Ownership Diversity, Patenting, and Venture Capital Success among R&D Performing Microbusinesses

Gary Anderson, Audrey Kindlon, Timothy Wojan[1]

January 3, 2025 | Society of Government Economists Session, Allied Social Science Associations Annual Meeting, San Francisco, CA

NATIONAL CENTER FOR SCIENCE AND ENGINEERING STATISTICS
U.S. NATIONAL SCIENCE FOUNDATION

[1] Oak Ridge Institute for Science and Education

Disclaimers and Acknowledgements

Disclaimer: This presentation provides results of exploratory research sponsored in part by the National Center for Science and Engineering Statistics (NCSES) within the U.S. National Science Foundation (NSF). This information is being shared to inform interested parties of ongoing activities and to encourage further discussion. Any views expressed are those of the authors and not necessarily those of NCSES or NSF.

Any views expressed are those of the author and not those of the Census Bureau. The Census Bureau has reviewed this data product to ensure appropriate access, use, and disclosure avoidance protection of the confidential source data used to produce this product. This research was performed at a Federal Statistical Research Data Center under FSRDC project number 2681 and at the NCSES Survey Sponsor Data Center under project number P-7504866 and Disclosure Review Board approval number CBDRB-FY25-0075. NCSES has reviewed this product for unauthorized disclosure of confidential information and approved its release (NCSES-DRN24-063).

Acknowledgments: This research was supported in part by NCSES within NSF and by the Oak Ridge Institute for Science and Education (ORISE) for the Department of Energy (DOE). ORISE is managed by Oak Ridge Associated Universities (ORAU) under DOE contract number DE-SC0014664. All opinions expressed in this paper are those of the authors and do not necessarily reflect the policies and views of DOE, ORAU, or ORISE.

Outline

- Strategies for decreasing researcher degrees of freedom and the probability of false discovery include
  - Axiomatic selection of a single ownership diversity measure
  - Results derived for all 127 unique combinations of age, educational level, education specialization, sex, ethnicity, race, and foreign-born status
  - Significant results passed through for *de novo* confirmatory analysis using holdout sample, with multiple comparison correction and using priors from exploratory analysis for Bayesian updating
- Protocol is applied to testing how ownership diversity is associated with patenting and venture capital (VC) funding among R&D-performing microbusinesses in the Annual Business Survey (ABS)
- Does diversity matter for radical innovation concentrated in high-tech start-ups?

The Annual Business Survey (ABS)

- Combination of former Survey of Business Owners, innovation module from Business R&D and Innovation Survey, and R&D module for microbusinesses (< 10 employees)
- Division of innovative labor explains radical innovation concentrating in R&D-performing microbusinesses (Baumol 2010)
- In the ABS, at least seven principal owner attributes (age, sex, ethnicity, educational level, education specialization, race, and foreign-born status) capture ownership diversity

Why focus on intermediate innovation outcomes?

- “Self-reported innovation” strongly associated with diversity measures in earlier 2018 ABS analysis (Wojan and Lambert 2024). Critique: only measures novelty, not whether innovation is good, bad, or mediocre.
- Affective conflict explanation: Incompatible attitudes or opinions on value of an innovation are launched in market to test. Possibly higher rate of unsuccessful innovations.
- Cognitive conflict explanation: Different attitudes or experiences increase combination of seemingly incongruent ideas, leading to better, more novel innovation.
- Diversity associated with increased probability of patenting/VC success would support latter explanation as intermediaries assess nonobviousness or potential for market/buyout success.

Selecting a diversity index axiomatically

- **Homophily Axiom:** All owners belonging to the same group must result in the lowest diversity-measure value.
- **Fractionalization Axiom:** An increase in the number of groups must increase the diversity-measure value.
- **Team Size Axiom:** Larger ownership teams not demonstrating homophily must increase the diversity-measure value relative to smaller ownership teams.
- **Concentration of Ownership Axiom:** Ownership concentrated in one member of the team must reduce the diversity-measure value relative to ownership that is more equally distributed among team members.

Ownership fractionalization (OF) index

Derived from the ethno-linguistic fractionalization index (ELF):

$$ELF = 1 - \sum_{i=1}^{n} p_i^2$$

where *p* is the population share of *n* groups.

Invariant to population size so **violates TEAM SIZE AXIOM**.

A minor modification of the ELF **satisfies all four axioms**

$$OF = 1 - \sum_{i=1}^{o} p_i^n$$

where *p* represents the ownership share of the i^{th} owner and *o* is the number of owners.

Split sample design

Restoring transparency to specification and hypothesis testing

1. Exploratory analysis used 35% of the 2018 ABS and the full 2021 ABS for **specification testing** to discover potentially useful models (Anderson & Magruder 2017; Wojan 2024)
2. *De novo* **hypothesis testing** applied to useful models passed through using 65% holdout of the 2018 ABS and the full 2022 ABS to generate valid test statistics after applying false discovery rate (FDR) and family-wise error rate (FWER) correction
3. Drastic reduction in power due to more stringent *p*-value threshold mitigated by using exploratory estimates as Bayesian informed priors
4. Bayesian estimates also provide probability of posterior distribution—Pr(effect | sample)—that is more informative than Pr(sample | null)

Frequentist and Bayesian specification

- Estimate Pr(*Patent Pending, Patent Owned, or VC Funding)* = *f(OF, Patent Intensive Industry (1/0), Family Business (1/0), Firm Age)* for R&D-performing microbusinesses.
- Frequentist estimation requires multiple comparison corrections as number of looks affects how one interprets a *p*-value. Family-wise error rate (FWER) = p/k = 0.05/126 = 0.0004.
- Bayesian estimation uses exploratory *OF* coefficient estimates as weakly informative priors with a standard error of 10. Priors for remaining coefficients are 0 with a standard error of 10.

Descriptive statistics

	2018 ABS 65% Sample R&D-Performing Microbusinesses		2022 ABS R&D-Performing Microbusinesses	
Variable	Mean	Range	Mean	Range
Age Diversity (A)	0.2406	0.9844	0.2806	0.9844
Educational Level Diversity (E)	0.2968	0.9844	0.2974	0.9844
Sex Diversity (G)	0.3029	0.7500	0.2516	0.7500
Ethnic Diversity (H)	0.0297	0.7500	0.0317	0.7500
Education Specialization Diversity (M)	0.1752	0.9844	0.2923	0.9844
Race Diversity (R)	0.0447	0.9375	0.0718	0.9375
Foreign-born Status Diversity (U)	0.0719	0.7500	0.1004	0.7500
Composite Diversity (AEGHMRU)	0.1660	0.7015	0.2051	0.7240
Patent Intensive Industry (0/1)	0.3561		0.6997	
Family/Jointly Owned (0/1)	0.6013		0.4615	
Firm Age	9.4880		10.300	
Patent Owned (0/1)	0.0960			
Patent Pending (0/1)	0.0883			
Venture/Angel Capital (0/1)			0.0605	

Source(s): 2018 ABS 35% exploratory sample and full 2021 ABS.

Selected patent pending confirmatory estimates

Diversity Measure	Prior Diversity Estimate	Frequentist Diversity Estimate	Bayesian Diversity Estimate	Bayesian Diversity Odds Ratio	Lower Credible Interval	Upper Credible Interval
AEGHMRU	3.731	3.6	3.615	***37.15***	7.790	177.3
AEHMRU	3.707	3.723	3.755	***42.74***	9.984	198.6
EHMRU	3.678	3.864	3.893	***49.01***	10.39	188.7
AHMRU	3.666	3.686	3.667	***39.15***	6.431	159.7
AEHRU	3.663	3.166	3.149	***23.32***	5.359	119.9
GR	1.141	0.171	0.131	***1.140***	0.306	4.721
GU	1.006	0.5704	0.567	***1.762***	0.464	8.632
H	0.6822	-0.2514	-0.287	***0.750***	0.083	3.711
G	-0.4249	-0.519	-0.530	***0.588***	0.196	1.788

Source(s): 2018 ABS 65% confirmatory sample.

Note(s): A = age, E = educational level, G = sex, H = ethnicity, M = education specialization, R = race, U = foreign-born status. Credible interval is 99.92% to represent FWER correction. Total of 126 out of 127 equations estimated.

Selected patent owned confirmatory estimates

Diversity Measure	Prior Diversity Estimate	Frequentist Diversity Estimate	Bayesian Diversity Estimate	Bayesian Diversity Odds Ratio	Lower Credible Interval	Upper Credible Interval
EHMRU	3.288	2.483	2.499	***12.18***	3.189	41.48
AEHMRU	3.241	2.434	2.447	***11.56***	2.45	44.65
EHMR	3.216	2.319	2.316	***10.14***	2.570	52.32
AEHMR	3.116	2.256	2.279	***9.771***	2.738	28.20
AHMRU	3.094	2.409	2.373	***10.73***	2.502	35.88
AGH	0.7309	0.9633	0.925	***2.522***	0.507	8.75
EG	0.6953	0.6722	0.7009	***2.016***	0.739	6.059
AG	0.556	0.6947	0.7203	***2.055***	0.816	4.367
G	-1.137	-0.3208	-0.2913	***0.747***	0.359	1.899
GH	-1.719	-0.3592	-0.3941	***0.674***	0.130	2.532

Source(s): 2018 ABS 65% confirmatory sample.

Note(s): A = age, E = educational level, G = sex, H = ethnicity, M = education specialization, R = race, U = foreign-born status. Credible interval is 99.92% to represent FWER correction. Total of 122 out of 127 equations estimate

Selected venture/angel capital confirmatory estimates

Diversity Measure	Prior Diversity Estimate	Frequentist Diversity Estimate	Bayesian Diversity Estimate	Bayesian Diversity Odds Ratio	Lower Credible Interval	Upper Credible Interval
AHR	2.305	1.263	1.297	***3.659***	0.527	20.33
AHRU	2.116	1.325	1.300	***3.670***	0.572	21.72
AH	2.014	0.0463	0.049	***1.050***	0.103	5.548
AHU	2.012	0.4549	0.486	***1.626***	0.367	7.013
AGHRU	1.877	1.149	1.172	***3.228***	0.576	27.85
EHM	0.3534	0.8077	0.818	***2.266***	0.525	8.298
AGM	0.348	0.5016	0.4848	***1.624***	0.400	5.526
E	0.3205	0.3188	0.3105	***1.364***	0.535	2.820
GM	-0.4759	0.3715	0.3488	***1.417***	0.467	4.333
G	-0.4937	-0.4144	-0.4227	***0.655***	0.161	2.342

Source(s): 2022 ABS.

Note(s): A = age, E = educational level, G = sex, H = ethnicity, M = education specialization, R = race, U = foreign-born status. Credible interval is 99.9% to represent FWER correction. Total of 107 out of 127 equations estimated.

Significant venture/angel capital confirmatory estimates

Diversity Measure	Prior Diversity Estimate	Frequentist Diversity Estimate	Bayesian Diversity Estimate	Bayesian Diversity Odds Ratio	Lower Credible Interval	Upper Credible Interval
HMRU	0.8998	1.921	1.913	***6.776***	1.137	50.89
EHMRU	1.0970	1.820	1.842	***6.311***	1.169	32.31
EMRU	0.6496	1.679	1.691	***5.423***	1.160	34.02
MRU	0.4016	1.685	1.691	***5.425***	1.282	27.52
AMRU	1.1550	1.562	1.576	***4.834***	1.006	24.79
EGMRU	0.3737	1.563	1.551	***4.715***	1.137	23.42
R	0.6187	1.429	1.430	***4.180***	1.294	11.65
ER	0.7822	1.378	1.396	***4.039***	1.237	14.47
AEMR	0.9824	1.330	1.360	***3.897***	1.059	15.50
M	0.6613	1.279	1.301	***3.672***	1.664	9.178

Source(s): 2022 ABS.

Note(s): A = age, E = educational level, G = sex, H = ethnicity, M = education specialization, R = race, U = foreign-born status. Credible interval is 99.9% to represent FWER correction. Total of 107 out of 127 equations estimated.

Regression decomposition of log odds by diversity dimension

	Patent Pending		Patent Owned		Venture/Angel Capital Funding	
Diversity Dimension	Estimate	Standard Error	Estimate	Standard Error	Estimate	Standard Error
Age	0.4642	0.06336	0.3244	0.05003	-0.1121	0.04697
Educational Level	0.5631	0.06336	0.3665	0.05003	0.06196[ns]	0.04697
Sex	-0.2619	0.06336	-0.1710	0.05003	-0.1869	0.04697
Ethnicity	0.4471	0.06336	0.3685	0.05003	0. 02001[ns]	0.04697
Education Specialization	1.1550	0.06336	0.9882	0.05003	0.4203	0.04697
Race	0.5544	0.06336	0.0238[ns]	0.05003	0.8261	0.04697
Foreign-born Status	0.5088	0.06336	0.3035	0.05003	0.2533	0.04697
Intercept	0.6856	0.09444	0.4292	0.07541	0.3128	0.06756

Source(s): 2018 ABS 65% exploratory sample, and full 2022 ABS.

Note(s): All coefficient estimates significant at <0.001 level except for Age in Venture Capital equation (0.05 level) and estimates designated *ns* (not significant).

Does diversity matter for radical innovation concentrated in high-tech start-ups?

- Yes, it appears to matter a lot, even with stringent FWER corrections:
  - Maximally diverse ownership teams up to eight times more likely to own a patent and five times more likely to get VC funding than homophilic teams.
  - Education specialization strongly associated with increased likelihood of patent ownership as well as increasing likelihood of VC funding.
  - Race diversity is most strongly associated with increased likelihood of VC funding, which may be capturing the racial composition of places where VC is most prevalent.
- Sex diversity tends to decrease the likelihood intermediate innovation outcomes, which is consistent with lower patenting and VC funding rates of female-owned businesses (Cook and Kongcharoen 2010; Gompers et al. 2022).
- Homophily may be good for social capital and trust (Putnam 2007) but it appears to be serious drag on innovation and creativity.

Thank you!

Contact: twojan@nsf.gov

https://ncses.nsf.gov

References

- Anderson ML, Magruder J. 2017. *Split-Sample Strategies for Avoiding False Discoveries*. Working Paper 23544. Cambridge, MA: National Bureau of Economic Research.
- Baumol WJ. 2010. *The Microtheory of Innovative Entrepreneurshi*p. Princeton, NJ: Princeton University Press.
- Cook LD, Kongcharoen C. 2010. *The Idea Gap in Pink and Black*. Working Paper 16331. Cambridge, MA: National Bureau of Economic Research.
- Gompers PA, Mukharlyamov V, Weisburst E, Xuan Y. 2022. Gender Gaps in Venture Capital Performance. *Journal of Financial and Quantitative Analysis* 57(2):485–513.
- Putnam RD. 2007. E pluribus unum: Diversity and community in the twenty-first century the 2006 Johan Skytte Prize Lecture. *Scandinavian Political Studies* 30(2):137-174.
- Wojan TR. 2024. Exploratory Report: Annual Business Survey Ownership Diversity and Its Association with Patenting and Venture Capital Success. Center for Economic Studies Working Paper No. 24-62.
- Wojan TR, Lambert DM. 2024. A Novel Framework for Increasing Research Transparency: Exploring the Connection Between Diversity and Innovation. *PLoS ONE.* doi:10.1371/journal.pone.0313826.

Comments regarding research transparency on contentious topics

- Null hypothesis significance testing fuels false discovery and the credibility crisis
  - Number of looks matters but rarely corrected in testing
  - Search for nonsignificance is statistically absurd
- Split sample/dual method protocol to the rescue
  - Frequentist exploration—calibration, specification search
  - Bayesian replication provides answer to the research question being asked: probability of magnitude of effect
  - Publication of all results guards against cherry picking