Learning in Non-routine Tasks: Productivity and Quality Improvement on a Digital Platform

Jiatong Zhong

University of Alberta

Motivation

Learning-by-doing in a complex and non-routine task: how does individual productivity improve with the accumulation of experience?

Challenges:

- How to observe individual-task level outcomes overtime?
- How to measure the outcomes of the task?
- How to disentangle individual learning from organizational learning and other confounding factors?

What this paper does:

- Use granulated firm data set that tracks the performance of authors on a for-profit online publishing platform
- Empirically investigate learning-by-doing by self-publishing novel writers (with an emphasis on quality improvement).

Literature

- Empirical investigation on learning-by-doing using detailed data:
  Taxi drivers [Haggag et al., 2017];automobile assembly [Levitt et al., 2013]; steel mill [Hendel and Spiegel, 2014]
  Contribution: evidence in creative and non-routine tasks.
- Individual and organizational learning in skilled and non-routine tasks:
  Surgeons [Stith, 2018], entrepreneurs [Rocha et al., 2015, Lafontaine and Shaw, 2016], researchers in biomedical fields [Yu et al., 2022]
  Contribution: use detailed data and focus on individual learning-by-doing.

The market of online publishing

- Free entry
  - Anyone can upload their stories anonymously, free of charge.
  - No barrier of entry in terms of age, educational attainment, or prior experience.
    - Authors <18 yo can sign the contract under the supervision of their guardians.
  - Only authors who have signed a contract with the platform can receive payment.
    - All fiction, with the exception of short stories (<30,000 characters), must be published exclusively on the platform
    - The contract lasts for 5, 10, or 20 years and can be renewed.

Pricing and Profit sharing

- Flat pricing for the book subscription
  - First 70,000 characters free (20 to 30 chapters)
  - 0.03 yuan/1000 characters after that.
  - Authors can decide if and when to charge the readers.
- Profit sharing between authors and the platform
  - 50% of all book subscriptions and readers contribute to the platform.
  - The website provides: platform, editorial service, legal service;

Why is this a clean context to study LBD?

Some common challenges in empirical studies of individual LBD:

- Need to disentangle productivity improvement from factors like improvement of inputs, shifts in input prices, changes in capital $\longrightarrow$this industry is labor intensive and capital plays a small role.
- Price is fixed, observable, and the same across individuals.
- Usually hard to separate individual LBD from organizational learning: organizations can improve organizational-level productivity by reallocating workers or displacing the least efficient workers$\longrightarrow$ authors make independent entry-exit decisions at any point in time.

Data

- Records from the online platform covering all authors with over 1000 followers as of November 10, 2022
- February 2005 to October 2022
- 57,190 novels, novellas, and short stories written by 6,561 authors
- **Author level**: number of followers, earliest active date, last active date, number of titles;
- **Book level**: genre, length, initial and final update date, completion status, number of comments and comment scores, number of patrons and patronage, clicks of the first 10 free chapters and first 10 pay-to-read chapters, number of subscribers.

Measures for book performance

- Ideal: stream of income
- Issue: not available. In addition, income does not necessarily reflect quality.
- Quality measure: reader retention rate.
  - VIP chapter 10 click/chapter 1 click ratio
  - Subscribers/chapter 1 click ratio
  - Projected subscription income

Parametric Specification

$$\ln y_{it} = \alpha + g(E_{it}; \beta) + X_{it} + \eta_t + \xi_i + \epsilon_{it} \tag{1}$$

where $g(E_{it}; \beta) = \beta \ln(e_{it})$,

- y_{it} is a measure of book performance.
- Controls: length of the current book, author fixed-effects, upload year-month fixed-effects, and genre fixed-effects.

Follower-neutral improvement with experience

Dependent variable Specification	ln(Vchp10/chp1 click) (1)	ln(Vchp10/chp1 click) (2)	ln(Vchp10/chp1 click) (3)	ln(Vchp10/chp1 click) (4)
New		-0.0207*		-0.0159
		(0.0123)		(0.0124)
New×ln(characters)		-0.0402**		-0.0344**
		(0.0167)		(0.0169)
ln(characters)	-0.0131	0.0419***	0.0174**	0.0309**
	(0.0149)	(0.0141)	(0.00767)	(0.0145)
New author only	Yes	No	No	No
Length	No	No	Yes	Yes
Genre FE	No	No	Yes	Yes
Observations	16818	51738	50567	50567
R^2	0.014	0.012	0.385	0.386

Note: Robust standard errors in parentheses, clustered at the author level. Author and upload year-month FE included in all specifications. New authors are authors who have written less than 1 million characters. * $p < 0.10, **p < 0.05, ***p < 0.01$

Nonparametric Specification

In this specification,

$$g(E_{it};\beta) = \beta_1 1\{0 \leq e_{it} \leq 1\} + \beta_2 1\{1 < e_{it} \leq 2\}$$
$$+ \dots + \beta_{10} 1\{10 < e_{it} \leq 11\}$$

- In this specification, authors who have written more than 11 million characters are the excluded category.
- Controls: length of the current book, author fixed-effects, upload year fixed-effects and genre fixed-effects.
- Cluster at the author level

Duration of Learning

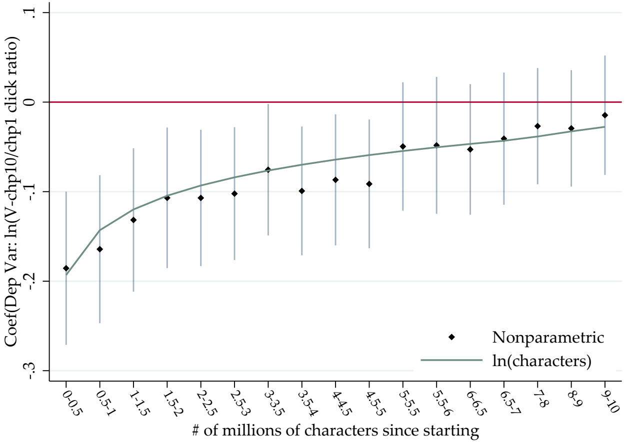


Figure: Parametric (log) versus Nonparameteric (dummy) Approaches

Learning by initial performance

	Full Sample	Lowest	Low	High	Highest
	Panel One: VIP Chapter 10 to First Chapter Clicks Ratio				
ln(characters)	0.0428***	0.0928***	0.0444***	0.0222*	-0.00315
	(0.00589)	(0.0123)	(0.0112)	(0.0119)	(0.0121)
Observations	46415	14107	12704	10821	8572
	Panel Two: Subscription to First Chapter Clicks Ratio				
ln(characters)	0.0384***	0.0858***	0.0518***	0.0214*	-0.0118
	(0.00658)	(0.0140)	(0.0119)	(0.0129)	(0.0140)
Observations	47008	13889	11548	10535	11036

Note: This table shows the estimates from equation 1 when authors are divided into four quartiles based on the performance of their first book. Robust standard errors in parentheses clustered at the author level. Length, genre, author and upload year-month fixed effects included in all specifications. *
$p < 0.10, **p < 0.05, ***p < 0.01$

Alternative Mechanisms

1. Survival bias
   - Test: Does learning persist if we exclude all inactive authors?
   - Inactive: authors who have not written or revised anything since Jan 2021 (data collected in Nov 2022).
2. Learning about consumer preference
   - Test: Do authors follow literary tropes?

Alternative Mechanism 1: Survival Bias

Specification	(1)	(2)	(3)	(4)
Panel One: VIP Chapter 10 to First Chapter Clicks Ratio				
ln(characters)	0.0515***	0.0537***	0.0455***	0.0475***
	(0.00627)	(0.00653)	(0.00639)	(0.00670)
Panel Two: Subscription to First Chapter Clicks Ratio				
ln(characters)	0.0275***	0.0235***	0.0174**	0.0133*
	(0.00753)	(0.00777)	(0.00767)	(0.00796)
Panel Three: Projected Income				
ln(characters)	0.0444***	0.0431***	0.0393***	0.0350***
	(0.0127)	(0.0132)	(0.0129)	(0.0134)
Active authors only	No	Yes	No	Yes
Length	No	No	Yes	Yes
Genre FE	No	No	Yes	Yes

Note: Robust standard errors in parentheses, clustered at the author level. Author and upload year-month fixed effects included in all specifications.

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

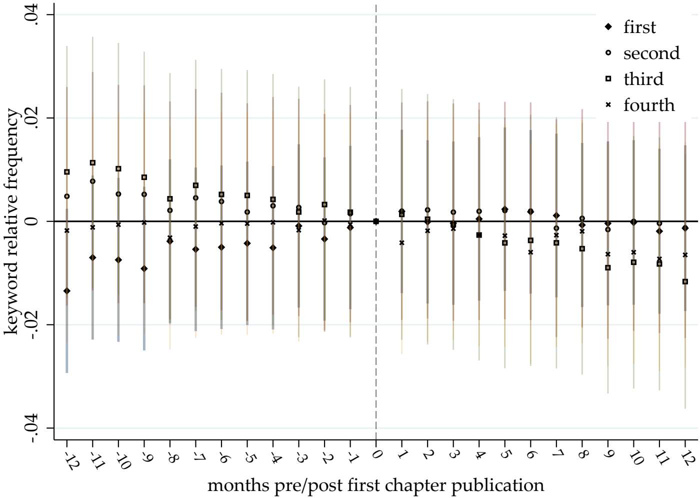
Alternative Mechanism 2: Learning about consumer preference

- Does the frequency of keywords increase after a superstar title reveals consumer preference?
- Superstar: the top 1% title within a broad genre in a quarter if that genre has at least 100 titles published in that quarter.

$$\mathbb{1}\{\text{keyword}\}_{it}^{n} = \sum_{\tau=-12}^{\tau=12} \beta_\tau \mathbb{1}\{t - t^* = \tau\} + \eta_t + \epsilon_{it} \qquad (2)$$

  - t^*: publication time of a superstar
  - η_t: upload year-month fixed effects
  - $\mathbb{1}\{\text{keyword}\}_{it}^{n}$: an indicator that takes value 1 if book i is uploaded in month t and includes the nth keyword of a superstar in its description.

Alternative Mechanism 2: Learning about consumer preference



Conclusion

New evidence of learning-by-doing in creative and non-routine tasks:

- Learning-by-doing improves quality, but the magnitude is small
- Quality improvement lasts for much longer than what the literature has found in the manufacturing context and other routine tasks.
- Learning primarily occurs among authors with a weaker start.

What's next:

- Authors who switch genres: are they developing specific or transferable skills?
- Quantify the income impact of reader accumulation versus learning

References I

Haggag, K., McManus, B., and Paci, G. (2017).
Learning by Driving: Productivity Improvements by New York City Taxi Drivers.
American Economic Journal: Applied Economics.

Hendel, B. I. and Spiegel, Y. (2014).
Small Steps for Workers , a Giant Leap for Productivity.
American Economic Journal: Applied Economics, 6(1):73–90.

Lafontaine, F. and Shaw, K. (2016).
Serial Entrepreneurship: Learning by Doing?
Technical Report 2.

Levitt, S. D., List, J. A., and Syverson, C. (2013).
Toward an Understanding of Learning by Doing: Evidence from an Automobile Assembly Plant.
Technical Report 4.

References II

Rocha, V., Carneiro, A., and Amorim Varum, C. (2015).
Serial entrepreneurship, learning by doing and self-selection.
International Journal of Industrial Organization, 40:91–106.

Stith, S. S. (2018).
Organizational learning-by-doing in liver transplantation.
International Journal of Health Economics and Management, 18(1):25–45.

Yu, H., Marschke, G., Ross, M. B., Staudt, J., and Weinberg, B. (2022).
Publish or Perish: Selective Attrition as a Unifying Explanation for Patterns in Innovation over the Career, volume 58.