Something interesting

Fu Zixuan*

July 4, 2024

Abstract

Something interesting

*Last compiled on June 19, 2024

Contents

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 [1]. We only focus on the employment level of nurses for (reasons) [2] 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.

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

[1] We refer the audience to xxx for detailed reasons of adopting this approach.

[2]

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 $\mathbb{E}\Delta x_{i,t-1}\varepsilon_{i,t}$, 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 β parameters. It also provides us with a noisy estimate of the underlying unobserved heterogeneity term denoted as θ_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 θ_i, denoted as G_θ. If the prior distribution G is known, having observed an estimate $\hat{\theta}_i$ of θ_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 θ_i. For example, in [Gu and Koenker, 2023], we are given the task of selecting the top 20% out of the population of θ_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

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

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 [**?**] 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-parmetric 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 $\hat{\theta}_i|\theta_i, \sigma_i^2$. To illustrate, in the case of the estimate of fixed effect, we have

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

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

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

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 ($N >> T$).

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. It then applies the classical panel data estimators to the same specification, distinguishing between whether strict exogeneity is assumed. In section 3, we introduce the compound decision framework and specifically/xxx define the selection problem, we then . Section 4 follows with a comparison of the different selection outcome as a result of imposing varying constraints and assumptions. We try to draw preliminary conclusion on the comparative performance of public and private hospitals. Section 5 discusses potential issues and concludes.

	Year	Teaching	Normal Public	Private For Profit	Private Non Profit	Total
1	2013	198	1312	1305	1382	4197
2	2014	201	1274	1293	1349	4117
3	2015	211	1275	1297	1349	4132
4	2016	212	1266	1297	1313	4088
5	2017	211	1249	1297	1306	4063
6	2018	214	1247	1296	1288	4045
7	2019	214	1236	1287	1281	4018
8	2021	219	1222	1293	1264	3998
9	2022	220	1220	1296	1259	3995

2 Data and Estimation

2.1 Data

The data we used is called *The Annual Statistics of Health Establishments (SAE)*[3]. 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 number of healthcare units is rather stable. The SAE data only distinguishes 3 types of units based on legal status.

1. *Ordinary public hospitals*
2. *Private for-profit hospitals*
3. *Private non-profit hospitals*

Following [Croiset and Gary-Bobo, 2024], we further single out the *public teaching hospitals* because its main objective consists not only of patient treatment, but also teaching and research by and large.

Summarize the output categories

	Output	Normal Public	Private Non Profit	Private For Profit	Teaching
1	STAC inpatient	8.08%	5.66%	16.3%	7.9%
2	STAC oupatient	2.26%	4.02%	22.61%	3.59%
3	Sessions	4.34%	23.31%	27.17%	4.8%
4	Outpatient Consultations	58.23%	43.55%	0.8%	69.18%
5	Emergency	21.14%	6.78%	17.3%	12.64%
6	Follow-up care and Long-term care	1.67%	11.26%	12.16%	1.09%
7	Home hospitalization	0.06%	0.76%	0.17%	0.08%
8	Psychiatry stays	4.22%	4.66%	3.49%	0.72%

Summarize the input: explain the reason why we use nurses instead of medical doctors.

[3] La Statistique annuelle des établissements (SAE)

2.2 Estimation (Pooling)

[4]

A simple counterfactual to perform:

2.3 Estimation (Fixed effect)

Though [Croiset and Gary-Bobo, 2024] has mentioned that identification (and significance level) is mainly attributed to *between-group* variation, there's still some level of *within-group* variation that guarantee the estimation of parameters.

The strict exogeneity assumption is not always realistic, especially in the case of production or factor demand function estimation. Therefore, we resort to a series of seminal work in panel data estimation [Anderson and Hsiao, 1982, Arellano and Bond, 1991, Arellano and Bover, 1995, Blundell and Bond, 1998].

It is worth mentioning that the standard first difference GMM where lagged level $x_{i,t-2}$ is used as instruments for the first difference equation $\Delta\varepsilon = \Delta y_{i,t} - \Delta x_{i,t}\beta$ does not produce results of pure noise. This weak instrument issue was pointed out by [Blundell and Bond, 1998].

In the rest of the article, I proceed with the estimation results of system GMM as in col 3 of Table **??**.

3 Compound Decision: The Selection Problem

3.1 Compound Decision

Leaving the estimation above aside, I will first introduce the idea of compound decision pioneered by [Robbins, 1956] before entering into the ranking and selection problem.

Now there are N units, each has an unobserved parameter labelled as $(\theta_1, \dots, \theta_n)$. We are given N estimates $(\hat{\theta}_1, \dots, \hat{\theta}_n)$ of the underlying heterogeneous parameters. And each estimate $\hat{\theta}_i$ conditioned on θ_i follows a distribution $\hat{\theta}_i|\theta_i \sim P_{\theta_i}$. No matter what the specific task is, I care about the collective performance of my decision. That being said, I will explicitly define the loss function to reflect my attention to the so-called collective performance.

First the decision rule is defined as a vector of individual decisions

$$\delta(Y) = (\delta_1(\hat{\theta}), \dots, \delta_n(\hat{\theta}))$$

where each $\delta_i(\cdot)$ is a decision rule for the estimate of θ_i with the vector $\hat{\theta}$ as input. The

[4] On a side note, we filtered the panel such that

1. the number of nurses is positive,
2. at least one of STAC inpatient, STAC outpatient, Sessions is positive,
3. the number of observations is larger than 6

Dependent Variable:	log(ETP_INF)			
	Pool	Pool IV	Dummy	Dummy IV
Model:	(1)	(2)	(3)	(4)
Variables				
Constant	1.38***	0.987***	1.60***	1.02***
	(0.059)	(0.056)	(0.020)	(0.063)
log(SEJHC_MCO)	0.282***	0.235***	0.277***	0.237***
	(0.013)	(0.012)	(0.003)	(0.011)
log(SEJHP_MCO)	0.048***	0.076***	0.057***	0.070***
	(0.008)	(0.007)	(0.002)	(0.008)
log(SEANCES_MED)	0.064***	0.085***	0.066***	0.082***
	(0.004)	(0.004)	(0.002)	(0.004)
log(CONSULT_EXT)	0.053***	0.050***	0.024***	0.033***
	(0.003)	(0.004)	(0.002)	(0.005)
log(PASSU)	0.020***	0.020***	0.022***	0.028***
	(0.003)	(0.003)	(0.002)	(0.003)
log(ENTSSR)	0.078***	0.051***	0.068***	0.048***
	(0.005)	(0.005)	(0.002)	(0.005)
log(SEJ_HAD)	0.018**	0.017**	0.027***	0.028***
	(0.009)	(0.008)	(0.004)	(0.007)
log(SEJ_PSY)	0.071***	0.082***	0.062***	0.079***
	(0.009)	(0.010)	(0.003)	(0.009)
STJR2			-0.300***	-0.021
			(0.021)	(0.059)
STJR3			-0.215***	0.014
			(0.019)	(0.053)
STJR0			0.718***	0.703***
			(0.022)	(0.051)
Fit statistics				
Observations	15,317	11,634	15,317	11,634
R^2	0.817	0.818	0.835	0.832

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Dependent Variable:	log(ETP_INF)			
	Teaching	Public	Forprofit	Nonprofit
Model:	(1)	(2)	(3)	(4)
Variables				
Constant	3.28^{a}	1.38^{a}	1.40^{a}	1.00^{a}
	(0.328)	(0.262)	(0.095)	(0.149)
log(SEJHC_MCO)	0.108^{b}	0.331^{a}	0.261^{a}	0.344^{a}
	(0.042)	(0.048)	(0.015)	(0.034)
log(SEJHP_MCO)	0.132^{a}	0.078^{a}	0.048^{a}	0.046^{c}
	(0.032)	(0.013)	(0.011)	(0.027)
log(SEANCES_MED)	0.060^{a}	0.051^{a}	0.075^{a}	0.094^{a}
	(0.020)	(0.007)	(0.006)	(0.016)
log(CONSULT_EXT)	0.017	0.025^{a}	-0.003	0.001
	(0.014)	(0.008)	(0.011)	(0.012)
log(PASSU)	0.049^{a}	-0.009	0.033^{a}	0.025^{b}
	(0.011)	(0.008)	(0.005)	(0.010)
log(ENTSSR)	0.058^{a}	0.052^{a}	0.057^{a}	0.118^{a}
	(0.013)	(0.008)	(0.008)	(0.019)
log(SEJ_HAD)	0.022	0.028^{a}	0.049^{a}	-0.011
	(0.027)	(0.007)	(0.018)	(0.022)
log(SEJ_PSY)	0.026^{b}	0.070^{a}	0.084^{a}	0.045
	(0.011)	(0.010)	(0.018)	(0.046)
Fit statistics				
Observations	1,123	5,260	4,415	2,604
R^2	0.779	0.860	0.742	0.754

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

Dependent Variable:		log(ETP_INF)	
	Within Group	First Difference	System GMM
Model:	(1)	(2)	(3)
Variables			
log(SEJHC_MCO)	0.10***	0.07***	0.54***
	(0.00)	(0.01)	(0.02)
log(SEJHP_MCO)	0.02***	0.01***	0.02
	(0.00)	(0.00)	(0.02)
log(SEANCES_MED)	0.02***	0.02***	0.06***
	(0.00)	(0.00)	(0.01)
log(CONSULT_EXT)	0.00	0.00	0.05***
	(0.00)	(0.00)	(0.01)
log(PASSU)	0.02***	0.01**	−0.04
	(0.00)	(0.00)	(0.02)
log(ENTSSR)	0.01***	0.01***	0.04**
	(0.00)	(0.00)	(0.02)
log(SEJ_HAD)	0.01***	0.02***	−0.00
	(0.00)	(0.00)	(0.02)
log(SEJ_PSY)	0.00	0.00	0.07***
	(0.00)	(0.00)	(0.01)
Fit statistics			
Observations	15335	13502	15335
R^2	0.07	0.03	
n			1833
T			9

*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1* individual loss for a decision is $L(\theta_i, \delta_i(\hat{\theta}_i))$, giving rise to the aggregated **compound loss**

$$L_n(\theta, \delta(\hat{\theta})) = \sum_{i=1}^{n} L(\theta_i, \delta_i(\hat{\theta})).$$

While the compound risk, defined as the expectation of compound loss can be expressed as

$$\begin{aligned}
R_n(\theta, \delta(\hat{\theta})) &= \mathbb{E}[L_n(\theta, \delta(\hat{\theta}))] \\
&= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}[L(\theta_i, \delta_i(\hat{\theta}))] \\
&= \frac{1}{n}\sum_{i=1}^{n} \int \dots \int L(\theta_i, \delta_i(\hat{\theta}_1, \dots, \hat{\theta}_n)) dP_{\theta_1}(\hat{\theta}_1) \dots dP_{\theta_n}(\hat{\theta}_n).
\end{aligned}$$

Given the objective function which is the compound risk, our goal is to find a decision rule $\delta(\cdot)$ that minimizes it. This is the optimal compound decision rule for a given vector $\hat{\theta}$.

$$\delta^*(\hat{\theta}) = \arg\min_{\delta} R_n(\theta, \delta(\hat{\theta}))$$

If $\delta^*(\hat{\theta})$ is separable [5], which means that $\delta^*(\hat{\theta}) = \{t(\hat{\theta}_1), \dots, t(\hat{\theta}_n)\}$, the compound risk can be written as

$$\begin{aligned}
R_n(\theta, \delta(\hat{\theta}))f &= \frac{1}{n}\sum \int \dots \int L(\theta_i, \delta(\hat{\theta}_1, \dots, \hat{\theta}_n)) dP_{\theta_1}(\hat{\theta}_1) \dots dP_{\theta_n}(\hat{\theta}_n) \\
&= \int_{\theta} \int L(\theta_i, t(\hat{\theta}_i)) dP_{\theta_i}(\hat{\theta}_i) dG_n(\theta) \\
&= \mathbb{E}_{G_n}\mathbb{E}_{\theta}\left[L(\theta_i, \delta_i(\hat{\theta}))\right]
\end{aligned}$$

where $G_n(\theta)$ is the empirical distribution of θ.[6]

It is worth mentioning that up til now we treat each θ_i as fixed unknown parameters [7]. However, if we take a Bayesian view on the vector $\boldsymbol{\theta}$ by assuming that each θ_i is an *i.i.d.* draw from an underlying common distribution G. The Bayesian risk is

$$\mathbb{E}_G\mathbb{E}_{\theta}\left[L(\theta_i, \delta_i(\hat{\theta}))\right]$$

The two views are closely linked via the G_n and G.

Compound Risk is equivalent to the Bayes risk with prior G_n.

[5] The linear shrinkage class belongs to this class as well. See appendix.

[6] $\mathbb{E}_{G_n}[(f(x))] = 1/n \sum_i^n f(x_i)$

[7] By abuse of terminology, we called this *fixed effect view* while the other assumption *random effect view*. Yet the two terms have nothing to do with whether θ_i is correlated with x_{it}

3.2 Selection Problem

Our task at hand is to select the top $\alpha\%$ (e.g. 20%) of the hospitals in terms of labor use efficiency. If θ_i represents a measure of *inefficiency* which is the fixed effect term in the linear input demand function specified in Section 2. We want to select the top θ_is that is below than the α quantile of the population $\theta_i < G_n^{-1}(\alpha)$. Moreover, we want to subject the selection to another constraint which is the False Discovery Rate constraint at level γ, that is

$$\begin{aligned}\mathbb{P}\left[\theta_i > \theta_\alpha | \delta_i = 1\right] &= \frac{\mathbb{P}\left[\theta_i > \theta_\alpha, \delta_i = 1\right]}{\mathbb{P}\left[\delta_i = 1\right]} \\ &= \frac{\mathbb{E}_G\left[1\left\{\theta_i > \theta_\alpha, \delta_i = 1\right\}\right]}{\mathbb{E}_G\left[\delta_i\right]} \\ &\leq \gamma\end{aligned}$$

Now we are in the position to write down the selection problem subject to the capacity constraint at α and FDR constraint at γ level, with multiplier τ_1 and τ_2 respectively. We denote $\delta_i = 1$ when unit i is selected and $h_i = 1\{\theta_i < \theta_\alpha\} = 1$ when unit i is truly below the threshold $\theta_\alpha = G^{-1}(\alpha)$. The compound loss function is defined as

$$L(\delta, \theta) = \sum h_i(1-\delta_i) + \tau_1\left(\sum(1-h_i)\delta_i - \gamma\delta_i\right) + \tau_2\left(\sum \delta_i - \alpha n\right)$$

To minimize the compound risk is thus

$$\begin{aligned}\min_\delta \mathbb{E}_G\mathbb{E}_{\theta|\hat{\theta}}\left[L(\delta,\theta)\right] &= \mathbb{E}_G\sum \mathbb{E}(h_i)(1-\delta_i) + \tau_1\left(\sum(1-\mathbb{E}(h_i))\delta_i - \gamma\delta_i\right) + \tau_2\left(\sum\delta_i - \alpha n\right) \\ &= \mathbb{E}_G\sum v_\alpha(\hat{\theta})(1-\delta_i) + \tau_1\left(\sum(1-v_\alpha(\hat{\theta}))\delta_i - \gamma\delta_i\right) + \tau_2\left(\sum\delta_i - \alpha n\right)\end{aligned}$$

where $v_\alpha(\hat{\theta}) = \mathbb{P}(\theta < \theta_\alpha|\hat{\theta})$, which we called **posterior tail probability**. This is in contrast to the posterior mean widely used to shrink the estimate $\hat{\theta}_i$. For the moment, it may not immediately obvious how important the prior distribution G is. I will further illustrate it in the next section.

3.3 Posterior tail probability

For each θ_i, I observe a sequence of Y_{it} coming from a longitudinal model

$$Y_{it} = \theta_i + \varepsilon_{it} \quad \varepsilon_{it} \sim \mathcal{N}(0, \sigma_i^2) \quad (\theta_i, \sigma_i^2) \sim G$$

Neither θ_i nor σ_i^2 is known to us. But there are two sufficient statistics for (θ_i, σ_i^2).

$$Y_i = \frac{1}{T_i}\sum_{t=1}^{T_i} Y_{it} \quad \text{where} \quad Y_i|\theta_i, \sigma_i^2 \sim \mathcal{N}(\theta_i, \sigma_i^2/T_i)$$
$$S_i = \frac{1}{T_i - 1}\sum_{t=1}^{T_i}(Y_{it} - Y_i)^2 \quad \text{where} \quad S_i|\sigma_i^2 \sim \Gamma(r_i = (T_i - 1)/2, 2\sigma_i^2/(T_i - 1))$$

The tail probability $v_\alpha(y_i, s_i)$ given the two sufficient statistics is defined as

$$\begin{aligned} v_\alpha(Y_i, S_i) &= P(\theta_i > \theta_\alpha | Y_i, S_i) \\ &= \frac{\int_{-\infty}^{\theta_\alpha} \Gamma(s_i|r_i, \sigma_i^2) f(y_i|\theta_i, \sigma_i^2) dG(\theta_i, \sigma_i^2)}{\int_{-\infty}^{\infty} \Gamma(s_i|r_i, \sigma_i^2) f(y_i|\theta_i, \sigma_i^2) dG(\theta_i, \sigma_i^2)} \end{aligned}$$

Write out the capacity constraint

Write out the FDR constraint If our specification and assumptions on exogeneity are correct, the consistency of $\hat{\beta}$ is guaranteed by N's asymptotic. However, our estimate of the fixed effect is

$$\begin{aligned} \hat{\theta}_i &= \frac{1}{T}\sum(\theta_i + \varepsilon_{it} + x_{it}(\beta - \hat{\beta})) \\ &\xrightarrow{N\to\infty} \theta_i + \frac{1}{T}\sum_t \varepsilon_{it} \end{aligned}$$

When T is relatively small (or even fixed), I am not in a good position to use central limit theorem to claim that $\hat{\theta}_i \xrightarrow{d} \mathcal{N}(\theta_i, \frac{\sigma_i^2}{T})$. A bold assumption that $\varepsilon_{it} \sim \mathcal{N}(0, \sigma_i^2)$ will save me from the T issue, which I will impose for the rest of the section (and abstract from whether that for each i is a testable/reasonable/feasible assumption).

4 Selection results

Although [Gu and Koenker, 2023] has presented the decision rule when (θ_i, σ_i^2) are unknown. The application assumes that σ_i^2 is known/observable. In this section, I will compare the selection results under known variance σ_i^2 and estimated variance S_i.

Known variance, and 4 rules

Unknown variance, and 4 rules

5 Conclusion

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.

[Anderson and Hsiao, 1982] Anderson, T. W. and Hsiao, C. (1982). Formulation and estimation of dynamic models using panel data. *Journal of econometrics*, 18(1):47–82.

[Arellano and Bond, 1991] Arellano, M. and Bond, S. (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *The review of economic studies*, 58(2):277–297.

[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, 1956] Robbins, H. E. (1956). An empirical bayes approach to statistics. In *Proceedings of the third berkeley symposium on mathematical statistics and probability*, volume 1, pages 157–163.