Decomposing the Dispersion of Earnings to Analyze the Impact of On-the-Job Training

Jacques Silber,

Audrey Dumas,

and

Saïd Hanchane

Interdisciplinary Congress on Research in Vocational Education and Training
Bern/Zollikofen, Switzerland
25th – 27th March 2009

Research Questions

- Job training ⇒ Human capital ⇒ Wage level
- Selection of trainees: the most educated (OECD (1999))

Job training ⇒ Wage inequality

⬇

- What is the contribution of wage variance between trainees and non trainees to the wage inequality ?

- What are the sources of wage inequality between trainees and non trainees?

- What are the sources of wage inequality when we eliminate the between variance ?

⇒ **Methodology of Shapley Decomposition**

Plan

- The Shapley Decomposition
  - The concept of Shapley value
  - The contribution of the Between and Within variance to the earning's dispersion
  - The components of the Between and Within contribution
- Our data
- Our evaluation strategy
- The results
- Conclusion

The Shapley Decomposition

- *I* is a function of two variables *a* and *b:* $I(a \neq 0; b \neq 0)$
- The contribution of *a* to *I:*

Two cases: *a* is eliminated *(a=0)* before or after *b* is eliminated

If *a* is eliminated first: $I(a \neq 0; b \neq 0) - I(a=0; b \neq 0)$

If *a* is eliminated second: $I(a \neq 0; b=0) - I(a=0; b=0) = I(a \neq 0; b=0)$

- The two sequences have the same probability: *1/2*
- The total contribution of *a* to *I* is:

$$C(a) = (1/2)\,[\, I(a \neq 0; b \neq 0) - I(a=0; b \neq 0) \,] + (1/2)\,[\, I(a \neq 0; b=0) \,]$$

- *I* is wage variance
- Population is composed of *k* groups (trainees and non trainees)
- *I* is a function of two variables : $I(\bar{y}_k \neq \bar{y}; y_{ik} \neq \bar{y}_k)$

- Within Variance *(a):* $I(\bar{y}_k = \bar{y}; y_{ik} \neq \bar{y}_k)$

- Between Variance (*b*): $I(\bar{y}_k \neq \bar{y}; y_{ik} = \bar{y}_k)$

$$C_{WITH} = (1/2)[I(\bar{y}_k \neq \bar{y}; y_{ik} \neq \bar{y}_k) - I(\bar{y}_k \neq \bar{y}; y_{ik} = \bar{y}_k)] + (1/2)I(\bar{y}_k = \bar{y}; y_{ik} \neq \bar{y}_k)$$

$$C_{BET} = (1/2)[I(\bar{y}_k \neq \bar{y}; y_{ik} \neq \bar{y}_k) - I(\bar{y}_k = \bar{y}; y_{ik} \neq \bar{y}_k)] + (1/2)I(\bar{y}_k \neq \bar{y}; y_{ik} = \bar{y}_k)$$

$$C_{BET} = (1/2)[I(\bar{y}_k \neq \bar{y}; y_{ik} \neq \bar{y}_k) - I(\bar{y}_k = \bar{y}; y_{ik} \neq \bar{y}_k)] + (1/2)I(\bar{y}_k \neq \bar{y}; y_{ik} = \bar{y}_k)$$

- Wage :

$$y_{ik} = (\sum_{h=1}^{H} b_{hk} x_{ihk}) + u_{ik}$$

- The contribution of the Between and Within variance to the overall variance are functions of four variables

$$C_{BET} = f(b_{hk} \neq b_h, \bar{x}_{hk} \neq \bar{x}_h, x_{ihk} \neq \bar{x}_{hk}, u_{ihk} \neq 0)$$

- Coefficients
- Between characteristics
- Within characteristics
- Unobservables

Example :

The contribution of the coefficient to the Between contribution to the overall variance

$$
\begin{aligned}
C_{coef,BET} = {} & (6/24)\{C_{BET}[b_{hk} \neq b_h; x_{ihk} \neq \bar{x}_{hk}; \bar{x}_{hk} \neq \bar{x}_h; u_{ik} \neq 0] - C_{BET}[b_{hk} = b_h; x_{ihk} \neq \bar{x}_{hk}; \bar{x}_{hk} \neq \bar{x}_h; u_{ik} \neq 0]\} \\
& +(2/24)\{C_{BET}[b_{hk} \neq b_h; x_{ihk} = \bar{x}_{hk}; \bar{x}_{hk} \neq \bar{x}_h; u_{ik} \neq 0] - C_{BET}[b_{hk} = b_h; x_{ihk} = \bar{x}_{hk}; \bar{x}_{hk} \neq \bar{x}_h; u_{ik} \neq 0]\} \\
& +(2/24)\{C_{BET}[b_{hk} \neq b_h; x_{ihk} \neq \bar{x}_{hk}; \bar{x}_{hk} = \bar{x}_h; u_{ik} \neq 0] - C_{BET}[b_{hk} = b_h; x_{ihk} \neq \bar{x}_{hk}; \bar{x}_{hk} = \bar{x}_h; u_{ik} \neq 0]\} \\
& +(2/24)\{C_{BET}[b_{hk} \neq b_h; x_{ihk} \neq \bar{x}_{hk}; \bar{x}_{hk} \neq \bar{x}_h; u_{ik} = 0] - C_{BET}[b_{hk} = b_h; x_{ihk} \neq \bar{x}_{hk}; \bar{x}_{hk} \neq \bar{x}_h; u_{ik} = 0]\} \\
& +(2/24)\{C_{BET}[b_{hk} \neq b_h; x_{ihk} = \bar{x}_{hk}; \bar{x}_{hk} = \bar{x}_h; u_{ik} \neq 0] - C_{BET}[b_{hk} = b_h; x_{ihk} = \bar{x}_{hk}; \bar{x}_{hk} = \bar{x}_h; u_{ik} \neq 0]\} \\
& +(2/24)\{C_{BET}[b_{hk} \neq b_h; x_{ihk} = \bar{x}_{hk}; \bar{x}_{hk} \neq \bar{x}_h; u_{ik} = 0] - C_{BET}[b_{hk} = b_h; x_{ihk} = \bar{x}_{hk}; \bar{x}_{hk} \neq \bar{x}_h; u_{ik} = 0]\} \\
& +(2/24)\{C_{BET}[b_{hk} \neq b_h; x_{ihk} \neq \bar{x}_{hk}; \bar{x}_{hk} = \bar{x}_h; u_{ik} = 0] - C_{BET}[b_{hk} = b_h; x_{ihk} \neq \bar{x}_{hk}; \bar{x}_{hk} = \bar{x}_h; u_{ik} = 0]\} \\
& +(6/24)\{C_{BET}[b_{hk} \neq b_h; x_{ihk} = \bar{x}_{hk}; \bar{x}_{hk} = \bar{x}_h; u_{ik} = 0] - C_{BET}[b_{hk} = b_h; x_{ihk} = \bar{x}_{hk}; \bar{x}_{hk} = \bar{x}_h; u_{ik} = 0]\}
\end{aligned}
$$

Our data

- French databases : *2006 Adult Education Survey*, *2005* and *2006 Labour Force Survey*
- Individual aged 18-65, private sector
- Sample 1: 3975 individuals
- Trainees : 38,2%

- Definition more restrictive of trainees

(Training programs: Financed by firms, not certified, stage or at work, were completed, whose objective is to make workers more comfortable and/or more efficient in their job)

- Sample 2: 3161 individuals
- Trainees H: 22,4%

The Evaluation Strategy

- OLS: - individual characteristics (age, sex, tenure, education, nationality, number of children),
  - job characteristics (working hours, type of contract, position, profession, living in Paris, working days and hours),
  - firm characteristics (industry, size, new work organisation, new equipment)

- Two-step Heckman's procedure: problem instruments

- First differences method: problem elimination time invariant variables

⇒ OLS + quintiles: dummies indicating the position of individuals in the wage distribution before training
(1th, 5th,10th, 25th, 50th, 75th, 90th, 95th, 99th)

The Results

Contribution of the Between variance to the overall variance

Contribution of Between Variance	Contribution of Within Variance
Sample 1 – Trainees	
5,4%* **	**94,6%*** **
Sample 2 – Trainees H	
5,3%* **	**94,7%*** **

⇒ Training is not a good criteria of market segmentation

The Results

Decomposition of the contribution of the Between variance

Coefficients	Within Characteristics	Between Characteristics	Unobservables
Sample 1			
OLS			
4,3%	2,1%	**93,5%*** **	0,1%
OLS + QUANTILES			
16,3%	5,7%	**77,7%*** **	0,3%
Sample 2			
OLS			
1,8%	0,8%	**98,2%*** **	-0,8%
OLS + QUANTILES			
7,4%	3,2%	**89,5%*** **	-0,09%

⇒ The heterogeneity of training returns does not affect wage inequality between trainees and non trainees

The Results

Decomposition of the contribution of Within variance

Sample 1			
OLS			
Coefficients	**Within Characteristics**	**Between Characteristics**	**Unobservables**
- 4,8%***	**65,9%*****	17,3%***	21,6%***
OLS + QUANTILES			
- 6,6%***	**75,5%*****	19,7%***	11,4%***
Sample 2			
OLS			
- 3,9%***	**60,6%*****	16,2%***	27,2%***
OLS + QUANTILES			
- 6,1%***	**74,1%*****	19,3%***	12,7%***

⇒ All sources have a significant and positive effect on wage dispersion excepted coefficients that reduce it

Conclusion

- Training participation is a variable that explain significantly 5,3% of wage dispersion
- Observable characteristics are the only source of wage inequality between trainees and non trainees
- Application of this methodology to other subgroups (firm size, sex, etc…)

Decomposing the Dispersion of Earnings to Analyze the Impact of On-the-Job Training

Jacques Silber,

Audrey Dumas,

Saïd Hanchane

Thanks for attention

Interdisciplinary Congress on Research in Vocational Education and Training
Bern/Zollikofen, Switzerland
25th – 27th March 2009