handle higher levels of variability effectively. Throughout all scenarios, Poisson consistently maintains the lowest error levels.

A similar pattern is observed in Figure IA.3 for the simulations based on continuous outcomes. Poisson consistently achieves the lowest error rates across all scenarios, while PRL Poisson demonstrates strong performance, particularly under high overdispersion or significant between-group variation. In contrast, FE Poisson shows limited robustness, with its error rates increasing sharply as overdispersion or between-group variation rises.

Overall, Poisson regressions with fixed effects are associated with lower rejection rates and higher error rates in comparison with PRL Poisson and Poisson regressions without fixed effects, especially in high-overdispersion and high between-group variation scenarios. These findings align with the results from the simulations for linear estimators in Section 4.4 and Section 5.4, and highlight the advantage of the machine learning methods in handling persistent explanatory variables.

7. Conclusion

In this paper, we examine an intriguing and puzzling phenomenon that R&D input cannot explain patent output in the emerging literature of corporate innovation. Our literature review suggests that such a missing link can be attributed to firm fixed effects in regression models. As these studies aim to discover new explanatory factors for firm-level patent outputs, the absence of positive R&D coefficients requires us to be cautious in interpreting the effects of these new factors.

In the first set of our empirical analysis, we examine the R&D-patent relation using OLS/Poisson regressions both without and with firm fixed effects. We show that it is indeed the existence of firm fixed effects that eliminates the explanatory power of R&D input for patent output. Further analyses suggest that R&D and patent variables' within heterogeneity (i.e., time-series variation) is smaller than their between heterogeneity (i.e., cross-sectional variation), which is consistent with the arguments made in the prior literature. We further propose the adjusted Hausman-Taylor estimator to address firm-specific unobservable effects, which delivers coefficients on R&D input that are much closer to the point estimates of OLS/Poisson models without firm fixed effects than to those that include these effects. All these findings confirm our proposition that firm fixed effects regressions underestimate the role of R&D input (and likely other persistent variables as well).

We further propose two advanced machine learning methods, the post-regularization LASSO (PRL) estimator and its extension based on the double machine learning (DML) method, to solve

the firm-specific unobservable effects and to select informative firm dummies. These advanced methods select 10% to 20% of individual firms dummies, suggesting that many firm dummies do not have distinct explanatory power for patent output. More importantly, estimation based on these two methods supports a substantial R&D effect with point estimates that are much closer to OLS/Poisson without firm fixed effects than to those that include these effects. Thus, adding all firm dummies in models may sometimes bias statistical inferences more than entirely ignoring firm fixed effects.

These empirical results are also supported in our simulations that highlight between- and within-firm variations in panels for our analyses of the efficacy and efficiency of different estimators. As between-firm variation increases relative to within-firm variation, PRL performs the best and fixed effects estimators perform the worst in rejection rates and error rates. These results confirm the advantage of machine learning in handling persistent explanatory variables.

This study suggests that researchers should use caution in including and selecting firm fixed effects in regression models. Researchers are also advised to (i) report estimation results of models both with and without firm fixed effects and (ii) demonstrate both within and between variations for a comprehensive assessment of the sources of explanatory power. Researchers are also recommended to consider using the adjusted HT or the advanced machine learning techniques we propose to make appropriate statistical inferences in empirical economics research.

References

Acharya, V.V., Baghai, R.P., Subramanian, K.V., 2014. Wrongful discharge laws and innovation. The Review of Financial Studies 27, 301–346.

Aghion, P., Van Reenen, J., Zingales, L., 2013. Innovation and institutional ownership. American Economic Review 103, 277–304.

Amore, M.D., Schneider, C., Žaldokas, A., 2013. Credit supply and corporate innovation. Journal of Financial Economics 109, 835–855.

Angrist, J.D., Pischke, J.S., 2009. Mostly harmless econometrics: An empiricist's companion. Princeton University Press.

Arora, A., Belenzon, S., Sheer, L., 2017. Back to basics: Why do firms invest in research? NBER Working Paper .

Atanassov, J., Liu, X., 2020. Can corporate income tax cuts stimulate innovation? Journal of Financial and Quantitative Analysis 55, 1415–1465.

Balsmeier, B., Fleming, L., Manso, G., 2017. Independent boards and innovation. Journal of Financial Economics 123, 536–557.

Belloni, A., Chernozhukov, V., Wei, Y., 2016. Post-selection inference for generalized linear models with many controls. Journal of Business & Economic Statistics 34, 606–619.

Bellstam, G., Bhagat, S., Cookson, J.A., 2021. A text-based analysis of corporate innovation. Management Science 67, 4004–4031.

Bena, J., Ferreira, M.A., Matos, P., Pires, P., 2017. Are foreign investors locusts? The long-term effects of foreign institutional ownership. Journal of Financial Economics 126, 122–146.

Blanco, I., Wehrheim, D., 2017. The bright side of financial derivatives: Options trading and firm innovation. Journal of Financial Economics 125, 99–119.

Bloom, N., Schankerman, M., Van Reenen, J., 2013. Identifying technology spillovers and product market rivalry. Econometrica 81, 1347–1393.

Blundell, R., Griffith, R., Reenen, J.V., 1995. Dynamic count data models of technological innovation. The Economic Journal 105, 333–344.

Blundell, R., Griffith, R., Van Reenen, J., 1999. Market share, market value and innovation in a panel of British manufacturing firms. The Review of Economic Studies 66, 529–554.

Breuer, M., DeHaan, E., 2024. Using and interpreting fixed effects models. Journal of Accounting Research 62, 1183–1226.

Celikyurt, U., Sevilir, M., Shivdasani, A., 2014. Venture capitalists on boards of mature public firms. The Review of Financial Studies 27, 56–101.

Chan, F., Mátyás, L., 2022. Linear econometric models with machine learning, in: Econometrics with Machine Learning. Springer Cham, pp. 1–39.

Chang, X., Chen, Y., Wang, S.Q., Zhang, K., Zhang, W., 2019. Credit default swaps and corporate innovation. Journal of Financial Economics 134, 474–500.

Chang, X., Fu, K., Low, A., Zhang, W., 2015. Non-executive employee stock options and corporate innovation. Journal of Financial Economics 115, 168–188.

Chemmanur, T.J., Kong, L., Krishnan, K., Yu, Q., 2019. Top management human capital, inventor mobility, and corporate innovation. Journal of Financial and Quantitative Analysis 54, 2383–2422.

Chemmanur, T.J., Loutskina, E., Tian, X., 2014. Corporate venture capital, value creation, and innovation. The Review of Financial Studies 27, 2434–2473.

Chemmanur, T.J., Tian, X., 2018. Do antitakeover provisions spur corporate innovation? A regression discontinuity analysis. Journal of Financial and Quantitative Analysis 53, 1163–1194.

Chen, M.A., Wu, Q., Yang, B., 2019. How valuable is FinTech innovation? The Review of Financial Studies 32, 2062–2106.

Chen, Y., Podolski, E.J., Rhee, S.G., Veeraraghavan, M., 2014. Local gambling preferences and corporate innovative success. Journal of Financial and Quantitative Analysis 49, 77–106.

Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C.B., Newey, W., Robins, J., 2018. Double/debiased machine learning for treatment and structural parameters. The Econometrics Journal 21, C1–C68.

Chernozhukov, V., Hansen, C.B., Spindler, M., 2015. Post-selection and post-regularization inference in linear models with many controls and instruments. American Economic Review Papers and Proceedings 105, 486–490.

Chinco, A., Clark-Joseph, A.D., Ye, M., 2019. Sparse signals in the cross-section of returns. The Journal of Finance 74, 449–492.

Chu, Y., Tian, X., Wang, W., 2019. Corporate innovation along the supply chain. Management Science 65, 2445–2466.

Cohen, L., Diether, K., Malloy, C., 2013. Misvaluing innovation. The Review of Financial Studies 26, 635–666.

Cohen, W.M., Levinthal, D.A., 1989. Innovation and learning: the two faces of R&D. The Economic Journal 99, 569–596.

Cohn, J.B., Liu, Z., Wardlaw, M.I., 2022. Count (and count-like) data in finance. Journal of Financial Economics 146, 529–551.

Coles, J.L., Lemmon, M.L., Meschke, J.F., 2012. Structural models and endogeneity in corporate finance: The link between managerial ownership and corporate performance. Journal of Financial Economics 103, 149–168.

Correia, S., 2016. Linear models with high-dimensional fixed effects: An efficient and feasible estimator. Working Paper.

Correia, S., Guimarães, P., Zylkin, T., 2020. Fast Poisson estimation with high-dimensional fixed effects. The Stata Journal 20, 95–115.

Custódio, C., Ferreira, M.A., Matos, P., 2019. Do general managerial skills spur innovation? Management Science 65, 459–476.

Custódio, C., Metzger, D., 2014. Financial expert CEOs: CEOs work experience and firm s financial policies. Journal of Financial Economics 114, 125–154.

Ederer, F., Manso, G., 2011. Incentives for innovation: Bankruptcy, corporate governance, and compensation systems. Handbook of Law, Innovation, and Growth , 90–111.

Erel, I., Stern, L.H., Tan, C., Weisbach, M.S., 2021. Selecting directors using machine learning. The Review of Financial Studies 34, 3226–3264.

Faleye, O., Hoitash, R., Hoitash, U., 2011. The costs of intense board monitoring. Journal of Financial Economics 101, 160–181.

Faleye, O., Kovacs, T., Venkateswaran, A., 2014. Do better-connected CEOs innovate more? Journal of Financial and Quantitative Analysis 49, 1201–1225.

Fang, V.W., Tian, X., Tice, S., 2014. Does stock liquidity enhance or impede firm innovation? The Journal of Finance 69, 2085–2125.

Feng, G., Giglio, S., Xiu, D., 2020. Taming the factor zoo: A test of new factors. The Journal of Finance 75, 1327–1370.

Flammer, C., Kacperczyk, A., 2016. The impact of stakeholder orientation on innovation: Evidence from a natural experiment. Management Science 62, 1982–2001.

Frisch, R., Waugh, F.V., 1933. Partial time regressions as compared with individual trends. Econometrica 1, 387–401.

Fu, R., Kraft, A., Tian, X., Zhang, H., Zuo, L., 2020. Financial reporting frequency and corporate innovation. The Journal of Law and Economics 63, 501–530.

Galasso, A., Simcoe, T.S., 2011. CEO overconfidence and innovation. Management Science 57, 1469–1484.

Gao, H., Zhang, W., 2017. Employment nondiscrimination acts and corporate innovation. Management Science 63, 2982–2999.

Gomes-Casseres, B., Hagedoorn, J., Jaffe, A.B., 2006. Do alliances promote knowledge flows? Journal of Financial Economics 80, 5–33.

Griffin, D., Li, K., Xu, T., 2021. Board gender diversity and corporate innovation: International evidence. Journal of Financial and Quantitative Analysis 56, 123–154.

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

Griliches, Z., 1998. Patent statistics as economic indicators: A survey, in: R&D and productivity: The econometric

evidence. University of Chicago Press, pp. 287–343.

Griliches, Z., Hausman, J.A., 1986. Errors in variables in panel data. Journal of Econometrics 31, 93–118.

Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. The Review of Financial Studies 33, 2223–2273.

Hall, B.H., Griliches, Z., Hausman, J.A., 1986. Patents and R and D: Is there a lag? International Economic Review 27, 265–283.

Hall, B.H., Jaffe, A., Trajtenberg, M., 2005. Market value and patent citations. RAND Journal of Economics 36, 16–38.

Hall, B.H., Thoma, G., Torrisi, S., 2007. The market value of patents and R&D: Evidence from European firms., in: Academy of Management Proceedings, Academy of Management Briarcliff Manor, NY 10510. pp. 1–6.

Hansen, B., 2022. Econometrics. Princeton University Press.

Hausman, J., Hall, B.H., Griliches, Z., 1984. Econometric models for count data with an application to the patents-R&D relationship. Econometrica 52, 909–938.

Hausman, J.A., Taylor, W.E., 1981. Panel data and unobservable individual effects. Econometrica: Journal of the Econometric society , 1377–1398.

He, J., Tian, X., 2018. Finance and corporate innovation: A survey. Asia-Pacific Journal of Financial Studies 47, 165–212.

He, J.J., Tian, X., 2013. The dark side of analyst coverage: The case of innovation. Journal of Financial Economics 109, 856–878.

Hirshleifer, D., Low, A., Teoh, S.H., 2012. Are overconfident CEOs better innovators? The Journal of Finance 67, 1457–1498.

Huang, Q., Yuan, T., 2021. Does political corruption impede firm innovation? Evidence from the United States. Journal of Financial and Quantitative Analysis 56, 213–248.

Imbens, G.W., Wooldridge, J.M., 2009. Recent developments in the econometrics of program evaluation. Journal of economic literature 47, 5–86.

Islam, E., Zein, J., 2020. Inventor CEOs. Journal of Financial Economics 135, 505–527.

Jennings, J., Kim, J.M., Lee, J., Taylor, D., 2024. Measurement error, fixed effects, and false positives in accounting research. Review of Accounting Studies 29, 959–995.

Kamien, M.I., Schwartz, N.L., 1975. Market structure and innovation: A survey. Journal of Economic Literature 13, 1–37.

Kaplan, S.N., Zingales, L., 1997. Do investment-cash flow sensitivities provide useful measures of financing constraints? The Quarterly Journal of Economics 112, 169–215.

Kock, A.B., Tang, H., 2019. Uniform inference in high-dimensional dynamic panel data models with approximately sparse fixed effects. Econometric Theory 35, 295–359.

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

Koh, P.S., Reeb, D.M., 2015. Missing R&D. Journal of Accounting and Economics 60, 73–94.

Kong, L., 2020. Government spending and corporate innovation. Management Science 66, 1584–1604.

Kortum, S., 1993. Equilibrium R&D and the patent–R&R ratio: US evidence. The American Economic Review 83, 450–457.

Lerner, J., Seru, A., 2022. The use and misuse of patent data: Issues for finance and beyond. The Review of Financial Studies 35, 2667–2704.

Lerner, J., Wulf, J., 2007. Innovation and incentives: Evidence from corporate R&D. The Review of Economics and Statistics 89, 634–644.

Lovell, M.C., 1963. Seasonal adjustment of economic time series and multiple regression analysis. Journal of the American Statistical Association 58, 993–1010.

Luong, H., Moshirian, F., Nguyen, L., Tian, X., Zhang, B., 2017. How do foreign institutional investors enhance firm innovation? Journal of Financial and Quantitative Analysis 52, 1449–1490.

Nickell, S., 1981. Biases in dynamic models with fixed effects. Econometrica 49, 1824–1851.

Noel, M., Schankerman, M., 2013. Strategic patenting and software innovation. The Journal of Industrial Economics 61, 481–520.

Ovtchinnikov, A.V., Reza, S.W., Wu, Y., 2020. Political activism and firm innovation. Journal of Financial and Quantitative Analysis 55, 989–1024.

Pakes, A., Griliches, Z., 1984. Patents and R&D at the firm level: A first look, in: R&D, Patents, and Productivity. University of Chicago Press, pp. 55–72.

Reeb, D.M., Zhao, W., 2021. Disregarding the shoulders of giants: Inferences from innovation research. David M. Reeb, and Wanli Zhao .

Roberts, M.R., Whited, T.M., 2013. Endogeneity in empirical corporate finance, in: Handbook of the Economics of Finance. volume 2, pp. 493–572.

Semenova, V., Goldman, M., Chernozhukov, V., Taddy, M., 2023. Inference on heterogeneous treatment effects in high-dimensional dynamic panels under weak dependence. Quantitative Economics 14, 471–510.

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

Silva, J.S., Tenreyro, S., 2006. The log of gravity. The Review of Economics and Statistics 88, 641–658.

Sunder, J., Sunder, S.V., Zhang, J., 2017. Pilot CEOs and corporate innovation. Journal of Financial Economics 123, 209–224.

Tian, X., Wang, T.Y., 2014. Tolerance for failure and corporate innovation. The Review of Financial Studies 27, 211–255.

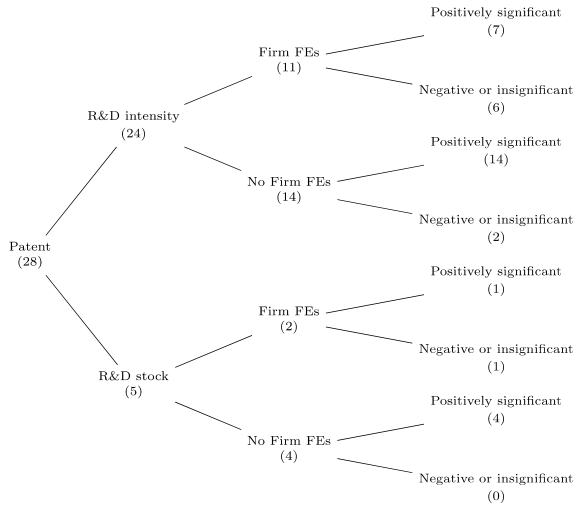
Tibshirani, R., 1996. Regression shrinkage and selection via the LASSO. Journal of the Royal Statistical Society. Series B (Methodological) 58, 267–288.

Trajtenberg, M., 1990. A penny for your quotes: Patent citations and the value of innovations. The Rand Journal of Economics 21, 172–187.

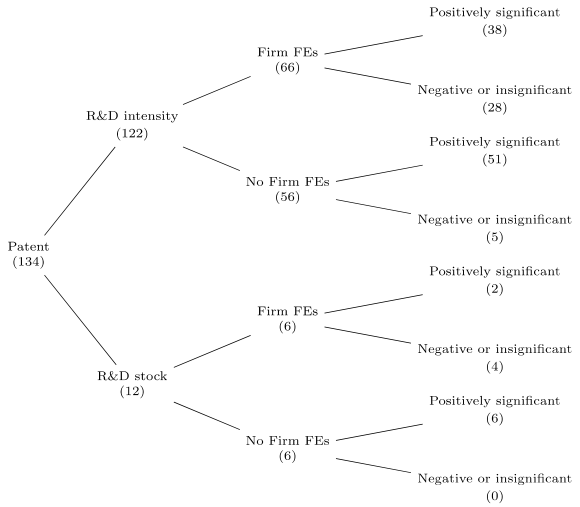
Wooldridge, J.M., 2002. Econometric analysis of cross section and panel data mit press. Cambridge, ma 108, 245–254.

Zhong, R.I., 2018. Transparency and firm innovation. Journal of Accounting and Economics 66, 67–93.

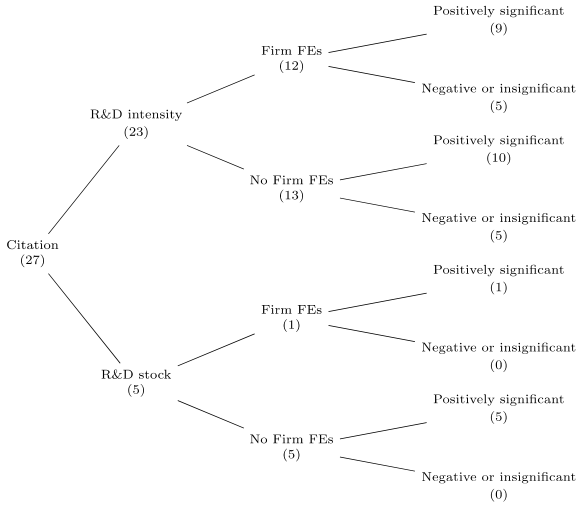
Zhou, X., 2001. Understanding the determinants of managerial ownership and the link between ownership and performance: Comment. Journal of Financial Economics 62, 559–571.



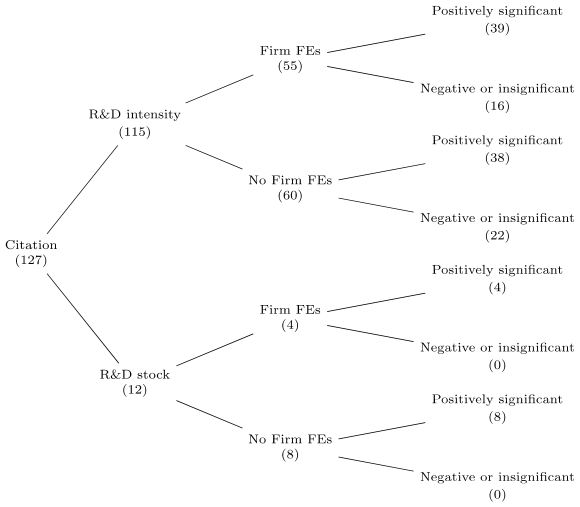
(a) Patent OLS model: article number



(b) Patent OLS model: regression number

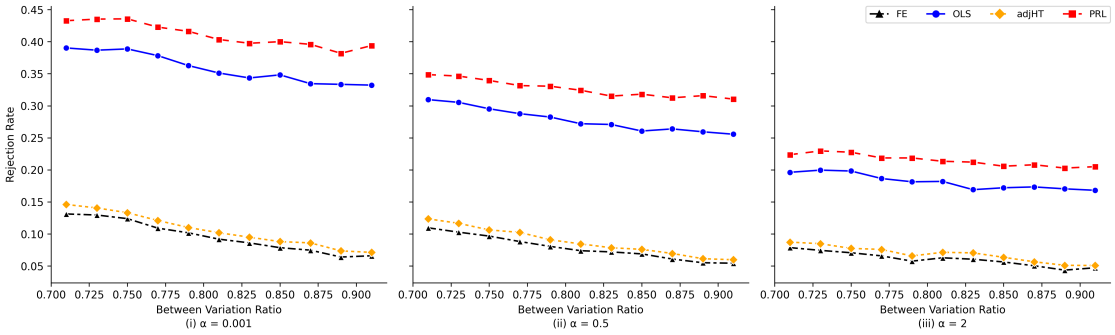


(c) Citation OLS model: article number

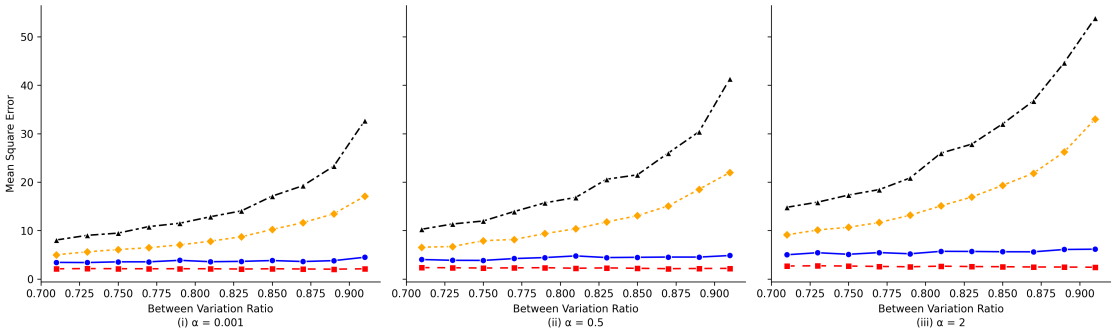


(d) Citation OLS model: regression number

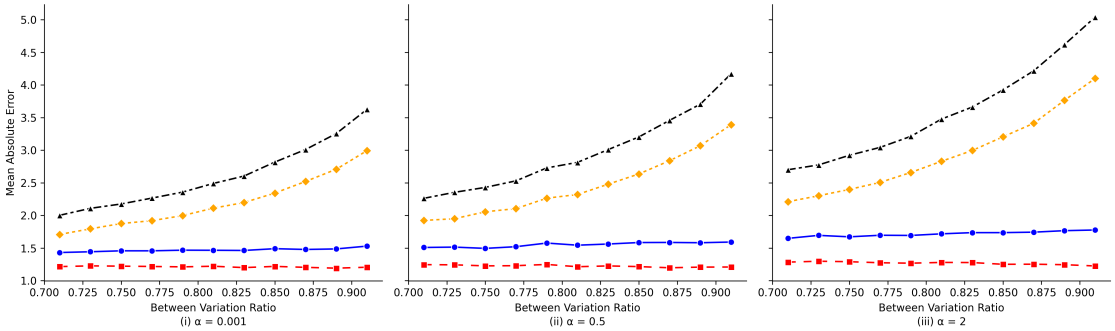
Figure 1: Patent and citation OLS studies



(a) Rejection rate

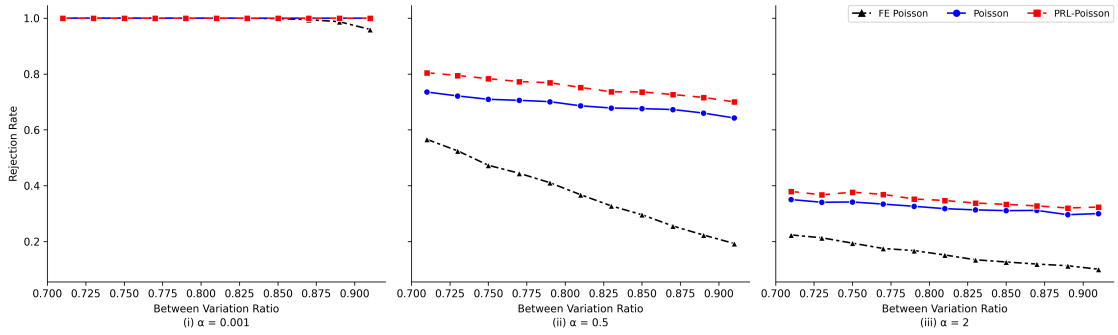


(b) Mean square error

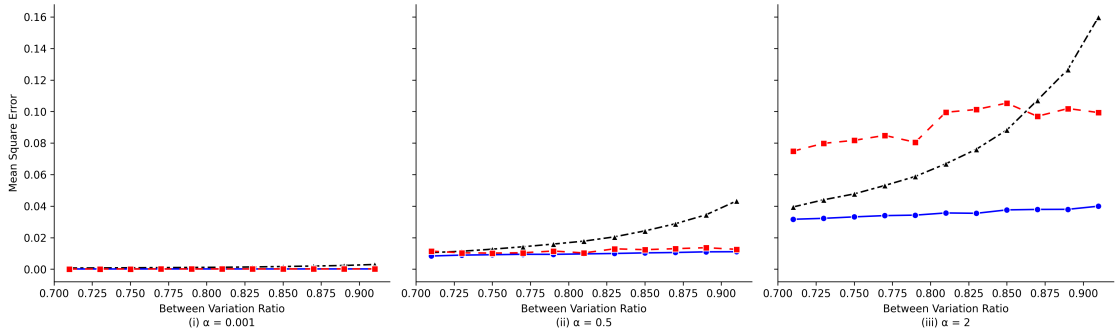


(c) Mean absolute error

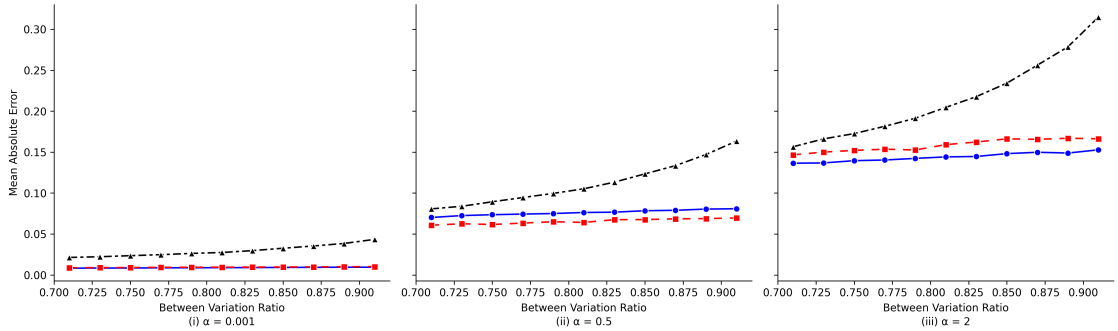
Figure 2: Simulation rejection rates and error rates for discrete outcomes: FE, OLS, adjHT, and PRL. This figure compares the rejection and error rates for four estimators: FE (regression with full set of firm dummies), OLS (regression without any firm dummies), adjHT (adjusted Hausman-Taylor), and PRL (post-regularization LASSO). The discrete outcome $y_{i,t}$ is generated from the negative binominal with conditional mean $\exp(0.2z_{i,t} + x_{i,t})$, and $z_{i,t} = 0.5\eta_i + 0.5\nu_{i,t}$. η_i and $\nu_{i,t}$ are independently drawn from normal distributions truncated at the 1st and 99th percentiles, with standard deviations σ_η and σ_ν. The between variation ratio is defined as $\sigma_\eta^2/(\sigma_\eta^2+\sigma_\nu^2)$. Subplots (i), (ii), and (iii) represent low ($\alpha = 0.001$), moderate ($\alpha = 0.5$), and high overdispersion ($\alpha = 2$) levels of the negative binominal outcomes, respectively.



(a) Rejection rate



(b) Mean square error



(c) Mean absolute error

Figure 3: Simulation rejection rates and error rates for discrete outcomes: FE Poisson, Poisson, and PRL Poisson. This figure compares the rejection and error rates for three estimators: FE Poisson (Poisson regression with full set of firm dummies), Poisson (Poisson regression without any firm dummies), PRL Poisson (post-regularization LASSO Poisson regression). The discrete outcome $y_{i,t}$ is generated from the negative binominal with conditional mean $\exp(0.2z_{i,t} + x_{i,t})$, and $z_{i,t} = 0.5\eta_i + 0.5\nu_{i,t}$. η_i and $\nu_{i,t}$ are independently drawn from normal distributions truncated at the 1st and 99th percentiles, with standard deviations σ_η and σ_ν. The between variation ratio is defined as $\sigma_\eta^2/(\sigma_\eta^2 + \sigma_\nu^2)$. Subplots (i), (ii), and (iii) represent low ($\alpha = 0.001$), moderate ($\alpha = 0.5$), and high overdispersion ($\alpha = 2$) levels of the negative binominal outcomes, respectively.

Table 1: R&D-Innovation literature overview

Article	Journal	Estimation	Innovation	R&D	Firm FEs	No Firm FEs	Pos-Sig	Neg-Insig
Acharya et al. (2014)	RFS	OLS	Patent/Citation	Intensity	V		V	
Aghion et al. (2013)	AER	OLS/Poisson/NB	Patent/Citation	Stock	V*	V	V	V
Amore et al. (2013)	JFE	Poisson	Patent/Citation	Stock	V		V	V
Atanassov and Liu (2020)	JFQA	OLS	Patent/Citation	Intensity	V		V	
Balsmeier et al. (2017)	JFE	OLS	Patent/Citation	Stock	V	V	V	V
Bellstam et al. (2021)	MS	OLS	Citation	Intensity	V	V		V
Bena et al. (2017)	JFE	OLS/Poisson	Patent	Stock	V*		V	
Blanco and Wehrheim (2017)	JFE	OLS/NB	Patent/Citation	Stock	V*	V	V	
Bloom et al. (2013)	ECTA	NB	Citation	Stock	V*	V	V	
Chang et al. (2015)	JFE	OLS	Patent/Citation	Intensity		V	V	
Chang et al. (2019)	JFE	OLS	Patent/Citation	Intensity	V			V
Chemmanur et al. (2014)	RFS	OLS	Patent/Citation	Intensity		V	V	V
Chemmanur and Tian (2018)	JFQA	OLS	Patent/Citation	Intensity	V		V	
Chemmanur et al. (2019)	JFQA	OLS	Patent/Citation	Intensity		V	V	V
Chen et al. (2014)	JFQA	OLS	Patent/Citation	Stock		V	V	
Chen et al. (2019)	RFS	Poisson	Patent	Stock	V		V	V
Chu et al. (2019)	MS	OLS	Patent/Citation	Intensity		V	V	V
Custódio and Metzger (2014)	JFE	NB	Patent/Citation	Intensity	V*		V	
Custódio et al. (2019)	MS	OLS	Patent/Citation	Intensity/Stock		V	V	
Faleye et al. (2014)	JFQA	OLS	Patent/Citation	Intensity		V	V	
Fang et al. (2014)	JF	OLS	Patent/Citation	Intensity	V		V	V
Flammer and Kacperczyk (2016)	MS	OLS	Patent/Citation	Intensity	V		V	
Fu et al. (2020)	JLE	OLS	Patent/Citation	Intensity	V		V	
Galasso and Simcoe (2011)	MS	Poisson	Patent/Citation	Stock	V*	V	V	V
Gao and Zhang (2017)	MS	OLS	Patent/Citation	Intensity	V		V	V
He and Tian (2013)	JFE	OLS	Patent/Citation	Intensity	V		V	V
Hirshleifer et al. (2012)	JF	OLS	Patent/Citation	Intensity		V	V	
Huang and Yuan (2021)	JFQA	OLS	Patent/Citation	Intensity		V	V	
Islam and Zein (2020)	JFE	OLS	Patent/Citation	Intensity		V	V	
Kong (2020)	MS	OLS	Patent/Citation	Intensity	V		V	V
Lerner and Wulf (2007)	REStat	OLS	Patent	Intensity		V	V	
Luong et al. (2017)	JFQA	OLS	Patent/Citation	Intensity	V	V	V	V
Ovtchinnikov et al. (2020)	JFQA	OLS	Patent/Citation	Intensity		V	V	
Sunder et al. (2017)	JFE	Poisson	Patent/Citation	Intensity		V	V	V
Tian and Wang (2014)	RFS	OLS	Patent/Citation	Intensity		V	V	V
Zhong (2018)	JAE	OLS	Patent	Intensity		V	V	

This table summarizes the R&D and Patent/Citation research. Each column indicates at least one estimation result in tables (including tables in the main text and appendix) has the marked properties. Therefore, some columns are not mutually exclusive. The journal includes Econometrica (ECTA), Journal of Accounting and Economics (JAE), Journal of Financial Economics (JFE), Journal of Financial and Quantitative Analysis (JFQA), Journal of Law and Economics (JLE), Management Science (MS), Review of Financial Studies (RFS), The American Economic Review (AER), The Journal of Finance (JF), and The Review of Economics and Statistics (REStat). Estimation approaches include: OLS (pooled least square and panel regressions), Poisson (pooled Poisson and inflated Poisson model), and NB stand for negative binomial models. The dependent variable includes patent-type and citation-type innovation measures or both of them. The Patent includes: patent count, patent with truncated-bias adjusted, patents weighted by forward citations, logarithm values of patents, and others; and Citation includes: forward citation counts, truncation-bias adjusted citation, technology-time adjusted citation, logarithm values of citation, and others. Two forms of R&D measures are studied. R&D Intensity includes: R&D scaled by the total asset or sales, the logarithm of the scaled R&D values; and R&D stock have R&D level or logarithm of the R&D values with different weighting schemes on annual R&D expenses. Firm fixed effect include the within transformation and firm dummies and the asterisk marks the research uses the Blundell et al. (1999) presample mean scaling estimator. Pos-Sig indicates that at least one column of the related R&D-innovation estimates is positive and significant at a 5% (or less) significance level. Neg-Insig indicates that the coefficient is equal or less than zero or positively insignificant at the 5% level.

Table 2: Summary statistics and correlation

Panel A: Summary statistics

	Mean	Stddev	Min	Q1	Median	Q3	Max
ln(1+Patent)	0.53	1.11	0.00	0.00	0.00	0.69	8.39
ln(1+Citation)	1.25	2.27	0.00	0.00	0.00	2.20	12.02
ln(1+Adjcitation)	0.54	1.16	0.00	0.00	0.00	0.34	8.62
R&D/Asset	0.13	0.26	0.00	0.00	0.00	0.14	1.64
R&D/ME	0.15	0.50	0.00	0.00	0.00	0.14	28.05
ln(1+R&D)	4.61	4.86	0.00	0.00	2.56	9.18	14.02
ln(ME)	11.13	2.07	6.99	9.60	10.95	12.51	16.41
R&D missing	0.43	0.50	0.00	0.00	0.00	1.00	1.00
ln(1+Age)	2.48	0.75	0.69	1.95	2.48	3.09	3.85
ln(K/L)	10.00	1.29	6.89	9.18	9.85	10.64	14.10
Tobin Q	1.76	1.66	0.53	0.94	1.22	1.85	12.64
ROA	0.10	0.18	-0.88	0.06	0.13	0.19	0.40
Leverage	0.23	0.18	0.00	0.07	0.21	0.35	0.74
Cash/Asset	0.14	0.17	0.00	0.02	0.07	0.18	0.88
KZ index	-3.42	10.77	-79.54	-3.48	-0.54	1.03	6.79
InstOwn	0.23	0.24	0.00	0.01	0.15	0.40	0.84
(1-HHI)	0.75	0.20	0.00	0.70	0.80	0.89	0.99
(1-HHI) square	0.61	0.23	0.00	0.49	0.65	0.79	0.98

Panel B: Correlations

	ln(1+Patent)	ln(1+Citation)	ln(1+Adjcitation)
ln(1+Patent)	1.00		
ln(1+Citation)	0.93***	1.00	
ln(1+Adjcitation)	0.97***	0.95***	1.00
R&D/Asset	0.17***	0.22***	0.17***
R&D/ME	0.08***	0.09***	0.08***
ln(1+R&D)	0.55***	0.57***	0.53***
ln(ME)	0.48***	0.44***	0.47***
R&D missing	-0.31***	-0.34***	-0.30***
ln(1+Age)	0.30***	0.25***	0.27***
ln(K/L)	0.11***	0.09***	0.10***
Tobin Q	0.07***	0.11***	0.09***
ROA	0.09***	0.07***	0.09***
Leverage	-0.08***	-0.11***	-0.09***
Cash/Asset	0.01***	0.06***	0.02***
KZ index	-0.01***	-0.03***	-0.01***
InstOwn	0.26***	0.26***	0.26***
(1-HHI)	0.01***	0.02***	0.01***
(1-HHI) square	0.00	0.01***	0.00

Panel A reports the pooled mean, median, standard deviation(Stddev), minimum (Min), 25th percentile (Q1), 75th percentile (Q3), and maximum (Max) for the entire sample. All variables except innovation measures (i.e., patent, citation, and adjusted citation) are winsorized at the 1st and 99th percentiles. ln(1+Patent) denotes the natural logarithmic value of 1 plus the patent count of firms filling in year $t+1$, and ln(1+Citation) (ln(1+Adjcitation)) denotes the natural logarithmic value of 1 plus (adjusted) the citation count of firm in year $t+1$. R&D/Asset is the last five years of accumulated R&D expenses divided by total assets in the fiscal year ending in year t. ln(ME) is the natural logarithmic value of market equity at the end of year t. R&D missing is a dummy variable that equals one for firms with no R&D expenses. ln(1+Age) is the natural logarithmic value of one plus the number of years since a firm's first appearance in Compustat. ln(K/L) is the natural logarithmic value of net property, plant, and equipment divided by the number of employees in Compustat. Tobin Q is the market value of equity plus the book value of assets minus the book value of equity minus balance sheet deferred taxes, divided by the book value of assets. ROA is the operating income before depreciation divided by the book value of total assets. Leverage is the long-term debt plus current debt, divided by total assets. Cash/Asset is the ratio of cash to total assets. KZ index is calculated as -1.002×(cash flow)+ 0.28×(Tobin Q)+ 3.139×(leverage)- 39.368×(Dividends)- 1.315×(Cash holdings). This calculation follows Kaplan and Zingales (1997). InstOwn is the percentage of institutional ownership defined as total institutional ownership from 13F filings divided by the total number of shares outstanding. (1-HHI) and (1-HHI) square are one minus the Herfindahl index and square of one minus the Herfindahl index. Herfindahl index is the sum of squared sales-based market shares of all firms in the same industry where industry is defined using the 3-digit SIC code. Panel B reports the pooled correlations between innovation measures and firm characteristics, and *** indicates statistical significance at the 1% level.

Table 3: Within and between variations

		Stddev	Between/Within	Within/Overall
Panel A: Full sample				
ln(1+Patent)	overall	1.105		
	between	0.764	1.775	
	within	0.430		0.389
ln(1+Citation)	overall	2.267		
	between	1.650	1.517	
	within	1.088		0.480
ln(1+Adjcitation)	overall	1.164		
	between	0.814	1.676	
	within	0.485		0.417
R&D/Asset	overall	0.263		
	between	0.265	2.259	
	within	0.117		0.446
R&D/ME	overall	0.292		
	between	0.243	1.342	
	within	0.181		0.621
ln(1+R&D)	overall	4.858		
	between	4.507	3.391	
	within	1.329		0.274
Panel B: Patenting firms				
ln(1+Patent)	overall	1.356		
	between	1.049	1.778	
	within	0.590		0.435
ln(1+Citation)	overall	2.656		
	between	2.050	1.374	
	within	1.492		0.562
ln(1+Adjcitation)	overall	1.440		
	between	1.131	1.698	
	within	0.666		0.462
R&D/Asset	overall	0.332		
	between	0.358	2.179	
	within	0.164		0.495
R&D/ME	overall	0.344		
	between	0.277	1.185	
	within	0.234		0.679
ln(1+R&D)	overall	4.637		
	between	4.094	2.618	
	within	1.563		0.337

This table reports the overall, between, and within variations of the innovation measures and R&D intensity and stock. Panel A consists of a sample of 86,341 public firm-year observations (11,544 unique firms). Panel B consists of a sample of patenting firms. Overall Stddev is the standard deviation of the overall pooled observations. Between Stddev first calculates the firm-level averages for each variable of interest, and then calculates the standard deviation for these averages. Within Stddev measures the within-firm variation of each variable. It calculates the standard deviations of the variable of interest with a firm cluster, and then averages these standard deviations.

Table 4: Innovation to R&D intensity regression

	(1) FE (Firm, Year)	(2) OLS (Year)	(3) adjHT (Year)	(4) PRL (Firm, Year)	(5) DML (Firm, Year)
Panel A: Patent regression					
R&D/Asset	0.041	0.593***	0.220***	0.199***	0.213***
	(0.028)	(0.042)	(0.027)	(0.018)	(0.014)
Number of FEs				11,570	11,570
Number of Selected FEs				1,241	1,737
Selected ratio				10.73%	15.01%
Control variables	Y	Y	Y	Y	Y
Firm FEs	Y	N	N	Y	Y
Year FEs	Y	Y	Y	Y	Y
Observations	84,620	86,341	86,341	86,341	86,341
Panel B: Citation regression					
R&D/Asset	−0.051	1.396***	0.551***	1.397***	1.364***
	(0.068)	(0.084)	(0.065)	(0.083)	(0.050)
Number of FEs				11,570	11,570
Number of Selected FEs				525	947
Selected ratio				4.54%	8.18%
Control variables	Y	Y	Y	Y	Y
Firm FEs	Y	N	N	Y	Y
Year FEs	Y	Y	Y	Y	Y
Observations	84,620	86,341	86,341	86,341	86,341
Panel C: Adjusted citation regression					
R&D/Asset	0.033	0.590***	0.231***	0.210***	0.198***
	(0.031)	(0.045)	(0.030)	(0.019)	(0.015)
Number of FEs				11,570	11,570
Number of Selected FEs				1,194	1,882
Selected ratio				10.32%	16.27%
Control variables	Y	Y	Y	Y	Y
Firm FEs	Y	N	N	Y	Y
Year FEs	Y	Y	Y	Y	Y
Observations	84,620	86,341	86,341	86,341	86,341

This table presents the regression results of innovation on R&D intensity using different estimation methods. Column (1) reports fixed effects (FE) estimates with firm and year fixed effects. Column (2) presents OLS estimates with year fixed effects. Column (3) reports adjusted Hausman-Taylor (adjHT) estimates with year fixed effects. Columns (4) and (5) present Post-Regularization LASSO (PRL) and Double Machine Learning (DML) estimates with firm and year fixed effects, respectively. The standard errors are clustered at the firm level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The selected ratio is the number of selected year and firm dummies divided by the total number of year and firm dummies.

Table 5: Innovation to alternative R&D measures regression

	(1) FE (Firm, Year)	(2) OLS (Year)	(3) adjHT (Year)	(4) PRL (Firm, Year)	(5) DML (Firm, Year)
Panel A: Patent regression					
A.1 R&D scaled by market equity					
R&D/ME	0.026	0.502***	0.139***	0.112***	0.123***
	(0.018)	(0.035)	(0.018)	(0.015)	(0.012)
Selected ratio				10.73%	15.01%
A.2 R&D stock					
ln(1+R&D)	0.023***	0.140***	0.071***	0.032***	0.035***
	(0.003)	(0.004)	(0.003)	(0.001)	(0.001)
Selected ratio				10.73%	15.00%
Panel B: Citation regression					
B.1 R&D scaled by market equity					
R&D/ME	−0.024	1.040***	0.286***	1.015***	1.001***
	(0.040)	(0.064)	(0.040)	(0.064)	(0.038)
Selected ratio				4.54%	8.18%
B.2 R&D Stock					
ln(1+R&D)	0.036***	0.286***	0.176***	0.285***	0.285***
	(0.007)	(0.006)	(0.006)	(0.006)	(0.003)
Selected ratio				4.54%	8.18%
Panel C: Adjusted Citation Regression					
C.1 R&D scaled by market equity					
R&D/ME	0.021	0.506***	0.135***	0.119***	0.117***
	(0.019)	(0.037)	(0.019)	(0.016)	(0.012)
Selected ratio				10.32%	16.27%
C.2 R&D stock					
ln(1+R&D)	0.022***	0.143***	0.071***	0.035***	0.035***
	(0.003)	(0.004)	(0.003)	(0.001)	(0.001)
Selected ratio				10.32%	16.26%
Control variables	Y	Y	Y	Y	Y
Firm FEs	Y	N	N	Y	Y
Year FEs	Y	Y	Y	Y	Y
Observations	84,620	86,341	86,341	86,341	86,341

This table presents the regression results of innovation on alternative R&D measures using different estimation methods. Column (1) reports fixed effects (FE) estimates with firm and year fixed effects. Column (2) presents OLS estimates with year fixed effects. Column (3) reports adjusted Hausman-Taylor (adjHT) estimates with year fixed effects. Columns (4) and (5) present Post-Regularization LASSO (PRL) and Double Machine Learning (DML) estimates with firm and year fixed effects, respectively. The standard errors are clustered at the firm level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The selected ratio is the number of selected year and firm dummies divided by the total number of year and firm dummies.

Table 6: Innovation to R&D intensity Poisson regression

	(1) FE Poisson (Firm, Year)	(2) Poisson (Year)	(3) PRL Poisson (Firm, Year)	(4) DML Poisson (Firm, Year)
Panel A: Patent count model				
R&D/Asset	−0.248	2.305***	2.407***	2.312***
	(0.255)	(0.187)	(0.161)	(0.120)
Number of FEs			11,570	11,570
Number of Selected FEs			1,218	2,484
Selected ratio			10.53%	21.47%
Control variables	Y	Y	Y	Y
Firm FEs	Y	N	Y	Y
Year FEs	Y	Y	Y	Y
Observations	84,620	86,341	86,341	86,341
Panel B: Citation count model				
R&D/Asset	−0.125	2.094***	2.299***	2.262***
	(0.252)	(0.219)	(0.256)	(0.113)
Number of FEs			11,570	11,570
Number of Selected FEs			1,216	2,426
Selected ratio			10.51%	20.97%
Control variables	Y	Y	Y	Y
Firm FEs	Y	N	Y	Y
Year FEs	Y	Y	Y	Y
Observations	84,620	86,341	86,341	86,341

This table presents the Poisson regression results of innovation on R&D intensity using different estimation methods. Column (1) reports fixed effects (FE) Poisson estimates with firm and year fixed effects. Column (2) presents Poisson estimates with year fixed effects. Columns (3) and (4) present Post-Regularization LASSO (PRL) Poisson and Double Machine Learning (DML) Poisson estimates with firm and year fixed effects, respectively. The standard errors are clustered at the firm level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The selected ratio is the number of selected year and firm dummies divided by the total number of year and firm dummies.

Table 7: Innovation to alternative R&D measure Poisson regression

	(1) FE Poisson (Firm, Year)	(2) Poisson (Year)	(3) PRL Poisson (Firm, Year)	(4) DML Poisson (Firm, Year)
Panel A: Patent count model				
A.1 R&D scaled by market equity				
R&D/ME	0.177	1.547***	1.742***	1.723***
	(0.148)	(0.179)	(0.105)	(0.081)
Selected ratio			9.89%	20.21%
A.2 R&D stock				
ln(1+R&D)	0.054*	0.790***	0.775***	0.779***
	(0.029)	(0.080)	(0.085)	(0.040)
Selected ratio			9.27%	20.48%
Control variables	Y	Y	Y	Y
Firm FEs	Y	N	Y	Y
Year FEs	Y	Y	Y	Y
Observations	84,620	86,341	86,341	86,341
Panel B: Citation count model				
B.1 R&D scaled by market equity				
R&D/ME	0.209	1.622***	1.872***	1.886***
	(0.161)	(0.173)	(0.151)	(0.103)
Selected ratio			10.35%	21.58%
B.2 R&D stock				
ln(1+R&D)	0.090**	0.784***	0.711	0.779***
	(0.043)	(0.115)	(1.225)	(0.056)
Selected ratio			9.91%	21.14%
Control variables	Y	Y	Y	Y
Firm FEs	Y	N	Y	Y
Year FEs	Y	Y	Y	Y
Observations	84,620	86,341	86,341	86,341

This table presents the Poisson regression results of innovation on alternative R&D measures using different estimation methods. Column (1) reports fixed effects (FE) Poisson estimates with firm and year fixed effects. Column (2) presents Poisson estimates with year fixed effects. Columns (3) and (4) present Post-Regularization LASSO (PRL) Poisson and Double Machine Learning (DML) Poisson estimates with firm and year fixed effects, respectively. The standard errors are clustered at the firm level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The selected ratio is the number of selected year and firm dummies divided by the total number of year and firm dummies.

Internet Appendix For "Limitation of Firm Fixed Effects Models and the Missing R&D-Patent Relation: New Methods and Evidence"

Hui-Ching Chuang, Po-Hsuan Hsu, Chung-Ming Kuan, Jui-Chung Yang

Internet Appendix A: Asymptotic Normality of PRL and DML

In this Appendix, we show the asymptotic normality of the post-regularization LASSO estimator (PRL, Chernozhukov et al., 2015) the double machine learning estimator (DML,Chernozhukov et al., 2018) used in our paper.

Consider the linear specification as in Models (10) and (11) of our paper. To simplify notation, we use $y_{i,t}$ to denote $Innov_{i,t+1}$, $\mathbf{X}_{i,t}$ to denote the $(k+1)\times 1$ vector $\left[R\&D_{i,t}, \mathbf{x}_{i,t}^{\top}\right]^{\top}$, $\mathbf{D}_{i,t}$ to denote $\left[\sqrt{N}\text{dummy}_{1,i}, \sqrt{N}\text{dummy}_{2,i}, \ldots, \sqrt{N}\text{dummy}_{N,i}\right]^{\top}$, and $\mathbf{v}_{i,t}$ to denote $\left[v_{1,i,t}, \mathbf{v}_{2,i,t}^{\top}\right]^{\top}$. We also use $\boldsymbol{\beta}$ to denote $\left[\beta_1, \boldsymbol{\beta}_2^{\top}\right]^{\top}$, $\boldsymbol{\alpha}$ to denote $[\alpha_1, \alpha_2, \ldots, \alpha_N]$, $\boldsymbol{\gamma}_0$ to denote $\left[\gamma_{1,0}, \boldsymbol{\gamma}_{2,0}^{\top}\right]^{\top}$, and $\boldsymbol{\Gamma}$ to denote the $N \times (k+1)$ matrix of $\left\{\gamma_{1,s}\right\}$ and $\left\{\boldsymbol{\gamma}_{2,s}\right\}$. Model (10) can be written as:

$$y_{i,t} = \beta_0 + \mathbf{X}_{i,t}^{\top}\boldsymbol{\beta} + \mathbf{D}_{i,t}^{\top}\boldsymbol{\alpha} + u_{i,t}, \quad \text{where} \quad \mathbb{E}\left(u_{i,t} \left| \mathbf{X}_{i,t}, \mathbf{D}_{i,t}\right.\right) = 0. \tag{A.1}$$

And Model (11) can be written as:

$$\mathbf{X}_{i,t} = \boldsymbol{\gamma}_0 + \boldsymbol{\Gamma}\mathbf{D}_{i,t} + \mathbf{v}_{i,t}, \quad \text{where} \quad \mathbb{E}\left(\mathbf{v}_{i,t} \left| \mathbf{D}_{i,t}\right.\right) = \mathbf{0}. \tag{A.2}$$

In models (A.1) and (A.2), we consider not only the important firm dummies in $\mathcal{S}$, which is the set of important firm dummies, but also consider all the firm dummies, since the set of important firm dummies are unknown to econometricians. Also, the firm dummies are standardized by pre-multiplying $\sqrt{N}$.

Let $\boldsymbol{\Gamma}^{\top} = \left[\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \ldots, \boldsymbol{\gamma}_{k+1}\right]$, i.e., let $\boldsymbol{\gamma}_r^{\top}$ denote the rth row of $\boldsymbol{\Gamma}$. Let $s_{NT} = \left\|\boldsymbol{\alpha} + \sum_{r=1}^{k+1}\boldsymbol{\gamma}_r\right\|_0$ denote the number of important firm dummies (the number of dummies with non-zero coefficients).

Condition A.1 As $N, T \to \infty$, $s_{NT}^2 \left(\ln N\right)^2 \Big/ (NT) = o(1)$.

Condition A.1 is the same as the corresponding assumption in Belloni et al. (2016) when the number of nuisance variables equals to N, and the sample size equals to NT. While allowing for the number of coefficients to grow as $N \to \infty$, Condition A.1 requires that the coefficients $\boldsymbol{\alpha}$ and $\boldsymbol{\Gamma}$ are sparse in the sense that the number of non-zero coefficients grows more slowly than $\sqrt{NT}$.

Substituting equation (A.2) into the rewritten equation (A.1) gives a system for $y_{i,t}$ and $\mathbf{X}_{i,t}$ that depends only on $\mathbf{D}_{i,t}$:

$$\begin{aligned} y_{i,t} &= \theta_0 + \mathbf{D}_{i,t}^\top \boldsymbol{\theta} + e_{i,t}, \text{ and} \\ \mathbf{X}_{i,t} &= \boldsymbol{\gamma}_0 + \boldsymbol{\Gamma}\mathbf{D}_{i,t} + \mathbf{v}_{i,t}, \end{aligned}$$

in which $\theta_0 = \beta_0 + \boldsymbol{\gamma}_0^\top \boldsymbol{\beta}$, $\quad \boldsymbol{\theta} = \boldsymbol{\alpha} + \boldsymbol{\Gamma}^\top \boldsymbol{\beta}$, and $e_{i,t} = u_{i,t} + \mathbf{v}_{i,t}^\top \boldsymbol{\beta}$.

Let $\widetilde{\boldsymbol{\beta}}$ denote the infeasible estimator:

$$\widetilde{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{v}_{i,t} \mathbf{v}_{i,t}^\top \right)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{v}_{i,t} e_{i,t}.$$

$\widetilde{\boldsymbol{\beta}}$ satisfies that:

$$\sqrt{NT}\left(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) = \left(\frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{v}_{i,t} \mathbf{v}_{i,t}^\top \right)^{-1} \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{v}_{i,t} u_{i,t},$$

which is asymptotic Gaussian distributed under regularity conditions.

$\widetilde{\boldsymbol{\beta}}$ is infeasible because θ_0, $\boldsymbol{\theta}$, $\boldsymbol{\gamma}_0$, and $\boldsymbol{\Gamma}$ are unknown to econometricians. We let $\widehat{e}_{i,t} = y_{i,t} - \widehat{\theta}_0 - \mathbf{D}_{i,t}^\top \widehat{\boldsymbol{\theta}}$, and $\widehat{\mathbf{v}}_{i,t} = \mathbf{X}_{i,t} - \widehat{\boldsymbol{\gamma}}_0 - \widehat{\boldsymbol{\Gamma}}\mathbf{D}_{i,t}$, in which $\widehat{\theta}_0$, $\widehat{\boldsymbol{\theta}}$, $\widehat{\boldsymbol{\gamma}}_0$, and $\widehat{\boldsymbol{\Gamma}}$ are post-LASSO estimates for θ_0, $\boldsymbol{\theta}$, $\boldsymbol{\gamma}_0$, and $\boldsymbol{\Gamma}$. We then let:

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{N} \sum_{t=1}^{T} \widehat{\mathbf{v}}_{i,t} \widehat{\mathbf{v}}_{i,t}^\top \right)^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} \widehat{\mathbf{v}}_{i,t} \widehat{e}_{i,t}.$$

Condition A.2 With a probability approaching one as $N, T \to \infty$, the post-LASSO estimates $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\gamma}}_r$ with $r = 1, 2, \ldots, k+1$ satisfy the following standard performance bounds:

$$\begin{aligned} \left\| \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right\|_1 &= O_p\left(\sqrt{\frac{s_{NT}^2}{NT} \ln N} \right), \quad \|\widehat{\boldsymbol{\gamma}}_r - \boldsymbol{\gamma}_r\|_1 = O_p\left(\sqrt{\frac{s_{NT}^2}{NT} \ln N} \right), \text{ and} \\ \left\| \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right\|_2 &= O_p\left(\sqrt{\frac{s_{NT}}{NT} \ln N} \right), \quad \|\widehat{\boldsymbol{\gamma}}_r - \boldsymbol{\gamma}_r\|_2 = O_p\left(\sqrt{\frac{s_{NT}}{NT} \ln N} \right). \end{aligned}$$

Condition A.2 is the same as the corresponding assumption in Chernozhukov et al. (2015) with the number of variable N and the sample size NT. We note that Condition A.2 only regulates $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\Gamma}}$, the estimated coefficients of the firm dummies. For the intercepts, the estimates have the

standard $\sqrt{NT}$ rate, i.e., $\widehat{\theta}_0 - \theta_0 = O_p\left(1 \Big/ \sqrt{NT}\right)$, and $\widehat{\boldsymbol{\gamma}}_0 - \boldsymbol{\gamma}_0 = O_p\left(1 \Big/ \sqrt{NT}\right)$. Condition A.2 is standard in sparsity-based estimators. One key requirement for Condition A.2 is the restricted eigenvalue condition. After standardization: $(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{D}_{i,t}\mathbf{D}_{i,t}^{\top} = \mathbf{I}_N$, and the eigenvalues are 1, such that the restricted eigenvalue condition is satisfied.

Proposition A.1 Suppose equations (A.1) and (A.2), and Conditions A.1 and A.2 hold. Then

$$\sqrt{NT}\left(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) = \sqrt{NT}\left(\widetilde{\boldsymbol{\beta}} - \boldsymbol{\beta}\right) + o_p(1).$$

Proof. Since

$$\begin{aligned}
&\widehat{\mathbf{v}}_{i,t} = \mathbf{X}_{i,t} - \widehat{\boldsymbol{\gamma}}_0 - \widehat{\boldsymbol{\Gamma}}\mathbf{D}_{i,t} = \mathbf{v}_{i,t} - (\widehat{\boldsymbol{\gamma}}_0 - \boldsymbol{\gamma}_0) - \left(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\right)\mathbf{D}_{i,t}, \text{ and} \\
&\widehat{e}_{i,t} = y_{i,t} - \widehat{\theta}_0 - \mathbf{D}_{i,t}^{\top}\widehat{\boldsymbol{\theta}} = e_{i,t} - \left(\widehat{\theta}_0 - \theta_0\right) - \mathbf{D}_{i,t}^{\top}\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right),
\end{aligned}$$

we have

$$\begin{aligned}
&\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{\mathbf{v}}_{i,t}\widehat{\mathbf{v}}_{i,t}^{\top} = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{v}_{i,t}\mathbf{v}_{i,t}^{\top} - \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{v}_{i,t}\right)(\widehat{\boldsymbol{\gamma}}_0 - \boldsymbol{\gamma}_0)^{\top} - \frac{\widehat{\boldsymbol{\gamma}}_0 - \boldsymbol{\gamma}_0}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{v}_{i,t}^{\top} \\
&- \left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{v}_{i,t}\mathbf{D}_{i,t}^{\top}\right)\left(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\right)^{\top} - \left(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\right)\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{D}_{i,t}\mathbf{v}_{i,t}^{\top} \\
&+ (\widehat{\boldsymbol{\gamma}}_0 - \boldsymbol{\gamma}_0)(\widehat{\boldsymbol{\gamma}}_0 - \boldsymbol{\gamma}_0)^{\top} - (\widehat{\boldsymbol{\gamma}}_0 - \boldsymbol{\gamma}_0)\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{D}_{i,t}^{\top}\right)\left(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\right)^{\top} \\
&- \left(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\right)\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{D}_{i,t}\right)(\widehat{\boldsymbol{\gamma}}_0 - \boldsymbol{\gamma}_0)^{\top} + \left(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\right)\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{D}_{i,t}\mathbf{D}_{i,t}^{\top}\right)\left(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma}\right)^{\top},
\end{aligned}$$

and

$$\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{\mathbf{v}}_{i,t}\widehat{e}_{i,t}=\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{v}_{i,t}e_{i,t}-\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{v}_{i,t}\right)\left(\widehat{\theta}_0-\theta_0\right)-\frac{\widehat{\boldsymbol{\gamma}}_0-\boldsymbol{\gamma}_0}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}e_{i,t}$$
$$-\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{v}_{i,t}\mathbf{D}_{i,t}^{\top}\right)\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)-\left(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}\right)\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{D}_{i,t}e_{i,t}\right)$$
$$+\sqrt{NT}\left(\widehat{\boldsymbol{\gamma}}_0-\boldsymbol{\gamma}_0\right)\left(\widehat{\theta}_0-\theta_0\right)-\left(\widehat{\boldsymbol{\gamma}}_0-\boldsymbol{\gamma}_0\right)\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{D}_{i,t}^{\top}\right)\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)$$
$$-\left(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}\right)\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{D}_{i,t}\right)\left(\widehat{\theta}_0-\theta_0\right)+\left(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}\right)\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{D}_{i,t}\mathbf{D}_{i,t}^{\top}\right)\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right).$$

Then for $(NT)^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{\mathbf{v}}_{i,t}\widehat{\mathbf{v}}_{i,t}^{\top}$,

$$\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{v}_{i,t}\mathbf{D}_{i,t}^{\top}\right)\left(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}\right)^{\top}=o_p\left(1\right)\cdot O_p\left(\sqrt{\frac{s_{NT}^2}{NT}\ln N}\right)=o_p\left(\frac{1}{\sqrt{\ln N}}\right)=o_p\left(1\right),$$
$$\left(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}\right)\left(\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{D}_{i,t}\mathbf{D}_{i,t}^{\top}\right)\left(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}\right)^{\top}=\left(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}\right)\left(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}\right)^{\top}.$$

By the Cauchy-Schwarz inequality, for any q and $r=1,2,\ldots,k+1$:

$$\left(\widehat{\boldsymbol{\gamma}}_q-\boldsymbol{\gamma}_q\right)^{\top}\left(\widehat{\boldsymbol{\gamma}}_r-\boldsymbol{\gamma}_r\right)\leq\left\|\widehat{\boldsymbol{\gamma}}_q-\boldsymbol{\gamma}_q\right\|_2\left\|\widehat{\boldsymbol{\gamma}}_r-\boldsymbol{\gamma}_r\right\|_2=O_p\left(\frac{s_{NT}}{NT}\ln N\right)=o_p\left(\frac{1}{\sqrt{NT}}\right).$$

Since all the other terms except for the first term clearly have an even smaller order, we have:

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{\mathbf{v}}_{i,t}\widehat{\mathbf{v}}_{i,t}^{\top}=\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{v}_{i,t}\mathbf{v}_{i,t}^{\top}+o_p\left(1\right).$$

For $(NT)^{-1/2}\sum_{i=1}^{N}\sum_{t=1}^{T}\widehat{\mathbf{v}}_{i,t}\widehat{e}_{i,t}$,

$$\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{v}_{i,t}\mathbf{D}_{i,t}^{\top}\right)\left(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)=O_p\left(1\right)\cdot O_p\left(\sqrt{\frac{s_{NT}^2}{NT}\ln N}\right)=o_p\left(\frac{1}{\sqrt{\ln N}}\right)=o_p\left(1\right),$$
$$\left(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}\right)\left(\frac{1}{\sqrt{NT}}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{D}_{i,t}e_{i,t}\right)=O_p\left(1\right)\cdot O_p\left(\sqrt{\frac{s_{NT}^2}{NT}\ln N}\right)=o_p\left(\frac{1}{\sqrt{\ln N}}\right)=o_p\left(1\right),$$

and by the Cauchy-Schwarz inequality, for $r = 1, 2, \ldots, k+1$,

$$
\begin{aligned}
& (\widehat{\boldsymbol{\gamma}}_r - \boldsymbol{\gamma}_r)^\top \left(\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{D}_{i,t} \mathbf{D}_{i,t}^\top \right) \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right) \\
= \; & \sqrt{NT} \, (\widehat{\boldsymbol{\gamma}}_r - \boldsymbol{\gamma}_r)^\top \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right) \leq \sqrt{NT} \, \|\widehat{\boldsymbol{\gamma}}_r - \boldsymbol{\gamma}_r\|_2 \left\| \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right\|_2 = \sqrt{NT} O_p \left(\frac{s_{NT}}{NT} \ln N \right) = o_p(1) .
\end{aligned}
$$

Since all other terms except for the first term clearly have an even smaller order, we have:

$$
\frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \widehat{\mathbf{v}}_{i,t} \widehat{e}_{i,t} = \frac{1}{\sqrt{NT}} \sum_{i=1}^{N} \sum_{t=1}^{T} \mathbf{v}_{i,t} e_{i,t} + o_p(1) . \square
$$

For the DML estimator, since $\widehat{\boldsymbol{\theta}}$ is computed using the auxiliary fold, and $\mathbf{v}_{i,t}$ and $\mathbf{D}_{i,t}$ are from the main fold, $\widehat{\boldsymbol{\theta}}$ is independent from $\mathbf{v}_{i,t}\mathbf{D}_{i,t}$, and $\left[(NT/2)^{-1/2} \sum_i \sum_t \mathbf{v}_{i,t} \mathbf{D}_{i,t}^\top \right] \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right)$ has a zero mean and a variance:

$$
\mathbb{E} \left[\left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right)^\top \left(\frac{2}{NT} \sum_i \sum_t \mathbf{D}_{i,t} \mathbf{v}_{i,t}^\top \mathbf{v}_{i,t} \mathbf{D}_{i,t}^\top \right) \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right) \right] = O(1) \cdot \left\| \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right\|_2^2 = O \left(\frac{s_{NT}}{NT} \ln N \right) .
$$

Therefore:

$$
\left(\frac{1}{\sqrt{NT/2}} \sum_i \sum_t \mathbf{v}_{i,t} \mathbf{D}_{i,t}^\top \right) \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta} \right) = O_p \left(\sqrt{\frac{s_{NT}}{NT} \ln N} \right) ,
$$

which is smaller than the one for PRL. We can also show that $\left(\widehat{\boldsymbol{\Gamma}} - \boldsymbol{\Gamma} \right) (NT/2)^{-1/2} \sum_i \sum_t \mathbf{D}_{i,t} e_{i,t} = O_p \left(\sqrt{s_{NT} \ln N / NT} \right)$. That is, the deviation of the asymptotic distribution of the DML estimator from the Gaussian approximation is smaller in magnitude than the deviation of the PRL estimator.

Internet Appendix B: Estimation of PRL Poisson

The post-regularization LASSO Poisson regression (PRL Possion) algorithm is based on the section 2.1 in Belloni et al. (2016). Consider the conditional mean function as: $E(y_{i,t+1} | d_{i,t-1}, \mathbf{x}_{i,t-1}) = G \left(\beta_{1,0} d_{i,t-1} + \mathbf{x}'_{i,t-1 2} \right)$, in which $G(\cdot) = \exp(\cdot)$, and $d_{i,t-1}$ is the policy variable or a treatment variable. The PRL Poisson method proceeds in three main steps:

1. Perform the Post-LASSO-Poisson of $y_{i,t}$ on $d_{i,t-1}$ and $\mathbf{x}_{i,t-1}$ to estimate the nuisance function $\mathbf{x}'_{i,t-1 2}$.
   (a) Run the LASSO-Poisson of $y_{i,t}$ on $d_{i,t-1}$ and $\mathbf{x}_{i,t-1}$. $\tilde{\mathbf{x}}_{i,t-1}$ denote the selected control.

(b) Run the Poisson regression of $y_{i,t}$ on $d_{i,t-1}$ and $\tilde{\mathbf{x}}_{i,t-1}$. Denote estimates as $\tilde{\beta}_1$, and $\tilde{\gamma}_2$, and keep the observation-level weights $w_i = G'\left(\tilde{\beta}_1 d_{i,t-1} + \tilde{\mathbf{x}}'_{i,t-1}\tilde{\gamma}_2\right)$, in which $G'(t) = \partial G(t)/\partial t$.

2. Perform Post-LASSO-OLS of $\sqrt{w_i}d_{i,t-1}$ on $\sqrt{w_i}\mathbf{x}_{i,t-1}$ to estimate the instrument $z_{0,i}$.
   (a) Perform a linear LASSO of $\sqrt{w_i}d_{i,t-1}$ on $\sqrt{w_i}\mathbf{x}_{i,t-1}$. $\check{\mathbf{x}}_{i,t-1}$ denote the selected control.
   (b) Perform the linear regression of $\sqrt{w_i}d_{i,t-1}$ on $\sqrt{w_i}\check{\mathbf{x}}_{i,t-1}$. Denote the estimates as $\check{\boldsymbol{\theta}}$. and use the unweighted residual $z_{i,t-1} := d_{i,t-1} - \check{\mathbf{x}}'_{i,t-1}\check{\boldsymbol{\theta}}$ as the instrument variable.
3. Combine the previous two steps to estimate $\beta_{1,0}$ using the instrumental Poisson regression with estimates of $\mathbf{x}'_{i,t-1}\tilde{\gamma}_2$ from step 1 and the instrument variable $z_{i,t-1}$ from step 2.
   (a) Compute $\hat{\beta}_1$ by solving the following sample moment equations:

$$\frac{1}{NT}\sum_i\sum_t\left\{\left[y_{i,t+1} - G\left(\beta_1 d_{i,t-1} + \tilde{\mathbf{x}}'_{i,t-1}\tilde{\gamma}_2\right)\right] z_{i,t-1}\right\} = 0.$$

In our empirical analysis, we perform the cross-validation with 10-folds in step 1 and 2 to select the optimal penalty parameters on the LASSO methods and use the cluster by firm standard errors to account for the within-firm heterogeneity in step 3.

Internet Appendix C: Asymptotic Normality of PRL Poisson

In this Appendix, we show the asymptotic normality of the post-regularization LASSO Poisson estimator in Belloni et al. (2016). We consider the Poisson setting in Models (13) and (14) of our paper, and to simplify the notation, we write these models as:

$$\mu_{i,t} = \mathbb{E}\left(y_{i,t}\left|\mathbf{X}_{i,t}, \mathbf{D}_{i,t}\right.\right) = \exp\left(\beta_0 + \mathbf{X}_{i,t}^\top\boldsymbol{\beta} + \mathbf{D}_{i,t}^\top\boldsymbol{\alpha}\right), \tag{C.1}$$

$$\mathbf{X}_{i,t} = \boldsymbol{\gamma}_0 + \boldsymbol{\Gamma}\mathbf{D}_{i,t} + \mathbf{v}_{i,t}. \tag{C.2}$$

$\boldsymbol{\beta}$ is estimated by GMM using the following moment condition:

$$g\left(\boldsymbol{\beta}\right) = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{X}_{i,t} - \widehat{\boldsymbol{\gamma}}_0 - \widehat{\boldsymbol{\Gamma}}\mathbf{D}_{i,t}\right)\left[y_{i,t} - \exp\left(\widehat{\beta}_0 + \mathbf{X}_{i,t}^\top\boldsymbol{\beta} + \mathbf{D}_{i,t}^\top\widehat{\boldsymbol{\alpha}}\right)\right] = \mathbf{0},$$

in which $\widehat{\beta}_0$, $\widehat{\boldsymbol{\alpha}}$, $\widehat{\boldsymbol{\gamma}}_0$, and $\widehat{\boldsymbol{\Gamma}}$ are post-Poisson-LASSO or post-LASSO estimates for β_0, $\boldsymbol{\alpha}$, $\boldsymbol{\gamma}_0$, and $\boldsymbol{\Gamma}$.

Let $\widetilde{\boldsymbol{\beta}}$ be the infeasible estimator using the moment condition:

$$\begin{aligned} h(\boldsymbol{\beta}) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{X}_{i,t}-\boldsymbol{\gamma}_0-\boldsymbol{\Gamma}\mathbf{D}_{i,t}\right)\left[y_{i,t}-\exp\left(\beta_0+\mathbf{X}_{i,t}^{\top}\boldsymbol{\beta}+\mathbf{D}_{i,t}^{\top}\boldsymbol{\alpha}\right)\right] \\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{v}_{i,t}\left[y_{i,t}-\exp\left(\beta_0+\mathbf{X}_{i,t}^{\top}\boldsymbol{\beta}+\mathbf{D}_{i,t}^{\top}\boldsymbol{\alpha}\right)\right]=\mathbf{0}. \end{aligned}$$

The conditions are similar to the ones for the linear specification. Let $s_{NT}=\left\|\boldsymbol{\alpha}+\sum_{r=1}^{k+1}\boldsymbol{\gamma}_r\right\|_0$ denote the number of important firm dummies (i.e., the number of dummies with non-zero coefficients).

Condition C.1 As $N,T\to\infty$, $s_{NT}^2(\ln N)^2\Big/(NT)=o(1)$.

Condition C.2 With a probability approaching one as $N,T\to\infty$, the post-LASSO estimates $\widehat{\boldsymbol{\theta}}$ and $\widehat{\boldsymbol{\gamma}}_r$ with $r=1,2,\ldots,k+1$ satisfy the following standard performance bounds:

$$\begin{aligned} \left\|\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}\right\|_1 &= O_p\left(\sqrt{\frac{s_{NT}^2}{NT}\ln N}\right), \quad \left\|\widehat{\boldsymbol{\gamma}}_r-\boldsymbol{\gamma}_r\right\|_1 = O_p\left(\sqrt{\frac{s_{NT}^2}{NT}\ln N}\right), \\ \left\|\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}\right\|_2 &= O_p\left(\sqrt{\frac{s_{NT}}{NT}\ln N}\right), \quad \left\|\widehat{\boldsymbol{\gamma}}_r-\boldsymbol{\gamma}_r\right\|_2 = O_p\left(\sqrt{\frac{s_{NT}}{NT}\ln N}\right). \end{aligned}$$

Proposition C.1 Suppose equations (C.1) and (C.2), and Conditions C.1 and C.2 hold. Then:

$$\sqrt{NT}\left(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)=\sqrt{NT}\left(\widetilde{\boldsymbol{\beta}}-\boldsymbol{\beta}\right)+o_p(1).$$

Proof. It suffices to show that $h(\boldsymbol{\beta})=g(\boldsymbol{\beta})+o_p\left((NT)^{-1/2}\right)$, i.e., the difference of $h(\boldsymbol{\beta})$, the moment condition for the infeasible estimator, and $g(\boldsymbol{\beta})$, the moment condition for the GMM, is $o_p\left((NT)^{-1/2}\right)$. $g(\boldsymbol{\beta})$ can be written as:

$$\begin{aligned} g(\beta) &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\mathbf{X}_{i,t}-\widehat{\boldsymbol{\gamma}}_0-\widehat{\boldsymbol{\Gamma}}\mathbf{D}_{i,t}\right)\left[y_{i,t}-\exp\left(\widehat{\beta}_0+\mathbf{X}_{i,t}^{\top}\boldsymbol{\beta}+\mathbf{D}_{i,t}^{\top}\widehat{\boldsymbol{\alpha}}\right)\right] \\ &= \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\mathbf{v}_{i,t}-\left(\widehat{\boldsymbol{\gamma}}_0-\boldsymbol{\gamma}_0\right)-\left(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}\right)\mathbf{D}_{i,t}\right] \\ &\quad \left\{y_{i,t}-\exp\left(\beta_0+\mathbf{X}_{i,t}^{\top}\boldsymbol{\beta}+\mathbf{D}_{i,t}^{\top}\boldsymbol{\alpha}\right)\left[1+\left(\widehat{\beta}_0-\beta_0\right)+\mathbf{D}_{i,t}^{\top}\left(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}\right)+(s.o.)\right]\right\} \\ &= h(\boldsymbol{\beta})+A_1+A_2+A_3+(s.o.), \end{aligned}$$

in which

$$A_1 = -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{v}_{i,t}\left[\left(\widehat{\beta}_0-\beta_0\right)+\mathbf{D}_{i,t}^{\top}\left(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}\right)\right]$$

$$A_2 = -\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\left(\widehat{\boldsymbol{\gamma}}_0-\boldsymbol{\gamma}_0\right)+\left(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}\right)\mathbf{D}_{i,t}\right]\left[y_{i,t}-\exp\left(\beta_0+\mathbf{X}_{i,t}^{\top}\boldsymbol{\beta}+\mathbf{D}_{i,t}^{\top}\boldsymbol{\alpha}\right)\right]$$

$$A_3 = \frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left[\left(\widehat{\boldsymbol{\gamma}}_0-\boldsymbol{\gamma}_0\right)+\left(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}\right)\mathbf{D}_{i,t}\right]\left[\left(\widehat{\beta}_0-\beta_0\right)+\mathbf{D}_{i,t}^{\top}\left(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}\right)\right],$$

and $(s.o.)$ denotes terms with smaller order. Using Conditions C.1 and C.2,

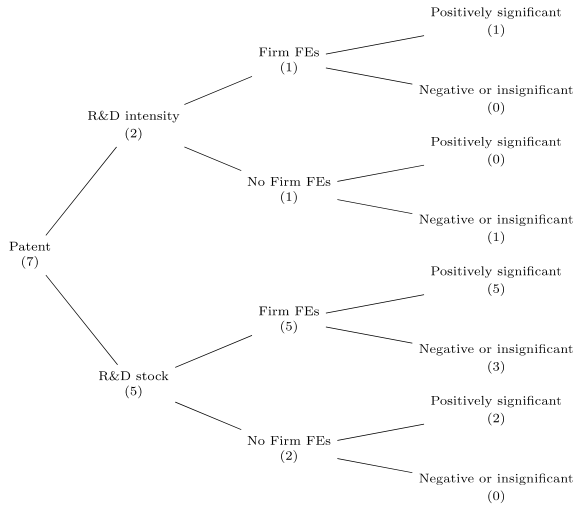
$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{v}_{i,t}\mathbf{D}_{i,t}^{\top}\left(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}\right) = O_p\left(\frac{1}{\sqrt{NT}}\right)\cdot O_p\left(\sqrt{\frac{s_{NT}^2}{NT}\ln N}\right) = o_p\left(\frac{1}{\sqrt{NT\ln N}}\right),$$

$$\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\left(\widehat{\boldsymbol{\Gamma}}-\boldsymbol{\Gamma}\right)\mathbf{D}_{i,t}\left[y_{i,t}-\exp\left(\beta_0+\mathbf{X}_{i,t}^{\top}\boldsymbol{\beta}+\mathbf{D}_{i,t}^{\top}\boldsymbol{\alpha}\right)\right] = o_p\left(\frac{1}{\sqrt{NT\ln N}}\right),$$

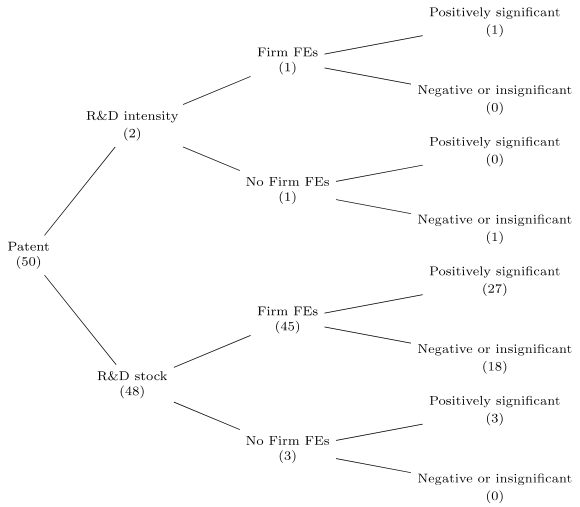
and for any $r=1,2,\ldots,k+1$, by the Cauchy-Schwarz inequality,

$$\begin{aligned}&\left(\widehat{\boldsymbol{\gamma}}_r-\boldsymbol{\gamma}_r\right)^{\top}\frac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T}\mathbf{D}_{i,t}\mathbf{D}_{i,t}^{\top}\left(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}\right)\\ =\;&\left(\widehat{\boldsymbol{\gamma}}_r-\boldsymbol{\gamma}_r\right)^{\top}\left(\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}\right)\leq\left\|\widehat{\boldsymbol{\gamma}}_r-\boldsymbol{\gamma}_r\right\|_2\left\|\widehat{\boldsymbol{\alpha}}-\boldsymbol{\alpha}\right\|_2 = O_p\left(\frac{s_{NT}}{NT}\ln N\right)=o_p\left(\frac{1}{\sqrt{NT}}\right).\end{aligned}$$

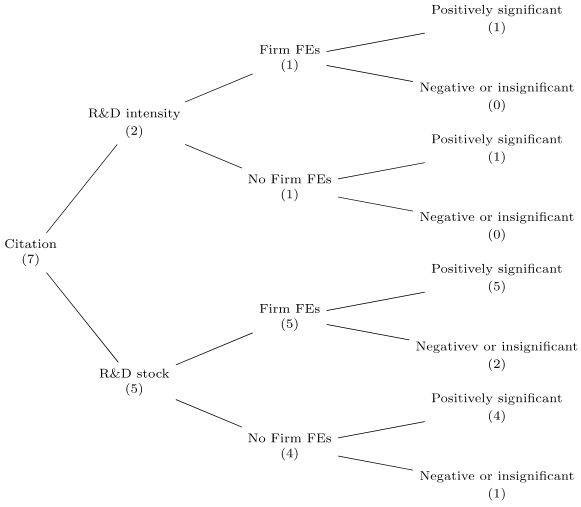
Since all the other terms except for the first term clearly have an even smaller order, we have: $h\left(\boldsymbol{\beta}\right)=g\left(\boldsymbol{\beta}\right)+o_p\left(\left(NT\right)^{-1/2}\right)$. □



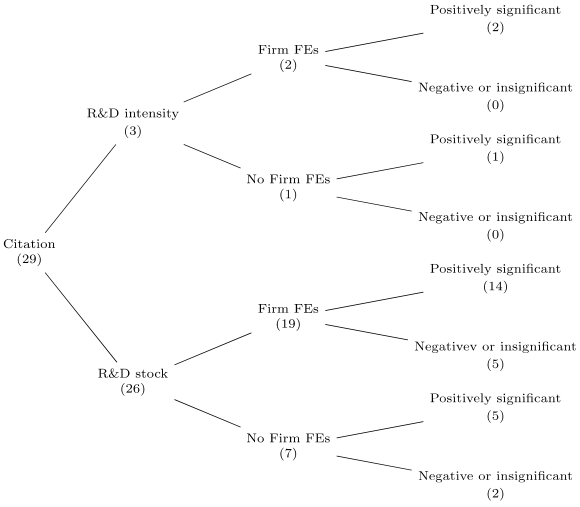
(a) Patent count model: article number



(b) Patent count model: regression number

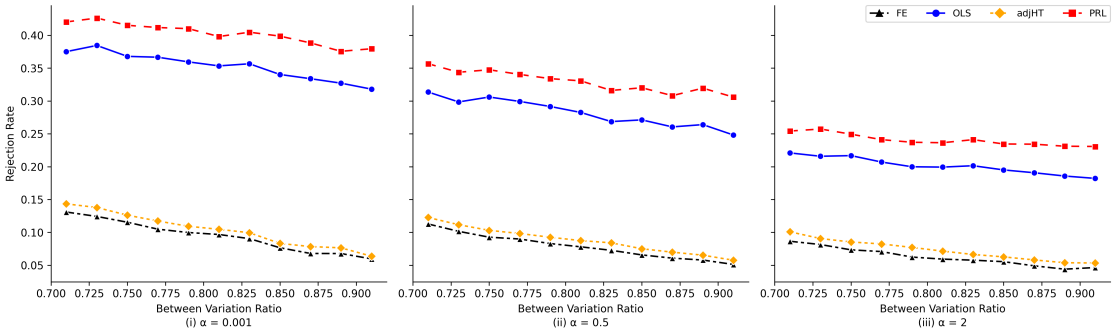


(c) Citation count model: article number

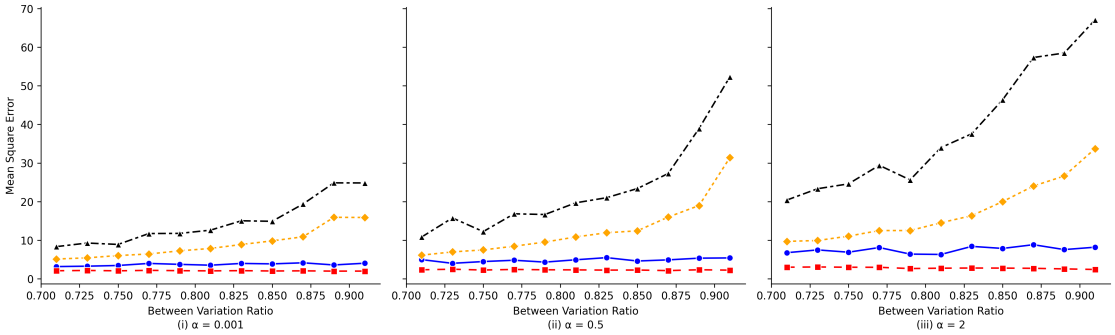


(d) Citation count model: regression number

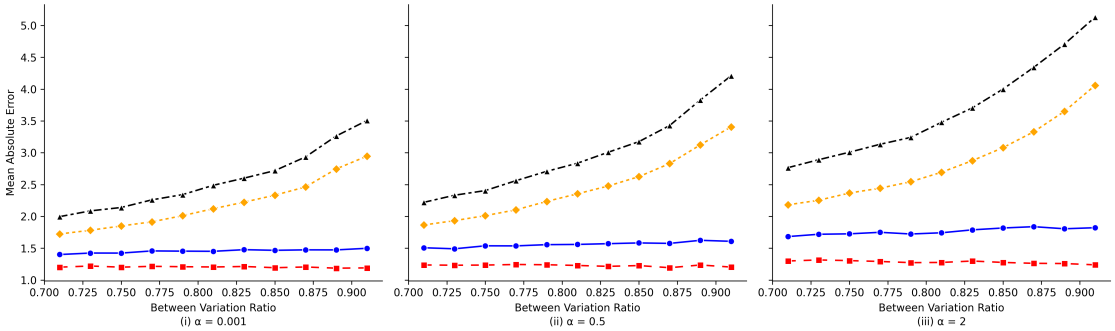
Figure IA.1: Patent and citation count-model studies



(a) Rejection rate

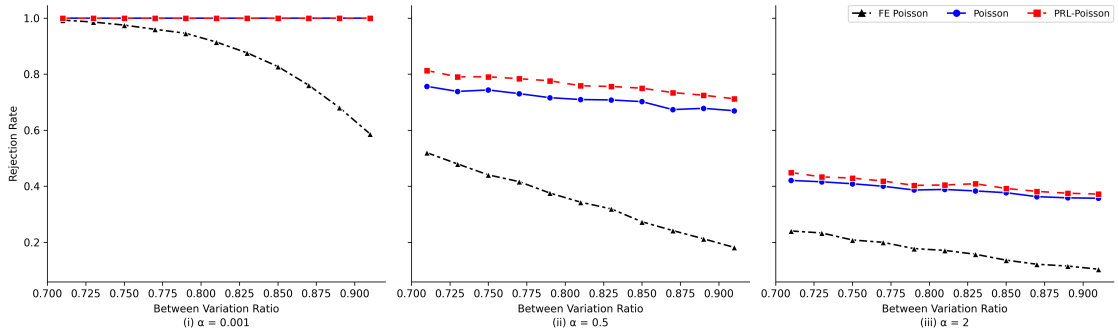


(b) Mean square error

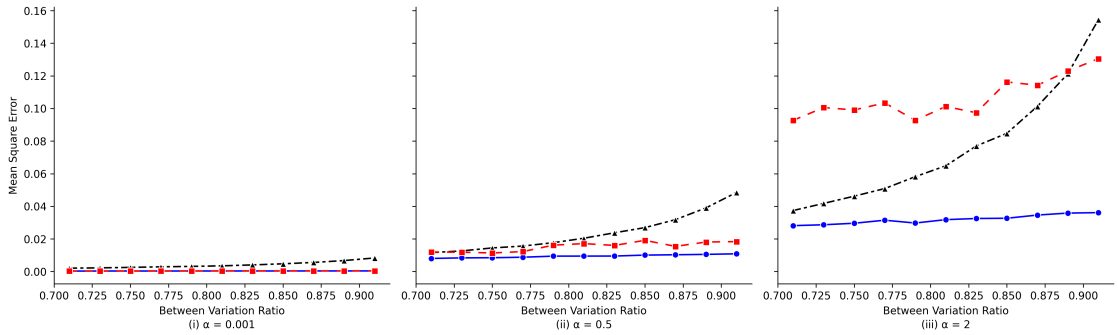


(c) Mean absolute error

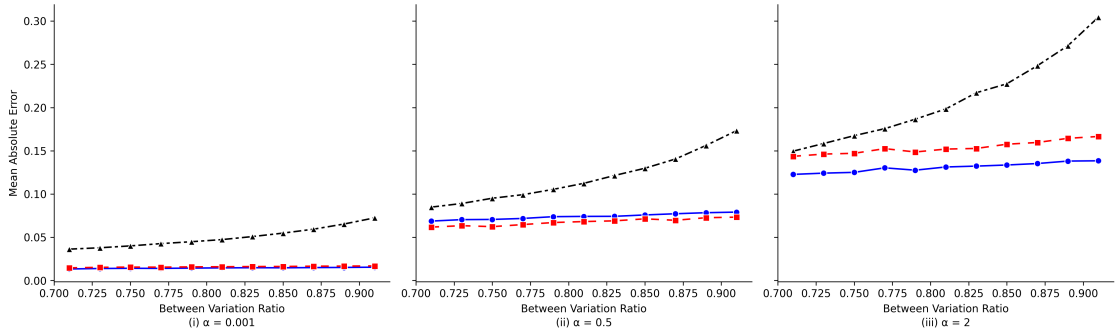
Figure IA.2: Simulation rejection rates and error rates for continuous outcomes: FE, OLS, adjHT, and PRL. This figure compares the rejection and error rates for four estimators: FE (regression with full set of firm dummies), OLS (regression without any firm dummies), adjHT (adjusted Hausman-Taylor), and PRL (post-regularization LASSO). The continuous outcome $y_{i,t}$ is generated from a mixture of $\chi^2(1)$ distributions. The number of mixture components is the product of $\exp(0.2z_{i,t} + x_{i,t})$, and a log-normal random variable with mean one and variance controlled by α, where $z_{i,t} = 0.5\eta_i + 0.5\nu_{i,t}$. η_i and $\nu_{i,t}$ are independently drawn from normal distributions truncated at the 1st and 99th percentiles, with standard deviations σ_η and σ_ν. The between variation ratio is defined as $\sigma_\eta^2/(\sigma_\eta^2 + \sigma_\nu^2)$. Subplots (i), (ii), and (iii) represent low ($\alpha = 0.001$), moderate ($\alpha = 0.5$), and high overdispersion ($\alpha = 2$) of the continuous outcomes, respectively.



(a) Rejection rate



(b) Mean square error



(c) Mean absolute error

Figure IA.3: Simulation rejection rates and error rates for continuous outcomes: FE Poisson, Poisson, and PRL Poisson. This figure compares the rejection and error rates for three estimators: FE Poisson (Poisson regression with full set of firm dummies), Poisson (Poisson regression without any firm dummies), PRL Poisson (post-regularization LASSO Poisson regression). The continuous outcome $y_{i,t}$ is generated from a mixture of $\chi^2(1)$ distributions. The number of mixture components is the product of $\exp(0.2z_{i,t} + x_{i,t})$, and a log-normal random variable with mean one and variance controlled by α, where $z_{i,t} = 0.5\eta_i + 0.5\nu_{i,t}$. η_i and $\nu_{i,t}$ are independently drawn from normal distributions truncated at the 1st and 99th percentiles, with standard deviations σ_η and σ_ν. The between variation ratio is defined as $\sigma_\eta^2/(\sigma_\eta^2 + \sigma_\nu^2)$. Subplots (i), (ii), and (iii) represent low ($\alpha = 0.001$), moderate ($\alpha = 0.5$), and high overdispersion ($\alpha = 2$) of the continuous outcomes, respectively.

Table IA.1: Innovation to R&D intensity regression (Full table)

	(1) FE (Firm, Year)	(2) OLS (Year)	(3) adjHT (Year)	(4) PRL (Firm, Year)	(5) DML (Firm, Year)
Panel A: Patent regression					
R&D/Asset	0.041	0.593***	0.220***	0.199***	0.213***
	(0.028)	(0.042)	(0.027)	(0.018)	(0.014)
ln(ME)	0.112***	0.275***	0.155***	0.034***	0.038***
	(0.007)	(0.013)	(0.007)	(0.004)	(0.003)
ln(1+Age)	0.108***	0.177***	0.085***	0.056***	0.055***
	(0.019)	(0.012)	(0.009)	(0.006)	(0.004)
ln(K/L)	0.023***	0.037***	0.028***	0.006**	0.010***
	(0.007)	(0.007)	(0.005)	(0.003)	(0.002)
R&D missing	−0.026*	−0.539***	−0.359***	−0.151***	−0.162***
	(0.014)	(0.020)	(0.014)	(0.009)	(0.006)
Tobin Q	−0.018***	−0.049***	−0.026***	0.004	0.001
	(0.004)	(0.006)	(0.003)	(0.003)	(0.002)
ROA	−0.085***	−0.131***	−0.130***	0.029	0.010
	(0.023)	(0.046)	(0.022)	(0.022)	(0.016)
Leverage	0.003	−0.065*	−0.078***	−0.055***	−0.065***
	(0.025)	(0.038)	(0.023)	(0.018)	(0.013)
Cash/Asset	0.050	−0.008	0.149***	0.075***	0.051***
	(0.031)	(0.048)	(0.027)	(0.024)	(0.018)
KZ index	0.001**	0.001	0.001***	0.000	−0.001**
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
InstOwn	−0.052*	−0.329***	−0.124***	0.005	0.007
	(0.030)	(0.065)	(0.032)	(0.025)	(0.019)
(1-HHI)	−0.185**	0.443***	0.093	0.289***	0.262***
	(0.081)	(0.171)	(0.082)	(0.055)	(0.040)
$(1\text{-HHI})^2$	0.171**	−0.473***	−0.060	−0.250***	−0.227***
	(0.070)	(0.143)	(0.071)	(0.049)	(0.036)
Firm FEs	Y	N	N	Y	Y
Year FEs	Y	Y	Y	Y	Y
Observations	84,620	86,341	86,341	86,341	86,341
Panel B: Citation regression					
R&D/Asset	−0.051	1.396***	0.551***	1.397***	1.364***
	(0.068)	(0.084)	(0.065)	(0.083)	(0.050)
ln(ME)	0.196***	0.486***	0.325***	0.478***	0.476***
	(0.014)	(0.019)	(0.013)	(0.019)	(0.010)
ln(1+Age)	0.190***	0.295***	0.157***	0.297***	0.300***
	(0.036)	(0.023)	(0.019)	(0.024)	(0.013)
ln(K/L)	0.040***	0.065***	0.055***	0.074***	0.075***
	(0.014)	(0.013)	(0.010)	(0.012)	(0.007)
R&D missing	−0.090***	−1.168***	−0.937***	−1.171***	−1.175***
	(0.031)	(0.037)	(0.032)	(0.037)	(0.020)
Tobin Q	−0.030***	−0.057***	−0.039***	−0.063***	−0.061***

Table IA.1 – continued from previous page

	(1) FE (Firm, Year)	(2) OLS (Year)	(3) adjHT (Year)	(4) PRL (Firm, Year)	(5) DML (Firm, Year)
	(0.008)	(0.011)	(0.007)	(0.011)	(0.007)
ROA	−0.150***	−0.071	−0.309***	−0.072	−0.092*
	(0.058)	(0.090)	(0.056)	(0.089)	(0.056)
Leverage	−0.027	−0.188**	−0.186***	−0.181**	−0.179***
	(0.058)	(0.076)	(0.053)	(0.075)	(0.047)
Cash/Asset	0.215***	0.334***	0.566***	0.291***	0.251***
	(0.074)	(0.101)	(0.067)	(0.096)	(0.061)
KZ index	0.001	0.002	0.002**	−0.001	−0.002**
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
InstOwn	0.052	−0.271**	−0.081	−0.227**	−0.228***
	(0.063)	(0.112)	(0.065)	(0.112)	(0.064)
(1-HHI)	−0.247	1.334***	0.460***	1.451***	1.306***
	(0.171)	(0.291)	(0.173)	(0.258)	(0.157)
$(1\text{-HHI})^2$	0.227	−1.298***	−0.399***	−1.372***	−1.227***
	(0.155)	(0.250)	(0.153)	(0.223)	(0.136)
Firm FEs	Y	N	N	Y	Y
Year FEs	Y	Y	Y	Y	Y
Observations	84,620	86,341	86,341	86,341	86,341
Panel C: Adjusted citation regression					
R&D/Asset	0.033	0.590***	0.231***	0.210***	0.198***
	(0.031)	(0.045)	(0.030)	(0.019)	(0.015)
ln(ME)	0.109***	0.282***	0.156***	0.037***	0.040***
	(0.008)	(0.013)	(0.007)	(0.004)	(0.003)
ln(1+Age)	0.104***	0.158***	0.067***	0.049***	0.054***
	(0.019)	(0.013)	(0.009)	(0.006)	(0.005)
ln(K/L)	0.020***	0.035***	0.025***	0.007**	0.010***
	(0.007)	(0.007)	(0.005)	(0.003)	(0.002)
R&D missing	−0.027*	−0.555***	−0.364***	−0.168***	−0.168***
	(0.015)	(0.020)	(0.015)	(0.009)	(0.006)
Tobin Q	−0.018***	−0.043***	−0.026***	0.003	0.001
	(0.004)	(0.006)	(0.004)	(0.003)	(0.002)
ROA	−0.089***	−0.100**	−0.121***	0.047**	0.010
	(0.025)	(0.049)	(0.024)	(0.024)	(0.018)
Leverage	0.006	−0.062	−0.055**	−0.061***	−0.057***
	(0.028)	(0.040)	(0.026)	(0.020)	(0.015)
Cash/Asset	0.090**	0.048	0.199***	0.084***	0.050**
	(0.035)	(0.052)	(0.031)	(0.026)	(0.020)
KZ index	0.001	0.001**	0.001***	0.000	−0.001***
	(0.000)	(0.001)	(0.000)	(0.000)	(0.000)
InstOwn	−0.048	−0.318***	−0.108***	0.035	0.016
	(0.032)	(0.069)	(0.033)	(0.028)	(0.021)
(1-HHI)	−0.214**	0.446**	0.095	0.301***	0.268***
	(0.085)	(0.175)	(0.086)	(0.059)	(0.044)
$(1\text{-HHI})^2$	0.194***	−0.479***	−0.071	−0.259***	−0.232***

Table IA.1 – continued from previous page

	(1) FE (Firm, Year)	(2) OLS (Year)	(3) adjHT (Year)	(4) PRL (Firm, Year)	(5) DML (Firm, Year)
	(0.075)	(0.147)	(0.075)	(0.053)	(0.039)
Firm FEs	Y	N	N	Y	Y
Year FEs	Y	Y	Y	Y	Y
Observations	84,620	86,341	86,341	86,341	86,341

This table presents the full regression results of innovation on R&D intensity using different estimation methods. Column (1) reports fixed effects (FE) estimates with firm and year fixed effects. Column (2) presents OLS estimates with year fixed effects. Column (3) reports adjusted Hausman-Taylor (adjHT) estimates with year fixed effects. Columns (4) and (5) present Post-Regularization LASSO (PRL) and Double Machine Learning (DML) estimates with firm and year fixed effects, respectively. The standard errors are clustered at the firm level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table IA.2: Innovation to R&D regression for patenting firms

	(1) FE (Firm, Year)	(2) OLS (Year)	(3) adjHT (Year)	(4) PRL (Firm, Year)	(5) DML (Firm, Year)
Panel A: Patent regression					
A.1 R&D scaled by total asset					
R&D/Asset	0.035	0.448***	0.161***	0.406***	0.399***
	(0.031)	(0.045)	(0.030)	(0.044)	(0.025)
Selected ratio				3.34%	6.32%
A.2 R&D scaled by market equity					
R&D/ME	0.102***	0.499***	0.244***	0.468***	0.472***
	(0.024)	(0.041)	(0.024)	(0.041)	(0.023)
Selected ratio				3.34%	6.80%
A.3 R&D stock					
ln(1+R&D)	0.032***	0.225***	0.090***	0.222***	0.221***
	(0.005)	(0.009)	(0.005)	(0.009)	(0.005)
Selected ratio				3.34%	6.32%
Panel B: Citation regression					
B.1 R&D scaled by total asset					
R&D/Asset	−0.061	0.777***	0.317***	0.738***	0.740***
	(0.079)	(0.087)	(0.075)	(0.084)	(0.054)
Selected ratio				3.34%	6.32%
B.2 R&D scaled by market equity					
R&D/ME	0.120**	0.818***	0.451***	0.779***	0.775***
	(0.054)	(0.073)	(0.053)	(0.072)	(0.045)
Selected ratio				3.34%	6.80%
B.3 R&D stock					
ln(1+R&D)	0.051***	0.390***	0.193***	0.385***	0.384***
	(0.010)	(0.013)	(0.009)	(0.013)	(0.007)
Selected ratio				3.34%	6.32%
Panel C: Adjusted citation regression					
C.1 R&D scaled by total asset					
R&D/Asset	0.033	0.419***	0.175***	0.369***	0.377***
	(0.036)	(0.049)	(0.034)	(0.047)	(0.028)
Selected ratio				3.34%	6.32%
C.2 R&D scaled by market equity					
R&D/ME	0.098***	0.508***	0.248***	0.471***	0.483***
	(0.026)	(0.044)	(0.026)	(0.044)	(0.025)

Table IA.2 – continued from previous page

	(1) FE (Firm, Year)	(2) OLS (Year)	(3) adjHT (Year)	(4) PRL (Firm, Year)	(5) DML (Firm, Year)
Selected ratio				3.34%	6.80%
C.3 R&D stock					
ln(1+R&D)	0.030***	0.231***	0.089***	0.228***	0.227***
	(0.005)	(0.010)	(0.005)	(0.009)	(0.005)
Selected ratio				3.34%	6.32%
Control variables	Y	Y	Y	Y	Y
Firm FEs	Y	N	N	Y	Y
Year FEs	Y	Y	Y	Y	Y
Observations	45,679	45,913	45,913	45,913	45,913

This table presents the regression results of innovation on R&D measures for patenting firms using different estimation methods. Column (1) reports fixed effects (FE) estimates with firm and year fixed effects. Column (2) presents OLS estimates with year fixed effects. Column (3) reports adjusted Hausman-Taylor (adjHT) estimates with year fixed effects. Columns (4) and (5) present Post-Regularization LASSO (PRL) and Double Machine Learning (DML) estimates with firm and year fixed effects, respectively. The standard errors are clustered at the firm level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The selected ratio is the number of selected year and firm dummies divided by the total number of year and firm dummies.

Table IA.3: Selected 100 company list

Rank	CONM	TIC	SIC	NAICS	CITY	STATE
1	INTL BUSINESS MACHINES CORP	IBM	7370	519130	Armonk	NY
2	LUCENT TECHNOLOGIES INC	LU.1	7373	541512	Plano	TX
3	GENERAL ELECTRIC CO	GE	9997	999977	Boston	MA
4	APTIV PLC	APTV	3714	336320	Dublin	
5	EASTMAN KODAK CO	KODK	3577	333316	Rochester	NY
6	MOTOROLA SOLUTIONS INC	MSI	3663	334220	Chicago	IL
7	GENERAL MOTORS CO	GM	3711	336110	Detroit	MI
8	DU PONT (E I) DE NEMOURS	DD.2	2820	3252	Indianapolis	IN
9	AT&T CORP	T.2	4813	517110	Bedminster	NJ
10	DUPONT DE NEMOURS INC	DD	2860	325199	Wilmington	DE
11	XEROX HOLDINGS CORP	XRX	3577	333316	Norwalk	CT
12	AT&T INC	T	4812	517312	Dallas	TX
13	TEXAS INSTRUMENTS INC	TXN	3674	334413	Dallas	TX
14	3M CO	MMM	2670	322220	Saint Paul	MN
15	RCA CORP	RCA.1	3600	335	New York	NY
16	BROADCOM CORP	BRCM	3674	334413	San Jose	CA
17	NORTH AMERICAN PHILIPS CORP	PHG1	3640	3351	Andover	MA
18	EXXON MOBIL CORP	XOM	2911	324110	Irving	TX
19	HP INC	HPQ	3570	33411	Palo Alto	CA
20	MERCK & CO	MRK	2834	325412	Kenilworth	NJ
21	FORD MOTOR CO	F	3711	33611	Dearborn	MI
22	HONEYWELL INTERNATIONAL INC	HON	9997	336413	Charlotte	NC
23	CBS CORP -OLD	CBS.2	4833	513120	New York	NY
24	RAYTHEON TECHNOLOGIES CORP	RTX	3724	336412	Arlington	VA
25	PROCTER & GAMBLE CO	PG	2840	325611	Cincinnati	OH
26	SUN MICROSYSTEMS INC	JAVA.1	3571	334111	Redwood City	CA
27	QUALCOMM INC	QCOM	3674	334413	San Diego	CA
28	MOBIL CORP	MOB.2	2911	324110	Fairfax	VA
29	CONOCOPHILLIPS	COP	1311	2111	Houston	TX
30	MICRON TECHNOLOGY INC	MU	3674	334413	Boise	ID
31	JOHNSON & JOHNSON	JNJ	2834	325412	New Brunswick	NJ
32	BRISTOL-MYERS SQUIBB CO	BMY	2834	325412	New York	NY
33	BOEING CO	BA	3721	336411	Chicago	IL
34	CATERPILLAR INC	CAT	3531	333120	Irving	TX
36	SHELL OIL CO	4108A	1311	211112	Houston	TX
37	LEXMARK INTL INC -CL A	LXK	3577	334118	Lexington	KY
38	HONEYWELL INC	HON.Z	3822	334512	Minneapolis	MN
39	MICROSOFT CORP	MSFT	7372	513210	Redmond	WA
40	PPG INDUSTRIES INC	PPG	2851	325510	Pittsburgh	PA
41	PHARMACIA CORP	PHA.1	2834	325412	Kalamazoo	MI
42	PFIZER INC	PFE	2834	325412	New York	NY
43	SPERRY CORP	SY.2	3570	3341	New York	NY
44	ROCKWELL AUTOMATION	ROK	3620	335314	Milwaukee	WI
45	LILLY (ELI) & CO	LLY	2834	325412	Indianapolis	IN
46	ITT INC	ITT	3561	333914	White Plains	NY
47	UNISYS CORP	UIS	7373	541512	Blue Bell	PA
48	JUNIPER NETWORKS INC	JNPR	3576	334118	Sunnyvale	CA
49	GOODYEAR TIRE & RUBBER CO	GT	3011	326211	Akron	OH
50	ABBOTT LABORATORIES	ABT	3845	334510	North Chicago	IL
51	AMERICAN CYANAMID CO	ACY.1	2800	325	Madison	NJ
52	BAKER HUGHES INC	BHI	1381	213111	Houston	TX
52	BAKER HUGHES CO	BKR	1389	213112	Houston	TX
53	CORNING INC	GLW	3679	334419	Corning	NY
54	INTEL CORP	INTC	3674	334413	Santa Clara	CA
55	PITNEY BOWES INC	PBI	3579	333318	Stamford	CT
56	EASTMAN CHEMICAL CO	EMN	2800	325	Kingsport	TN
57	UNION CARBIDE CORP	DOW2	2860	325199	Seadrift	TX
58	AMKOR TECHNOLOGY INC	AMKR	3674	334413	Tempe	AZ
59	PALM INC	PALM	3571	334111	Sunnyvale	CA
60	CHEVRON CORP	CVX	2911	324110	San Ramon	CA
61	TEXACO INC	TX.2	2911	324110	White Plains	NY
62	TRW INC	TRW.1	3714	336330	Cleveland	OH
63	AMOCO CORP	AN.2	2911	324110	Houston	TX

Continued on next page

Rank	CONM	TIC	SIC	NAICS	CITY	STATE
64	DEERE & CO	DE	3523	333111	Moline	IL
65	MILLENNIUM PHARMACEUTICALS	MLNM	2835	325413	Cambridge	MA
66	NATIONAL SEMICONDUCTOR CORP	NSM.2	3674	334413	Santa Clara	CA
67	NORTHROP GRUMMAN CORP	NOC	3812	334511	Falls Church	VA
68	RAYTHEON CO	RTN	3812	334511	Waltham	MA
69	VERIZON COMMUNICATIONS INC	VZ	4812	517312	New York	NY
70	EMERSON ELECTRIC CO	EMR	3823	334513	Saint Louis	MO
71	FMC CORP	FMC	2870	325320	Philadelphia	PA
72	BAXTER INTERNATIONAL INC	BAX	2834	325412	Deerfield	IL
73	ADVANCED MICRO DEVICES	AMD	3674	334413	Santa Clara	CA
74	AIR PRODUCTS & CHEMICALS INC	APD	2810	325120	Allentown	PA
75	LINDE PLC	LIN	2810	325120	Woking	
76	OWENS CORNING	OC	3290	327993	Toledo	OH
77	L3HARRIS TECHNOLOGIES INC	LHX	3812	334511	Melbourne	FL
78	UOP INC	UOP.1	3350		Des Plaines	IL
79	XILINX INC	XLNX	3674	334413	San Jose	CA
80	KIMBERLY-CLARK CORP	KMB	2621	322121	Dallas	TX
81	LEAR CORP	LEA	2531	336360	Southfield	MI
82	ILLINOIS TOOL WORKS	ITW	3560	33399	Glenview	IL
83	ALLERGAN INC	AGN.2	2834	325412	Irvine	CA
84	COLGATE-PALMOLIVE CO	CL	2844	325611	New York	NY
85	GRACE (W R) & CO	GRA	2810	325998	Columbia	MD
86	CIENA CORP	CIEN	3661	334210	Hanover	MD
87	TEKTRONIX INC	TEK.1	3825	334515	Beaverton	OR
88	GTE CORP	GTE.1	4813	513310	Irving	TX
89	ROHM AND HAAS CO	ROH	2821	325211	Philadelphia	PA
90	BECTON DICKINSON & CO	BDX	3841	339112	Franklin Lakes	NJ
91	DELL TECHNOLOGIES INC	DELL	3571	334111	Round Rock	TX
92	MEDTRONIC PLC	MDT	3845	334510	Dublin	
93	LOCKHEED MARTIN CORP	LMT	3760	336414	Bethesda	MD
94	ALCOA INC	AA.3	3720	3364	New York	NY
95	LSI CORP	LSI.1	3674	334413	San Jose	CA
96	ATLANTIC RICHFIELD CO	ARC.3	2911	324110	Wilmington	DE
97	AUTOLIV INC	ALV	3714	336390	Stockholm	
98	IONIS PHARMACEUTICALS INC	IONS	2834	325412	Carlsbad	CA
99	HUMAN GENOME SCIENCES INC	HGSI	2835	325413	Rockville	MD
100	WYETH	WYE	2834	325412	New York	NY

Table IA.4: PRL Poisson and DML: R&D intensity (Full table)

	(1) FE Poisson (Firm, Year)	(2) Poisson (Year)	(3) PRL Poisson (Firm, Year)	(4) DML Poisson (Firm, Year)
Panel A: Patent count model				
R&D/Asset	−0.248	2.305***	2.407***	2.312***
	(0.255)	(0.187)	(0.161)	(0.120)
ln(ME)	0.331***	1.054***	0.937***	0.895***
	(0.063)	(0.052)	(0.081)	(0.082)
ln(1+Age)	0.745***	0.335**	0.469***	0.563***
	(0.287)	(0.142)	(0.182)	(0.162)
ln(K/L)	−0.072	−0.038	0.062	0.080*
	(0.146)	(0.061)	(0.044)	(0.044)
R&D missing	−0.252*	−2.081***	−2.099***	−2.111***
	(0.152)	(0.273)	(0.251)	(0.154)
Tobin Q	0.011	−0.168***	−0.161***	−0.127***
	(0.018)	(0.037)	(0.030)	(0.032)
ROA	−0.648*	0.870**	1.073***	1.024***
	(0.359)	(0.384)	(0.358)	(0.300)
Leverage	0.374	0.706*	0.595	0.536*
	(0.270)	(0.416)	(0.433)	(0.310)
Cash/Asset	0.074	0.676	0.305	0.244
	(0.167)	(0.445)	(0.486)	(0.428)
KZ index	0.009	0.024**	0.015	0.015*
	(0.008)	(0.011)	(0.011)	(0.008)
InstOwn	−0.808***	−1.084***	−0.905***	−0.888***
	(0.245)	(0.405)	(0.287)	(0.223)
(1-HHI)	−3.489***	0.842	1.420	0.452
	(0.985)	(0.704)	(1.355)	(1.451)
$(1\text{-HHI})^2$	3.206***	−1.245*	−1.582	−0.702
	(0.953)	(0.671)	(1.190)	(1.295)
Firm FEs	Y	N	Y	Y
Year FEs	Y	Y	Y	Y
Observations	84,620	86,341	86,341	86,341
Panel B: Citation count model				
R&D/Asset	−0.125	2.094***	2.299***	2.262***
	(0.252)	(0.219)	(0.256)	(0.113)
ln(ME)	0.317***	1.107***	0.984***	0.987***
	(0.071)	(0.070)	(0.149)	(0.044)
ln(1+Age)	0.593**	0.094	0.150	0.165*
	(0.295)	(0.146)	(0.243)	(0.092)
ln(K/L)	−0.055	−0.081	0.055	0.036
	(0.139)	(0.070)	(0.096)	(0.031)
R&D missing	−0.161	−2.027***	−2.660***	−2.486***
	(0.172)	(0.404)	(0.872)	(0.175)

Table IA.4 – continued from previous page

	(1) FE Poisson (Firm, Year)	(2) Poisson (Year)	(3) PRL Poisson (Firm, Year)	(4) DML Poisson (Firm, Year)
Tobin Q	0.005	−0.106**	−0.128**	−0.103***
	(0.019)	(0.047)	(0.059)	(0.028)
ROA	−0.549	1.282***	1.371***	1.595***
	(0.353)	(0.358)	(0.480)	(0.242)
Leverage	0.321	0.455	0.598	0.672***
	(0.321)	(0.451)	(0.549)	(0.255)
Cash/Asset	0.211	1.072**	0.471	0.651*
	(0.216)	(0.547)	(0.869)	(0.334)
KZ index	−0.000	0.022**	0.010	0.013*
	(0.006)	(0.010)	(0.017)	(0.007)
InstOwn	−0.909***	−1.410***	−1.096**	−1.145***
	(0.230)	(0.468)	(0.469)	(0.184)
(1-HHI)	−3.482***	1.083	1.053	2.004**
	(0.943)	(0.853)	(2.107)	(0.855)
$(1\text{-HHI})^2$	3.272***	−1.229	−1.053	−1.856**
	(0.897)	(0.780)	(1.888)	(0.746)
Firm FEs	Y	N	Y	Y
Year FEs	Y	Y	Y	Y
Observations	84,620	86,341	86,341	86,341

This table presents the full Poisson regression results of innovation on R&D intensity using different estimation methods, including Post-Regularization LASSO (PRL) Poisson and Double Machine Learning (DML) Poisson estimates. Column (1) reports fixed effects (FE) Poisson estimates with firm and year fixed effects. Column (2) presents Poisson estimates with year fixed effects. Columns (3) and (4) present PRL Poisson and DML Poisson estimates with firm and year fixed effects, respectively. The standard errors are reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively.

Table IA.5: Innovation to R&D Poisson regression for patenting firms

	(1) FE Poisson (Firm, Year)	(2) Poisson (Year)	(3) PRL Poisson (Firm, Year)	(4) DML Poisson (Firm, Year)
Panel A: Patent count model				
A.1 R&D scaled by total asset				
R&D/Asset	−0.208	1.663***	1.937	1.752***
	(0.238)	(0.172)	(1.320)	(0.089)
Selected ratio			10.44%	20.86%
A.2 R&D scaled by market equity				
R&D/ME	0.222	1.254***	1.429***	1.440***
	(0.166)	(0.156)	(0.089)	(0.070)
Selected ratio			8.53%	18.35%
A.3 R&D stock				
ln(1+R&D)	0.092***	0.746***	0.604**	0.667***
	(0.033)	(0.068)	(0.280)	(0.065)
Selected ratio			9.29%	21.60%
Control variables	Y	Y	Y	Y
Firm FEs	Y	N	Y	Y
Year FEs	Y	Y	Y	Y
Observations	45,679	45,913	45,913	45,913
Panel B: Citation count model				
B.1 R&D scaled by total asset				
R&D/Asset	−0.111	1.480***	1.575***	1.590***
	(0.231)	(0.188)	(0.144)	(0.082)
Selected ratio			10.30%	19.89%
B.2 R&D scaled by market equity				
R&D/ME	0.240	1.311***	1.449***	1.508***
	(0.174)	(0.156)	(0.110)	(0.117)
Selected ratio			9.73%	18.49%
B.3 R&D stock				
ln(1+R&D)	0.136***	0.746***	0.654***	0.587**
	(0.050)	(0.099)	(0.112)	(0.276)
Selected ratio			8.53%	21.32%
Control variables	Y	Y	Y	Y
Firm FEs	Y	N	Y	Y
Year FEs	Y	Y	Y	Y
Observations	45,679	45,913	45,913	45,913

This table presents the Poisson regression results of innovation on R&D measures for patenting firms using different estimation methods. Column (1) reports fixed effects (FE) Poisson estimates with firm and year fixed effects. Column (2) presents Poisson estimates with year fixed effects. Columns (3) and (4) present Post-Regularization LASSO (PRL) Poisson and Double Machine Learning (DML) Poisson estimates with firm and year fixed effects, respectively. The standard errors are clustered at the firm level and reported in parentheses. *, **, and *** denote statistical significance at the 10%, 5%, and 1% levels, respectively. The selected ratio is the number of selected year and firm dummies divided by the total number of year and firm dummies.