

A study of the correlation between number of classification symbols and patent citation count

Max Kuan

National Taiwan University of Science
and Technology



Outlines

- Background
- Methodology
- Result
- Conclusion



Background



Patent Classification

Current U.S. Class:	715/863; 345/173; 345/179
Current CPC Class:	G06F 3/04883 (20130101); G06F 21/36 (20130101); H04M 1/663 (20130101); G06F 3/0488 (20130101); G06F 3/017 (20130101); G06F 3/0484 (20130101); G06F 3/04842 (20130101); H04M 1/67 (20130101); H04M 1/575 (20130101); H04M 2250/22 (20130101)
Current International Class:	G06F 3/033 (20060101)

- Every patent is assigned **one or more classification symbols** during its application process by examiner according to the patent's disclosed invention and **a standard scheme** such as IPC (international patent classification), CPC (cooperative patent classification), etc.



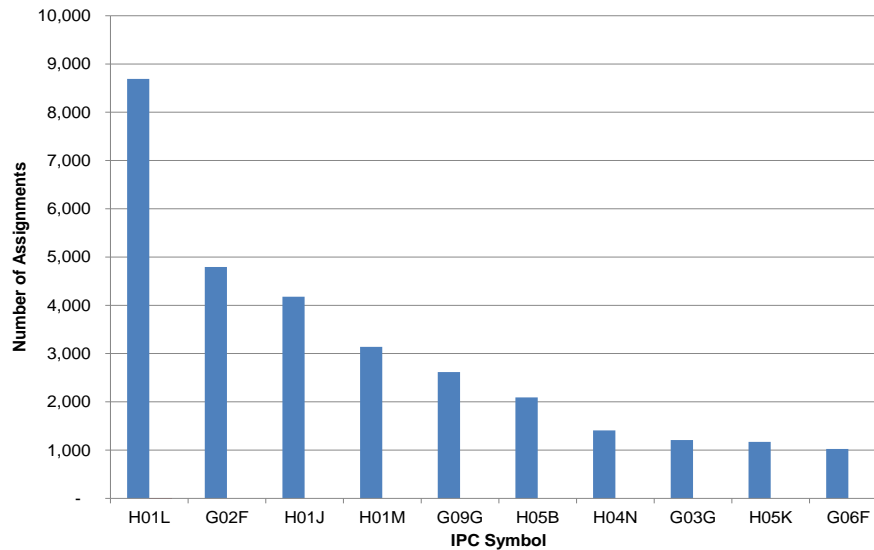
Importance of Patent Classification

- Patents' classification symbols are a valuable source of information
 - determined by **professional examiner**
 - representative of the patents' **technical contents**
 - based on a **common standard**



Various of Applications of Patent Classification

- Detecting the **R&D focus** of assignees, institutions, ...



- Measuring and evaluating **similarity or relatedness** between patents, portfolios, assignees, institutions, ...



Patent Technology Scope (PTS)

Current U.S. Class:	713/189
Current CPC Class:	G06F 21/78 (20130101)
Current International Class:	G06F 21/00 (20130101)
Field of Search:	;713/189

Current U.S. Class:	715/863; 345/173; 345/179
Current CPC Class:	G06F 3/04883 (20130101); G06F 21/36 (20130101); H04M 1/663 (20130101); G06F 3/0488 (20130101); G06F 3/017 (20130101); G06F 3/0484 (20130101); G06F 3/04842 (20130101); H04M 1/67 (20130101); H04M 1/575 (20130101); H04M 2250/22 (20130101)
Current International Class:	G06F 3/033 (20060101)

- PTS=the number of **different classification symbols**



Conflicting Ideas

- Some consider that PTS **reflects the diversity** of the patent's content
- Some consider that the existing patent classification schemes are “never intended to provide conceptual delineations of technology areas, but instead identify inventions by function at very low levels of abstraction in order to **serve as aids to prior art searching** (Allison et al.)



PTS is indeed dubious

Current U.S. Class:	715/863; 345/173; 345/179
Current CPC Class:	G06F 3/04883 (20130101); G06F 21/36 (20130101); H04M 1/663 (20130101); G06F 3/0488 (20130101); G06F 3/017 (20130101); G06F 3/0484 (20130101); G06F 3/04842 (20130101); H04M 1/67 (20130101); H04M 1/575 (20130101); H04M 2250/22 (20130101)
Current International Class:	G06F 3/033 (20060101)

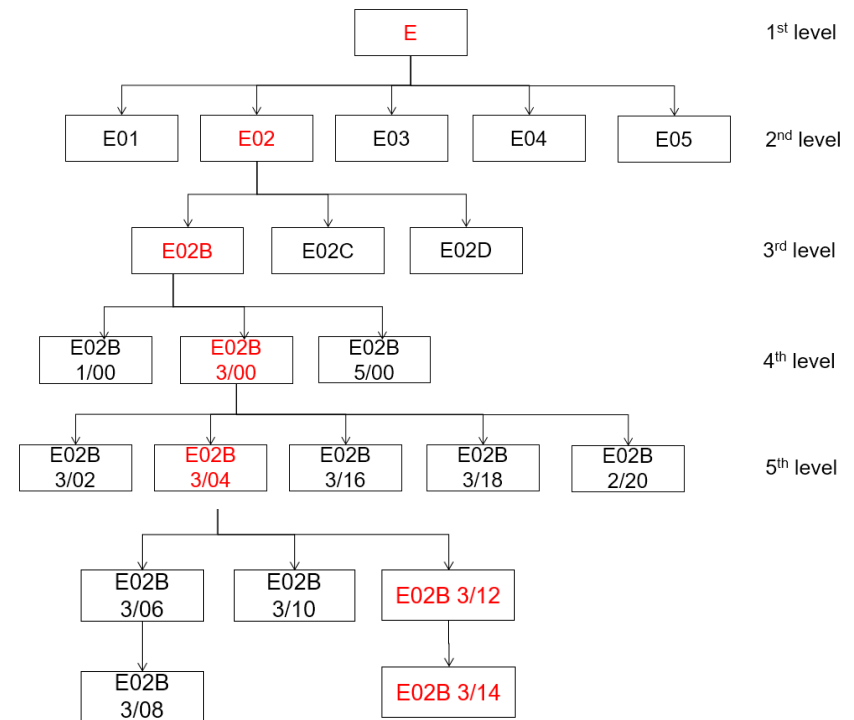
- **Dependent on the scheme used**
 - Using IPC, PTS=1; Using CPC, PTS=10
- **Dependent on the granularity of classification**
 - if the CPC symbols are considered to the more coarse sub-class level (i.e., the 3rd level),
 - PTS is 2 as there are two distinct symbols “G06F” and “H04M”
 - if the CPC symbols are considered to the finer sub-class level (i.e., the 4th level and the digits before the “/”),
 - PTS becomes 4 as there are 4 distinct symbols “G06F 3,” “G06F 21,” “H04M 1,” and “H04M 2250”



PTS is indeed dubious

- Symbols may have **superordinate/subordinate relationship** and **different degrees of relatedness**

– So counting each different symbol as 1 may not be fair



So the question is:
Is PTS really a valid indicator?



Why the answer is important?

- PTS is often used
 - alone or integrated with other indicators
 - to reflect the value of a patent
- Patents provide delayed information
 - The only way to shorten the delay is to use the so called ***published patent application*** published after 18 months of filing
 - ***Published patent application*** also has classification symbols
 - So, PTS may be applied to ***Published patent application***



Methodology



We investigate

- Not whether PTS really reflects diversity
- But **whether PTS has a positive correlation with the patent citation count**
 - Patent citation count = how many times a patent is cited (i.e., *forward citations*)
 - Widely accepted as an indicator to the patent's impact, influence, quality, value, ...
 - If PTS has a positive correlation with the patent citation count -> PTS indeed reflects some nice quality of the patent



Empirical data

- U.S. utility patents issued in the years 2007, 2009, and 2011
- Their citation counts up to Dec. 31st, 2013
 - Patent data from three different years are used because patent citation counts need time to accumulate.
- Their CPC symbols in levels 3, 4, 5
 - But we ignore their superordinate/subordinate relationship, and relatedness



Result

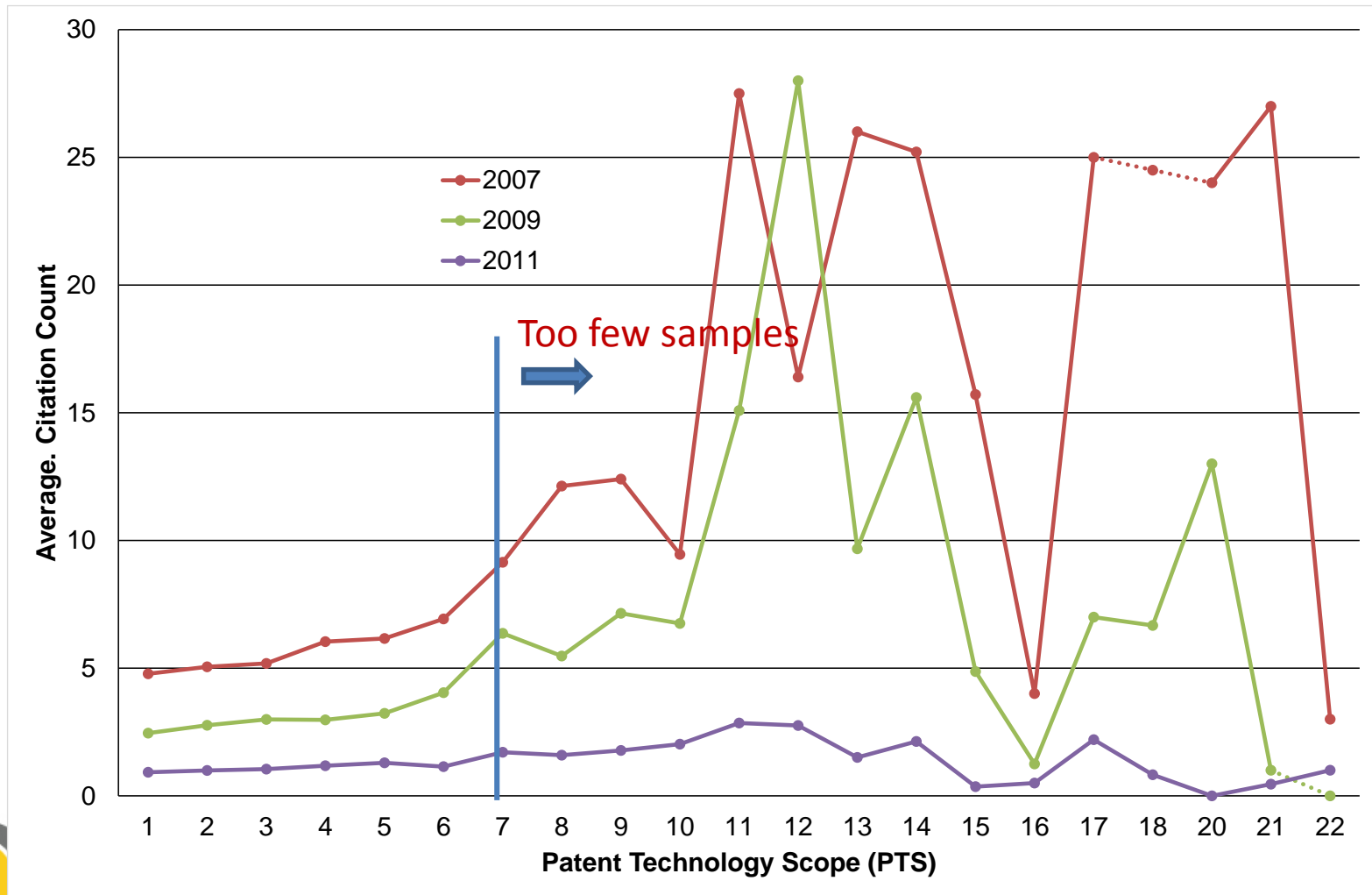


Summary of Data

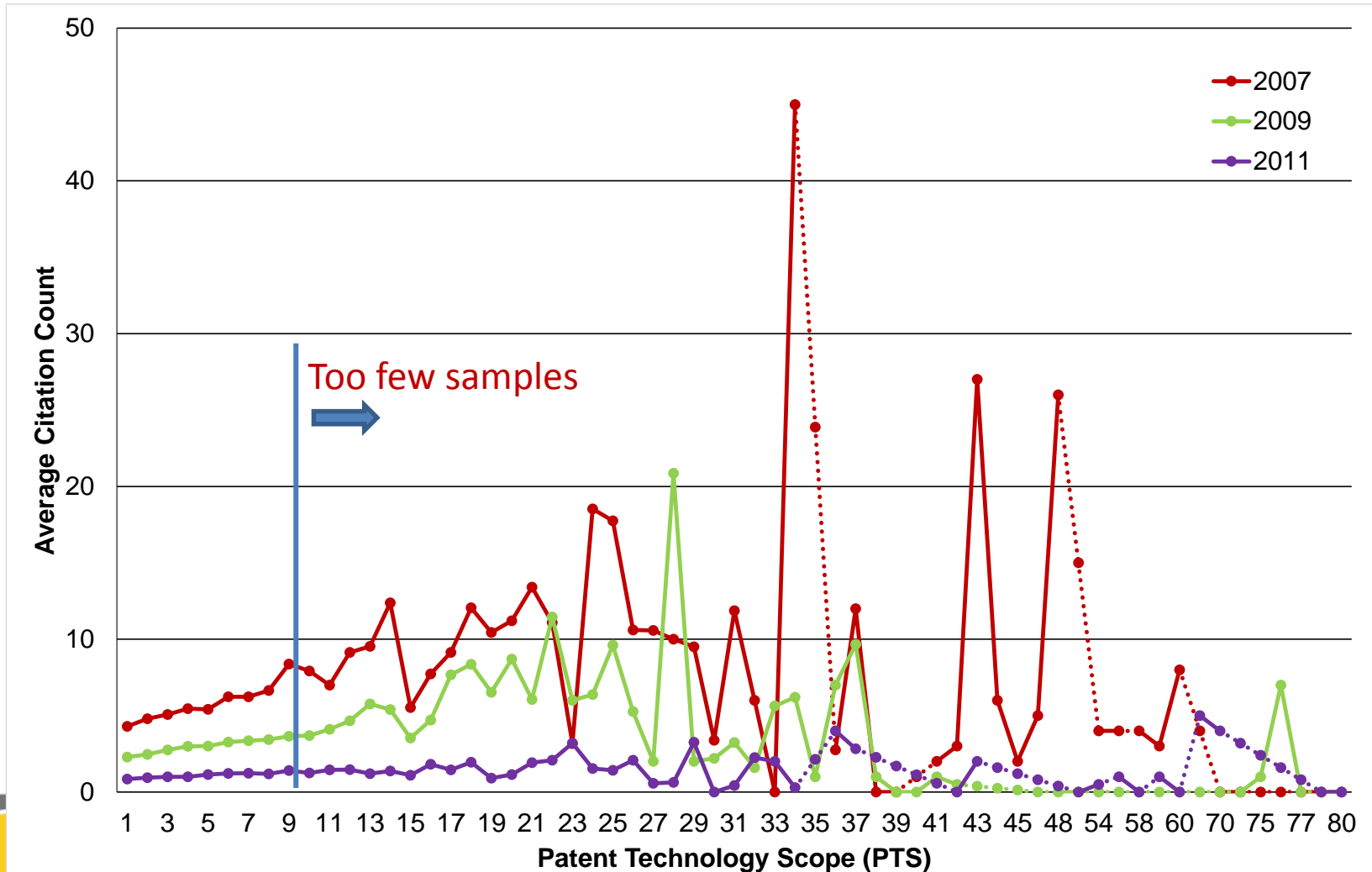
Year	2007	2009	2011
Patents	137,720	152,280	208,124
Citation count – Mean	5.055	2.672	0.978
Citation count – Max	392	352	90
Citation count – Min	0	0	0
Citation count – Std. Dev.	10.73	6.39	2.62
CPC – Mean (Level 3/4/5)	1.84/3.00/6.17	1.78/2.93/5.99	1.82/3.04/6.30
CPC – Max (Level 3/4/5)	22/70/389	21/77/390	22/80/1095
CPC – Min (Level 3/4/5)	1	1	1
CPC – Std. Dev. (Level 3/4/5)	1.17/2.45/8.56	1.13/2.4/8.49	1.18/2.46/9.17



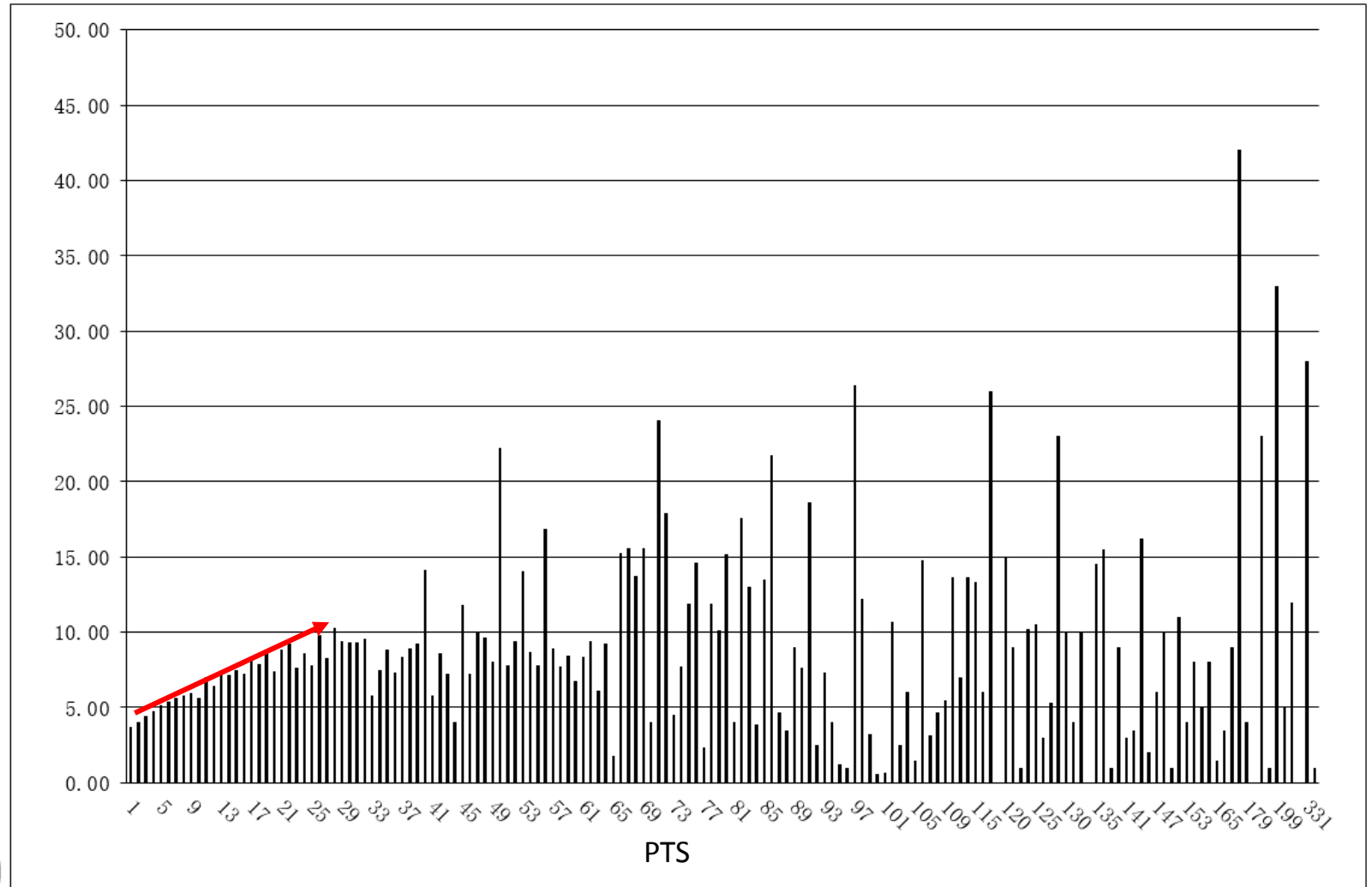
Observation: CPC 3rd level



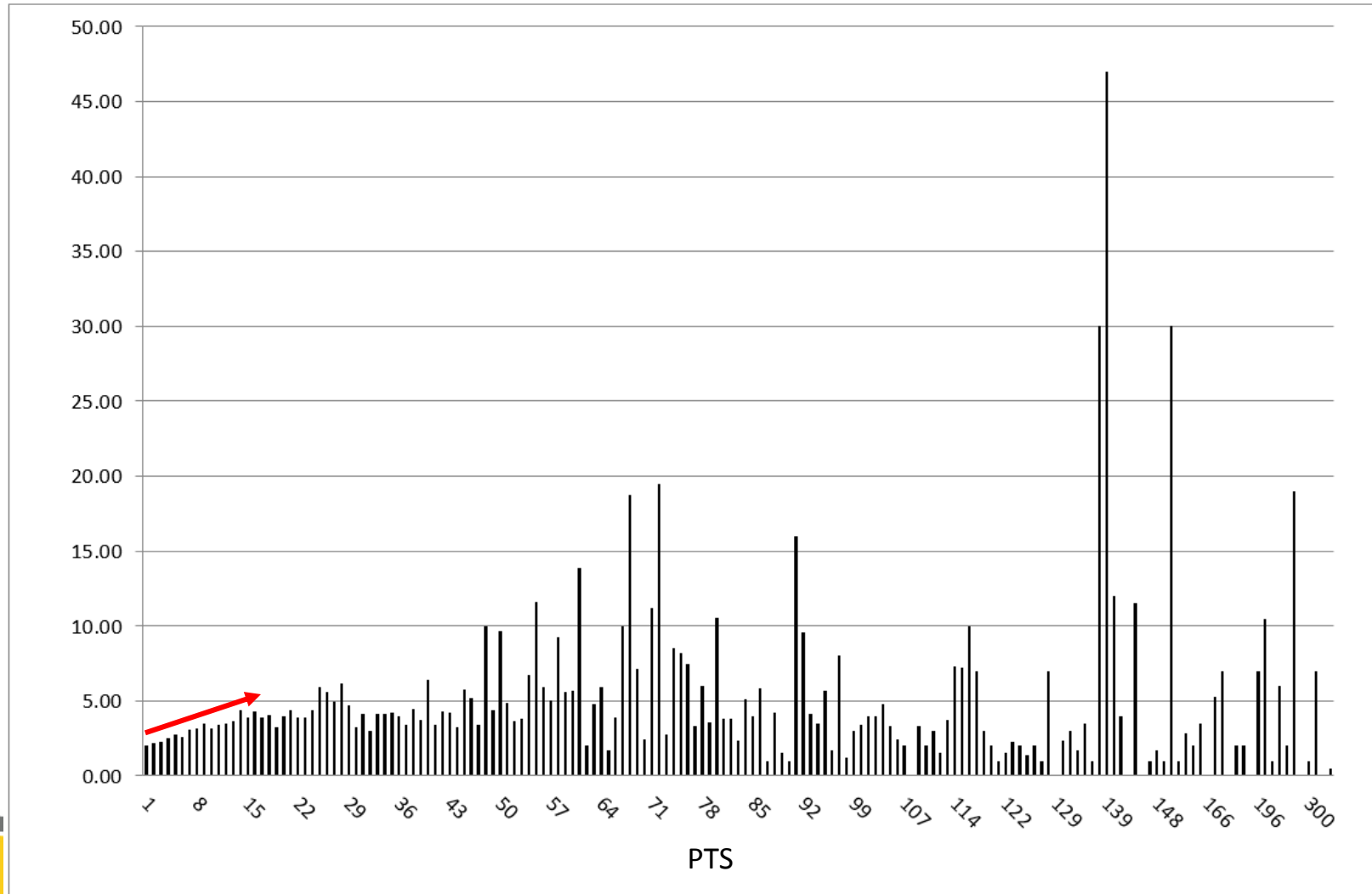
Observation: CPC 4th level



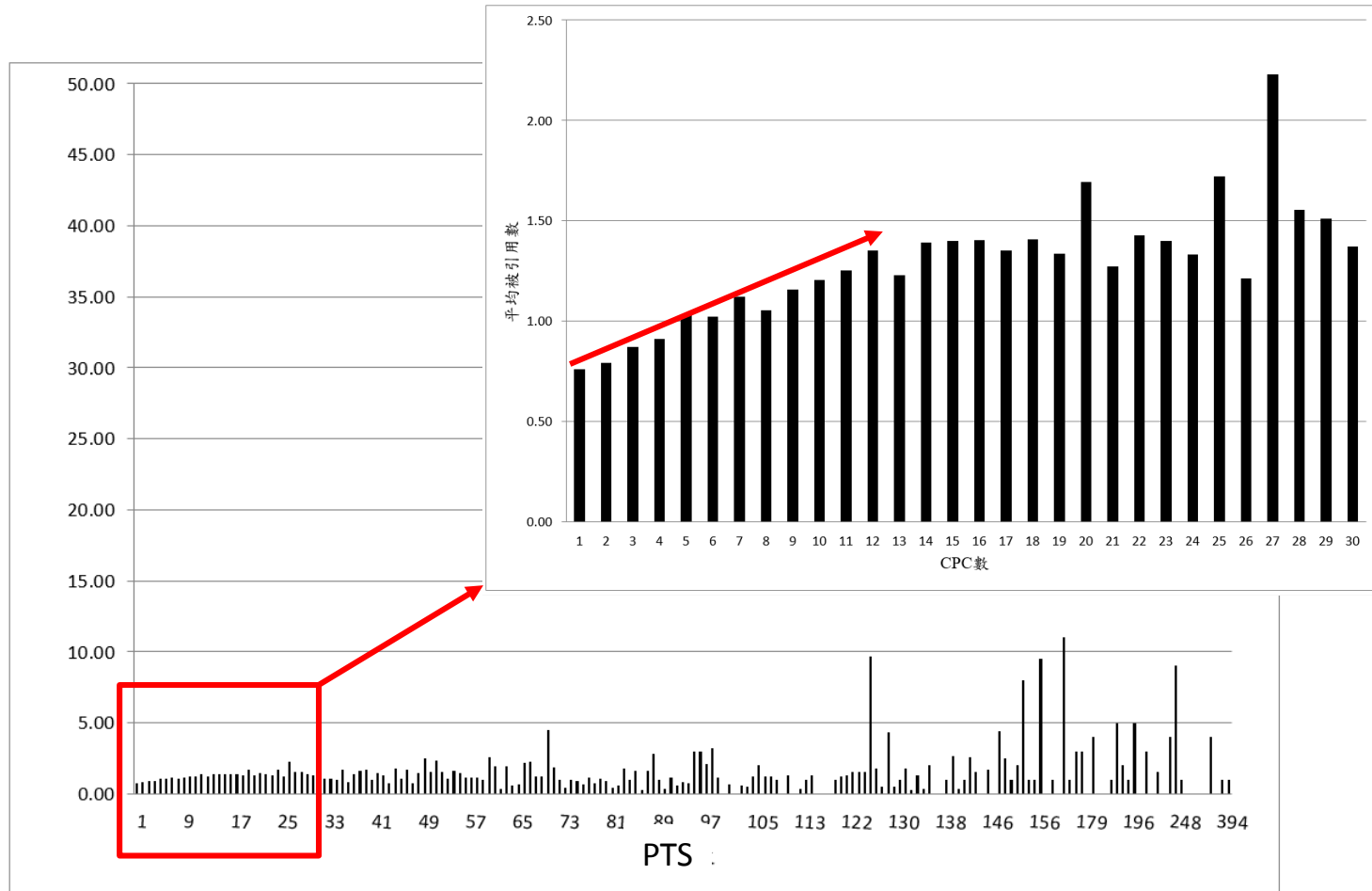
Observation: CPC 5th level, 2007



Observation: CPC 5th level, 2009



Observation: CPC 5th level, 2011



Pearson's correlation coefficient

	2007	2009	2011
For PTS using 3rd- level symbols	0.053**	0.059**	0.037**
For PTS suing 4th- level symbols	0.077**	0.071**	0.047**
For PTS suing 5th- level symbols	0.098**	0.087**	0.051**

**p-value less than 0.01 significance level

PTS is indeed positively correlated to the citation count



Conclusion

- PTS indeed captures some characteristics of patents.
- For PTSes greater than a **threshold**, there is actually not enough statistical evidence to support that higher PTSes imply more citation counts due to too few samples.



For PTS to be statistically reliable

CPC	3 rd Level	4 th Level	5 th Level
2007	≤ 8	≤ 10	≤ 18
2009	≤ 8	≤ 11	≤ 22
2011	≤ 6	≤ 7	≤ 10



Conclusion

- Due to the limited correlation coefficient values, the **two patents' PTSES should have a greater difference** (e.g, one is 6 and one is 2) so that a more confident conclusion may be drawn.
- Our observation would be especially helpful **when evaluating young patents** that are issued for only a limited period of time
 - As they are too young to accumulate meaningful citation counts, and as such they cannot be differentiated reliably using citation counts



Implication

- A patent has a higher number of citations because
 - It is really more influential? or
 - Its content is more diversified, and therefore has a higher chance being cited by more patents from different technical areas?
- Perhaps a **modified patent citation count** may be developed



Thank You

