

Analyzing the Innovative Activity of Young Businesses: Perspectives and Difficulties using Administrative Data

Claire Lelarge (STI/EAS)

**21st Meeting of the Wiesbaden Group
on Business Registers**

Session 2b: The user perspective from analytical needs

November 24th, 2008

Outline

- **Background: the OECD WPIA micro-data project about Entrepreneurship**

Description of the preliminary experiment using French data (CREST-OECD):

Linking Patenting Information to Business Registers

- **Preliminary decisions regarding relevant concepts**
- **Description of the matching method**
- **Quality checks**
- **First analytical insights**
- **Concluding remarks**

**Analyzing the Innovative Activity of Young Businesses:
Perspectives and Difficulties using Administrative Data**

Background:

**The OECD-WPIA Project about
Entrepreneurship**

WPIA Projects about Entrepreneurship and the Growth of New Firms

Participating countries: Finland, France, Germany, Spain, Sweden, Switzerland, US

Possibly participating: Australia, (Austria), Canada, Czech. Rep., Israel, Japan, Netherlands, New Zealand, Poland, Portugal, UK

Precise Analytical Question	What is the novelty?	Data requirements
Inventing and patenting activities of new firms	New thematic, new data, new information	- (Longitudinal) database including new ventures since their birth
Specificities of the financing structure of new firms	Harmonized information across OECD countries	- Detailed firm-level financial information - Matched information about firm age

“Inventing and patenting activities of new firms”

Main requirement: longitudinal database including start-ups since their birth, to be matched with Patent information

- France (INSEE-CREST): Exhaustive business registers
- US: Integrated Longitudinal Business-Level Database at the Census Bureau (Davis, Haltiwanger, Jarmin, Krizan, Miranda, Nucci, Sandusky)
- Sweden (Lööf and Andersson): use of a panel of 20 000 manufacturing firms with 1-25 employees over 2000-2004.

Complementary:

Accounting / R&D / Skills information?

Alternatively: any data source containing firm-level information (in particular about innovation) together with information about firm age, **provided the population of young firms is covered**

**Analyzing the Innovative Activity of Young Businesses:
Perspectives and Difficulties using Administrative Data**

**First Experiments Using French Data
(*OECD-CREST*)**

Preliminary decisions regarding relevant concepts

- **Files to be matched:**
 - **OEB applications / with SIREN ID numbers**
(based on previous work at Insee)
 - **Exhaustive business registers:** date of creation, employment
- **Files to be matched:**
 - **No mapping between the statistical units** of each file...
 - Decision to try to construct a version of the EPO file **at the firm level...** (driven by the analytical needs)
 - **But not straightforward; shortcomings**
(e.g. group strategies, etc.)

Description of the matching procedure

- Information used to guess the identifier:
 - **Names (of applicants / companies)**
 - **Complete address (city, zipcode, street)**
- **Cleaning** / harmonization and “**Sirenage**” procedure
(Service offered by the French NSI, the INSEE)

Main issue with this strategy:

The **1993-2003** applicants were compared to the **2006** version of the business register

-> Impossible to take account of important phenomenon (**restructuring**, etc.)

-> **Solution** (cf INPI strategy): perform the operation **year after year** (instead of at one single point in time)

Result of the Matching Procedure

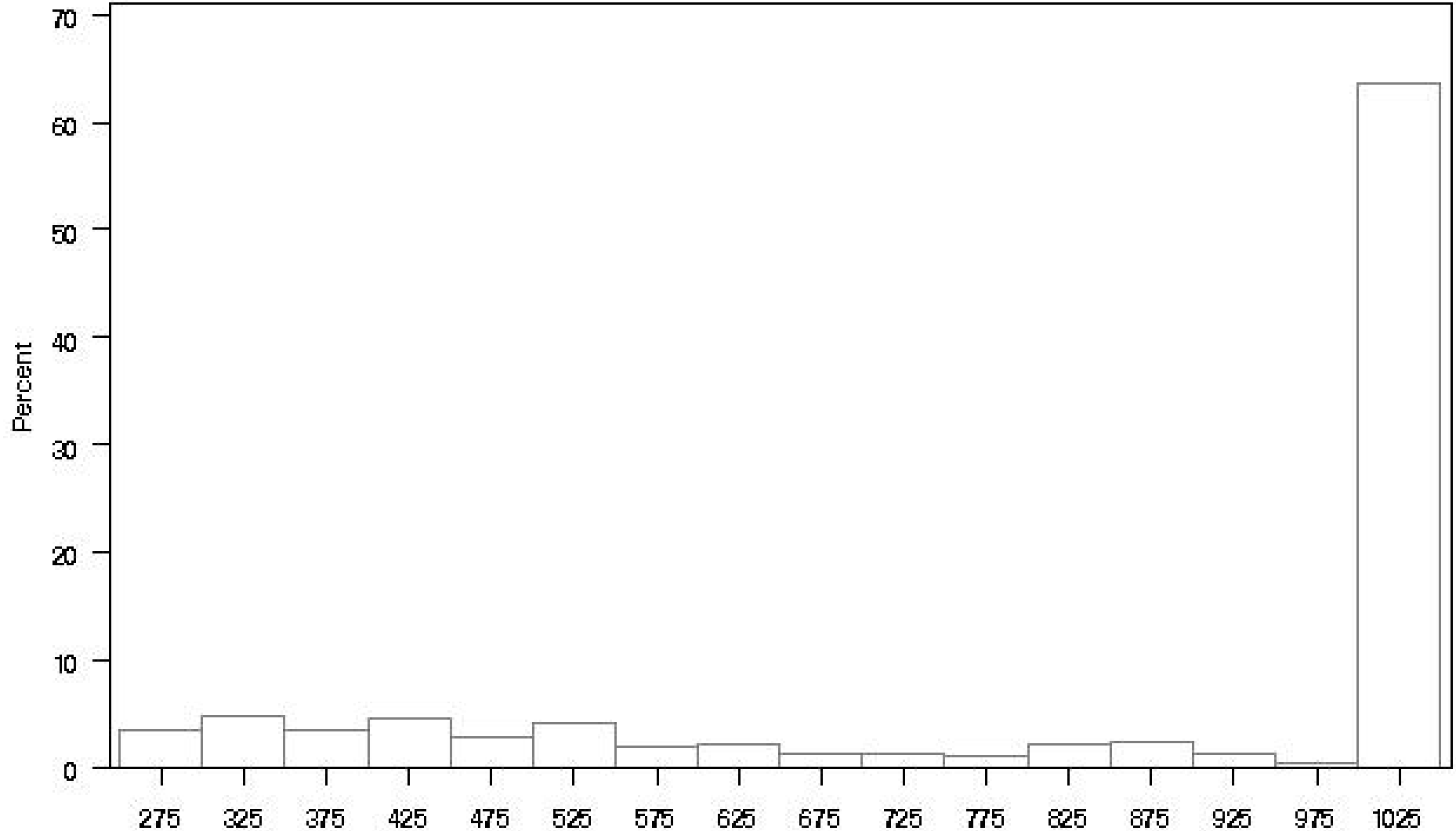
Source	# Different EPO Identification Codes	Corresponding # of Patent Applications	Quality	
Initial EPO File to be “Identified”	14 495	61 683	-	
Automatic Treatment (“Sirenage”)	9 011	44 462	To be Checked (see below)	Total Processed: 95% Applicant Codes (Applicants) 96% Patent Applications
Manual Treatment (Web, BR Files inc. Historical Files)	1 613	10 881	Good	
Identified Individuals (as opposed to Firms, i.e. not relevant)	3 092	4 026	Good	
No Match Proposed (neither Automatic, nor Manual)	779	2 314	-	Residual



Quality Check (1): Use of the Matching Scores

SCORE = 1,000 * (# matched characters / total # of characters)

Source: C. Lelarge, OECD-CREST



Quality Check (2): Use of “Out-of-Sample” Information

The age information available in BR provides a quality check of the matching between the OEB file and the French BR

-> Restriction to the 1997/2000 period

Source: C. Lelarge, OECD-CREST

Number of Years	Number of Firms in OEB file	Firms Lost in Matching with BR	%	Age is In-consistent	... After Firm Birth	Final Dataset
1	2923	208	7.12	364	254	2461
2	655	64	9.77	73	50	541
3	274	29	10.58	27	22	223
4	300	39	13	26	22	239
# Firms	4152	340	8.19	490	348	3464

**Analyzing the Innovative Activity of Young Businesses:
Perspectives and Difficulties using Administrative Data**

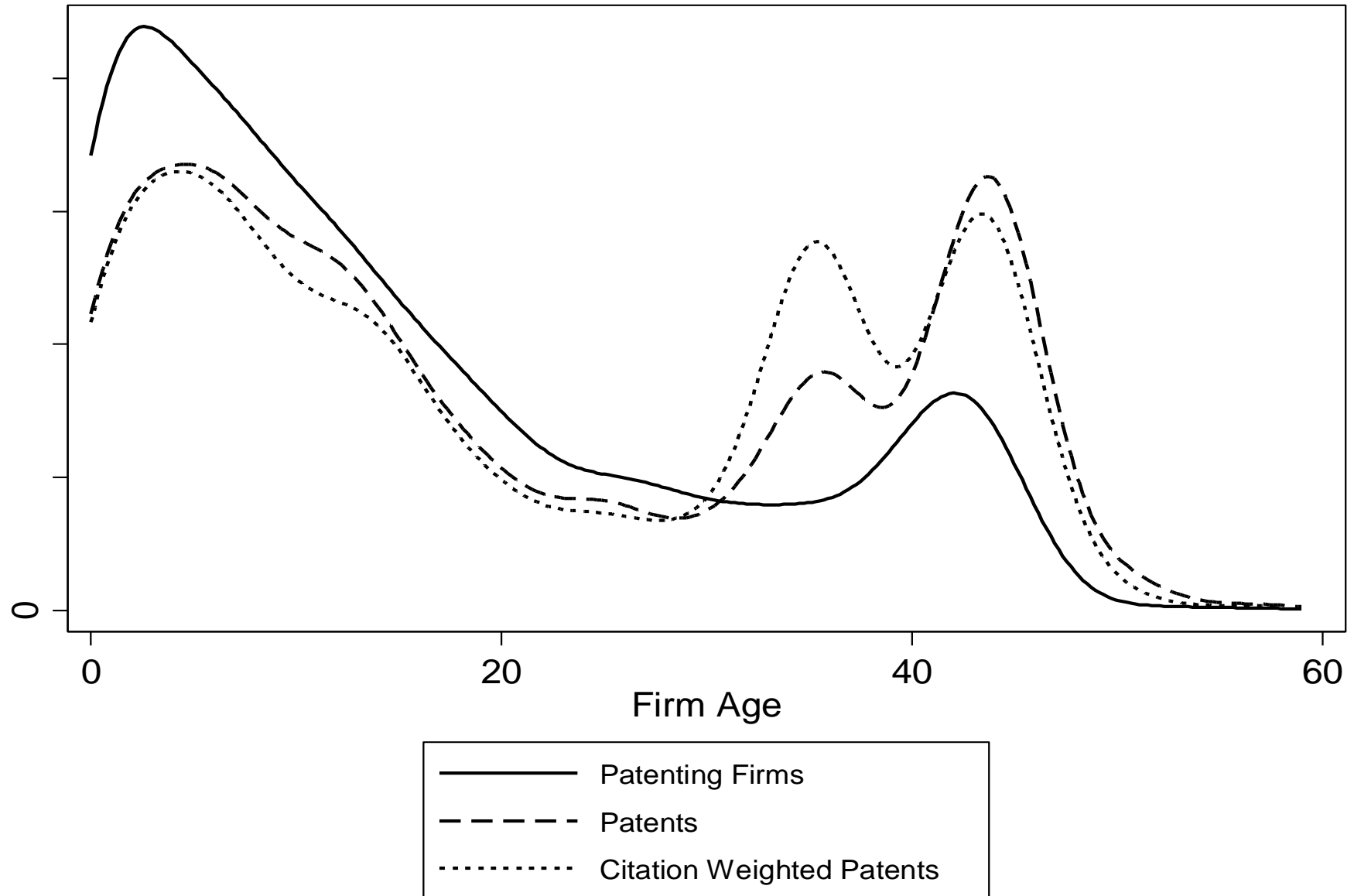
Preliminary Analytical Insights (*OECD-CREST*)

Analyzing the

Inventing and patenting activities of new firms

- **Description of new firms' innovation through patent information:**
 - Distribution of patent filings by age
 - Description of the characteristics of the inventions: scope, radicalness, etc.
- **Further steps: the timing of inventive activities and the life cycle of the firm:**
 - Determinants of patenting / inventive activities in relation with firm age
 - Inventive performance and firm survival (duration analysis)

EPO Applications (based on priority date) over the 1997-2000 period

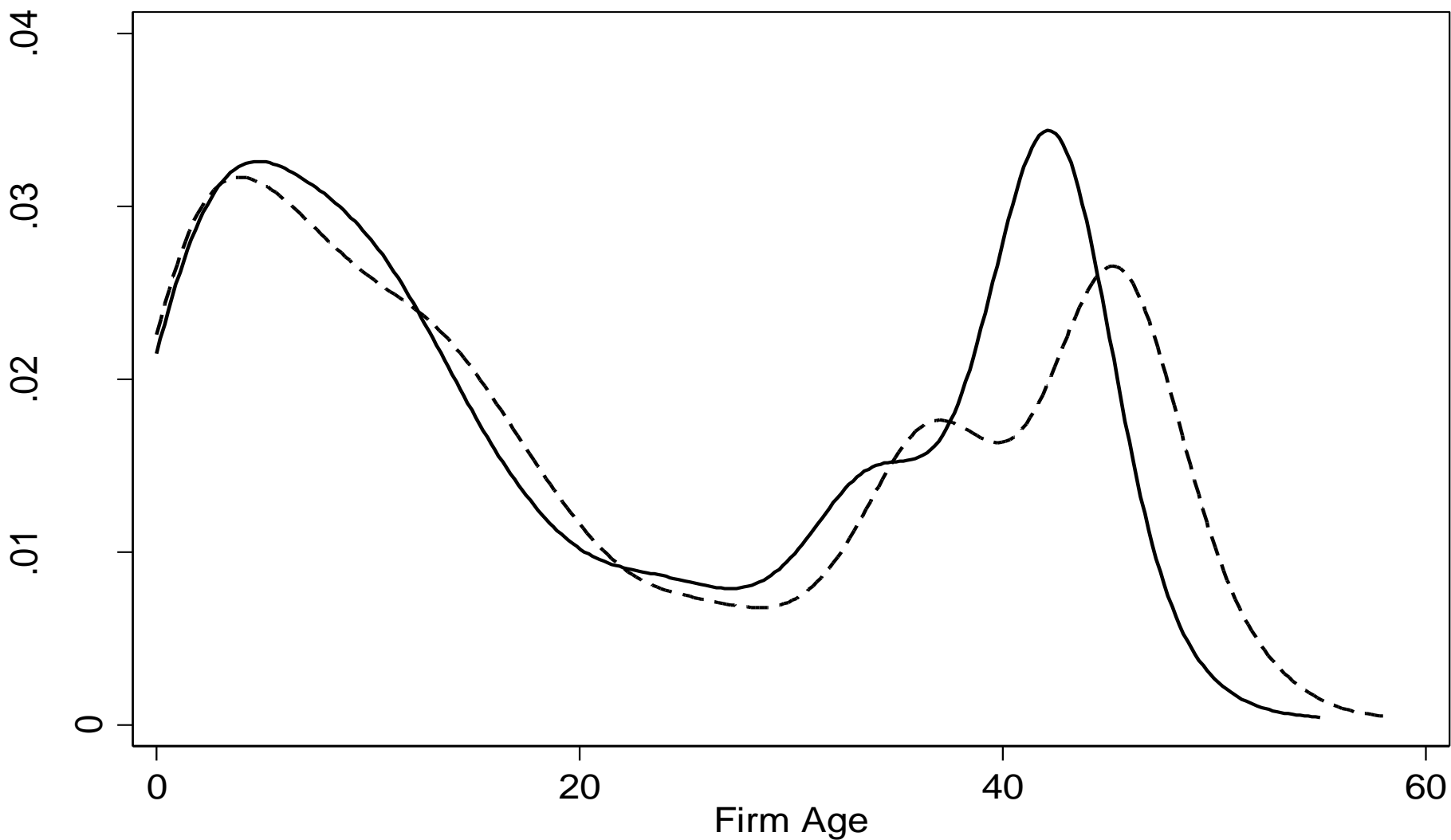


Longitudinal Analysis

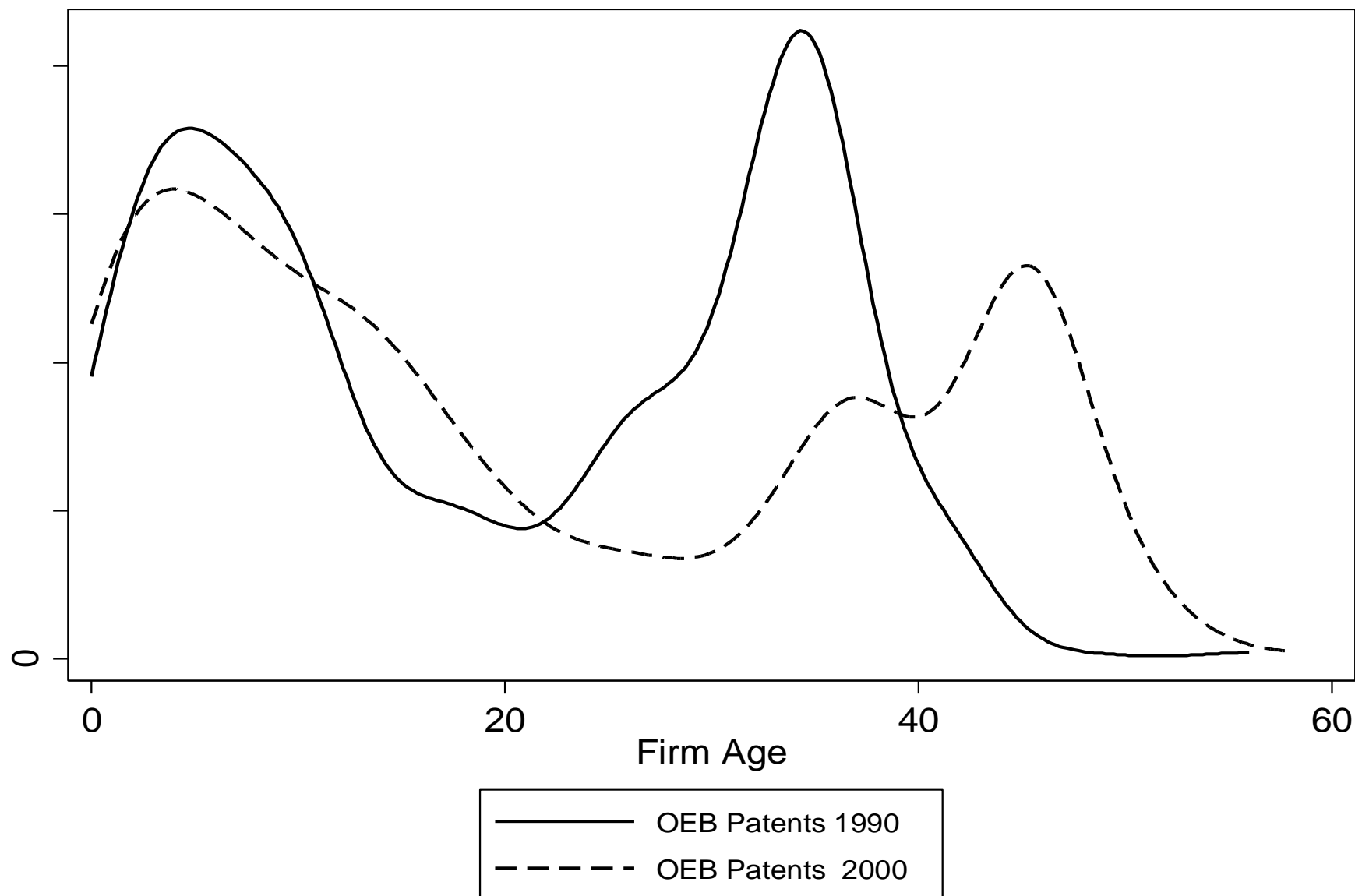
Relying on an additional dataset relating to the 1990's,

It is possible to compare the patenting behaviour between 1990 and 2000

EPO Patent Applications (based on priority date) 1997 versus 2000



EPO Patent Applications (based on priority date) 1990 versus 2000



Concluding Remarks

Experience from the matching between two independent administrative datasets:

- French business registers
- File containing all French patent applications to the European Patent Office for the year 1993 to 2000.

Main difficulties arising from the fact that:

- Observation units in each file are not identical
- Large size of each dataset
- Limited information available for identification and matching (name, address).

Special care should be devoted to the implementation of the matching procedure:

Possible selection (e.g. stock sampling) ¹⁸ biases

Thank you!

Analyzing the Innovative Activity of Young Businesses: Perspectives and Difficulties using Administrative Data

Claire Lelarge (STI/EAS)

**21st Meeting of the Wiesbaden Group
on Business Registers**

Session 2b: The user perspective from analytical needs

November 24th, 2008