

“Revisiting the Role of Collaboration in Creating Breakthrough Inventions”

(Online Supplement)

I) Identifying number of chunks in a utility patent

What follows is pseudo-code for the algorithm used to identify the number of chunks in a utility patent. The algorithm uses Stanford CoreNLP (9 December 2015 version) with Perl (5.14.2.1 v7) as a wrapper.

1. For each claim, identify whether it is independent or dependent by using a regular-expression search for the term “claim #”, where # signifies any number.
2. If the claim is independent:
 - a. Break down the claim into individual words.
 - b. Tag each word in the claim with its type: noun, adjective, verb, etc.
 - c. Identify the root for each word (e.g., reduce the word “surfaces” to “surface”).
 - d. Identify the subject matter for the claim based on first-occurring noun phrase with adjectives (e.g., “outer free end”, “elastic strip”, “bioprosthetic mitral valve replacement”).
3. If the claim is dependent:
 - a. Take the sentence starting after “claim #”—for example, from the text “A mitral valve replacement as in claim 1, wherein the base is...” use only “wherein the base is...”.
 - b. Return to steps a.–d. of Step 2.
4. Count the number of distinct subject matter in the patent.

II) Additional analyses:

This supplement also provides tables that summarize results of the following analyses:

- Interactions with invention “size” and using alternative measures of decomposability: Table 6
- Models (logit with random effects, linear probability, and quantile) that use the full sample: Table 7.
- *Effects on different quantiles*
 - Examining the effects across the entire distribution of outcomes using quantile regressions: utility patents, Table 8; design patents, Table 9
 - Redefining “breakthrough” as the top 1st percentile (in citations): Table 10, Models 38 and 39
 - Predicting failures using the bottom 5th percentile (in citations): Table 10, Models 40 and 41

- *Models used for additional robustness tests*
 - Pooling design data and utility data: Table 11, Models 42 and 43
 - Controlling for team characteristics: Table 11, Models 44 and 45
 - Omitting all collinear variables: Table 11, Models 46 and 47
 - Inclusion of the interaction between *Sole* and *Fees*: Table 11, Models 48 and 49
 - Incorporating product class (resp., firm heterogeneity): Table 12, Models 50 and 51 (resp., Models 52 and 53)
 - Using the alternative inventor ID of Li et al. (2014): Table 12, Model 54
 - Using a random-inventor instead of a lead-inventor benchmark: Table 12, Models 55 and 56
 - First-stage regression results for instrumental variable estimation: Table 13, Models 57 and 58

Table 6: Interactions with invention “size” and using alternative measures of decomposability

| Dataset Model | Utility | Utility | Utility |
|--|--|-----------------------------------|-------------------------------|
| | Interaction with “size” Model 19 | Gini-Simpson index Model 20 | Modularity Q Model 21 |
| <i>Sole</i> | -0.15*** (0.02) | -0.13*** (0.02) | -0.15*** (0.02) |
| <i>LogPastCollab(dm)</i> | 0.02 (0.02) | -0.04* (0.02) | 0.02 (0.02) |
| <i>LogExp</i> | -0.52*** (0.03) | -0.34*** (0.03) | -0.52*** (0.03) |
| <i>LogExpDiversity</i> | 0.27*** (0.03) | 0.10** (0.04) | 0.27*** (0.03) |
| <i>Assigned</i> | 0.03 (0.03) | 0.04 (0.03) | 0.03 (0.03) |
| <i>LogChunks(dm)</i> | 0.19*** (0.02) | | |
| <i>GSI(dm)</i> | | 0.09+ (0.05) | |
| <i>Q(dm)</i> | | | 0.47*** (0.08) |
| <i>LogClaims</i> | 0.17*** (0.02) | 0.28*** (0.01) | 0.27*** (0.01) |
| <i>LogWords</i> | 0.08*** (0.01) | 0.09*** (0.01) | 0.09*** (0.01) |
| <i>LogFigures</i> | 0.30*** (0.01) | 0.29*** (0.01) | 0.30*** (0.01) |
| <i>LogAssignedClasses</i> | 0.33*** (0.01) | 0.35*** (0.01) | 0.33*** (0.01) |
| <i>LogCitedClasses</i> | 0.02 (0.02) | -0.06** (0.02) | 0.01 (0.02) |
| <i>LogPatentCites</i> | 0.09*** (0.01) | 0.13*** (0.01) | 0.09*** (0.01) |
| <i>LogNonpatentCites</i> | 0.05*** (0.01) | 0.04*** (0.01) | 0.04*** (0.01) |
| <i>Fees</i> | 0.38*** (0.03) | 0.33*** (0.03) | 0.37*** (0.03) |
| <i>Sole</i> × <i>LogChunks(dm)</i> | -0.08* (0.03) | | |
| <i>Sole</i> × <i>LogClaims(dm)</i> | 0.01 (0.03) | | |
| <i>Sole</i> × <i>LogWords(dm)</i> | 0.03 (0.02) | | |
| <i>Sole</i> × <i>GSI(dm)</i> | | -0.17* (0.07) | |
| <i>Sole</i> × <i>Q(dm)</i> | | | -0.36** (0.13) |
| <i>Sole</i> × <i>LogPastCollab(dm)</i> | 0.06*** (0.01) | 0.07*** (0.01) | 0.06*** (0.01) |
| Filing-year FE | Yes | Yes | Yes |
| Lead-inventor FE | Yes | Yes | Yes |
| <i>N</i> | 368,899 | 415,249 | 368,899 |
| Log-likelihood | -93,899 | -107,873 | -93,944 |
| χ^2 | 12,288 | 11,712 | 12,280 |

Notes: Standard errors (in parentheses) are clustered by lead inventor. (*dm*) = de-meaned; FE = fixed effects.

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 7: Approaches using the entire dataset

| Dataset | <i>Utility</i> | <i>Design</i> | <i>Utility</i> | <i>Design</i> | <i>Utility</i> | <i>Design</i> |
|---|--------------------|--------------------|-----------------------|-----------------------|--------------------|--------------------|
| Specification: | Logit with RE | Logit with RE | Linear probability | Linear probability | Quantile | Quantile |
| Variable | Model 22 | Model 23 | Model 24 | Model 25 | Model 26 | Model 27 |
| <i>Sole</i> | -0.16*** (0.02) | 0.06 (0.04) | -0.008*** (0.001) | 0.005 (0.003) | -0.05*** (0.01) | 0.01 (0.01) |
| <i>LogPastCollab(dm)</i> | 0.11*** (0.01) | 0.04 (0.04) | -0.002* (0.001) | -0.003 (0.004) | -0.03*** (0.00) | -0.06*** (0.01) |
| <i>LogExp</i> | -0.29*** (0.02) | -0.20*** (0.05) | -0.034*** (0.002) | -0.022*** (0.005) | -0.10*** (0.01) | 0.02* (0.01) |
| <i>LogExpDiversity</i> | 0.23*** (0.02) | 0.28** (0.09) | 0.025*** (0.002) | 0.026*** (0.007) | 0.32*** (0.01) | 0.29*** (0.02) |
| <i>Assigned</i> | 0.14*** (0.02) | 0.89*** (0.05) | -0.000 (0.001) | 0.007 (0.005) | 0.18*** (0.00) | 0.28*** (0.01) |
| <i>LogChunks(dm)</i> | 0.20*** (0.02) | | 0.010*** (0.001) | | 0.05*** (0.00) | |
| <i>LogClaims</i> | 0.24*** (0.01) | | 0.008*** (0.001) | | 0.12*** (0.00) | |
| <i>LogWords</i> | 0.10*** (0.01) | -0.40*** (0.11) | 0.004*** (0.000) | -0.039*** (0.008) | 0.03*** (0.00) | -0.28*** (0.03) |
| <i>LogFigures</i> | 0.35*** (0.01) | 0.12* (0.05) | 0.014*** (0.000) | 0.015*** (0.004) | 0.15*** (0.00) | 0.05*** (0.01) |
| <i>LogAssignedClasses</i> | 0.39*** (0.01) | 0.78*** (0.06) | 0.016*** (0.001) | 0.034*** (0.005) | 0.15*** (0.00) | 0.27*** (0.02) |
| <i>LogCitedClasses</i> | 0.10*** (0.01) | -0.52*** (0.05) | 0.001+ (0.001) | -0.018*** (0.004) | -0.06*** (0.00) | -0.14*** (0.01) |
| <i>LogPatentCites</i> | 0.15*** (0.01) | 0.43*** (0.03) | 0.001 (0.001) | 0.015*** (0.003) | 0.05*** (0.00) | 0.11*** (0.01) |
| <i>LogNonpatentCites</i> | 0.13*** (0.01) | 0.17*** (0.03) | 0.001*** (0.000) | 0.005 (0.003) | 0.07*** (0.00) | 0.06*** (0.01) |
| <i>Fees</i> | 0.46*** (0.02) | 1.54*** (0.16) | 0.022*** (0.001) | 0.011 (0.017) | 0.22*** (0.00) | 0.33*** (0.03) |
| <i>Sole</i> × <i>LogChunks(dm)</i> | -0.10*** (0.02) | | -0.006*** (0.001) | | -0.01* (0.00) | |
| <i>Sole</i> × <i>LogPastCollab(dm)</i> | 0.07*** (0.01) | 0.19*** (0.05) | 0.003*** (0.001) | 0.010** (0.004) | 0.03*** (0.00) | 0.02+ (0.01) |
| Log σ^2 | 0.93*** (0.01) | 1.12*** (0.04) | | | | |
| ρ | 0.44*** (0.003) | 0.48*** (0.01) | | | | |
| Filing-year FE | Yes | Yes | Yes | Yes | Yes | Yes |
| Lead-inventor effects | RE | RE | FE | FE | FE | FE |
| <i>N</i> | 1,603,970 | 198,265 | 1,603,970 | 198,265 | 1,603,970 | 198,265 |
| Log-likelihood | -272,670 | -34,970 | | | | |
| χ^2 | 26,023 | 1,495 | | | | |
| Within- R^2 | | | 0.03 | 0.01 | | |
| <i>F</i> | | | 246.60 | 9.72 | | |
| Quantile objective | | | | | 0.08 | 0.06 |

Notes: Standard errors (in parentheses) are clustered by lead-inventor. (*dm*) = de-measured; FE = fixed effects; RE = random effects.

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.00$

Table 8: Effect of the lone inventor assessed at different quantiles of the log-citation distribution—utility patents

| Variable | Quantile: | Q = 0.05 | Q = 0.25 | Q = 0.5 | Q = 0.75 | Q = 0.95 |
|--|-----------|-----------------|-----------------|-----------------|-----------------|-----------------|
| | | Model 28 | Model 29 | Model 30 | Model 31 | Model 32 |
| <i>Sole</i> | | -0.09*** (0.01) | -0.06*** (0.00) | -0.07*** (0.00) | -0.07*** (0.00) | -0.05*** (0.01) |
| <i>LogPastCollab(dm)</i> | | 0.03*** (0.00) | 0.00 (0.00) | -0.00*** (0.00) | -0.02*** (0.00) | -0.03*** (0.00) |
| <i>LogExp</i> | | -0.38*** (0.01) | -0.31*** (0.00) | -0.24*** (0.00) | -0.15*** (0.00) | -0.10*** (0.01) |
| <i>LogExpDiversity</i> | | -0.03*** (0.01) | -0.03*** (0.00) | 0.16*** (0.00) | 0.33*** (0.00) | 0.32*** (0.01) |
| <i>Assigned</i> | | -0.16*** (0.01) | -0.02*** (0.00) | 0.02*** (0.00) | 0.09*** (0.00) | 0.18*** (0.00) |
| <i>LogChunks(dm)</i> | | 0.12*** (0.01) | 0.09*** (0.00) | 0.09*** (0.00) | 0.09*** (0.00) | 0.05*** (0.00) |
| <i>LogClaims</i> | | 0.05*** (0.00) | 0.09*** (0.00) | 0.09*** (0.00) | 0.09*** (0.00) | 0.12*** (0.00) |
| <i>LogWords</i> | | 0.10*** (0.00) | 0.06*** (0.00) | 0.06*** (0.00) | 0.06*** (0.00) | 0.03*** (0.00) |
| <i>LogFigures</i> | | 0.16*** (0.00) | 0.16*** (0.00) | 0.16*** (0.00) | 0.16*** (0.00) | 0.15*** (0.00) |
| <i>LogAssignedClasses</i> | | 0.14*** (0.00) | 0.14*** (0.00) | 0.14*** (0.00) | 0.14*** (0.00) | 0.15*** (0.00) |
| <i>LogCitedClasses</i> | | -0.10*** (0.00) | -0.09*** (0.00) | -0.09*** (0.00) | -0.09*** (0.00) | -0.06*** (0.00) |
| <i>LogPatentCites</i> | | 0.18*** (0.00) | 0.12*** (0.00) | 0.12*** (0.00) | 0.12*** (0.00) | 0.05*** (0.00) |
| <i>LogNonpatentCites</i> | | -0.01*** (0.00) | 0.01*** (0.00) | 0.03*** (0.00) | 0.04*** (0.00) | 0.07*** (0.00) |
| <i>Fees</i> | | 0.08*** (0.00) | 0.13*** (0.00) | 0.14*** (0.00) | 0.14*** (0.00) | 0.22*** (0.00) |
| <i>Sole</i> × <i>LogChunks(dm)</i> | | -0.05*** (0.01) | -0.02*** (0.00) | -0.02*** (0.00) | -0.02*** (0.00) | -0.01* (0.00) |
| <i>Sole</i> × <i>LogPastCollab(dm)</i> | | -0.00 (0.00) | 0.02*** (0.00) | 0.01*** (0.00) | 0.01*** (0.00) | 0.03*** (0.00) |
| Filing-year FE | | Yes | Yes | Yes | Yes | Yes |
| Lead-inventor FE | | Yes | Yes | Yes | Yes | Yes |
| <i>N</i> | | 1,603,970 | 1,603,970 | 1,603,970 | 1,603,970 | 1,603,970 |
| Quantile objective | | 0.08 | 0.21 | 0.24 | 0.21 | 0.08 |

Table 9: Effect of the lone inventor assessed at different quantiles of the log-citation distribution—design patents

| Variable | Quantile: | Q = 0.05 | Q = 0.25 | Q = 0.5 | Q = 0.75 | Q = 0.95 |
|--|-----------|-----------------|-----------------|-----------------|-----------------|-----------------|
| | | Model 33 | Model 34 | Model 35 | Model 36 | Model 37 |
| <i>Sole</i> | | 0.01 (0.02) | 0.01 (0.01) | 0.01*** (0.00) | 0.01+ (0.01) | 0.01 (0.01) |
| <i>LogPastCollab(dm)</i> | | 0.09*** (0.01) | 0.05*** (0.01) | 0.01*** (0.00) | -0.03*** (0.01) | -0.06*** (0.01) |
| <i>LogExp</i> | | -0.24*** (0.01) | -0.19*** (0.01) | -0.08*** (0.00) | -0.00 (0.01) | 0.02* (0.01) |
| <i>LogExpDiversity</i> | | -0.26*** (0.02) | -0.16*** (0.01) | 0.04*** (0.01) | 0.26*** (0.01) | 0.29*** (0.02) |
| <i>Assigned</i> | | -0.24*** (0.01) | -0.05*** (0.00) | 0.04*** (0.00) | 0.16*** (0.00) | 0.28*** (0.01) |
| <i>LogWords</i> | | -0.35*** (0.03) | -0.33*** (0.00) | -0.33*** (0.00) | -0.33*** (0.00) | -0.28*** (0.03) |
| <i>LogFigures</i> | | 0.15*** (0.01) | 0.09*** (0.00) | 0.09*** (0.00) | 0.09*** (0.00) | 0.05*** (0.01) |
| <i>LogAssignedClasses</i> | | 0.43*** (0.02) | 0.35*** (0.00) | 0.35*** (0.00) | 0.35*** (0.00) | 0.27*** (0.02) |
| <i>LogCitedClasses</i> | | -0.01 (0.01) | -0.11*** (0.00) | -0.11*** (0.00) | -0.11*** (0.00) | -0.14*** (0.01) |
| <i>LogPatentCites</i> | | 0.16*** (0.01) | 0.15*** (0.00) | 0.15*** (0.00) | 0.15*** (0.00) | 0.11*** (0.01) |
| <i>LogNonpatentCites</i> | | 0.01+ (0.01) | 0.03*** (0.00) | 0.03*** (0.00) | 0.03*** (0.00) | 0.06*** (0.01) |
| <i>Fees</i> | | -0.04 (0.03) | 0.17*** (0.00) | 0.17*** (0.00) | 0.17*** (0.00) | 0.33*** (0.03) |
| <i>Sole</i> × <i>LogPastCollab(dm)</i> | | 0.01 (0.02) | 0.02* (0.01) | 0.02*** (0.00) | 0.02** (0.01) | 0.02+ (0.01) |
| Filing-year FE | | Yes | Yes | Yes | Yes | Yes |
| Lead-inventor FE | | Yes | Yes | Yes | Yes | Yes |
| <i>N</i> | | 198,265 | 198,265 | 198,265 | 198,265 | 198,265 |
| Quantile objective | | 0.07 | 0.17 | 0.20 | 0.17 | 0.06 |

Notes (Tables 8 and 9): Standard errors (in parentheses) are clustered by lead inventor; (*dm*) = de-meaned. Fixed effects (FE) are assumed to have a common mean effect on log-citations (cf. Canay 2011).

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 10: Effect of sole inventor at the top 1st percentile (*Top1*) and the bottom 5th percentile (*Bottom5*) of citations

| Data set | <i>Utility</i> | <i>Design</i> | <i>Utility</i> | <i>Design</i> |
|---------------------------------|---------------------------|-----------------|----------------------|-----------------|
| Dependent variable | <i>Top1</i> | <i>Top1</i> | <i>Bottom5</i> | <i>Bottom5</i> |
| Variable | Model 38 | Model 39 | Model 40 | Model 41 |
| <i>Sole</i> | -0.16*** (0.03) | 0.12 (0.12) | 0.14*** (0.01) | 0.01 (0.06) |
| <i>LogPastCollab(dm)</i> | -0.01 (0.04) | -0.04 (0.15) | -0.03 (0.02) | 0.02 (0.05) |
| <i>LogExp</i> | -0.58*** (0.05) | -0.39*** (0.12) | 0.48*** (0.03) | 0.21*** (0.05) |
| <i>LogExpDiversity</i> | 0.37*** (0.06) | 0.30 (0.23) | -0.30*** (0.03) | -0.01 (0.08) |
| <i>Assigned</i> | 0.01 (0.06) | 0.26 (0.17) | -0.03 (0.03) | 0.10 (0.08) |
| <i>LogChunks(dm)</i> | 0.18*** (0.04) | | -0.16*** (0.02) | |
| <i>LogClaims</i> | 0.19*** (0.03) | | -0.14*** (0.01) | |
| <i>LogWords</i> | 0.08** (0.03) | -1.09*** (0.33) | -0.13*** (0.01) | 0.66*** (0.15) |
| <i>LogFigures</i> | 0.31*** (0.02) | 0.19 (0.12) | -0.28*** (0.01) | -0.26*** (0.06) |
| <i>LogAssignedClasses</i> | 0.42*** (0.02) | 0.95*** (0.19) | -0.26*** (0.01) | -0.98*** (0.10) |
| <i>LogCitedClasses</i> | 0.13** (0.04) | -0.33** (0.12) | 0.28*** (0.02) | 0.24*** (0.06) |
| <i>LogPatentCites</i> | -0.03 (0.03) | 0.13 (0.08) | -0.33*** (0.01) | -0.37*** (0.04) |
| <i>LogNonpatentCites</i> | 0.05** (0.02) | 0.20** (0.08) | -0.06*** (0.01) | -0.12** (0.04) |
| <i>Fees</i> | 0.46*** (0.05) | 0.33 (0.61) | -0.19*** (0.03) | -0.47 (0.27) |
| <i>Sole × LogChunks(dm)</i> | -0.07 ⁺ (0.04) | | 0.03* (0.02) | |
| <i>Sole × LogPastCollab(dm)</i> | 0.08** (0.03) | 0.30** (0.12) | -0.02 (0.01) | -0.06 (0.06) |
| Filing-year FE | Yes | Yes | Yes | Yes |
| Lead-inventor FE | Yes | Yes | Yes | Yes |
| <i>N</i> | 136,056 | 18,711 | 470,958 ^a | 53,365 |
| Log-likelihood | -25,558 | -3,570 | -117,770 | -14,634 |
| χ^2 | 3,977 | 228 | 26,743 | 628 |

Notes: Because *Bottom5* models predict failure, we expect their signs to be reversed. Standard errors (in parentheses) are clustered by lead inventor. (*dm*) = de-meaned; FE = fixed effects.

^a In Model 40, identification requires that we observe the bottom tail of technology classes; hence classes with thick bottom tails (i.e., >20% of patents with no citations) are excluded.

⁺ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 11: Additional robustness models

| Data set | <i>Utility & Design</i> | | <i>Utility</i> | <i>Design</i> | <i>Utility</i> | <i>Design</i> | <i>Utility</i> | <i>Design</i> |
|--|-------------------------------|--------------------|--------------------------------------|--------------------|----------------------------------|--------------------|---|--------------------|
| Specification | Pooling design & utility data | | Controlling for team characteristics | | Omitting all collinear variables | | Interaction of <i>Sole</i> with <i>Fees</i> | |
| Variable | M.42 | M.43 | M.44 | M.45 | M.46 | M.47 | M.48 | M.49 |
| <i>Sole</i> | -0.17*** (0.01) | -0.14*** (0.01) | -0.10*** (0.02) | 0.10 (0.06) | -0.15*** (0.02) | 0.05 (0.06) | -0.15*** (0.02) | 0.04 (0.06) |
| <i>Design</i> | Absorbed in FE | | | | | | | |
| <i>Sole</i> × <i>Design</i> | 0.24*** (0.05) | 0.20*** (0.06) | | | | | | |
| <i>LogPastCollab(dm)</i> | 0.02 (0.01) | 0.01 (0.01) | -0.02 (0.03) | -0.16 (0.12) | -0.21*** (0.01) | -0.11 (0.07) | 0.02 (0.02) | 0.04 (0.07) |
| <i>LogExp</i> | -0.44*** (0.02) | -0.45*** (0.02) | -0.50*** (0.04) | -0.13 (0.10) | | | -0.52*** (0.03) | -0.30*** (0.07) |
| <i>LogExpDiversity</i> | 0.22*** (0.02) | 0.22*** (0.02) | 0.20*** (0.04) | 0.01 (0.17) | | | 0.27*** (0.03) | 0.21 (0.12) |
| <i>Assigned</i> | 0.04 (0.03) | 0.03 (0.02) | 0.03 (0.03) | 0.13 (0.11) | 0.06* (0.03) | 0.14 (0.12) | 0.03 (0.03) | 0.13 (0.11) |
| <i>LogChunks(dm)</i> | 0.17*** (0.02) | 0.19*** (0.02) | 0.19*** (0.02) | | 0.38*** (0.01) | | 0.19*** (0.02) | |
| <i>LogClaims</i> | 0.17*** (0.01) | 0.17*** (0.01) | 0.17*** (0.01) | | | | 0.17*** (0.01) | |
| <i>LogWords</i> | 0.08*** (0.01) | 0.08*** (0.01) | 0.09*** (0.01) | -0.86*** (0.17) | 0.07*** (0.01) | -0.91*** (0.17) | 0.09*** (0.01) | -0.86*** (0.17) |
| <i>LogFigures</i> | 0.30*** (0.01) | 0.30*** (0.01) | 0.30*** (0.01) | 0.27*** (0.07) | 0.31*** (0.01) | 0.27*** (0.07) | 0.30*** (0.01) | 0.27*** (0.07) |
| <i>LogAssignedClasses</i> | 0.34*** (0.01) | 0.34*** (0.01) | 0.33*** (0.01) | 0.70*** (0.09) | 0.34*** (0.01) | 0.60*** (0.09) | 0.33*** (0.01) | 0.69*** (0.09) |
| <i>LogCitedClasses</i> | -0.04* (0.02) | -0.04* (0.02) | 0.01 (0.02) | -0.38*** (0.07) | | | 0.02 (0.02) | -0.37*** (0.07) |
| <i>LogPatentCites</i> | 0.12*** (0.01) | 0.12*** (0.01) | 0.09*** (0.01) | 0.29*** (0.04) | 0.10*** (0.01) | 0.14*** (0.04) | 0.09*** (0.01) | 0.29*** (0.04) |
| <i>LogNonpatentCites</i> | 0.05*** (0.01) | 0.05*** (0.01) | 0.05*** (0.01) | 0.07* (0.04) | 0.05*** (0.01) | 0.07 (0.04) | 0.05*** (0.01) | 0.07 (0.04) |
| <i>Fees(dm)</i> | 0.38*** (0.02) | 0.38*** (0.02) | 0.38*** (0.03) | 0.41 (0.33) | | | 0.43*** (0.03) | 0.43 (0.36) |
| <i>LogTeamExp</i> | | | -0.03 (0.03) | -0.19* (0.10) | | | | |
| <i>LogTeamExpDiversity</i> | | | 0.09* (0.04) | 0.21 (0.14) | | | | |
| <i>LogTeamPastCollab</i> | | | 0.07** (0.03) | 0.24* (0.10) | | | | |
| <i>Sole</i> × <i>Fees(dm)</i> | | | | | | | -0.13*** (0.03) | -0.01 (0.31) |
| <i>Sole</i> × <i>LogChunks(dm)</i> | | -0.07*** (0.02) | -0.07*** (0.02) | | -0.07*** (0.02) | | -0.06** (0.02) | |
| <i>Sole</i> × <i>LogPastCollab(dm)</i> | | 0.07*** (0.01) | 0.04** (0.01) | 0.16* (0.06) | 0.03* (0.01) | 0.17** (0.06) | 0.07*** (0.01) | 0.18** (0.06) |
| Filing-year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Lead-inventor FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| <i>N</i> | 415,249 | 415,249 | 368,899 | 46,350 | 368,899 | 46,350 | 368,899 | 46,350 |
| Log-likelihood | -107,165 | -107,141 | -93,866 | -13,126 | -94,506 | -13,208 | -93,890 | -13,136 |
| χ^2 | 30,007 | 30,055 | 12,260 | 455 | 10,818 | 341 | 12,273 | 446 |

Notes: Standard errors (in parentheses) are clustered by lead inventor. (*dm*) = de-meanded; FE = fixed effects; **M.** = Model.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 12: Modeling technology/firm heterogeneity; testing for alternative ways to identify inventors

| Data set | <i>Utility</i> | <i>Design</i> | <i>Utility</i> | <i>Design</i> | <i>Utility</i> | <i>Utility</i> | <i>Design</i> |
|---|--------------------------------------|--------------------|-----------------------------|--------------------|------------------------|---------------------------|--------------------|
| Specification | Modeling product-class heterogeneity | | Modeling firm heterogeneity | | Alternative inventorID | Random-inventor benchmark | |
| Variable | M.50 | M.51 | M.52 | M.53 | M.54 | M.55 | M.56 |
| <i>Sole</i> | -0.17*** (0.02) | 0.00 (0.07) | -0.15*** (0.02) | 0.05 (0.07) | 0.39* (0.16) | -0.16*** (0.02) | 0.03 (0.06) |
| <i>LogPastCollab(dm)</i> | 0.02 (0.02) | 0.08 (0.08) | -0.02 (0.02) | 0.06 (0.08) | 0.02 (0.02) | 0.09*** (0.02) | 0.11+ (0.06) |
| <i>LogExp</i> | -0.44*** (0.04) | -0.26*** (0.08) | -0.44*** (0.03) | -0.28*** (0.08) | -0.53*** (0.03) | -0.56*** (0.03) | -0.28*** (0.06) |
| <i>LogExpDiversity</i> | -0.01 (0.05) | 0.07 (0.13) | 0.24*** (0.04) | 0.14 (0.13) | 0.30*** (0.03) | 0.26*** (0.03) | 0.19+ (0.10) |
| <i>Assigned</i> | 0.10* (0.04) | 0.11 (0.13) | | | 0.03 (0.03) | 0.02 (0.03) | 0.12 (0.11) |
| <i>LogChunks(dm)</i> | 0.23*** (0.02) | | 0.19*** (0.02) | | 0.18*** (0.02) | 0.20*** (0.02) | |
| <i>LogClaims</i> | 0.17*** (0.02) | | 0.17*** (0.02) | | 0.18*** (0.01) | 0.17*** (0.01) | |
| <i>LogWords</i> | 0.10*** (0.02) | -0.88*** (0.21) | 0.10*** (0.01) | -0.84*** (0.19) | 0.09*** (0.01) | 0.08*** (0.01) | -0.82*** (0.16) |
| <i>LogFigures</i> | 0.34*** (0.01) | 0.32*** (0.08) | 0.32*** (0.01) | 0.34*** (0.07) | 0.30*** (0.01) | 0.31*** (0.01) | 0.28*** (0.06) |
| <i>LogAssignedClasses</i> | 0.35*** (0.02) | 0.75*** (0.11) | 0.33*** (0.01) | 0.71*** (0.09) | 0.34*** (0.01) | 0.33*** (0.01) | 0.67*** (0.08) |
| <i>LogCitedClasses</i> | -0.03 (0.03) | -0.35*** (0.08) | -0.01 (0.02) | -0.38*** (0.07) | 0.02 (0.02) | 0.03 (0.02) | -0.32*** (0.06) |
| <i>LogPatentCites</i> | 0.08*** (0.02) | 0.24*** (0.05) | 0.09*** (0.01) | 0.29*** (0.05) | 0.09*** (0.01) | 0.09*** (0.01) | 0.30*** (0.04) |
| <i>LogNonpatentCites</i> | 0.03** (0.01) | 0.06 (0.05) | 0.05*** (0.01) | 0.04 (0.04) | 0.05*** (0.01) | 0.05*** (0.01) | 0.08* (0.03) |
| <i>Fees</i> | 0.33*** (0.04) | 0.56 (0.38) | 0.36*** (0.03) | 0.19 (0.41) | 0.38*** (0.03) | 0.36*** (0.03) | 0.23 (0.32) |
| <i>Sole</i> × <i>LogChunks(dm)</i> | -0.09*** (0.03) | | -0.08*** (0.02) | | -0.06** (0.02) | -0.08*** (0.02) | |
| <i>Sole</i> × <i>LogPastCollab(dm)</i> | 0.05** (0.02) | 0.18** (0.07) | 0.04* (0.02) | 0.15* (0.06) | 0.05*** (0.01) | 0.04** (0.01) | 0.11+ (0.06) |
| Filing-year FE | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Lead-inventor FE | Lead inventor × Product class | | Lead inventor × Firm | | Yes | Yes | Yes |
| <i>N</i> | 192,311 | 36,588 | 277,063 | 38,197 | 378,195 | 320,276 | 43,111 |
| Log-likelihood | -50,122 | -11,003 | -73,421 | -11,295 | -96,187 | -82,323 | -12,229 |
| χ^2 | 8,803 | 382 | 9,870 | 421 | 12,465 | 13,169 | 485 |

Notes: Standard errors (in parentheses) are clustered by lead inventor. (dm) = de-meanded; FE = fixed effects; M. = Model.

+ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 13: First-stage regressions predicting *Sole*

| Dataset | Utility | Design |
|-----------------------------|-----------------|-----------------|
| Specification | First-stage | First-stage |
| Variable | Model 57 | Model 58 |
| <i>LogLawyers</i> | -0.08***(0.02) | -0.13* (0.05) |
| <i>LogPastCollab(dm)</i> | 0.07***(0.02) | 0.23***(0.05) |
| <i>LogExp</i> | -0.06** (0.02) | -0.15***(0.04) |
| <i>LogExpDiversity</i> | -0.05 (0.02) | -0.23** (0.08) |
| <i>Assigned</i> | -0.88***(0.02) | -0.66***(0.09) |
| <i>LogChunks(dm)</i> | -0.05** (0.01) | |
| <i>LogClaims</i> | -0.04***(0.01) | |
| <i>LogWords</i> | -0.04***(0.01) | -0.30* (0.14) |
| <i>LogFigures</i> | -0.32***(0.01) | -0.32***(0.05) |
| <i>LogAssignedClasses</i> | -0.01 (0.01) | 0.03 (0.09) |
| <i>LogCitedClasses</i> | -0.05** (0.02) | -0.21***(0.05) |
| <i>LogPatentCites</i> | -0.05***(0.01) | -0.02 (0.03) |
| <i>LogNonpatentCites</i> | -0.04***(0.01) | -0.01 (0.03) |
| <i>Fees</i> | -0.39***(0.02) | -1.78***(0.23) |
| Filing Year Fixed Effects | Yes | Yes |
| Lead Inventor Fixed Effects | Yes | Yes |
| <i>N</i> | 368,899 | 46,350 |
| Log-likelihood | -115,894 | -10,719 |
| χ^2 | 6,579 | 770 |

Notes: Standard errors (in parentheses) are clustered by lead inventor; (dm) = demeaned

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$