

# Something interesting

Fu Zixuan\*

July 4, 2024

## **Abstract**

someting interesting

---

\*Last compiled on June 18, 2024

# Contents

1	Introduction	3
---	--------------	---

# 1 Introduction

It is almost of human nature to compare, rank and select. And competition, be it good or bad, emerges in the wake. As invidious as ranking and selection can be, in many cases it is one of the driving forces behind improvement in performances, not to mention the role of natural selection in the history of evolution. The society itself is constantly constructing league table. It rewards the meritorious and question or even punishes the unsatisfactory. The measure based on which rank is constructed ranges from teacher’s evaluation to communities’ mobility index.

The present article extends the practice to the health sectors. To be more specific, it studies the labor efficiency across all hospitals in France. By exploring a comprehensive database (SAE) of French hospitals, we first construct a measure of labor efficiency. Then based on the estimates, we compare the public and private hospitals by selecting the top-performing units. We borrow from the recent developments in Empirical Bayes method to achieve the comparison.

## Conclusion here

The article bridges two fields of interests. One is on productivity analysis. The most popular methods are Data Envelopment Analysis [Charnes et al., 1978] and Stochastic Frontier Analysis [Aigner et al., 1977, Meeusen and van Den Broeck, 1977]. Yet we abstract from both of them for the convenience of statistical inference. We use the simple method applied in [Croiset and Gary-Bobo, 2024] by estimating a conditional input demand function. To put it simply, we estimate a linear function of how much labor input is needed to produce a give list of outputs <sup>1</sup>. We only focus on the employment level of nurses for (reasons) <sup>2</sup> in the specification.

The second area of interests is the Empirical Bayes Methods. The name "Empirical Bayes" is self-telling in the sense that Bayes implies a prior distribution while Empirical means empirically estimate the prior from the data. Details on how it can be of use in ranking and selection will be presented in the rest of the section.

In [Croiset and Gary-Bobo, 2024], the authors argue that public hospital is less efficient than private counterpart in the sense that it would need a smaller size of personnel if it were to use the input demand function of the private hospital, which is the main result of their counterfactuals.

Having roughly replicated the results after doubling the length of the panel, the paper differentiates itself by including/adding the standard/classical panel data methods in input demand function estimation, specifically the fixed-effect within-group estimation and GMM.

With respect to the latter estimator, we may relax the assumption of strict exogeneity of the regressors, claiming that current regressors (output) may be endogeneous/affect current and future errors term.

$$y_{it} = x_{it}\beta + \alpha y_{i,t-1} + \theta_i + \varepsilon_{it} \quad \mathbb{E}[\varepsilon_{it}|x_{i1}, \dots, x_{it-1}] = 0$$

---

<sup>1</sup>We refer the audience to xxx for detailed reasons of adopting this approach.

<sup>2</sup>

Note that the panel used in my estimation has relatively high persistency in the variables. This kind of characteristics was pointed out in [Blundell and Bond, 1998] as well. It argues that when  $T$  is relatively small and the regressors exhibit relatively high auto correlation, the lagged levels of regressors  $x_{i,t-2}$  only serve as weak instruments for first differences equation  $\Delta\varepsilon_{i,t}$ . By imposing a reasonable assumption that  $\mathbb{E}\Delta x_{i,t-1}\varepsilon_{i,t} = 0$ , I instrument the current level with lagged first differences, implementing the so-called system GMM explained in [Arellano and Bover, 1995, Blundell and Bond, 1998].

Though the original focus of the panel data estimator is on the  $\beta$  parameters. It also provides us with a noisy estimate of the underlying unobserved heterogeneity term denoted as  $\theta_i$ . (It's important to note that this heterogeneity is not necessarily indicative of inefficiency). Now that we have set the stage for empirical bayes, it is time to bring about the prior distribution of  $\theta_i$ , denoted as  $G_\theta$ . If the prior distribution  $G$  is known, having observed an estimate  $\hat{\theta}_i$  of  $\theta_i$ , we can update our noisy estimate  $\hat{\theta}_i$  using or incorporating our knowledge of  $G$ .

The usefulness of a prior  $G$  is further exemplified/highlighted in the ranking and selection problem mentioned above, when the object of interests is the noisy estimate of  $\theta_i$ . For example, in [Gu and Koenker, 2023], we are given the task of selecting the top 20% out of the population of  $\alpha_i$ , that is to say selecting those  $i$  whose  $\alpha_i > G^{-1}(0.8)$ , while controlling for the overall false discovery rate ( $\frac{x}{y}$ ) at 5%. In [Gu and Koenker, 2023], the authors try to develop an optimal decision rule for the given task. To put it in the language of optimization, they want to have a decision rule that optimizes the performance of selection, equivalently, minimize the loss of selection

$$\delta^* = \min_{\delta} \text{Loss} \quad \text{subject to constraints}$$

where the loss function can take different forms, for example the expected number of total type 1 and 2 mistakes.

The task at hand falls naturally under the compound decision framework pioneered by [Robbins, 1992] if we define the loss function in such a way that takes into account the results of all the individual decisions  $\delta(Y_i)$ . For instance, summing all mistakes would be one way to aggregate/compound individual decisions,

It is obvious that in order to impose the two stated constraints (capacity and FDR) in formulating the optimization problem, we need to know the prior distribution  $G$ . Despite the importance of the  $G$ , it does not fall from heaven. Therefore, Empirical Bayes methods come to the rescue, as its name suggests, we will have to empirically estimating the unknown prior  $G$ .

Often times "empirically estimating  $G$ " is done with parametric assumption that  $G$  is normal. Notable use cases are found in teacher evaluation [Chetty et al., 2014], social mobility in communities [Chetty and Hendren, 2018] and job discrimination [Kline and Walters, 2021]. By assuming a Gaussian  $G$ , they shrink the estimated fixed effect linearly, thus giving the name "linear shrinkage". However, departure from normality may render the linear shrinkage rule as unhelpful. Thanks to the foundational work of [Kiefer and Wolfowitz, 1956] who has shown that non-parametric estimation is also feasible and consistent, it is preferable to

relax the normality assumption and estimate the prior  $G$  non-parametrically. In terms of computation method, [Koenker and Mizera, 2014] has formulated the non-parametric estimation as a convex optimization problem. Compared to other popular methods such as EM algorithm [Laird, 1978], recent advancements in convex optimization computation methods [Andersen and Andersen, 2010] has made the novel approach of [Koenker and Mizera, 2014] computationally more attractive.

It is worth mentioning here that though a discrete  $G$  with at most  $x$  atoms can be estimated using the REBayes package [Koenker and Gu, 2017], we are not free of imposing any assumptions, that is the distribution of estimate  $\hat{\alpha}_i|\alpha_i, \sigma_i^2$ . To illustrate, in the case of the estimate of fixed effect, we have

$$\begin{aligned}\hat{\alpha}_i &= \frac{1}{T} \sum_t (y_{it} - x_{it}\hat{\beta}) \\ &= \frac{1}{T} \sum_t (\alpha_i + \varepsilon_{it} + x_{it}(\beta - \hat{\beta})) \\ &\rightarrow^d \alpha_i + \frac{1}{T} \sum_t \varepsilon_{it}\end{aligned}$$

The asymptotic distribution follows from the consistency of  $\hat{\beta}$  when  $N \rightarrow \infty$ , a reasonable assumption in wide panels.

If we may boldly assume that the errors are *i.i.d.* normally distributed for each  $i$

$$\varepsilon_{it} \sim N(0, \sigma_i^2)$$

Then a fixed/small  $T$  won't do jeopardize/imperil xxx our argument too much since we do not need to invoke central limit theorem to have

$$\hat{\alpha}_i \rightarrow^d N(\alpha_i, \sigma_i^2/T)$$

However without the normality assumption on the error terms, we have to resort to the central limit theorem from the claim that  $T \rightarrow \infty$ , which may seem unrealistic for a wide panel ( $N \gg T$ ).

The rest of the paper is organized as follows. Section 2 briefly describes the data and lays out the reduced form estimation of the input demand function, treating the number of nurses as the dependent variable and a list of 9 output measures as the regressors. Section 3 applies the classical panel data estimators to the same specification, distinguishing between whether strict exogeneity is assumed. In section 4, we introduce the compound decision framework and specifically/xxx define the selection problem. Section 5 follows with a comparison of the different selection outcome as a result of imposing varying constraints and assumptions. In section 6, we try to draw preliminary conclusion on the comparative performance of public and private hospitals. Section 7 discusses potential issues and concludes.

## References

- [Aigner et al., 1977] Aigner, D., Lovell, C. K., and Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of econometrics*, 6(1):21–37.
- [Andersen and Andersen, 2010] Andersen, E. D. and Andersen, K. D. (2010). The mosek optimization tools manual, version 6.0.
- [Arellano and Bover, 1995] Arellano, M. and Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of econometrics*, 68(1):29–51.
- [Blundell and Bond, 1998] Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of econometrics*, 87(1):115–143.
- [Charnes et al., 1978] Charnes, A., Cooper, W. W., and Rhodes, E. (1978). Measuring the efficiency of decision making units. *European journal of operational research*, 2(6):429–444.
- [Chetty et al., 2014] Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American economic review*, 104(9):2593–2632.
- [Chetty and Hendren, 2018] Chetty, R. and Hendren, N. (2018). The impacts of neighborhoods on intergenerational mobility ii: County-level estimates. *The Quarterly Journal of Economics*, 133(3):1163–1228.
- [Croiset and Gary-Bobo, 2024] Croiset, S. and Gary-Bobo, R. (2024). Are public hospitals inefficient? an empirical study on french data.
- [Gu and Koenker, 2023] Gu, J. and Koenker, R. (2023). Invidious comparisons: Ranking and selection as compound decisions. *Econometrica*, 91(1):1–41.
- [Kiefer and Wolfowitz, 1956] Kiefer, J. and Wolfowitz, J. (1956). Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *The Annals of Mathematical Statistics*, pages 887–906.
- [Kline and Walters, 2021] Kline, P. and Walters, C. (2021). Reasonable doubt: Experimental detection of job-level employment discrimination. *Econometrica*, 89(2):765–792.
- [Koenker and Gu, 2017] Koenker, R. and Gu, J. (2017). Rebayes: an r package for empirical bayes mixture methods. *Journal of Statistical Software*, 82:1–26.
- [Koenker and Mizera, 2014] Koenker, R. and Mizera, I. (2014). Convex optimization, shape constraints, compound decisions, and empirical bayes rules. *Journal of the American Statistical Association*, 109(506):674–685.

- [Laird, 1978] Laird, N. (1978). Nonparametric maximum likelihood estimation of a mixing distribution. *Journal of the American Statistical Association*, 73(364):805–811.
- [Meeusen and van Den Broeck, 1977] Meeusen, W. and van Den Broeck, J. (1977). Efficiency estimation from cobb-douglas production functions with composed error. *International economic review*, pages 435–444.
- [Robbins, 1992] Robbins, H. E. (1992). An empirical bayes approach to statistics. In *Breakthroughs in Statistics: Foundations and basic theory*, pages 388–394. Springer.