Jack of Less Trades:
Evidence on Increasing Specialization in Research

김 진 영
KU and IZA

고 강 혁
UNIST

한양대학교
2017년 6월 8일

Motivation

- Knowledge accretion through scientific and industry research is well noted as an important factor for the improvement of the standard of living.
- Until the early 20th century, scientific discovery was widely considered the realm of individual masterminds such as Einstein and Edison (Lamoreaux and Sokoloff, 2009).
- But now, knowledge creation is increasingly accomplished by research teams with bigger size than ever.

Motivation

- Empirical evidence on rising research team size
  - 42% of US patents listed multiple inventors in 1975, over 70% today
  - Average number of inventors per patent (US): 1.6 in 1975, about 3 today
  - Scientific papers: the number of authors increased by about 50% over the period 1981–1999 (Adams et al., 2005).

- Several explanations for bigger teams in research or production
  - Market size (A. Smith, 1776)
  - Decreasing coordination costs (G. Becker and K. Murphy, 1992)
  - Accumulation of knowledge (B. Jones, 2009)
  - Racing against time (J. Kim, 2017)

Motivation

- Another prediction from these models
  - Increasing specialization among team members with a narrower range of tasks or expertise
- However, we have surprisingly little evidence for increasing specialization in production, not to mention in research.
- Goals of the paper
  - We investigate empirically whether the degree of specialization of researchers has risen over time. (*Are younger generations more specialized?*)
  - Also, we study how specialization evolves over the life cycle of a researcher. (*Do researchers become more or less specialized as they get older?*)

Introduction

- Inventor-level info in patent data: an ideal source for this topic
- However, patent data from patent offices around the world have not been well utilized because unique inventors cannot be identified across patents.
- Two approaches to overcome this problem

  (i) Matching inventor names, or disambiguation (Trajtenberg, Shiff, Melamed, 2006; Lai, D'Amour, Fleming, 2009)

  (ii) Inventor surveys (PatVal-EU I and II, PatVal-JP, PatVal-US)
- Caveats

  (i) accuracy in matching (esp. problematic for East Asian names)

    Under-matching error (Type I): likely for prolific inventors

    Over-matching error (Type II): likely for popular names (“John Smith” problem)

  (ii) response rate

Introduction

- Korean Intellectual Property Office (KIPO): collecting inventors' resident registration numbers (주민번호) since 1991 until 2005 (not for public use)
- We were able to acquire from KIPO unique ID's for inventors along with information on birth year and gender.
- Unique advantages of our Korean inventor data:

  (i) measuring patent productivity of inventors accurately at each age, by gender, and by birth-year cohort

  (ii) no need for disambiguation to construct the inventor-level panel data

Research Questions

- Three questions about researchers' specialization

  (1) whether younger cohorts of researchers are more specialized when they start their careers

  (2) how the degree of specialization evolves over the career

  (3) how far an inventor moves in terms of research fields over his career

Preliminary Findings

- Younger cohorts are increasingly more specialized through formal education (more specialized when starting careers at age 25).
  - About 20-30% more specialized for the birth-cohort of 1976 than the cohort of 1947
- Over the life cycle, inventors become more specialized until age 31~36 and then less specialized.
  - The age-specialization profile takes an inverted-U shape.
  - In most measures of specialization, specialization reaches a lower level at age 53 than at age 25.
- Inventors move neither farther nor nearer in the space of research fields at the later stage of career than at the earlier stage.

Outline

1. Estimation strategy
   a. Regression I: Cohort effect + Age effect
   b. Regression II: Drifting
2. Empirical specification
3. Regression results
4. Sensitivity analysis
5. Concluding remarks

Estimation Strategy I

- Our contention: Technology classes by International Patent Classification (IPC) appearing on patents of an inventor reflect his fields of expertise.
- We measure the degree of specialization at each age by evaluating how similar or close are the classes of his patents in one year.
  - Pick a patent (call patent T) in year t
  - Find another patent filed immediately after T
  - Pick the IPC class listed first on each patent (using the second-level class)
  - Using a distance measure between IPC classes (see next slide), calculate the similarity between the two fields of expertise as a measure of specialization.

Korean Patent Sample

(19) 대한민국특허청(KR)
(12) 등록특허공보(B1)

(45) 공고일자 2014년03월12일
(11) 등록번호 10-1365614
(24) 등록일자 2014년02월14일

(51) 국제특허분류(Int. Cl.)
H04N 21/4227 (2011.01) *H04Q 9/00* (2006.01)
(21) 출원번호 10-2012-0092446
(22) 출원일자 2012년08월23일
심사청구일자 2012년08월23일
(56) 선행기술조사문헌
KR1020090009814 A

(73) 특허권자
엘지전자 주식회사
서울특별시 영등포구 여의대로 128 (여의도동)
(72) 발명자
안오봉
경기 평택시 진위면 엘지로 222, 특허센터 (LG전자(주))
김건우
경기 평택시 진위면 엘지로 222, 특허센터 (LG전자(주))
조상현
경기 평택시 진위면 엘지로 222, 특허센터 (LG전자(주))
(74) 대리인
김용인, 박영복

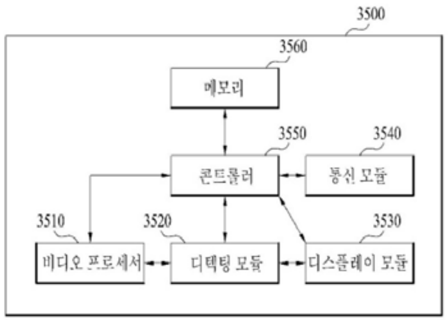
전체 청구항 수 : 총 17 항 심사관 : 박금옥

(54) 발명의 명칭 **외부 전자 디바이스와 커넥티트된 멀티미디어 디바이스 및 그 제어 방법**

(57) 요 약

리모트 컨트롤러 및 멀티미디어 디바이스의 제어 방법에 개시된다. 본 발명의 일실시예에 의한 외부 전자 디바이스와 커넥티트된 멀티미디어 디바이스는, 적어도 하나 이상의 문자열 및 대응하는 OSD 그래픽 데이터를 저장하고 있는 메모리와, 상기 외부 전자 디바이스로부터 수신된 비디오 이미지를 처리하는 비디오 프로세서와, 상기 비디오 이미지내 문자열을 검출 하는 디텍팅 모듈과, 상기 메모리에 액세스 하여, 상기 검출된 문자열에 대응하는 OSD 그래픽 데이터를 추출하는 컨트롤러와, 상기 추출된 OSD 그래픽 데이터를 출력하는 디스플레이 모듈과, 그리고 상기 OSD 그래픽 데이터내 특정 영역이 선택된 경우, 상기 특정 영역에 대응하는 커맨드 시그널을 상기 외부 전자 디바이스 또는 리모트 컨트롤러로 전송하는 통신 모듈을 포함한다.

대 표 도 - 도35



International Patent Classification (IPC)

(1) H04N 21/4227 (2011.01)

- Section H: Electricity
- Class 04: Electric Communication Technique
- Subclass N: Pictorial Communication, e.g. Television
- Group 21/4227: Remote input by a user located remotely from the client device, e.g. at work

(2) H04Q 9/00 (2006.01)

- Subclass Q: Selecting (switches, relays, selectors)
- Group 9/00: Arrangements in telecontrol or telemetry systems for selectively calling a substation from a main station

Distance Measures

- How to measure the distance between two tech classes?
- We try four distinct measures: Jaffe (1986), Bloom-Schankerman-Van Reenen (2013), Citation-based, Class-co-occurrence

I. Jaffe measure (1986)

- Distance between class i and j ≡ D(i,j) = 1, if i = j,
  = 0, otherwise.

II. BSV measure (2013): "Mahalanobis distance measure"

- Let $C_i = [a_{i1}, a_{i2}, \ldots, a_{iM}]$ for class i, where a_{ik} = share of patents in class i which are produced by firm k, and M = total number of firms in the patent data over period 1991-2005 (71,628 firms which have at least 2 patents).
- D(i,j) = cosine similarity of vectors C_i and C_j = $\frac{\sum_k C_{ik} C_{jk}}{\sqrt{\sum_k C_{ik}^2}\sqrt{\sum_k C_{jk}^2}}$

Distance Measures

III. Citation-based measure

- For patent s filed in year t, we construct $C_{st} = [b_{s1t}, b_{s2t}, \ldots, b_{sNt}]$, where b_{skt} = number of patents in class k that are cited by s, and N = total number of classes (340 at two-level classification).
- We calculate the cosine similarity of all possible pairs of patents filed in year t.
- D(i,j,t) = mean of the CS values from all the patent pairs in class i and j
- We then take the weighted average of D(i,j,t) over the whole sample period.
- We use both Korean and US patent citation data for constructing two measures (because citations in the Korean data are quite scarce until mid 2000's).

Distance Measures

IV. Class-co-occurrence measure (Engelsman, van Raan, 1994)

- For all patents in class i filed in year t, we construct $C_{it} = [c_{i1t}, c_{i2t}, \ldots, c_{iNt}]$, where c_{ikt} = number of patents with multiple classes which have IPC class k as well as class i ($i \neq k$), c_{iit} = number of patents which have class i, and N = total number of classes.
- D(i,j,t) = cosine similarity of vectors C_{it} and C_{jt}
- We then take the weighted average of D(i,j,t) over the whole sample period.
- We use both Korean and US patent citation data for constructing two measures.

Estimation Strategy II

- We measure the drift of an inventor's specialization over the life cycle as follows.
  - Pick a patent (call patent T) filed in year t
  - Find patents filed within the range of 3,650 (i.e. 10 years) +/- 180 days after patent T's filing date and select one closest to the 10 year anniversary of T's filing
  - Pick the IPC class listed first on each patent
  - Using a distance measure between IPC classes, we calculate the distance between the two classes

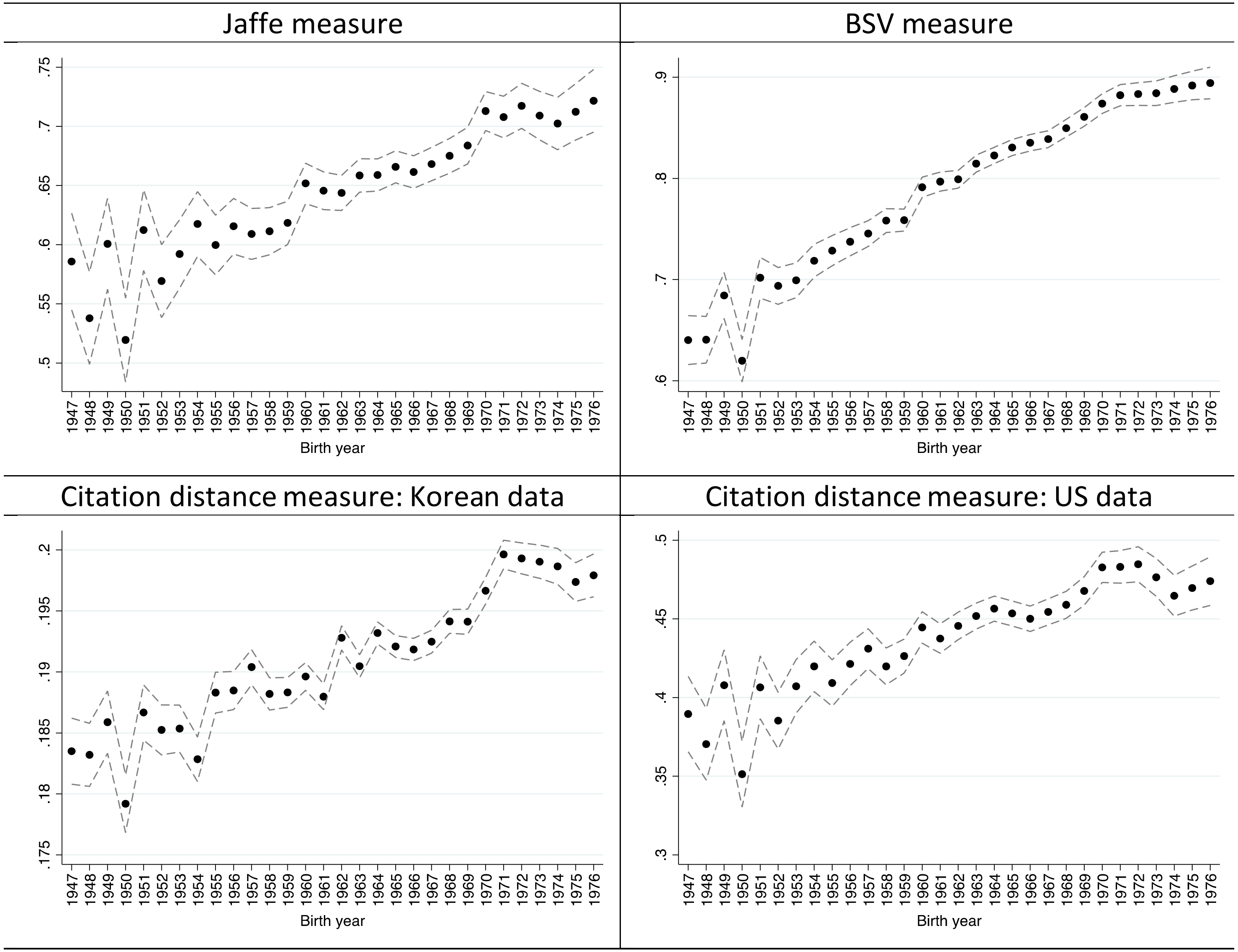
Regression Model I

- To investigate our questions 1 (cohort effects) and 2(age effects)
- Dependent var. = similarity of two patents by inventor k in year t
- Regressors: dummies for birth-year cohorts (1947 to 1976)

  dummy variables for inventor ages (25 to 53)

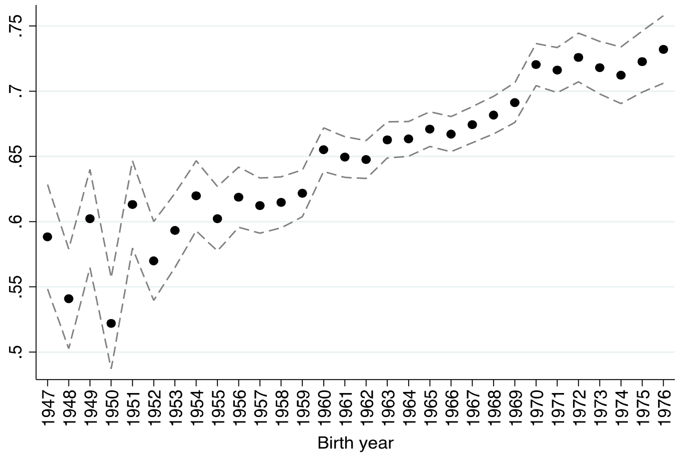
  time trend variable in linear, quadratic and cubic forms
- Sample: male only, inventors with the average annual number of patents $\geq$ 2 (47,951 Inventors)
- Interpretation
  - Cohort effects: degree of specialization at age 25 (market entry)
  - Age effects: age profile of specialization

Specialization by Cohort

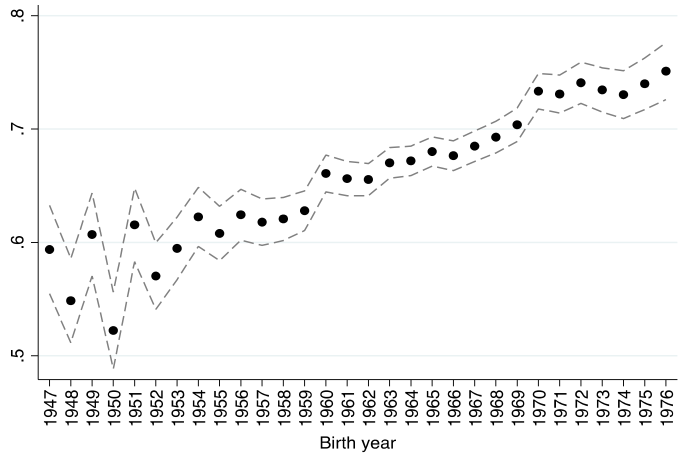


Specialization by Cohort

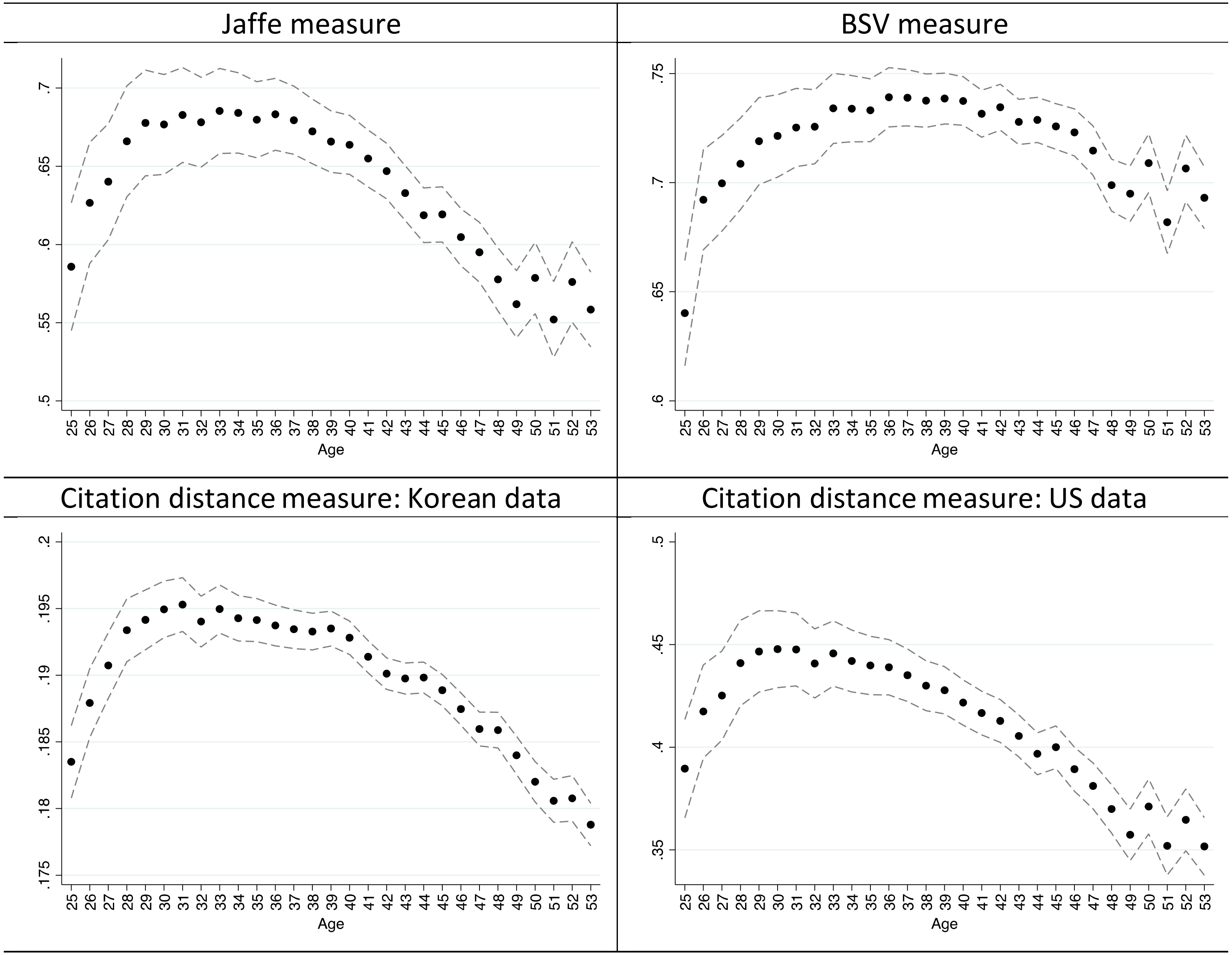
Class-co-occurrence measure: Korean data



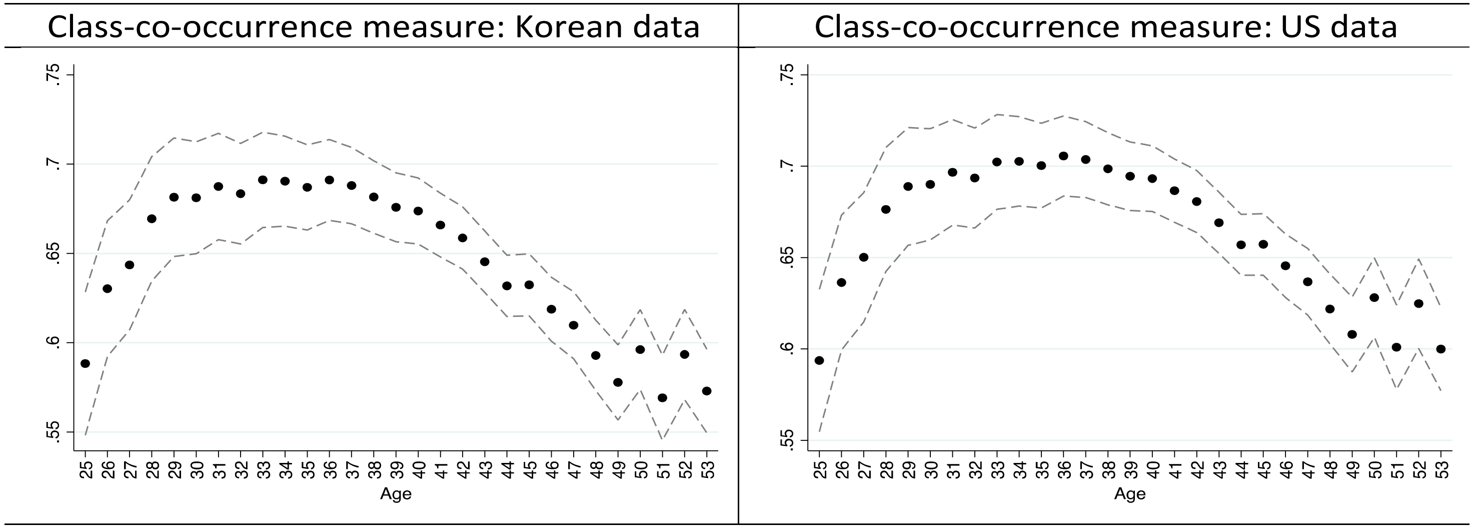
Class-co-occurrence measure: US data



Age Profile of Specialization



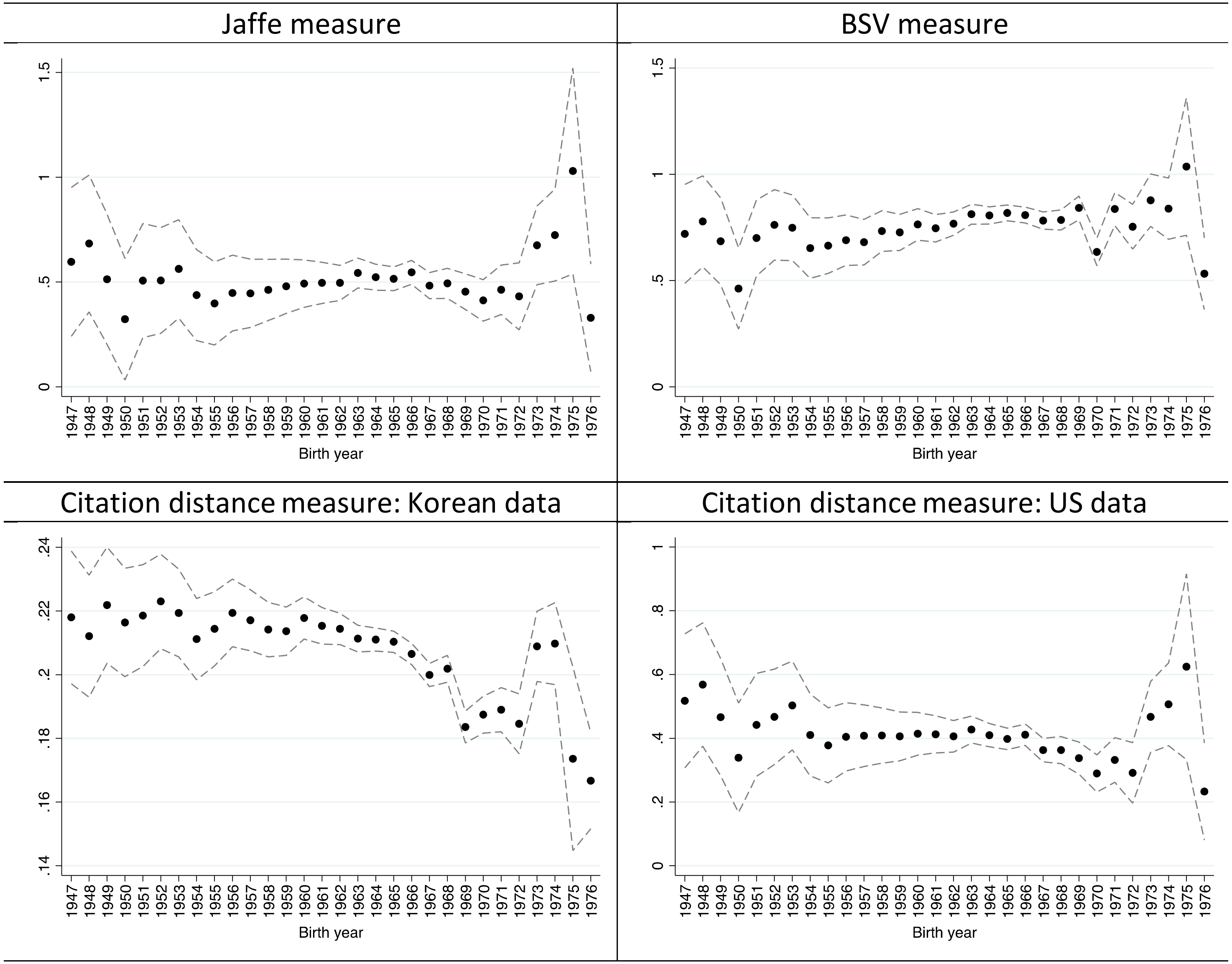
Age Profile of Specialization



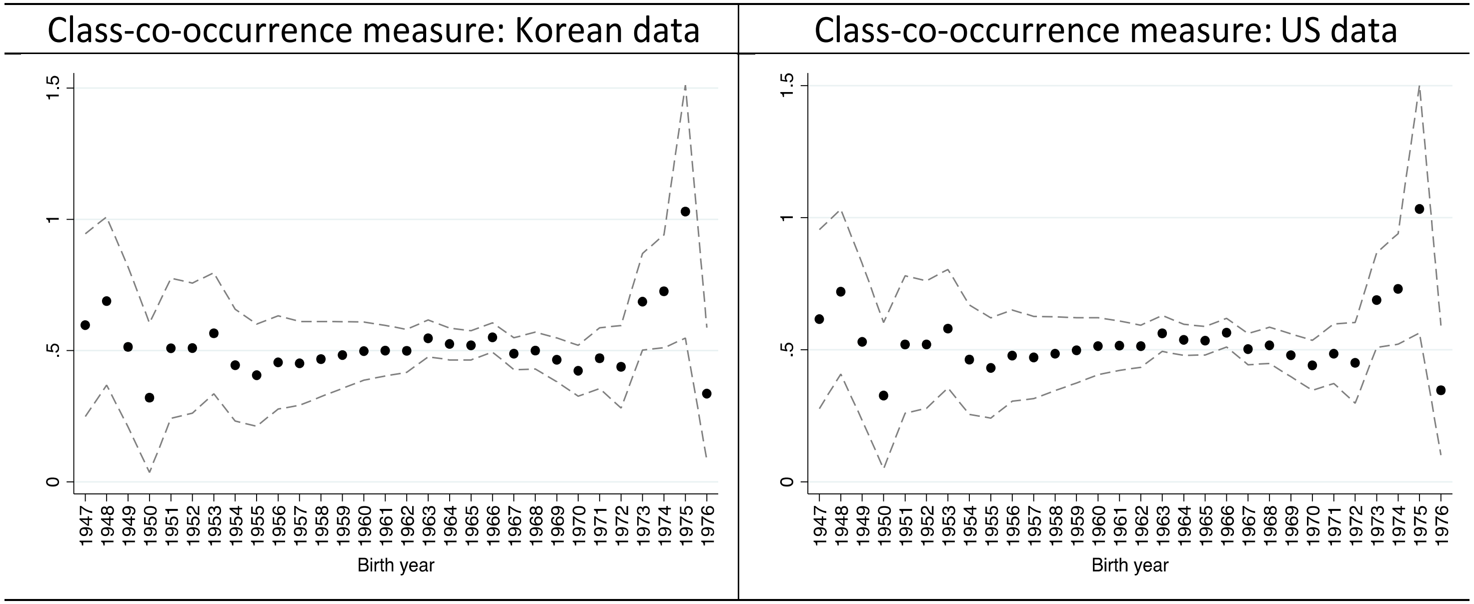
Regression Model II

- To investigate our question 3 (drifting in specialized fields)
- Dependent var. = distance between the classes of two patents, one in year t and the other in year t+10, with the time window of six month before and after
- Regressors: same as in Model I
- Sample: male only, NOT restricted to those with the average annual patent count ≥ 2
- Interpretation
  - how far an inventor moves in terms of research fields over his career

Drift in Specialized Fields

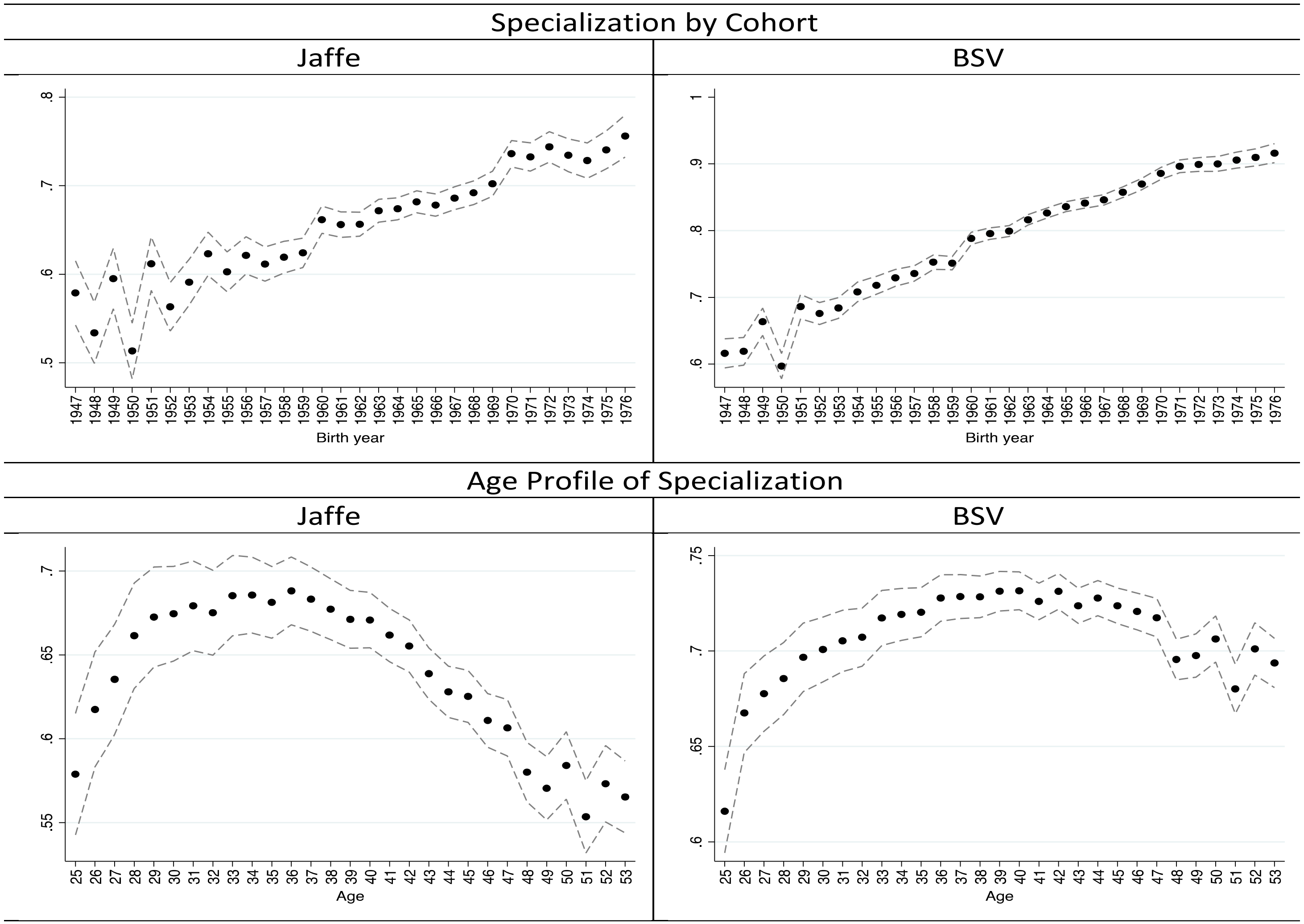


Drift in Specialized Fields



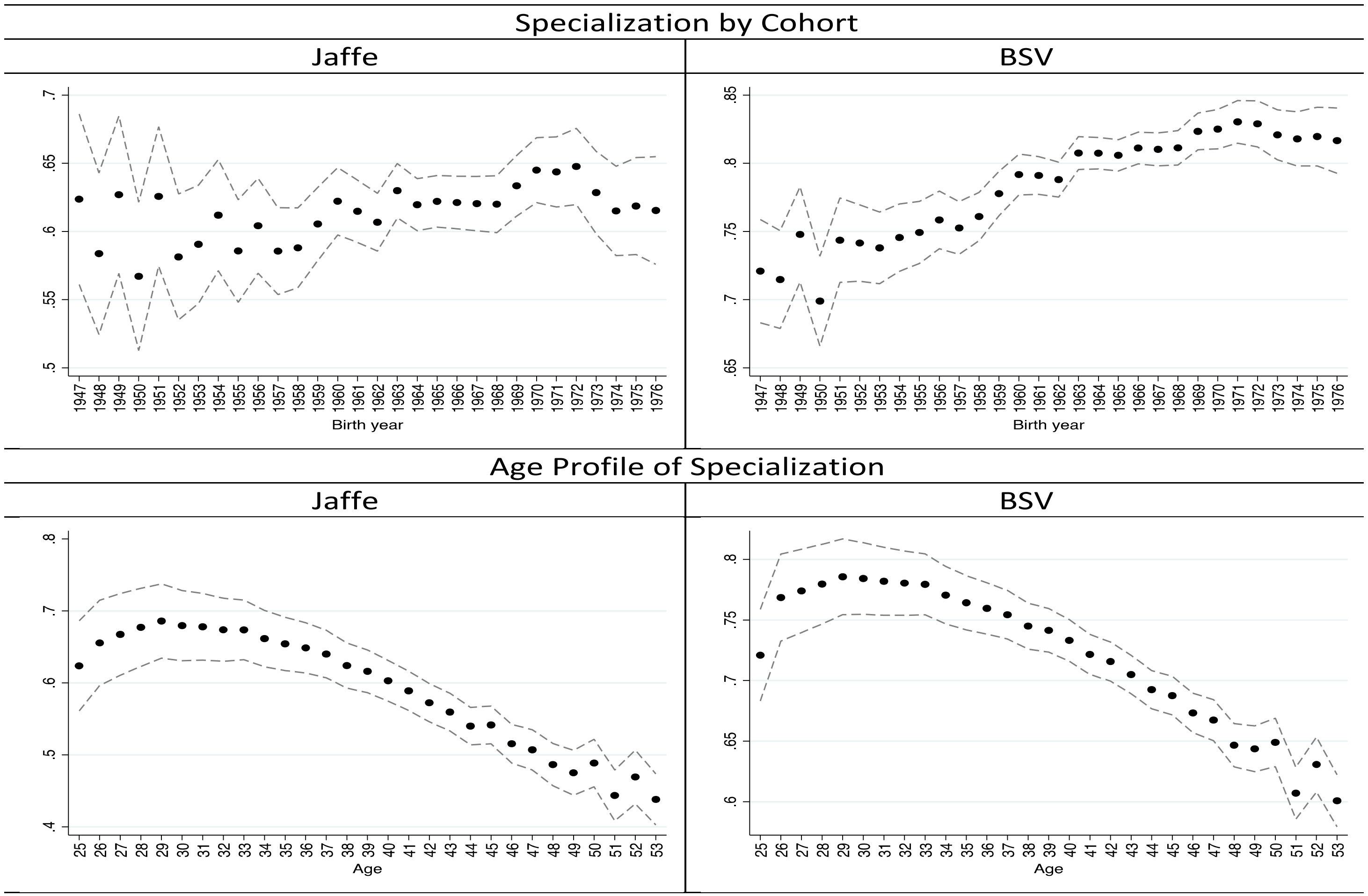
Sensitivity Analysis I

- Pick 3 or 5 patents instead of 2 in Model I



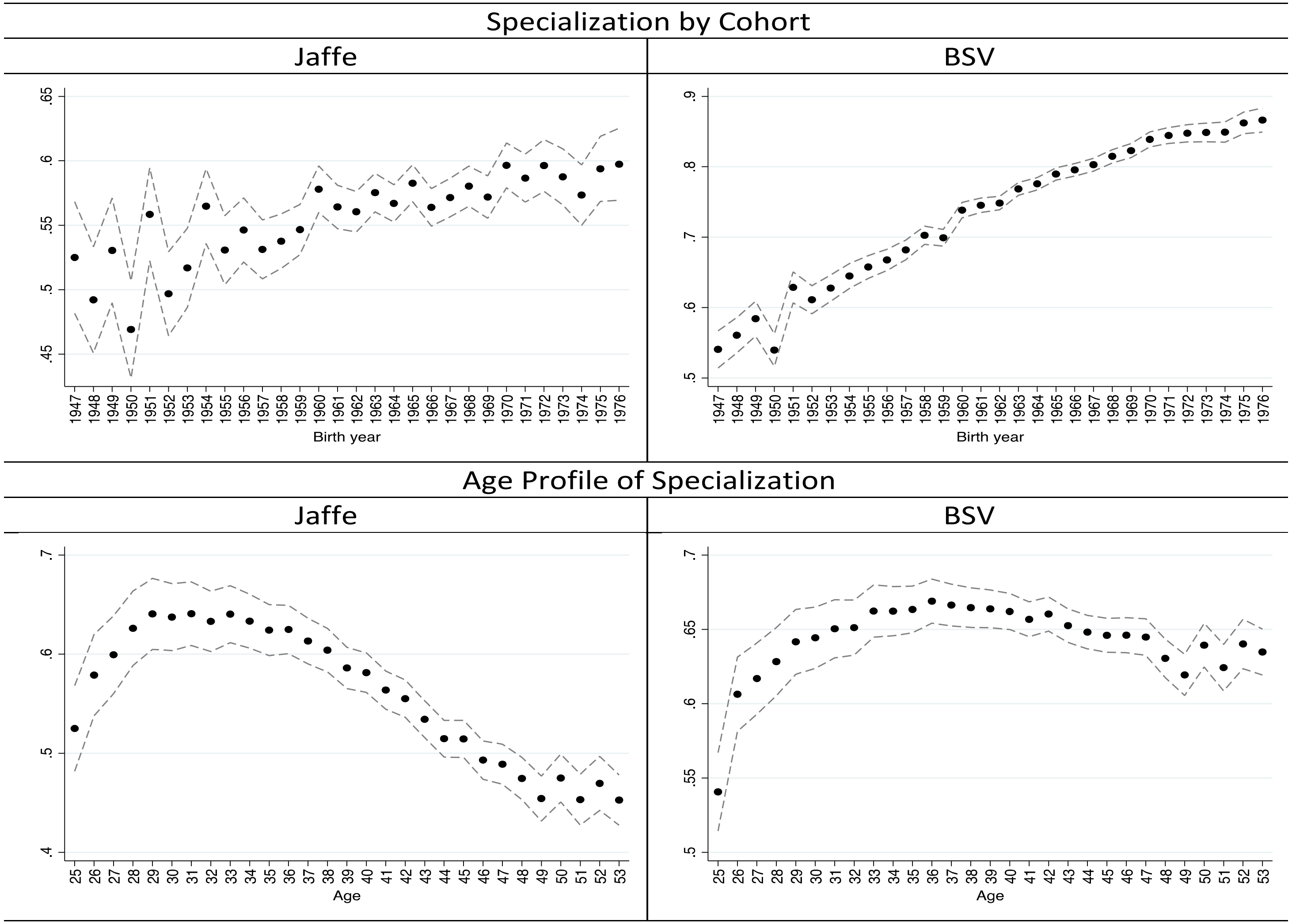
Sensitivity Analysis II

- Application bunching: exclude sample whose difference in application dates <= 16 (median) days among two consecutive patents



Sensitivity Analysis III

- IPC 3rd-level class instead of 2nd-level class



Sensitivity Analysis IV

- Additional regressor: inventor team size