L'Hopital's (Selection) Rule:
A Bayesian Application to French Hospitals

Fu Zixuan
Supervised by Thierry Magnac *

July 4, 2024

Abstract

Something interesting

*Last compiled on June 23, 2024

Contents

1 Introduction **3**

2 Data and Estimation **5**

2.1 Data 5

2.2 Estimation 7

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. The society itself is constantly constructing league table as well. It rewards the meritorious and question or even punishes the unsatisfactory. The measure based on which rank is constructed ranges from teacher's evaluation (Chetty et al., 2014), communities' mobility index (Chetty and Hendren, 2018) to firm discrimination (Kline et al., 2022).

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 called *The Annual Statistics of Health Establishments (SAE)* of French hospitals, I 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. I borrow from the recent developments in Empirical Bayes method to achieve the comparison.

I found that out of the top 20% best performing hospitals, there are roughly 5 times more private units than the public, adjusted by the number of hospitals in each category. The difference is more pronounced when I also control for the expected number of wrongly selected. The takeaway is that public hospitals are in general less efficient than private ones. While the conclusion is in line with that of Croiset and Gary-Bobo (2024) that, now we have a granular perspective on the performance comparison.

The article bridges two fields of interests. The first one is on productivity analysis. The most popular methods in the field are Data Envelopment Analysis (Charnes et al., 1978) and Stochastic Frontier Analysis (Aigner et al., 1977; Meeusen and van Den Broeck, 1977). Yet I abstract from both of them and use the *conditional input demand function* specification stated in Croiset and Gary-Bobo (2024).[1] To put it simply, we estimate a linear function of how much labor input is needed to produce a give list of 8 hospital outputs. I only focus on the employment level of nurses because unlike medical doctors, this is a category that do not suffer from a shortage of labor supply.

The second area of interests is the Empirical Bayes Methods. I lean on a series of work by Jiaying Gu and Roger Koenker, chiefly the following two papers. Gu and Koenker (2017) discussed the usefulness of estimating a prior distribution in baseball batting average prediction. And Gu and Koenker (2023) has formally defined the selection problem as a compound decision on which the estimated prior can be of help as well. Kiefer and Wolfowitz (1956) has shown that non parametric maximun likelihood estimation of the prior is feasible and consistent. The computation of NPMLE is greatly improved by Koenker and Mizera (2014) by leveraging the recent development in convex optimization (Andersen and Andersen, 2010). I will be using the `REBayes` package (Koenker and Gu, 2017) in the estimation, which is based on software `MOSEK` developed by Andersen and Andersen (2010).

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

[1] I refer the reader to Croiset and Gary-Bobo (2024) for detailed reasons of adopting such an approach.

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 standard fixed-effect estimation and GMM.

The benefit of the panel data estimator is that it gives us an estimate of the underlying heterogeneity, which opens door to individual comparisons. However, the fixed effect estimates are generally noisy, rendering the ensuing decision maker hand wavy in making choices. The EB methods are proposed in an attempt to rectify the situation by empirically estimating the prior distribution of the fixed effect.

For example, in Gu and Koenker (2023), we are given the task of selecting the top 20% fixed effect denoted by θ_i. If the θ_i follows a distribution G, this is to say we are selecting those $\theta_i > G^{-1}(0.8)$. The decision rule for individual i is an indicator function δ_i, determining whether i belongs to selection set. The task naturally falls under the compound decision framework pioneered by **?** if we define the loss function of the selection problem in such a way that takes into account the results of all the individual decisions δ_i.

$$\delta^* = \arg\min_{\delta} \mathbb{E}_G \mathbb{E}_{\theta|\hat{\theta}} \left(L_n \right).$$

Since we don't know the true value θ, we minimize the expected compound loss L_n over the distribution of θ given the observed $\hat{\theta}$.

In addition to the capacity constraint of the top 20%, Gu and Koenker (2023) further controls for the number of Type II mistakes made in the selection process. The false discovery rate (FDR) constraint is imposed to ensure that the expected number of wrongly selected units is below a certain level. The FDR constraint is a measure of the proportion of false positives among all the selected units defined as $\mathbb{P}(h_i = 0|\delta_i = 1) \leq \gamma$.

Being interested in the top performing French hospitals, I define my selection problem as *Left tail selection* because the goal is to choose the bottom 20% of the hospital fixed effect θ_i. A smaller θ_i indicates that less labor input is needed to produce the same amount of output, as compared to hospitals with higher θ_i.

It is worth mentioning that classical empirical Bayes method assumes a parametric form of the prior distribution G which is computationally more attractive. Yet thanks to fast convex optimization algorithms, the non-parametric maximum likelihood estimation is now both feasible and efficient. Nevertheless, we are completely free from imposing any parametric assumption. In fact, there are two *layers* of distribution. The lower hierarchy is the prior G with $\theta \sim G$ while the higher hierarchy is $\hat{\theta}|\theta \sim P_\theta$. It is when P_θ belongs to the exponential family that the Lindsay (1995) results hold. Usually in application, we need to impose assumptions or perform some transformation such that P_θ is normal. This kind of procedure is often questionable. Often times, researchers resort to asymptotics to justify the normality assumption, which may not be valid in small samples.

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. It then applies the classical panel data estimators to the same specification, distinguishing between whether strict exogeneity is assumed. In section 3, I introduce the compound decision framework and the method to non parametrically estimate G. In section 4, I specifically define the selection problem following the framework of Gu and Koenker (2023). Section 5 follows with a comparison of the different selection outcome. I try to draw preliminary conclusion on the comparative performance of public and private hospitals. Section 6 discusses potential issues and concludes.

2 Data and Estimation

2.1 Data

The data we used is called *The Annual Statistics of Health Establishments (SAE)*[2]. It is a comprehensive, mandatory administrative survey and the primary source of data on all health establishments in France. We primarily exploited the report of healthcare output (a list of 10 output measure) and labor input (registered and assistant nurses). The panel covers 9 years from 2013 to 2022, with 2020 missing due to the pandemic. The SAE data only distinguishes 3 types of units based on legal status.

1. *Public hospitals*
2. *Private for-profit hospitals*
3. *Private non-profit hospitals*

Following Croiset and Gary-Bobo (2024), I further single out/distinguish the *public teaching hospitals* from the public hospitals since it is intrinsically different from others in the French healthcare system.

As shown in Table **??**, The number of hospitals in normal public, private for-profit, private non-profit are roughly equal and stable over the years. With respect to the teaching hospitals, it is worth mentioning that they not only provide treatments like other types of hospitals but spend a significant amount of resources on doctor training and research as well. Since teaching hospitals have more missions on top of the regular healthcare provision, it is natural that they are in general larger in size. This latter point can be seen much more clearly we present the hospital's output share. Despite being relatively few in number, their share of output is quite substantial. The difference is more pronounced after being adjusted by the number of hospitals as shown in Table **??**.

Moreover, we see that each type of hospital differs in terms of the mix of services they provide. For example, emergency care is mostly taken care of by public hospitals and private hospitals are strong in medical sessions.

[2] La Statistique annuelle des établissements (SAE)

Year	Teaching	Normal Public	Private For Profit	Private Non Profit	Total
2013	198	1312	1305	1382	4197
2014	201	1274	1293	1349	4117
2015	211	1275	1297	1349	4132
2016	212	1266	1297	1313	4088
2017	211	1249	1297	1306	4063
2018	214	1247	1296	1288	4045
2019	214	1236	1287	1281	4018
2021	219	1222	1293	1264	3998
2022	220	1220	1296	1259	3995

Table 1: Number of hospitals in each category, 2013-2022

Output	Teaching	Normal Public	Private For Profit	Private Non Profit	Total
STAC inpatient	25.17%	43.09%	23.64%	8.1%	100%
STAC oupatient	18.4%	19.46%	52.95%	9.18%	100%
Sessions	14.49%	21.96%	34.4%	29.16%	100%
Outpatient Consultations	36.8%	52.45%	0.23%	10.52%	100%
Emergency	21.4%	60.06%	13.37%	5.17%	100%
Follow-up care and Long-term care	7.6%	19.47%	37.95%	34.98%	100%
Home hospitalization	13%	17.38%	12.4%	57.22%	100%
Psychiatry stays	6.53%	62.26%	12.93%	18.28%	100%

Table 2: Hospital share of output, 2013-2022

Output	Teaching	Normal Public	Private For Profit	Private Non Profit	Total
STAC inpatient	66.98%	19.29%	10.25%	3.48%	100%
STAC oupatient	57.91%	10.29%	27.13%	4.67%	100%
Sessions	50.12%	12.7%	20.18%	16.99%	100%
Outpatient Consultations	77.69%	18.64%	0.08%	3.59%	100%
Emergency	62.02%	29.26%	6.31%	2.41%	100%
Follow-up care and Long-term care	33.5%	14.37%	27.31%	24.82%	100%
Home hospitalization	47.83%	10.75%	7.46%	33.96%	100%
Psychiatry stays	29.65%	47.38%	9.6%	13.37%	100%

Table 3: Hospital share of output weighted by the number of hospitals, 2013-2022

For example, the value a_{ij} where i is STAC inpatient and j is teaching hospitals, is calculated by $a_{ij} = \frac{\text{Number of STAC inpatient in teaching hospitals}}{\text{Share of teaching hospitals} \times \text{Total number of STAC inpatient}}$.

2.2 Estimation

Let $\log(x_{it})$ be the number of nurses in hospital i at time t, and $\log(y_{it})$ denote a vector of output levels. I estimate

$$\log(x_{it}) = \beta_0 + \beta_1 \log(y_{it}) + \varepsilon_{it} \tag{1}$$

First, having performed the regression separately for each type of hospital, it is without surprise that teaching hospitals have very different coefficients, as shown in Table **??**. In addition to the differences in descriptive statistics from the last section, this intrinsic difference in input demand functions or equivalently in production function is another sign that teaching hospitals may not be directly comparable to other types of hospitals. For this reason, I will exclude teaching hospitals from the subsequent analysis.

Dependent Variable:	log(ETP_INF)			
	Teaching	Public	Forprofit	Nonprofit
Model:	(1)	(2)	(3)	(4)
Variables				
Constant	2.40^a	0.609^a	1.20^a	0.637^a
	(0.338)	(0.222)	(0.088)	(0.142)
log(SEJHC_MCO)	0.121^a	0.327^a	0.222^a	0.309^a
	(0.045)	(0.041)	(0.013)	(0.030)
log(SEJHP_MCO)	0.132^a	0.090^a	0.059^a	0.046^c
	(0.031)	(0.011)	(0.009)	(0.025)
log(SEANCES_MED)	0.081^a	0.062^a	0.080^a	0.121^a
	(0.019)	(0.007)	(0.005)	(0.015)
log(CONSULT_EXT)	0.019^b	0.024^a	-0.008	0.006
	(0.009)	(0.006)	(0.009)	(0.011)
log(PASSU)	0.053^a	0.003	0.034^a	0.031^a
	(0.011)	(0.008)	(0.005)	(0.011)
log(ENTSSR)	0.044^a	0.036^a	0.040^a	0.080^a
	(0.012)	(0.008)	(0.006)	(0.019)
log(SEJ_HAD)	0.028	0.028^a	0.047^a	-0.004
	(0.029)	(0.006)	(0.018)	(0.021)
log(SEJ_PSY)	0.034^a	0.085^a	0.082^a	0.073
	(0.011)	(0.012)	(0.021)	(0.051)
Fit statistics				
Observations	1,268	5,938	5,047	3,064
R^2	0.828	0.873	0.723	0.707

Clustered (FI) standard-errors in parentheses
Signif. Codes: a: 0.01, b: 0.05, c: 0.1

Table 4: Separate estimation of input demand function, 2013-2022

References

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, E. D. and Andersen, K. D. (2010). The mosek optimization tools manual, version 6.0.

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, 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, 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, S. and Gary-Bobo, R. (2024). Are public hospitals inefficient? an empirical study on french data.

Gu, J. and Koenker, R. (2017). Empirical bayesball remixed: Empirical bayes methods for longitudinal data. *Journal of Applied Econometrics*, 32(3):575–599.

Gu, J. and Koenker, R. (2023). Invidious comparisons: Ranking and selection as compound decisions. *Econometrica*, 91(1):1–41.

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, P., Rose, E. K., and Walters, C. R. (2022). Systemic discrimination among large us employers. *The Quarterly Journal of Economics*, 137(4):1963–2036.

Koenker, R. and Gu, J. (2017). Rebayes: an r package for empirical bayes mixture methods. *Journal of Statistical Software*, 82:1–26.

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.

Lindsay, B. G. (1995). Mixture models: theory, geometry, and applications. Ims.

Meeusen, W. and van Den Broeck, J. (1977). Efficiency estimation from cobb-douglas production functions with composed error. *International economic review*, pages 435–444.