Revisiting the Missing R&D-Patent Relation: Challenges and Solutions for Firm Fixed Effects Models

Hui-Ching Chuang	National Taipei University
Po-Hsuan Hsu	National Tsing Hua University
Chung-Ming Kuan	National Taiwan University
Jui-Chung Yang	National Taiwan University

The prevailing use of firm fixed effects

- A common practice in economics, finance, and accounting studies: firm fixed effects in regression models.
- FE regression absorbs the influences of individual-specific, unobservable, and time-invariant effects.
- We argue that, without theoretical modeling or appropriate econometric designs, prevailing use of fixed effects may
  - failing to identify the effect of the persistent variable of interest.
- The R&D-patent relation is perhaps the most intuitive relation in economics.
  - More R&D input, more patent output
  - We conduct a comprehensive literature from 200+ papers based on the survey papers of Ederer and Manso (2011), He and Tian (2018), Lerner and Seru (2022) and author's reading list.
  - Statistics suggests that 40% to 50% estimates are insignificant or even have negative values on R&D.

Example: Luong et al. (2017, JFQA)

TABLE 2
Baseline Regressions

Table 2 reports the regressions of firm innovation on institutional ownership. Columns 1 and 2 (3 and 4) show the pooled ordinary least squares (OLS) (Firm fixed effects) regression results. The dependent variable is shown as the column heading in columns 1–4. The main independent variable is foreign institutional ownership (FIO). All explanatory variables are lagged by 1 year. Variable definitions are in Appendix B. Standard errors are clustered at the firm level and reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Variables	ln(PATENT) 1	ln(CITEPAT) 2	ln(PATENT) 3	ln(CITEPAT) 4
FIO	0.010***	0.014***	0.008***	0.011***
	(0.003)	(0.004)	(0.003)	(0.004)
DIO	−0.010***	−0.012***	−0.001	−0.001
	(0.002)	(0.003)	(0.001)	(0.002)
INSIDE	−0.072	−0.054	0.062*	0.084*
	(0.063)	(0.070)	(0.032)	(0.044)
ln(AGE)	0.062**	0.062*	0.086**	0.118**
	(0.029)	(0.032)	(0.037)	(0.049)
HHI	0.396	0.400	−0.170	0.152
	(0.292)	(0.326)	(0.274)	(0.347)
HHI^2	−0.277	−0.280	0.152	0.050
	(0.277)	(0.307)	(0.241)	(0.290)
RD	2.267***	2.637***	0.054	−0.232
	(0.238)	(0.288)	(0.132)	(0.215)
CAPEX	2.313***	2.913***	0.378***	0.519***
	(0.269)	(0.308)	(0.134)	(0.184)
PPE	−0.232**	−0.192	−0.082	−0.095
	(0.116)	(0.129)	(0.086)	(0.114)
LEV	−0.365***	−0.453***	−0.132**	−0.185***
	(0.097)	(0.104)	(0.057)	(0.070)
ROA	−0.616***	−0.850***	−0.037	−0.169
	(0.144)	(0.165)	(0.079)	(0.108)
Year fixed effects	Yes	Yes	Yes	Yes
Firm fixed effects	No	No	Yes	Yes
Industry fixed effects	Yes	Yes	No	No
Country fixed effects	Yes	Yes	No	No

Dependent variables: innovation output

ln(PATENT): Natural logarithm of the number of patents filed by each firm in a year plus 1.

ln(CITEPAT): Natural logarithm of the number of citations received by each firm's patents in a year plus 1.

Our focus:

RD: Research and development expenditures scaled by total assets.

How come R&D does not explain patents?

Our explanations

- OLS allows us to understand R&D's explanatory power for total variations of patents (= <u>cross-sectional/between-firm</u> variations + <u>time-series/within-firm</u> variations)
- FE models absorb all cross-sectional/between-firm variations in patents
  - An analogy: a high (low) tech firm's R&D and patents are persistently high (low). Thus, cross-sectional variation could be more important than time-series variation (Hausman et al., 1984; Hall et al., 2005).
  - However, FE models eliminate all cross-sectional variations in firms' patents – so R&D role is missing
  - So, the estimation results of FE models only tell us R&D's explanatory power for a firm's time-series variations in patents

Simulation Study

- Innovation outcome (patent, citations) equation:
  - $y_{i,t}$ negative binominal distribution with conditional mean

$$E\left(y_{i,t} \middle| z_{i,t}, x_{i,t}\right) = \exp\left(\beta z_{i,t} + x_{i,t}\right)$$

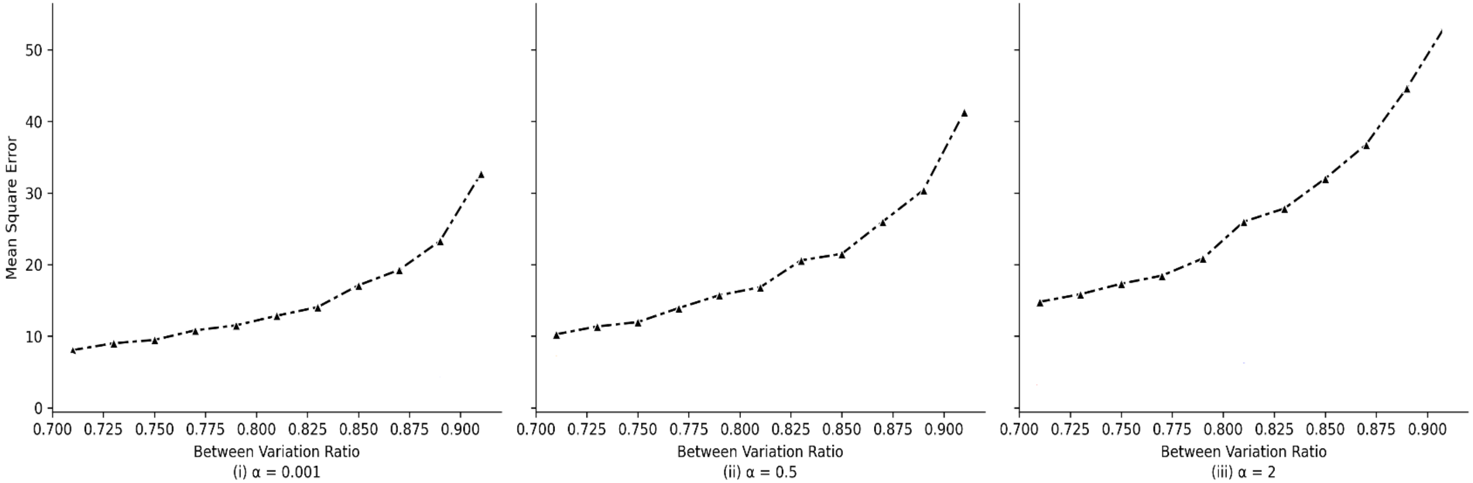
  - and a over-dispersion parameter α, larger value corresponds to a greater dispersion.
- R&D equation:

$$z_{i,t} = 0.5\, \eta_i + 0.5\, v_{i,t}$$

  - η_i : time-invariant firm-fixed effect drawn from $N\left(0, \sigma_\eta\right)$
  - $v_{i,t}$: firm and time varying component drawn from $N(0, \sigma_v)$
    - Between Ratio $= \frac{\sigma_\eta^2}{\sigma_\eta^2 + \sigma_v^2}$
- $x_{i,t}$ drawn from $N(0,2)$, and $N = 500, T = 10$, Monte Carlo replicate 10,000.

Simulation Study (cont'd)

– Mean Square Error (FE)



Econometric Tools

- A lack of appropriate econometric tools to address the issue for more reliable statistical inferences.
- Not to include firm fixed effects (Baltagi et al., 2000; Hall et al., 2005; Noel and Schankerman, 2013; Pesaran and Zhou, 2018) may introduce alternative biases.

- Our propositions and contributions:

1. Adjusted Hausman and Taylor ("adj-HT" 1981) method
2. Machine learning
   - Post-Regularization LASSO (**PRL**)
   - Double-machine learning (**DML**)

Overview: OLS, FE, adjHT, PRL, and DML

$$Innov_{i,t+1} = \beta_0 + \beta_{R\&D} R\&D_{i,t} + \boldsymbol{\beta}_2 \boldsymbol{X}_{i,t} + \sum_{s \in \boldsymbol{S}} \alpha_s dummy_{s,i} + u_{i,t}$$

- OLS includes none of the firm dummies, i.e., $S = \emptyset$.
- FE includes all of the firm dummies, i.e., $S = \{1, \cdots, N\}$.
- Adjusted HT uses the demeaned $X_{i,t}$ and demeaned R&D to construct the moment conditions in GMM estimation for $\beta_{R\&D}$
- PRL and DML select some of the firm dummies, i.e., $\boldsymbol{S} \in \{1, \cdots, N\}$ while keep the valid inference of $\beta_{R\&D}$.
  - Intuitively, since important dummies have been selected to control for, we prevent the omitted-variable bias.
  - On the other hand, since unimportant dummies are not selected, we have better power in identifying the role of persistent R&D.

Our recommendations

1. Instead of only reporting regressions **with** firm fixed effects, please also present the results **without** firm fixed effects and discuss why the coefficient estimates vary
2. We recommend to report **R-squared** and **within R-squared** results from regressions
3. If the results from 1 and 2 are inconsistent. Consider our adjusted Hausman and Taylor, PRL and DML methods as "second opinion".
   - easy to implement by STATA (or R/Python). We make our codes available online:
     - ✓ https://github.com/hcchuang/Revisiting-the-Missing-RD-Patent-Relation_Challenges-and-Solutions-for-Firm-Fixed-Effects-Models
     - ✓ handle omitted variable issues without strict assumptions
     - ✓ enable researchers to decide exactly which firm dummies should be added in regressions.

Our proposition-1
Adjusted Hausman-Taylor methods

- Consider the simplified HT (1981, Econometrica) model
$$Y_{i,t+1} = \beta \boldsymbol{Z}_i + \boldsymbol{\beta}_2 \boldsymbol{X}_{i,t} + \alpha_i + \epsilon_{i,t}.$$
- HT allow arbitrary correlation between Z_i and α_i, and use moment conditions:
$$E[(\boldsymbol{X}_{i,t} - \overline{\boldsymbol{X}}_i)'(Y_{i,t+1} - \beta Z_i - \boldsymbol{\beta}_2 \boldsymbol{X}_{i,t})] = \mathbf{0}.$$
$$E[\boldsymbol{X}_{i,t}{}'(Y_{i,t+1} - \beta Z_i - \boldsymbol{\beta}_2 \boldsymbol{X}_{i,t})] = \mathbf{0}.$$
- Treat rarely time-varying *R&D* as Z_i, and add an extra moment condition:
  - The correlation between firm fixed effects (FEs) and R&D mainly arises from the firm's population-level R&D
  - Deviations from this level are exogenous to the FEs.
$$E[(R\&D_{i,t} - \overline{R\&D_i})(Y_{i,t+1} - \beta_{R\&D} R\&D_{i,t} - \boldsymbol{\beta}_2 \boldsymbol{X}_{i,t})] = 0.$$
- Thus, similar to HT, we can identify $\beta_{R\&D}$ by GMM, using $(\boldsymbol{X}_{i,t} - \overline{\boldsymbol{X}}_i), \boldsymbol{X}_{i,t}$, and $(R\&D_{i,t} - \overline{R\&D_i})$ to construct the moment conditions.

Our proposition-2

- Unobserved heterogeneity exists in some firms but not others.
  - Some managers are aggressive in investing in R&D and pursuing patent output, but others are not.
  - Some firms have a strong, innovation-oriented culture, while others do not.
- A smarter methodology that can select which individual firm dummies to be included is called for.
- In this paper, we proposed the second advanced machine learning method:
  1. Post-regularization LASSO (PRL, Chernozhukov et al., 2015)
  2. Double machine learning (DML, Chernozhukov et al., 2018)
  - to select individual firm dummies (and explanatory variables) in explaining firm-level patent outputs.

Post-Regularization LASSO (PRL)

- **PRL** proceeds in the following 3 steps:

➢ **Step1:** LASSO of $Innov_{i,t+1}$ on firm dummies and force small coefficients of some dummies to 0. (estimate step) Then, Post LASSO: OLS of $Innov_{i,t+1}$ on selected firm dummies, obtain the residuals, $\hat{r}_y$. (get residual step)

➢ **Step2:**

a) LASSO of $R\&D_{i,t}$ on firm dummies and force small coefficients of some dummies to 0. Then, Post LASSO: OLS of $R\&D_{i,t}$ on selected firm dummies, obtain the residuals, $\hat{r}_{R\&D}$.

b) LASSO of $X_{i,t}$ on firm dummies and force small coefficients of some dummies to 0. Then, Post LASSO: OLS of $X_{i,t}$ on selected firm dummies, obtain the residuals, $\hat{r}_X$.

➢ **Step3**: OLS of $\hat{r}_y$ on $\hat{r}_{R\&D}$, $\hat{r}_X$ and obtain the coefficient $\hat{\beta}_{R\&D,\, PRL}$.

- If a firm dummy is selected in either Step 1 or Step 2 (partialing-out/residualizing), it is informative to $Innov_{i,t+1}$ and $R\&D_{i,t}$.

Double Machine Learning (DML)

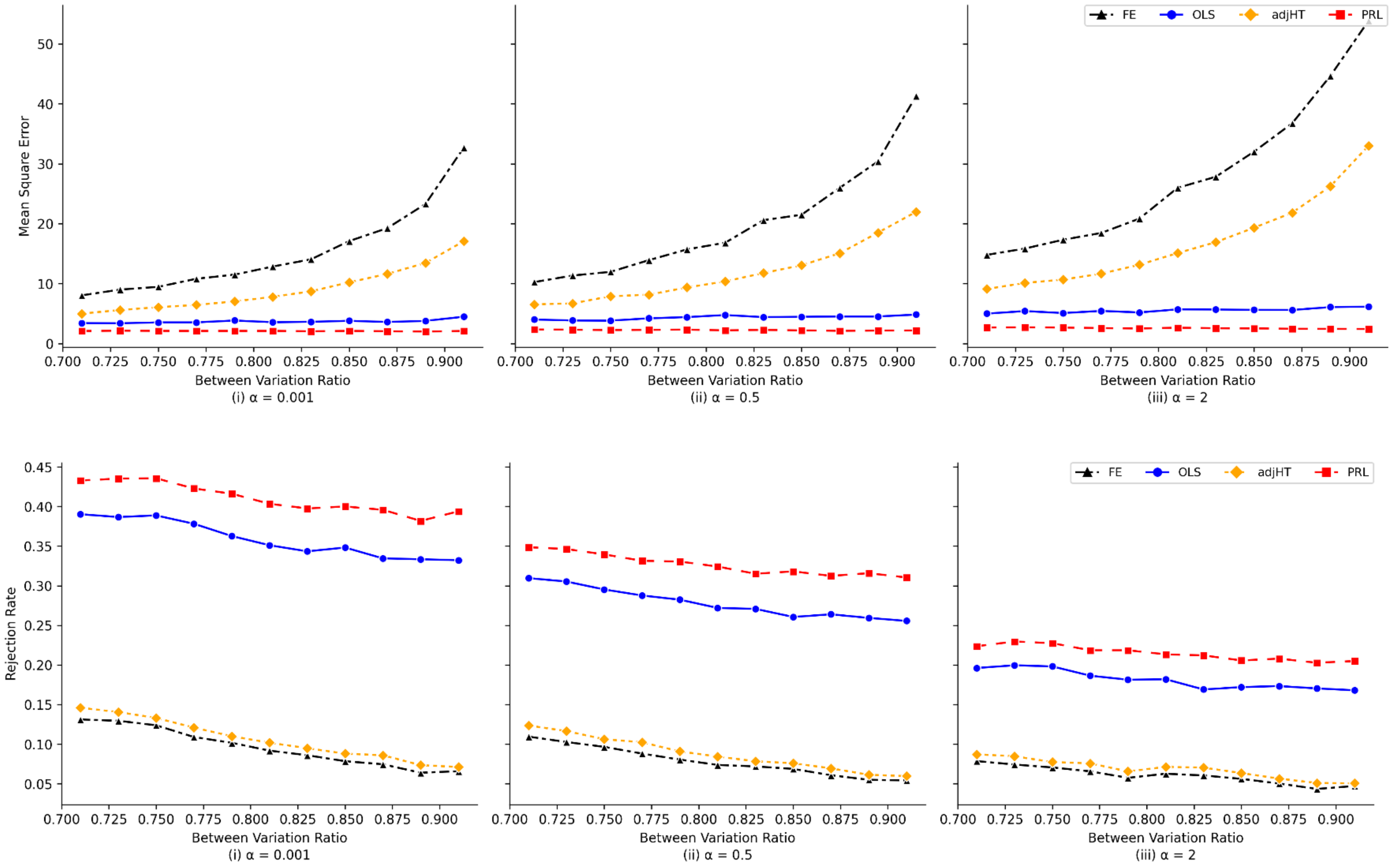
- Chernozhukov et al. (2018) propose the DML which generalizes the PRL to a general model selection (LASSO, random forests, gradient boosting, neural nets, etc.) and add the cross-fitting procedures to PRL.
- DML proceeds in the following steps:
  - splits sample into random K folds,
  - use leave-*k*-out sample in the estimate step 1&2
  - use the *k*th-fold sample to obtain the residuals for Y and R&D
  - stake all K folds residuals, use OLS to obtain $\hat{\beta}_{R\&D,DML}$.

- DML uses sample splitting to eliminates the dependence between the estimation steps, reduce the post-model-selection bias (or, errors in estimated variables) of PRL. However, as the cross-fit procedure reduces the sample size, DML also reduces the estimation efficiency.
- Yang, Chuang, and Kuan (2020, JoEcts) use DML to examine the Big N audit quality effect in the accounting literature.

PRL and DML benefits

- Both allow us to select an appropriate model that contains only important covariates, including separate firm dummies.
- PRL and DML estimator follow the standard asymptotic normal distributions which facilitate the empirical usage by assuming the sparsity condition holds (i.e., the number of strong dummies is bounded from above by an order of $\sqrt{NT}/\ln N.$)

Simulation Study (cont'd)



Sample

- We first collect the financial and accounting data of all publicly-listed firms in the U.S. from CRSP and Compustat.
- We exclude financial and utility firms (SIC in 6000-6999, and 4900-4999), and firms with negative and missing total asset and sales.
- We then collect the patent and citation data of all public firms from the PatentsView patent database that is organized by the USPTO.
- As a result, we have 86,341 firm-year observations during 1976-2000. (We also consider sample of firms with at least one patent during the sample period.)

Our baseline regressions

$$Innov_{i,t+1} = \beta_0 + \boldsymbol{\beta}_{\boldsymbol{R\&D}} R\&D_{i,t} + \boldsymbol{\beta}_2 \boldsymbol{X}_{i,t} + \sum_{s \in S} \alpha_s dummy_{s,i} + u_{i,t}$$

- $Innov_{i,t+1}$ is one of innovation measures: ln(1+Patent), ln(1+Citation), and ln(1+AdjCitation).
- $R\&D_{i,t}$ is the past five years R&D expenditures divide by total asset. We also consider R&D/ME, or ln(1+R&D) for five years for robustness.
- $\boldsymbol{X}_{i,t}$ denotes firm characteristic controls: *R&D missing dummy, capital level, ln(1+Firm age), ln(K/L), Tobin's Q, ROA, leverage, cash divide by the total asset, Institutional ownership ratio, KZ index, Herfindahl-Hirschman index, and Herfindahl-Hirschman index square.*
- We will discuss the Poisson regression later.

Diagnostic steps: Within R square and Adj-HT

	OLS (Year Dummies Only)	Fixed Effects (All Firm and Year Dummies)	Adj HT
R&D/Asset	0.593***	0.041	0.220***
	(0.042)	(0.028)	(0.027)
R square	0.424	0.822	
Adj R square	0.423	0.804	
Within R square		**0.049**	
Adj Within R square		**0.049**	

- FE models, **R square** reports the regression R square where model include raw y and raw x were x include all explanatory variables and all firm dummies.
- FE models : **Within R square** is the R square from the regression of the de-entity-meaned (i.e, within transformation) of y and de-meaned x.

Patent regression: PRL and DML results

PRL

	OLS (Year Dummies Only)	PRL (Firm and Year Dummies)	Fixed Effects (All Firm and Year Dummies)
R&D/Asset	0.593***	0.199***	0.041
	(0.042)	(0.018)	(0.028)
Number of dummies		11,570	
Number of selected dummies		1,241 (10.73%)	

DML

	OLS (Year Dummies Only)	DML (Firm and Year Dummies)	Fixed Effects (All Firm and Year Dummies)
R&D/Asset	0.593***	0.213***	0.041
	(0.042)	(0.014)	(0.028)
Number of dummies		1,1570	
Number of selected dummies		1,737 (15.01%)	

Citation regression: PRL and DML results

PRL

	OLS (Year Dummies Only)	PRL (Firm and Year Dummies)	Fixed Effects (All Firm and Year Dummies)
R&D/Asset	1.396***	1.397***	-0.051
	(0.084)	(0.083)	(0.068)
Number of dummies		11,570	
Number of selected dummies		525 (4.54%)	

DML

	OLS (Year Dummies Only)	DML (Firm and Year Dummies)	Fixed Effects (All Firm and Year Dummies)
R&D/Asset	1.396***	1.364***	-0.051
	(0.084)	(0.050)	(0.068)
Number of dummies		11,570	
Number of selected dummies		947 (8.18%)	

Adjusted-Citation regression: PRL and DML results

PRL

	OLS (Year Dummies Only)	PRL (Firm and Year Dummies)	Fixed Effects (All Firm and Year Dummies)
R&D/Asset	0.590***	0.210***	0.033
	(0.045)	(0.019)	(0.031)
Number of dummies		11,570	
Number of selected dummies		1,194 (10.32%)	

DML

	OLS (Year Dummies Only)	DML (Firm and Year Dummies)	Fixed Effects (All Firm and Year Dummies)
R&D/Asset	0.590***	0.198***	0.033
	(0.045)	(0.015)	(0.031)
Number of dummies		11,570	
Number of selected dummies		1,882 (16.27%)	

PRL and DML results

- The coefficients on R&D input are statistically significant and that their economic magnitude is much closer to those from OLS models without firm fixed effects (than those with firm fixed effects).
- PRL and DML select about 10% to 20% of firm dummies to be included in regression models -- the bias from adding all firm dummies overpowers the bias from not adding any at all (the consequence is an insignificant R&D coefficient)
- These results, together with prior analyses, suggest that most firm dummies do not play a crucial role.
- To recap: FE model = 0.041 (insig.)
- OLS = 0.593,
- adj-HT = 0.220,
- PRL = 0.199, DML = 0.213

STATA code

- To implement adjusted Hausman and Taylor:

```
ivregress gmm y z x (z = demean_z demean_x),
      wmatrix(cluster firmID)
```

- To implement PRL

```
poregress y z x, controls(i.firmID)
      vce(cluster firmID)
```

- To implement DML

```
xporegress y z x, controls(i.firmID)
      vce(cluster firmID) xfolds(#folds)
```

Robustness

- Alternative R&D measures
  – Tested R&D/ME, and Ln(1+R&D) in addition to R&D/AT,
- Patenting firms
  – Excluded firms without any patent for during its sample period.
- Handling missing R&D values
  – Remove firm-year observations with missing R&D
- Alternative specifications in HT, PRL, and DML methods
  – Different fold count from two to five in DML method

Poisson regression

$$E(Innov_{i,t+1} \,|R\&D_{i,t}, x_{i,t}) =$$
$$\exp(\beta_0 + \beta_{R\&D} R\&D_{i,t} + \boldsymbol{\beta}_2 X_{i,t} + \textstyle\sum_{s \in \boldsymbol{S}} \alpha_s dummy_{s,i})$$

- Poisson regression includes none of the firm dummies, i.e., $\boldsymbol{S} = \emptyset$.
- Poisson fixed effect regression includes all of the firm dummies, i.e., $\boldsymbol{S} = \{1, \cdots, N\}$.
- Adjusted Hausman-Taylor uses demeaned $X_{i,t}$ and demeaned $R\&D_{i,t}$ in GMM to identify $\beta_{R\&D}$ of the rarely time-varying R&D.
- PRL Poisson (Belloni, Chernozhukov and Wei, 2016, JBES) and DML select some of the firm dummies, i.e., $\boldsymbol{S} \in \{1, \cdots, N\}$.
  - PRL Poisson proceeds in the similar fashion as PRL, except it uses the post LASSO Poisson regression in Step 1 and use GMM in Step 3.
- DML follows the PRL Poisson steps with cross-fitting.

PRL Poisson and DML

Patent Counts

	Poisson	FE Poisson	PRL Poisson	DML
	Year Dummies only	All Firm and Year Dummies	Firm and Year Dummies	Firm and Year Dummies
R&D/Asset	2.305***	-0.248	2.407***	2.312***
	(0.187)	(0.255)	(0.161)	(0.120)
Number of dummies			11,570	11,570
Number of selected dummies			1,218 (10.53%)	2,484 (21.47%)

Citation Counts

	Poisson	FE Poisson	PRL Poisson	DML
	Year Dummies only	Firm and Year Dummies	Firm and Year Dummies	Firm and Year Dummies
R&D/Asset	2.094***	-0.125	2.299***	2.262***
	(0.219)	(0.252)	(0.256)	(0.113)
Number of dummies			11,570	11,570
Number of selected dummies			1,226 (10.51%)	2,426 (20.97%)

Our recommendations

1. Instead of only reporting regressions **with** firm fixed effects, please also present the results **without** firm fixed effects and discuss why the coefficient estimates vary
2. We recommend to report **within R-squared** and results from regressions without firm fixed effects
3. If the results from 1 and 2 are inconsistent. Consider our adjusted Hausman and Taylor, PRL and DML methods as "second opinion".
   - easy to implement by STATA (or R/Python). We make our codes available online:
     - ✓ https://github.com/hcchuang/Revisiting-the-Missing-RD-Patent-Relation_Challenges-and-Solutions-for-Firm-Fixed-Effects-Models
     - ✓ handle omitted variable issues without strict assumptions
     - ✓ enable researchers to decide exactly which firm dummies should be added in regressions.

Our contributions

- Corporate finance 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 as our lab.
  - More importantly, we offer two feasible and ready-to-use methodologies to enable corporate finance researchers 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).

Our contributions (Cont.)

- We add to modern machine learning techniques in corporate finance research, for the selection of relevant covariates (e.g., Chinco et al., 2019; Feng et al., 2020; Gu et al., 2020; Erel et al., 2021).
- This study also adds to the economics literature by supporting and 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).

Thank you!

Questions? Comments?

hcchuang@gm.ntpu.edu.tw (Hui-Ching Chuang)

Chuang, Hui-Ching and Hsu, Po-Hsuan and Kuan, Chung-Ming and Yang, Jui-Chung, Revisiting the Missing R&D-Patent Relation: Challenges and Solutions for Firm Fixed Effects Models
(2024). Available at SSRN: https://ssrn.com/abstract=4636846

https://github.com/hcchuang/Revisiting-the-Missing-RD-Patent-Relation_Challenges-and-Solutions-for-Firm-Fixed-Effects-Models

Selected Top 30 Firms

1	INTL BUSINESS MACHINES	11	XEROX HOLDINGS CORP	21	FORD MOTOR CO
2	LUCENT TECHNOLOGIES	12	AT&T INC	22	HONEYWELL INTERNATIONAL INC
3	GENERAL ELECTRIC	13	TEXAS INSTRUMENTS INC	23	CBS CORP -OLD
4	APTIV PLC	14	3M CO	24	RAYTHEON TECHNOLOGIES CORP
5	EASTMAN KODAK	15	RCA CORP	25	PROCTER & GAMBLE CO
6	MOTOROLA SOLUTIONS	16	BROADCOM CORP	26	SUN MICROSYSTEMS INC
7	GENERAL MOTORS CO	17	NORTH AMERICAN PHILIPS CORP	27	QUALCOMM INC
8	DU PONT (E I) DE NEMOURS	18	EXXON MOBIL CORP	28	MOBIL CORP
9	AT&T CORP	19	HP INC	29	CONOCOPHILLIPS
10	DUPONT DE NEMOURS INC	20	MERCK & CO	30	MICRON TECHNOLOGY INC

Alternative R&D measures: Patent regression

	Fixed Effects (Firm and Year Dummies)	OLS (Year Dummies)	adjHT (Year Dummies)	PRL (Firm and Year Dummies)	DML (Firm and Year Dummies)
R&D/ME	0.026	0.502***	0.139***	0.112***	0.123***
	(0.018)	(0.035)	(0.018)	(0.015)	(0.012)
Ln(R&D)	0.023***	0.140***	0.071***	0.032***	0.035***
	(0.003)	(0.004)	(0.003)	(0.001)	(0.001)

Firm cluster standard errors in parentheses. *p<0.1, **p<0.05, and ***p<0.01.
We suppress the year and firm characteristics variables to save space.

Alternative R&D measures: Citation regression

	Fixed Effects (Firm and Year Dummies)	OLS (Year Dummies)	adjHT (Year Dummies)	PRL (Firm and Year Dummies)	DML (Firm and Year Dummies)
R&D/ME	-0.024	1.040***	0.286***	1.015***	1.001***
	(0.040)	(0.064)	(0.040)	(0.064)	(0.038)
Ln(R&D)	0.036***	0.286***	0.176***	0.285***	0.285***
	(0.007)	(0.006)	(0.006)	(0.006)	(0.003)

Firm cluster standard errors in parentheses. *p<0.1, **p<0.05, and ***p<0.01.
We suppress the year and firm characteristics variables to save space.

Alternative R&D measures: AdjCitation regression

	Fixed Effects (Firm and Year Dummies)	OLS (Year Dummies)	adjHT (Year Dummies)	PRL (Firm and Year Dummies)	DML (Firm and Year Dummies)
R&D/ME	0.021	0.506***	0.135***	0.119***	0.117***
	(0.019)	(0.037)	(0.019)	(0.016)	(0.012)
Ln(R&D)	0.022***	0.143***	0.071***	0.035***	0.035***
	(0.003)	(0.004)	(0.003)	(0.001)	(0.001)

Firm cluster standard errors in parentheses. *p<0.1, **p<0.05, and ***p<0.01.
We suppress the year and firm characteristics variables to save space.

Patenting firms (Observation: 45,913)

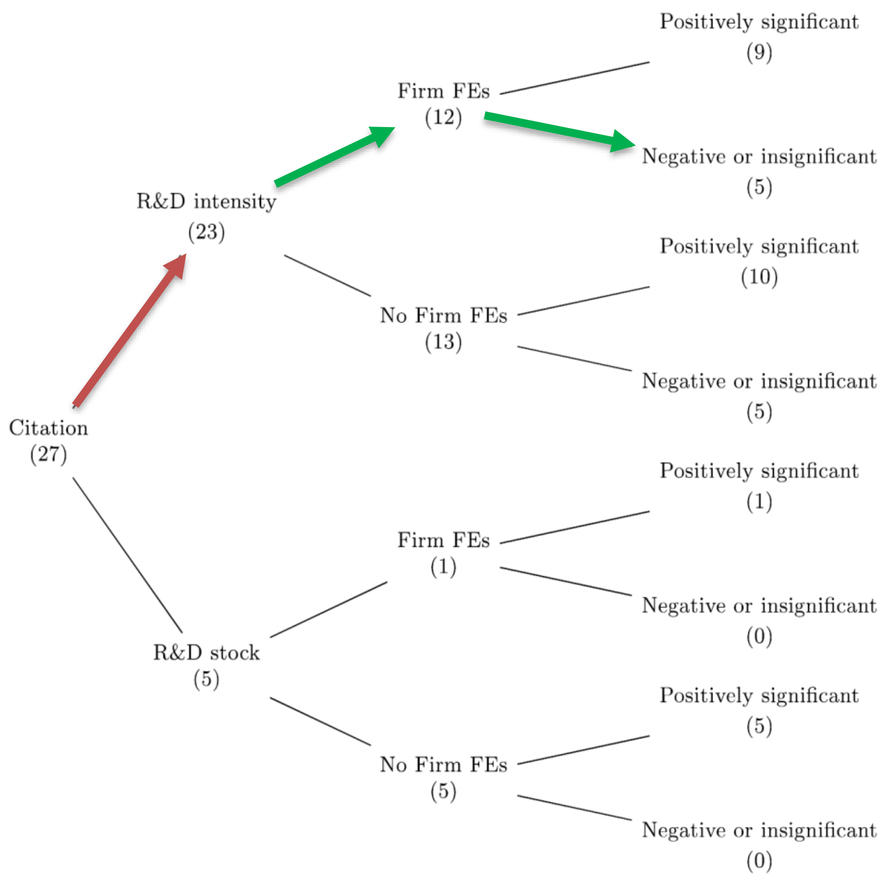
	Fixed Effects (All Firm and Year Dummies)	OLS (Year Dummies)	adjHT (Year Dummies)	PRL (Firm and Year Dummies)	DML (Firm and Year Dummies)
Patent regression					
R&D/AT	0.035	0.448***	0.161***	0.406***	0.399***
	(0.031)	(0.045)	(0.030)	(0.044)	(0.025)
Citation regression					
R&D/AT	-0.061	0.777***	0.317***	0.738***	0.740***
	(0.079)	(0.087)	(0.075)	(0.084)	(0.054)
AdjCitation regression					
R&D/AT	0.033	0.419***	0.175***	0.369***	0.377***
	(0.036)	(0.049)	(0.034)	(0.047)	(0.028)

Firm cluster standard errors in parentheses. *p<0.1, **p<0.05, and ***p<0.01.
We suppress the year and firm characteristics variables to save space.

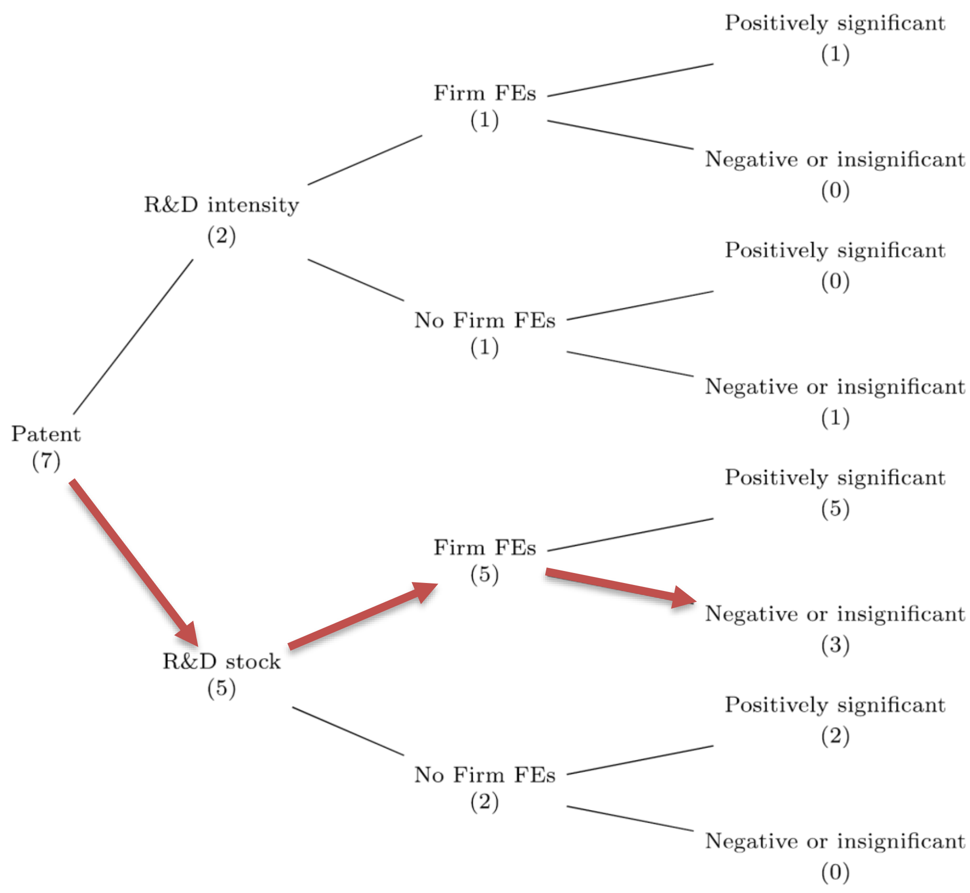
Summary statistics

	Obs.	Mean	Median	Std. dev.	P25	P75
ln(1+Patent)	86,341	0.53	0.00	1.11	0.00	0.69
ln(1+Citation)	86,341	1.25	0.00	2.27	0.00	2.20
ln(1+AdjCitation)	86,341	0.54	0.00	1.16	0.00	0.34
R&D/Asset	86,341	0.13	0.00	0.26	0.00	0.14
ln(ME)	86,341	11.13	10.95	2.07	9.60	12.51
R&D Missing Dummy	86,341	0.43	0.00	0.50	0.00	1.00
ln(1+Age)	86,341	2.48	2.48	0.75	1.95	3.09
ln(K/L)	86,341	10.00	9.85	1.29	9.18	10.64
Tobin Q	86,341	1.76	1.22	1.66	0.94	1.85
ROA	86,341	0.10	0.13	0.18	0.06	0.19
Leverage	86,341	0.23	0.21	0.18	0.07	0.35
Cash/Asset	86,341	0.14	0.07	0.17	0.02	0.18
KZ Index	86,341	-3.42	-0.54	10.77	-3.48	1.03
Institutional ownership	86,341	0.23	0.15	0.24	0.01	0.40
HH Index	86,341	0.25	0.20	0.20	0.11	0.30
HH Index Square	86,341	0.10	0.04	0.18	0.01	0.09

Our survey of the corporate innovation literature
Least square approach on Citation (paper #)



Our survey of the corporate innovation literature Poisson and negative binominal approach on Patent



Our survey of the corporate innovation literature Poisson and negative binominal approach on Citation