

# Faut-il choisir entre égalité et efficacité?

## L'effet différencié des lycées sur les résultats de leurs élèves

Pauline Givord<sup>1</sup> Milena Suarez<sup>1</sup>

<sup>1</sup>Insee

13 juin 2018

# Motivation

Peut-on faire plus complet qu'un "effet lycée" moyen?

- ▶ **Les écoles améliorent-elles les résultats de leurs élèves?**  
Littérature conséquente sur la question, centrée autour des modèles de *valeurs ajoutées*

*Que mesure-t-on?* [Raudenbush and Willms, 1995] *Ces indicateurs sont-ils biaisés?* [Chetty et al., 2013]

- ▶ **Des mesures largement appliquées/diffusées par les institutions**

*En France, les IVAL «Indicateurs de valeur ajoutée des lycées»*  
[Duclos and Murat, 2014]

- ▶ **Ne tiennent pas compte d'effet sur la distribution des résultats:** Les écoles réduisent-elles ou au contraire augmentent-elles les écarts de résultats entre leurs élèves?

*Un effet moyen peut cacher des situations diverses*

## Retour sur les modèles de valeur ajoutée

VA: effet **moyen** de l'école une fois tenu compte des observables de l'élève

$$\mathbb{E}(\underbrace{Y_{ij}}_{\text{Note finale}} \mid X_{ij}, \mathbf{1}\{i \in j\}) = \underbrace{\alpha_j}_{\text{VA}} + X_{ij}\beta$$

- ▶ Contrôle des effets de composition (par rapport aux simples taux bruts)
- ▶ Repose sur une hypothèse de sélection (dans les lycées) sur observables

## Et la dispersion intra-école?

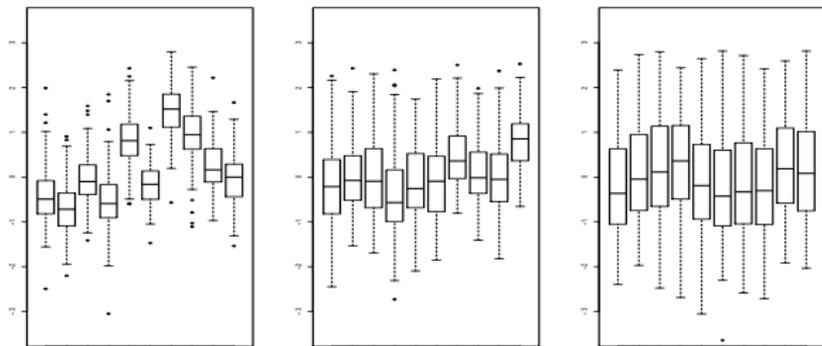


Figure: Distribution des notes au baccalauréat pour 30 lycées (10 plus petits interquartiles, 10 aléatoires, 10 plus grand interquartiles)

# Un effet fixe par quantile

## Interpretation

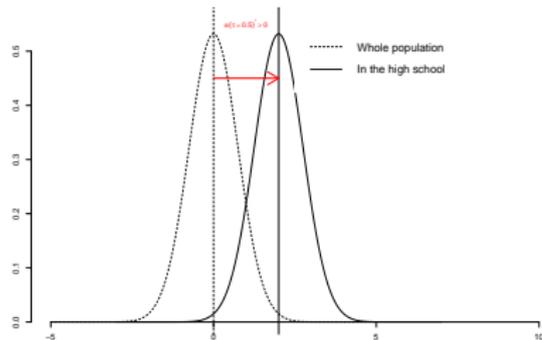
$$q_{\tau}(Y_{ij}|X_{ij}, 1\{i \in j\}) = \underbrace{m(\tau) + X_{ij}\beta(\tau)}_{\text{Attente}} + \underbrace{\alpha_j^*(\tau)}_{\text{Effet lycée}}$$

où  $\sum_j \alpha_j^*(\tau) = 0$

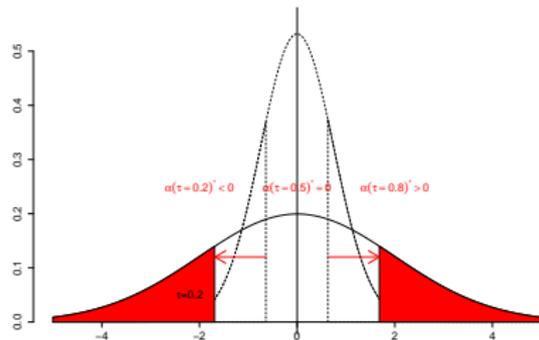
Interpretation  $\alpha_j^* : \tau \rightarrow \alpha_j^*(\tau)$

- ▶ **Signe** : performance
- ▶ **Monotonicit **: tendance   augmenter (fonction croissante) ou au contraire diminuer (fonction d croissante) la dispersion des r sultats des  l ves (caract ris s par  $X_{ij}$ )

$$\text{Quand } q_{\tau}(Y_{ij}|X_{ij}, 1\{i \in j\}) = \underbrace{m(\tau) + X_{ij}\beta(\tau)}_{\text{Attente}} + \underbrace{\alpha_j^*(\tau)}_{\text{Effet lycée}}$$

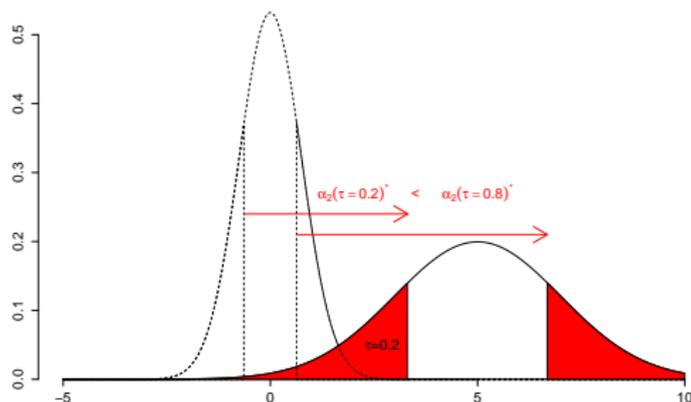


(a) Effet homogène et positif :  
 $\tau \rightarrow \alpha^*(\tau) = \alpha_0 > 0$



(b) Effet hétérogène (inégalitaire):  
 $\tau \rightarrow \alpha^*(\tau)$  est croissante

$$\text{Quand } q_{\tau}(Y_{ij}|X_{ij}, 1\{i \in j\}) = \underbrace{m(\tau) + X_{ij}\beta(\tau)}_{\text{Attente}} + \underbrace{\alpha_j^*(\tau)}_{\text{Effet lycée}}$$



(c) Effet hétérogène (inégalitaire) et positif :  
 $\tau \rightarrow \alpha^*(\tau)$  est croissante et positive

# Données

## FAERE et les résultats au baccalauréat 2015

**Variable "expliquée":** Note moyenne à la première session du baccalauréat 2015

- ▶ Un modèle par filière (générale, technologique), exclusion de la filière professionnelle (séries très fragmentées, effectifs par lycée faibles)

**Les covariables:** (importantes pour l'hypothèse de sélection sur observables)

- ▶ Niveau initial: note à l'écrit du brevet (avant harmonisation)
- ▶ Indice de position sociale de l'élève [Rocher, 2016]
- ▶ Indicateur pour les redoublants avant l'entrée au lycée
- ▶ Genre

**Un contrôle supplémentaire pour la notation commune au sein des séries:**

- ▶ Série de l'élève (e.g. en général, S,ES,L)

# Estimation

- ▶ Estimation d'une régression quantile à la médiane et aux deux quintiles extrêmes ( $\tau \in \{0.2, 0.5, 0.8\}$ ):  $\alpha_L, \alpha_M, \alpha_H$
- ▶ Restriction de l'échantillon pour avoir "suffisamment" d'élèves par lycée: seuil de 65 (filrière G) et 25 (filrière T) élèves dans chaque filière, ce qui correspond à garder 95% des élèves (mais à exclure les plus petits lycées)
- ▶ Définition des classes des lycées (approximation de la monotonie):
  - ▶  $\alpha_H > \alpha_L$  sur 90% des échantillons bootstraps : lycée inégalitaire
  - ▶  $\alpha_H < \alpha_L$  : lycée égalitaire
  - ▶ Sinon, lycée homogène
- ▶ Bootstrap: dans chaque lycée, tirage avec remise du même nombre d'élèves et réestimation sur l'échantillon bootstrap

## Avec/sans effets fixes

Table: Régressions quantiles: coefficients des covariables - Filière générale

| Quantile                 | Avec effets fixes |               |               | Sans effets fixes |               |               |
|--------------------------|-------------------|---------------|---------------|-------------------|---------------|---------------|
|                          | 20%               | 50%           | 80%           | 20%               | 50%           | 80%           |
| Prior score              | 0.593 (0.002)     | 0.632 (0.002) | 0.646 (0.002) | 0.622 (0.002)     | 0.655 (0.002) | 0.661 (0.002) |
| Prior score <sup>2</sup> | 0.107 (0.001)     | 0.105 (0.001) | 0.082 (0.001) | 0.108 (0.001)     | 0.105 (0.001) | 0.079 (0.001) |
| Social position          | 0.079 (0.002)     | 0.079 (0.002) | 0.079 (0.002) | 0.108 (0.002)     | 0.106 (0.002) | 0.102 (0.002) |
| Repeaters                | -0.271(0.008)     | -0.245(0.007) | -0.193(0.008) | -0.284(0.008)     | -0.253(0.008) | -0.201(0.008) |
| Girls                    | 0.080 (0.004)     | 0.052 (0.003) | 0.032 (0.004) | 0.080 (0.004)     | 0.051 (0.003) | 0.035 (0.004) |
| L (réf ES)               | 0.074 (0.005)     | 0.086 (0.005) | 0.088 (0.006) | 0.065 (0.005)     | 0.077 (0.005) | 0.070 (0.005) |
| S                        | -0.194(0.004)     | -0.172(0.004) | -0.147(0.004) | -0.214(0.004)     | -0.184(0.004) | -0.162(0.004) |

- Probable sélection sur observables dans les lycées

## Quelles proportions de lycées à effet hétérogène?

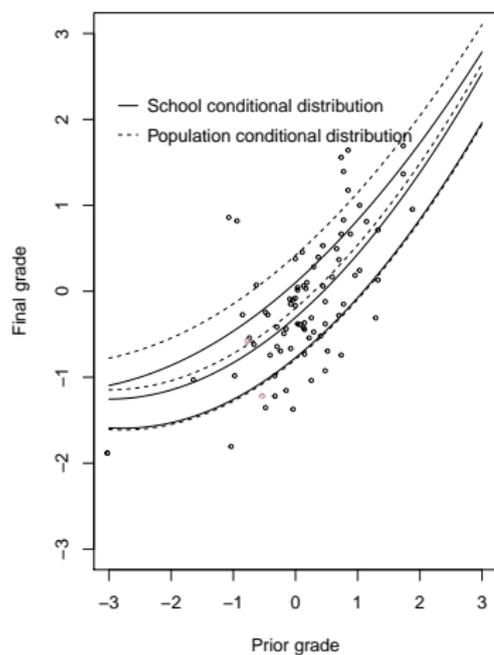
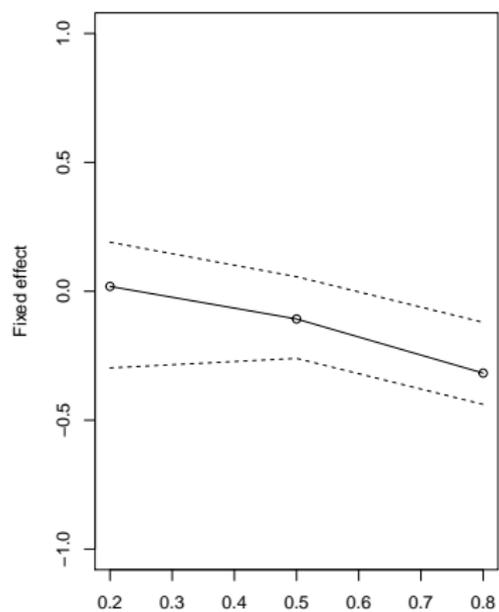
Table: Présence d'effets hétérogènes

|                        | Proportion de lycées (%) |                   |                   |
|------------------------|--------------------------|-------------------|-------------------|
|                        | <u>Inégalitaire</u>      | <u>Egalitaire</u> | <u>Heterogène</u> |
| <i>Général</i>         |                          |                   |                   |
| Sans correction        | 13.6                     | 15.2              | 28.8              |
| Correction MHT: BBSGoF | 8.2                      | 8.5               | 16.7              |
| <i>Technologique</i>   |                          |                   |                   |
| Sans correction        | 12.5                     | 14.5              | 27                |
| Correction MHT: BBSGoF | 6                        | 7.6               | 13.6              |

- ▶ BBSGoF: Correction "souple" pour le test d'hypothèses multiples ( $= N_H$ , nombre de lycées), à même de garder un certain pouvoir de détection quand  $N_H$  est grand

# Exemple d'un lycée

Monotonie de l'effet lycée par quantile et dispersion des résultats conditionnels



# Répartition des lycées: performance et "égalitarisme"

Table: Valeur ajoutée médiane et interquintile de valeur ajoutée

| <b>General</b>               | Inégalitaire        | Homogène            | Egalitaire          |     |
|------------------------------|---------------------|---------------------|---------------------|-----|
| Performance médiane négative | 15.6<br><i>30.9</i> | 76.1<br><i>29.2</i> | 8.3<br><i>15.6</i>  | 100 |
| Performance médiane n.s.     | 15.5<br><i>55.1</i> | 73.1<br><i>50.5</i> | 11.3<br><i>38.3</i> | 100 |
| Performance médiane positive | 8.4<br><i>14</i>    | 62.6<br><i>20.3</i> | 29.1<br><i>46.1</i> | 100 |
|                              | <i>100</i>          | <i>100</i>          | <i>100</i>          |     |

- Répartition qualitativement très proche dans la filière technologique

# Robustesse

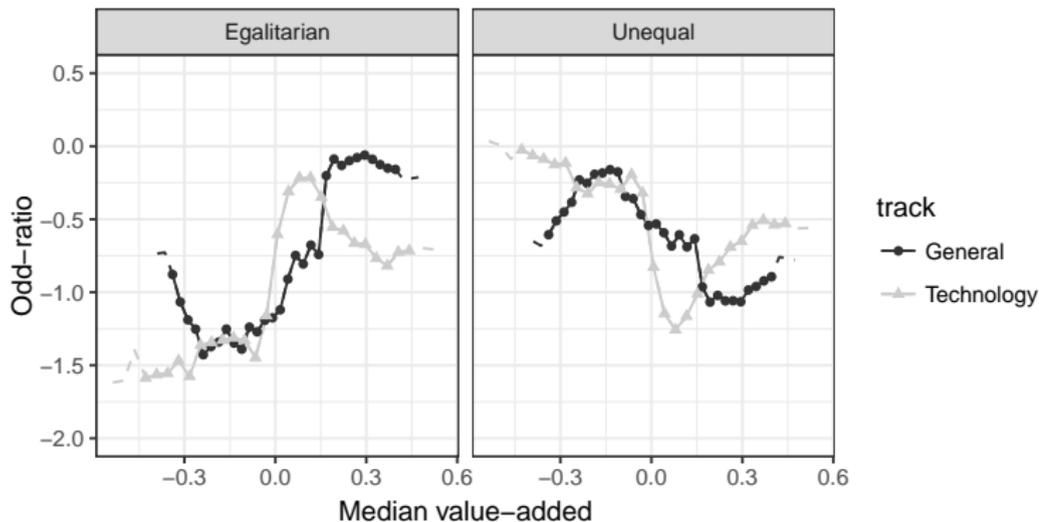
- ▶ Sélection en cours de scolarité d'élèves qui auraient peu de chance de bien réussir leur bac?
  - ▶ Estimation des effets lycées sur les résultats au bac de l'année N sur la population présente dans le lycée en N-1 et passant son bac en N.
  - ▶ Un lien significatif entre baisse de la mesure d'inégalité et faible taux d'accès dans le privé
  - ▶ Peu de changement de classification (égalitaire/homogène/inégalitaire) et pas de changement entre les extrêmes
  - ▶ Une manipulation possible mais avec un impact probablement faible sur la classification
- ▶ Paramètres incidents?
  - ▶ Estimer les coefficients sur covariables avec la méthode de [Canay, 2011], qui n'est pas soumis aux problèmes des paramètres incidents donnent des résultats très proches sur les coefficients des covariables

# Description des caractéristiques des lycées "égalitaire" et "inégalitaire"

- ▶ Approche descriptive: caractéristiques des lycées associées à un classement égalitaire/inégalitaire?
- ▶ Classification par forêt aléatoire à partir d'un ensemble de caractéristiques du lycée (élèves accueillis, moyens, caractéristiques des enseignants..) selon 3 modalités: lycée égalitaire, homogène, inégalitaire
- ▶ Avantage: pas d'idées préconçues sur la dépendance en les variables explicatives
- ▶ Graphes de dépendance partielle: comment évolue la prédiction en faveur d'une modalité plutôt qu'une autre en fonction d'une caractéristique du lycée (du niveau moyen des élèves accueillis ou encore en fonction de l'effectif de l'établissement)?

# Graphes de dépendance partielle

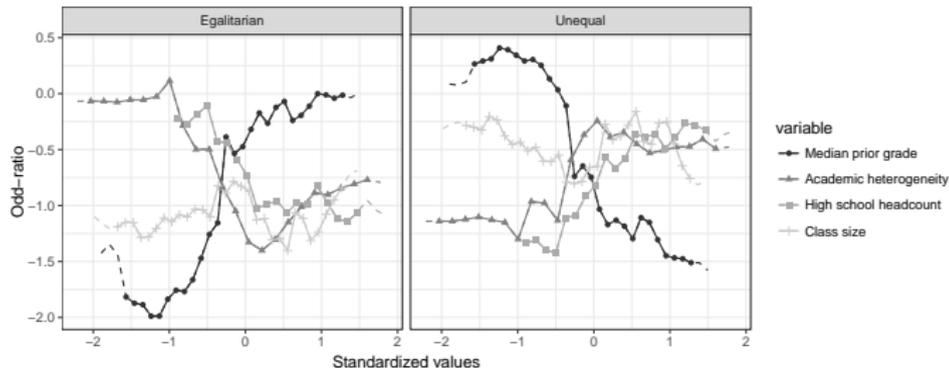
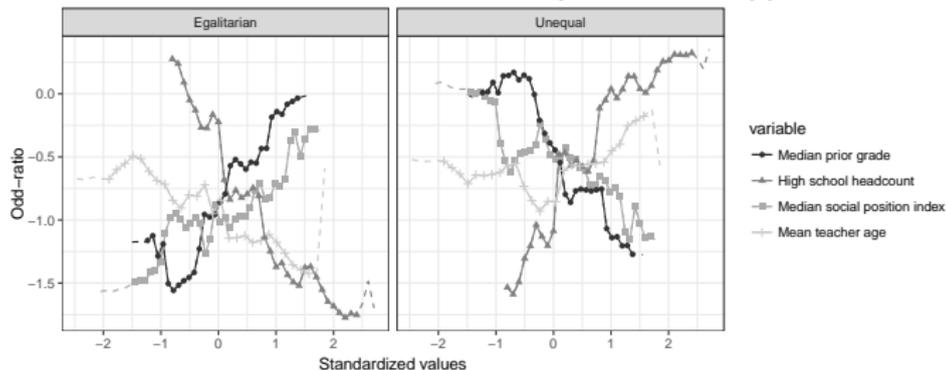
## Performance et égalité



- ▶ Une association performance/égalité globale, sauf pour les meilleurs lycées (en particulier en filière technologique)

# Graphes de dépendance partielle

Variables où les votes de la forêt varient le plus sur le support



- ▶ Des lycées plus petits, accueillants des élèves aux résultats au brevet meilleurs sont plus probablement "égalitaire".
- ▶ Possibilité aussi de décliner ces graphes par type d'élèves

# Conclusion

- ▶ Un nouvel indicateur, complémentaire de la performance moyenne du lycée, d'égalité de résultats parmi ses élèves
- ▶ Une proportion conséquente de lycées français sont concernés par des résultats "non homogènes", une fois contrôlé des observables de la population accueillie
- ▶ Point d'interprétation: valide pour le choix des familles, moins pour l'évaluation de l'action du lycée
- ▶ En effet, même sans remettre en cause l'hypothèse de sélection sur observables et la persistance de l'indicateur, l'effet lycée mesuré ici est une combinaison complexe entre action du lycée, effets de pairs, effets de l'environnement du lycée...

# Bibliography I

-  Canay, I. A. (2011).  
A simple approach to quantile regression for panel data.  
*The Econometrics Journal*, 14(3):368–386.
-  Chetty, R., Friedman, J. N., and Rockoff, J. E. (2013).  
Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates.  
Technical report, National Bureau of Economic Research.
-  Duclos, M. and Murat, F. (2014).  
Comment évaluer la performance des lycées.  
*Éducation & formations*, (85):74.
-  Raudenbush, S. W. and Willms, J. (1995).  
The estimation of school effects.  
*Journal of educational and behavioral statistics*, 20(4):307–335.

# Bibliography II



Rocher, T. (2016).

Construction d'un indice de position sociale des élèves.

*Éducation & formations*, 90.