Something interesting

Fu Zixuan*

July 4, 2024

Abstract

someting interesting

*Last compiled on June 17, 2024

Contents

1 Introduction

Selection is ubiquitous, whether it is the all-powerful natural selection in biological evolution or the artificial construction of a league table to identify (and improve) the best (and worst) performing schools, firms, and hospitals. In most cases, it serves as the driving force behind improvement and change. While the first is beyond human control, artificial selection is so significant that it must be approached with great caution.

In the health sector, it is financially and socially advantageous to study the labor and employment efficiency of each healthcare provider. Ranking and selection can be performed based on measures of labor efficiency. Naturally, learning from the best or investigating the worst are actions that may follow the results of such selections.

By exploring a comprehensive database (SAE) of French hospitals, the article attempts to compare public and private hospitals by selecting the top-performing units. It utilizes recent advances in the field of Empirical Bayes to achieve this comparison.

***Conclusion Here**

The article bridges two fields of interests. One is the on the productivity analysis. The other on Empirical Bayes Methods. In the area of productivity analysis, we use the simple method applied in [?] by estimating a conditional input demand function, that is to say, how much labor input is needed given a list of outputs [1]. We only focus on the employment level of nurses for (reasons) in the regression specification.

A bit about the standard productivity analysis literature

Having roughly replicated the results in [?], the paper differentiates itself by including the standard/classical panel data methods in estimation, specifically the fixed-effect within-group estimation and GMM [?].

A tiny bit about the literature of GMM

The two methods employed provide us with a noisy estimate of the underlying unobserved heterogeneity term (it's important to note that this heterogeneity is not necessarily indicative of inefficiency). Central to Bayesian philosophy is the prior distribution of α_i, denoted as G_α. If the prior distribution G is known, having observed an estimate $\hat{\alpha}_i$ of α_i, we can update our estimate using or incorporating our knowledge of G.

The usefulness of a prior G is further exemplified in the ranking and selection problem mentioned above, when the object of interests is the noisy estimate of α_i. For example,in [?], we are given the task of selecting the top 20

$$\delta^* = \min_\delta \text{Losssubject to contstraints}$$

where the loss function can take different forms, for example it can be 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 a way that takes into account the results of all the individual decisions $\delta(Y_i)$. For instance, summing all mistakes would be one aggregate/compound individual decision.

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

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, "empirically estimating G" is done with parametric assumption that G is normal. Notable use cases are found in [?] which focuses on teacher evaluation and social mobility in communities. 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 work of [?] who has shown that nonparametric estimation is also feasible and consistent, it is preferable to relax the normality assumption and non parametrically estimate the prior G. [?] has formulated the nonparmetric estimation as a convex optimization problem. Compared to the other estimation methods such as EM [?], recent advancements in convex optimization computation methods [?] has made the novel approach [?] computationally more attractive.

It is worth mentioning here that though G can be estimated non parametrically, we have to impose assumptions on the distribution of the estimate of $\hat{\alpha}_i|\alpha_i, \sigma_i^2$. To illustrate,

$$\hat{\alpha}_i = \frac{1}{T}\sum_t (y_{it} - x_{it}\hat{\beta}) = \frac{1}{T}\sum (\alpha_i + \varepsilon_{it} + x_{it}(\beta - \hat{\beta})) \to^d \alpha_i + \frac{1}{T}\sum_t \varepsilon_{it}$$

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

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

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

Then a fixed/small T won't do too much harm since we do not need to invoke central limit theorem to have

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

Without the normality assumption on the error terms, we have to resort to the central limit theorem from $T \to \infty$, which may seem unrealistic for a short panel.

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 nurses as the dependent variable and a list of 9 output measures as the regressors. In Section 3 applies the classical panel estimators to the same specification, distinguishing between whether strict exogeneity is assumed. In section 4, we introduce the compound decision framework and specifically define the selection problem. Section 5 follows with a comparison of different selection results as a result of imposing varying constraints and assumptions. In section 6, we try to draw conclusion on the comparative performance of public and private hospitals. Section 7 discusses potential issues and concludes.