ESTIMATING INVESTMENT IN TRAINING USING PIAAC DATA

Mariagrazia Squicciarini, Luca Marcolin, Peter Horvát

OECD Directorate for Science, Technology and Innovation
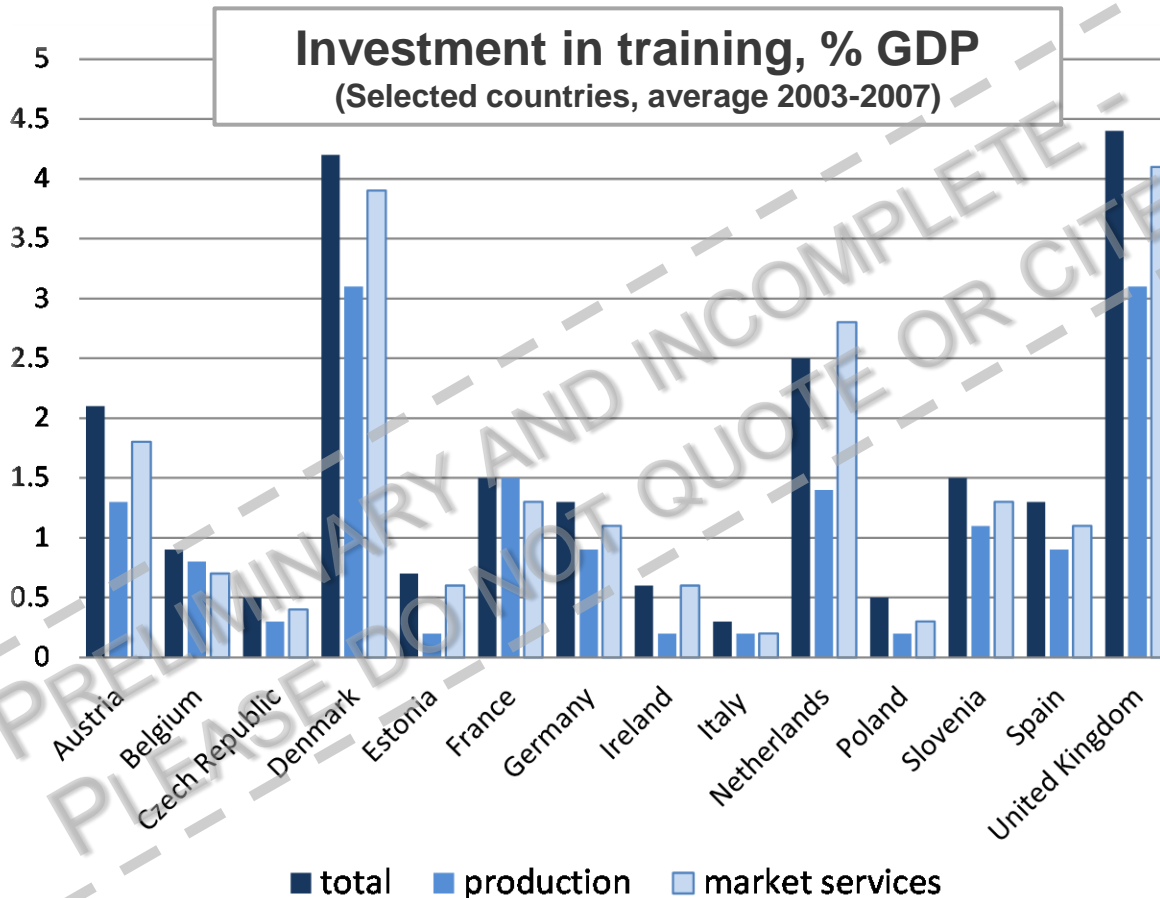Economic Analysis and Statistics Division

PRELIMINARY AND INCOMPLETE -
PLEASE DO NOT QUOTE OR CITE

SPINTAN meeting, London, 23-24 April 2015

OECD
BETTER POLICIES FOR BETTER LIVES

Training: upgrading human capital

.... And adapting workforce to technological change.

Investment in training, % GDP
(Selected countries, average 2003-2007)

Source: *O'Mahony (2012)*

This paper

- Training as investment in human capital.

- KBC approach:
  - Expenditure based;
  - Both opportunity cost and direct cost;
  - Goes beyond vocational training, and extends to formal and informal forms of training.

- Exploits survey-based micro-data and links to official sources (e.g. LFS) and to SNA:
  - PIAAC, Uniform methodology for 22 countries, with info about occupations, industry and skills.
  - Employment by occupation (LFS) and structural (SNA) data.

This paper (cont'd)

- Preliminary evidence provided:
  - 22 countries, average 2011-2012
    - Can get estimates of non-formal for > countries.
    - Non-formal only 40% of total investment in training.
    - Industry/country heterogeneity, but some patterns emerge:
      - Importance of formal education in subset of countries
      - Manufacturing vs services intensive countries ?
        - » Production more intensive in non-formal training than services, but not of business services only.
        - » Public services more intensive in formal training.
    - Coherent with previous estimates
  - Investment in training – no capitalization at present

Types of training considered

- Different types of training;
  - <u>Formal</u>: organised, outside work, yields certificate.
  - <u>Informal</u>: unstructured. Learning from peers, learning by doing.
  - <u>Non-formal</u>: organised, without certificate, both in- or outside work.
- Not perfectly overlapping with previous definitions (CEDEFOP 2014, Werquin 2007, O'Mahony 2012).
- Not necessarily "on the job" (formal), but in employment.
- Types can coexist. Overlapping.

Main hypotheses made

- Expenditure-based = hourly cost of training*incidence

- <u>General vs. firm-specific training</u>. Perfect labour markets?

  - Workers can be asked to pay for general training if unemployment, or employer pays if limited worker mobility.

- Who benefits from training? <u>Revealed preferences</u>: fraction of expenditure into investment determined by:

  - Who finances the training (employer, employee, both).

  - When training takes place (during working hours? Yes, no, in part)

  - Self-reported usefulness of training wrt current occupation.

Existing measures

Previous estimating methodologies (industry):

– **Corrado et al. (2014)**:

$$\left(\frac{Cost\ of\ training\ courses\ (CVTS)}{total\ labour\ cost}\right) * (employee\ compensations\ (NA)) +$$

$$\left(\frac{apprentices\ cost}{total\ labour\ cost}\right) * (employee\ compensations\ (NA))$$

– **Miyagawa & Hisa (2013)**: from Ooki (2003), opportunity cost / off-the-job training cost = 1.51

$$\left(\frac{Cost\ of\ f\ the\ job\ training}{total\ labour\ cost}(survey)\right) * (2.51))$$

– **O'Mahony (2012)**: CVT Survey

$$\left(\frac{Direct\ cost\ of\ training}{Opportunity\ cost\ of\ training}\right) * (avg\ compensation\ (EUKLEMS)) +$$

$$\left(\frac{\#\ trained\ employees\ by\ type\ \&\ industry}{total\ \#\ employees\ by\ type\ \&\ industry}\right) * (wage\ by\ type\ \&\ industry)$$

Methodology: Non-formal training (i)

$$I^N = \sum_{k,o,j} (p_{k,o,j} h^N_{k,o,j} w_{k,o,j} + q_{k,o,j} h^N_{k,o,j} C^N_j)$$

Opportunity Cost   Direct Cost

- $p_{k,o,j}$: weight. "Subjective usefulness" (1= very useful) ∩ "proportion of training in working hours" (1=only during working hours)

Allocation →				
Subjective Usefulness ↓	1	**0.667**	**0.333**	**0**
1	1	0.833	0.667	0.5
0.667	0.833	0.667	0.5	0.333
0.333	0.667	0.5	0.333	0.167
0	0.5	0.333	0.167	0

Intuition: revealed preferences approach. If training paid for or during working hours -> more likely to be investment. However, no 0 value if not:

- Usefulness for worker vs for employer
- Employer can pressure employee in bad cycles, or workers can decide to solve his own mismatch.

Methodology: Non-formal training (ii)

$$I^N = \sum_{k,o,j} (p_{k,o,j} h^N_{k,o,j} w_{k,o,j} + q_{k,o,j} h^N_{k,o,j} C^N_j)$$

<div align="center">Opportunity Cost   Direct Cost</div>

- $q_{k,o,j}$: weight. "Subjective usefulness" ∩ "proportion of training paid by employer" (1=totality).

- C^N_j: from Continuing Vocational Training Survey
  - Clustering for non-EU countries.
  - Market services used for non-market ones.
  - Account for firm size, correct for CVTS sampling.

Methodology: **Formal training**

$$I^{F} = \sum_{k,o,j} (p_{k,o,j} h_{k,o,j} w_{k,o,j}) + \sum_{k,e,o,j} (q_{k,e,o,j}) \frac{C_e}{S_e}$$

Opportunity Cost   Direct Cost

- $\frac{C_e}{S_e}$: public & private expenditure per student by ISCED97 category, in PPS (OECD). This is what is paid to participate in formal education.

- $q_{k,e,o,j}$: computed by education level of individuals.

- $h_{k,o,j}$: annual hours worked, from data restriction => assume entire work time devoted to training, proportionally to p (subjective usefulness).

Methodology: Informal training

- Three concepts involved in PIAAC: "peer learning", "learning by doing", "knowledge updating"

Methodology: Informal training

- Three concepts involved in PIAAC: "peer learning", "learning by doing", "knowledge updating".

  - PIAAC expresses in days. Double for peer learning.

  - Not mutually exclusive activities in the same time unit =>
    - *Upper bound* ($\bar{d}_{k,o,j}$): sum of days spent in all three activities.
    - *Lower bound* ($\underline{d}_{k,o,i}$): days having some peer learning.

$$I^{IN} = \{\underline{I}^{IN} \; ; \; \bar{I}^{IN}\} =$$

$$= \{\textstyle\sum_{k,o,j}(\underline{d}_{k,o,i} \qquad) ; \sum_{k,o,j}(\bar{d}_{k,o,j} \qquad)\}$$

Lower bound   Upper bound

Methodology: Informal training

- Three concepts involved in PIAAC: "peer learning", "learning by doing", "knowledge updating", expressed in days.
  - Double for peer learning.
  - Not mutually exclusive activities in the same time unit =>
    - *Upper bound* ($\bar{d}_{k,o,j}$): sum of days spent in all three activities.
    - *Lower bound* ($\underline{d}_{k,o,i}$): days having some peer learning.
  - From existing studies, obtain number of hours of training per day.
    - In the 90s: conservative approach
    - Average data for JP and U.S. . Employee's tenure >/< 1 year.

$$z_{k,o,j} = \{30m; 15m\}$$

$$I^{IN} = \{\underline{I}^{IN} \; ; \; \bar{I}^{IN}\} =$$

$$= \{\textstyle\sum_{k,o,j}(\underline{d}_{k,o,i} z_{k,o,j} \qquad); \sum_{k,o,j}(\bar{d}_{k,o,j}\, z_{k,o,j} \qquad)\}$$

Lower bound		Upper bound

Methodology: Informal training

- Three concepts involved in PIAAC: "peer learning", "learning by doing", "knowledge updating", expressed in days.

  – Double for peer learning.

  – Not mutually exclusive activities in the same time unit =>

    • *Upper bound* ($\bar{d}_{k,o,j}$): sum of days spent in all three activities.

    • *Lower bound* ($\underline{d}_{k,o,i}$): days having some peer learning.

  – From existing studies, obtain number of hours of training per day.

    – In the 90s: conservative approach

    – Average data for JP and U.S. . Employee's tenure >/< 1 year.

$$z_{k,o,j} = \{30m; 15m\}$$

$$I^{IN} = \{\underline{I}^{IN} ; \bar{I}^{IN}\} =$$

$$= \{\sum_{k,o,j}(\underline{d}_{k,o,i} z_{k,o,j} w_{k,o,j}) ; \sum_{k,o,j}(\bar{d}_{k,o,j} z_{k,o,j} w_{k,o,j})\}$$

Lower bound		Upper bound

Data: why PIAAC?

- Uniform questionnaire.

- Cross-country: not only EU. Recent (vs U.S. and Japan).

- Information on various types of training.

- Usefulness and repartition of costs of training between employer and employee

- Both occupational and industry dimension (vs country-wide).

- Not LEED.

Other data

- Direct input:
  - Education at a Glance – the incidence of attained education, direct cost of formal training.
  - Continuous Vocational Training Survey- direct cost of non-formal training.

- For calibration:
  - LFS-like surveys (gender by age structure, employment related info - participation rates, full-time and part-time related, etc.).
  - SNA- especially wages and salaries, industry structure-related information (e.g. employment by sector), GDP per capita, PPP.

- Recalibrate weights to align with SNA data

  GREG estimators minimize the deviation between the vector of original sampling weights (inverse value of a selection probability) and the calibrated weights using a linear regression model to fit the marginal totals of control variables.

- Outliers + adjustment by PPP.

Incidence of training

Incidence of training by type of training and skill level,
(% of total employment , Avg 2011-2012)
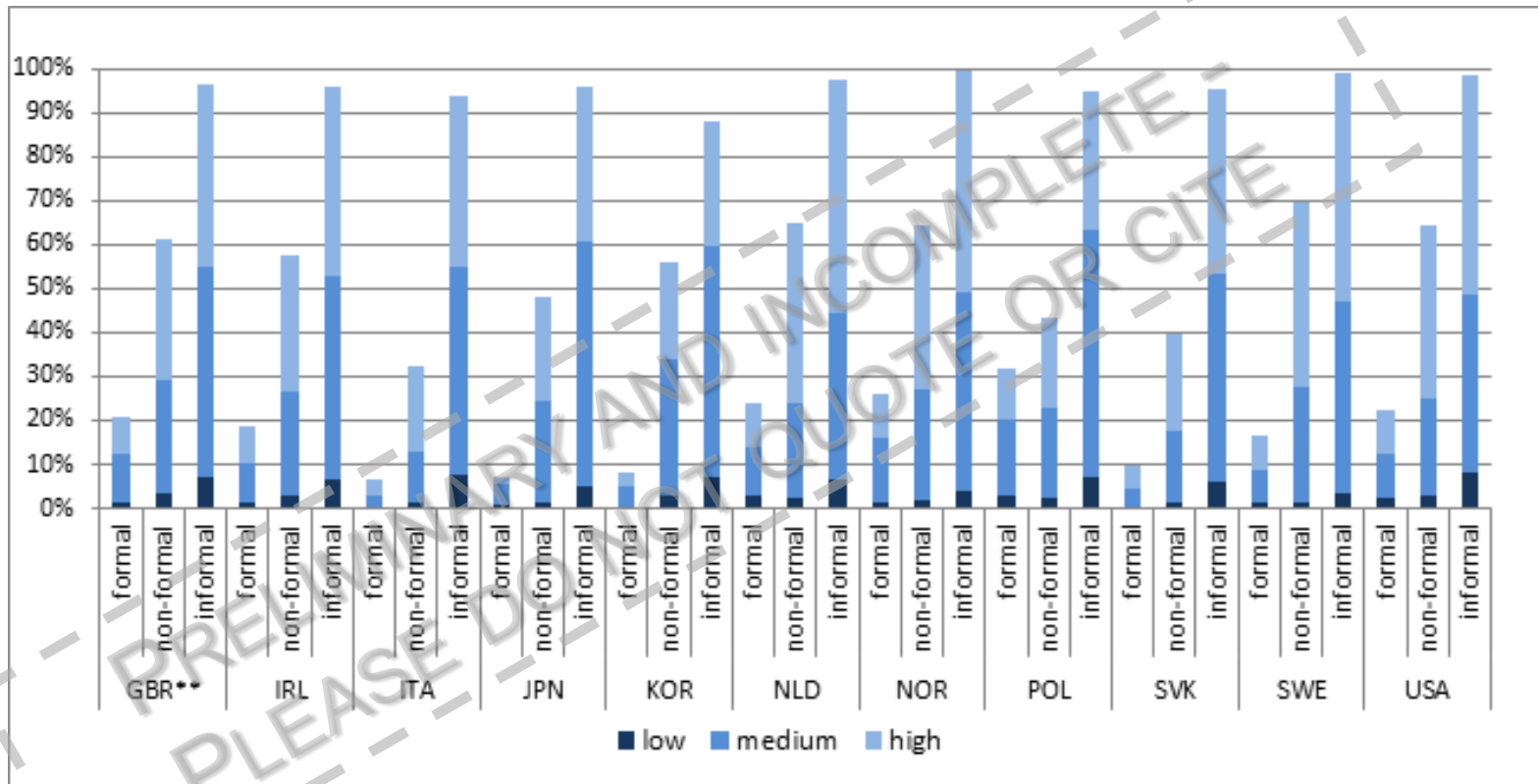
Driven by industry composition? Services more high skilled among the trained ones (ICT, finance, education, not retail/hotel).

Estimates of Investment (1)

Investment in training as % GVA

	Non-formal	Total		Non-formal	Total
Australia	3.37%	11.22%	*Italy*	1.06%	3.03%
Austria	2.63%	5.36%	*Japan*	2.28%	3.64%
Belgium	1.65%	4.86%	*Korea*	3.25%	4.76%
Canada	3.21%	9.75%	*Netherlands*	3.21%	10.06%
Czech Rep.	1.59%	3.88%	*Norway*	1.97%	8.13%
Denmark	2.93%	10.99%	*Poland*	1.91%	4.56%
Estonia	1.74%	5.89%	*Slovak Rep.*	1.41%	3.76%
Finland	2.46%	8.67%	*Spain*	2.56%	6.41%
France	1.36%	3.87%	*Sweden*	2.76%	6.54%
Germany	2.66%	5.12%	*UK*	2.84%	9.73%
Ireland	2.18%	7.29%	*USA*	3.12%	8.23%

- Median: non-formal: 2.51%, formal: 2.58%, informal: 1.04%, Total: 6.15%

- Two "models", for total? 4-5% vs 9-10%

- But not for non-formal (< dispersed than total and all others)

Estimates of Investment (2)

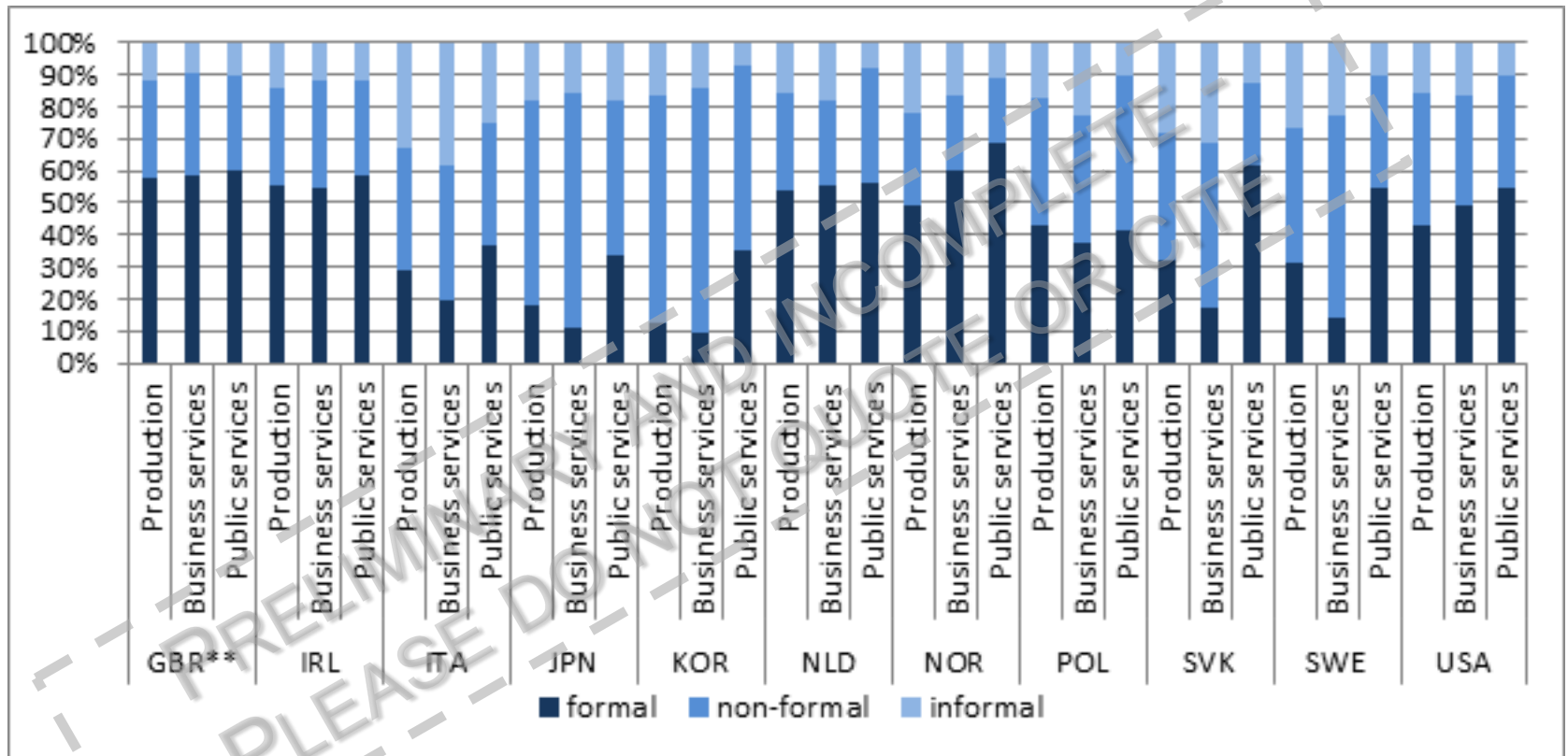Investment in non formal training as % output:
comparison with previous studies

	O'Mahony (2012)		Corrado et al. (2014)		Non-formal based on PIAAC	
	Total	Production	Total	Production	Total	Production
DEU	1.3%	0.9%	1.4%	1.5%	2.7%	2.2%
DNK	4.2%	3.1%	2.3%	1.6%	2.9%	2.0%
FRA	1.5%	1.5%	1.6%	1.3%	1.4%	1.8%
IRL	0.6%	0.2%	1.8%	4.7%	2.2%	1.2%
ITA	0.3%	0.2%	0.9%	0.8%	1.1%	1.0%
NLD	2.5%	1.4%	1.4%	1.0%	3.2%	2.0%

- No clear direction in "bias" wrt previous estimates
- Not all same denominator across studies

Estimates of Investment (3)

Investment in training by type and industry

Possible issues to address

- Ad-hoc choices:

  - Misclassification of individuals between work and study
    => ROBUSTNESS: use subjective status. Apprenticeship.

  - Weights based on allocation, usefulness, employer's quota questions.
    => ROBUSTNESS: recalculate p and q with different weights.

    - Informal training: for hours/day, no industry/country specific data available (but for JP, NL and US); when available, different years.
      => ALTERNATIVES?

Possible issues to address

- Ad-hoc choices:

  - Misclassification of individuals between work and study
    => ROBUSTNESS: use subjective status. Apprenticeship.

  - Weights based on allocation, usefulness, employer's quota questions.
    => ROBUSTNESS: recalculate p and q with different weights.

    - Informal training: for hours/day, no industry/country specific data available (but for JP, NL and US); when available, different years.
      => ALTERNATIVES?

- Labour cost based metrics:

  (i) translating reported time frequencies into exact hour equivalents;
  (ii) Double counting with other assets if a person is not only in training.

- One cross section (for the moment).

The way forward: your input

- How to improve this methodology?
  - Other similar studies ?
  - Country-specific data information we neglected?
  - Double counting of employees' time

The way forward: your input

- How to improve this methodology?
  - Other similar studies ?
  - Country-specific data information we neglected?
  - Double counting of employees' time

- Extensions:
  - Correlations with other industry-level outcomes (e.g. productivity)
  - Complementarity OC and training?
  - Capitalization.
    - Time series : issue of cyclicality of training.
    - Depreciation (separation rate?) and decay, deflation.