L'Hôpital's (Selection) Rule
An Empirical Bayes Application to French Hospitals

Fu Zixuan

Supervised by Prof.Thierry Magnac

July 1, 2024

Literature: Measuring efficiency of individual units

- *Productivity/Efficiency*: Factories, Schools, Hospitals etc.
- *Ownership*: Public (Teaching, Ordinary) vs. Private (For profit, Non-profit).
- *Methodology*: Following Croiset and Gary-Bobo (2024), we use the *conditional input demand function*. The smaller the θ_i, the less input is needed to produce the same amount of output, the more efficient the hospital is. Reasons

Literature: Invidious decision

- *League table mentality*: Ranking & Selection.(Gu and Koenker, 2023)
- *Noisy estimates*: Unobserved heterogeneity, fixed effect θ_i. (Chetty et al., 2014; Kline et al., 2022)
- *Compound Decision/ Empirical Bayesian*: Compound decision framework (Robbins, 1956), (Non-parametric) Estimation of the prior distribution of θ_i. (Koenker and Mizera, 2014; Gu and Koenker, 2017)

Questions

- Out of the top 20% hospitals in France in terms of labor employment efficiency, how many of them are public hospitals/private hospitals?
- What would be the selection outcome if I want to control the number of mistakes that I make?
- Does different selection rule produce different results? And to what degree?

Recipe

1. **Estimate the efficiency** with input demand function.
   - LHS X: Labor input (number of full time equivalent nurses).
   - RHS Y: Hospital output (e.g., inpatient/outpatient stays, medical sessions).

$$\log(x_{it,\text{nurses}}) = \log(y_{it,\text{output}})\beta + \theta_i + \varepsilon_{it} \quad \text{where} \quad \varepsilon_{it} \sim \mathcal{N}(0, \sigma_i^2)$$

2. **Estimate the prior distribution of θ_i** with NPMLE.

3. **Select the 20% most efficient hospitals.**

Hospital Types

The Annual Statistics of Health Establishments (SAE)[1], 2013-2022 [2].

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

[1] La Statistique annuelle des établissements (SAE)

[2] 2020 missing due to Covid-19

Output

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%

- Hospitals differ not only in efficiency but also in the mix of services they provide.
- Teaching hospitals may be innately very different from others (training, research).

Appendix

Compound Decision Framework

Observe:

$$\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \ldots, \hat{\theta}_n)$$

$$\text{where} \quad \hat{\theta}_i | \theta_i \sim P_{\theta_i}$$

Decision:

$$\delta(\hat{\boldsymbol{\theta}}) = (\delta_1(\hat{\boldsymbol{\theta}}), \ldots, \delta_n(\hat{\boldsymbol{\theta}}))$$

Compound Loss and Risk

Loss:

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

Risk (Expectation of loss):

$$\begin{aligned} R_n(\theta, \delta(\hat{\boldsymbol{\theta}})) &= \mathbb{E}[L_n(\theta, \delta(\hat{\boldsymbol{\theta}}))] \\ &= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}_{\theta_i|\hat{\theta}_i}[L(\theta_i, \delta_i(\hat{\boldsymbol{\theta}}))] \quad \text{Separable decision rule } \delta \\ &= \frac{1}{n} \sum_{i=1}^{n} \int L(\theta_i, \delta_i(\hat{\theta}_i)) dP_{\theta_i}(\hat{\theta}_i) \\ &= \int \int L(\theta_i, \delta(\hat{\theta}_i)) dP_{\theta_i}(\hat{\theta}_i) dG_n(\theta) \end{aligned}$$

where $G_n(\theta)$ is the empirical distribution (Frequentist View) of $\theta \sim G$.
$\rightarrow$ Bayesian view: replace G_n by a distribution G. $\rightarrow$ Empirical Bayes: Estimate the G.

Estimate G

Kiefer and Wolfowitz (1956) established the nonparametric maximum likelihood estimator (NPMLE)

$$\hat{G} = \underset{G \in \mathcal{G}}{\arg\min} \left\{ -\sum_{i=1}^{n} \log g(y_i) \middle| g(y_i) = \int \mathbb{P}(y_i|\theta) dG(\theta) \right\}$$

where $\mathbb{P}(y_i|\theta)$ is the probability density function of y_i conditional on the true parameter $\theta \longrightarrow g(y_i)$ is the marginal pdf of y_i.

Estimate G

This is an **infinite-dimensional** convex optimization problem with a strictly convex objective subject to linear constraints.

$$\min_{f=dG}\left\{-\sum_i \log g(y_i)\middle| g(y_i)=T(f),\ K(f)=1,\ \forall i\right\}$$

where $T(f)=\int \mathbb{P}(y_i|\theta) f d\theta$ and $K(f)=\int f d\theta$.
Consistency is proven by Kiefer and Wolfowitz (1956). Efficient computation method introduced by Koenker and Mizera (2014). Implemented with *Mosek* created by Andersen and Andersen (2010).

The Selection Task

- Select the bottom 20% (the smaller the θ_i, the more efficient) of the true θ_i. Since we assume that $\theta_i \sim G$, those i whose $\theta_i < G^{-1}(0.2)$
- Control the overall false discovery rate at 20%,

$$\frac{\mathbb{E}_G\left[1\left\{\theta_i > \theta_\alpha, \delta_i = 1\right\}\right]}{\mathbb{E}_G\left[\delta_i\right]} \leq \gamma$$

1. Nominator: Selected but whose true value $> G^{-1}(0.2)$.
2. Denominator: Selected.

Problem Formulation

The **loss** function is

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

where $h_i = 1\{\theta_i < \theta_\alpha = G^-1(\alpha)\}$. h_i is an indicator of whether the true value belong to the set. δ_i is an indicator of whether i is being selected. Therefore, the **problem** is to find δ such that

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

where $v_\alpha(\hat{\theta}) = \mathbb{P}(\theta < \theta_\alpha | \hat{\theta})$ is the **posterior tail probability**.

Derive tail probability v_α

Pick hospital i whose true efficiency value is θ_i, which we don't observe. We only observe a sequence of Y_{it} where

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

Neither θ_i nor σ_i^2 is known. But the sufficient statistics are

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

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

In our input demand function specification, we have $Y_{it} = \log(x_{it}) - \beta \log(y_{it})$. Appendix

TP and Constraints

Given the two sufficient statistics, the posterior tail probability is

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

We want to find a cutoff λ such that both constraints are satisfied [3]:

- Capacity: $\int \int P(v_\alpha(Y_i, S_i) > \lambda) dG(\theta_i, \sigma_i^2) \leq \alpha$
- FDR: $\int \int \frac{E[1\{v_\alpha(Y_i,S_i) > \lambda\}(1 - v_\alpha(Y_i,S_i))]}{E[1\{v_\alpha(Y_i,S_i) > \lambda\}]} dG(\theta_i, \sigma_i^2) \leq \gamma$

[3] Relaxed discrete optimization problem, following (Basu et al., 2018)

Recap

1. We have a $N \times T$ panel. Y_{it} is an observation of hospital i's efficiency term θ_i at time t. Say $Y_{it}|\theta_i, \sigma_i \sim \mathcal{N}(\theta_i, \sigma_i^2)$.
2. Given a panel of Y_{it}, perform NPMLE to get an estimate of $G(\theta)$ or $G(\theta, \sigma^2)$.
3. Given the estimated prior G, derive the explicit form of posterior tail probability $v_\alpha(Y_i, S_i)$ and the two constraints.
4. Solve the selection problem and find the optimal δ^*

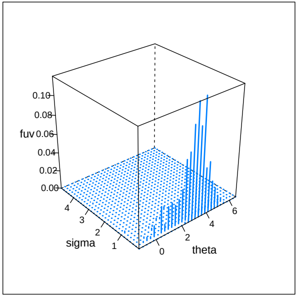
$$\min_{\delta} \mathbb{E}_G \sum v_\alpha(\hat{\theta})(1-\delta_i) + \tau_1 \left(\sum (1 - v_\alpha(\hat{\theta}))\delta_i - \gamma\delta_i\right) + \tau_2 \left(\sum \delta_i - \alpha n\right)$$

5. The decision rule is defined by the cutoff λ

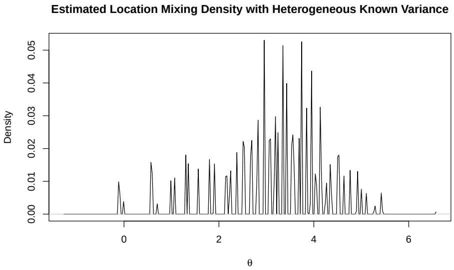
$$\delta^*(y_i, s_i) = 1\left\{v_\alpha(y_i, s_i) > \lambda^*\right\}$$

The estimated $\hat{G}$

Case 1: $G(\theta, \sigma)$ for $v_\alpha(Y_i, S_i)$

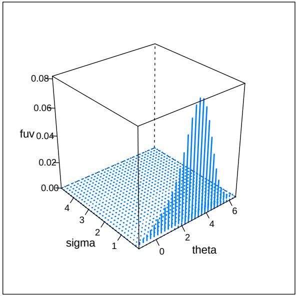


Case 2: $G(\theta)$ for $v_\alpha(Y_i)$

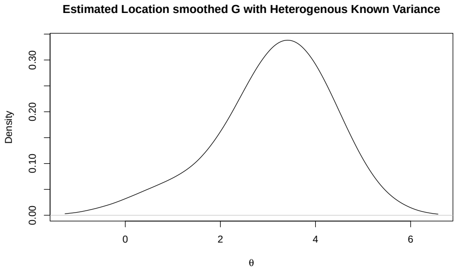


The estimated $\hat{G}$

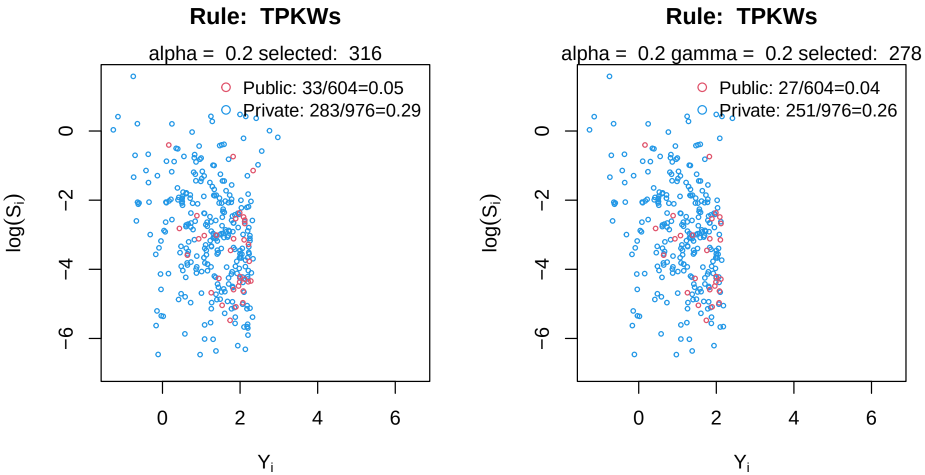
Case 1: $G(\theta, \sigma)$ for $v_\alpha(Y_i, S_i)$



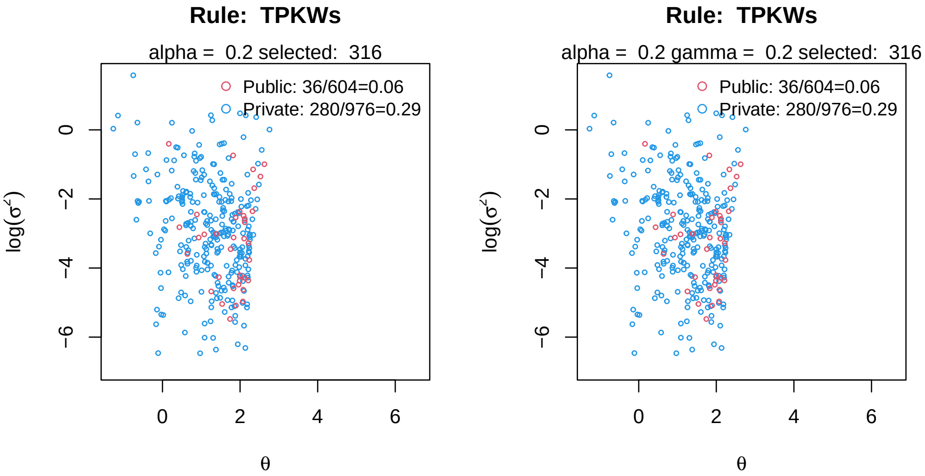
Case 2: $G(\theta)$ for $v_\alpha(Y_i)$



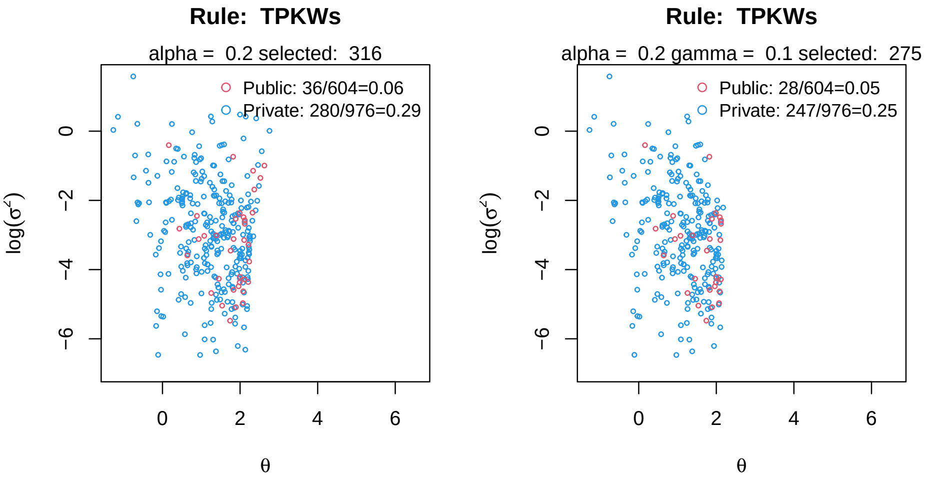
$G(\theta, \sigma)$: Posterior Tail probability (0.2,0.2)



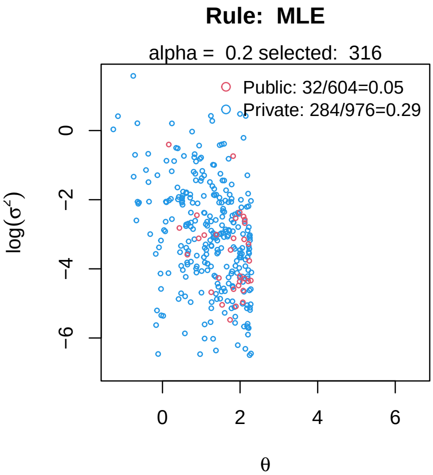
$G(\theta)$: Posterior Tail probability (0.2,0.2)

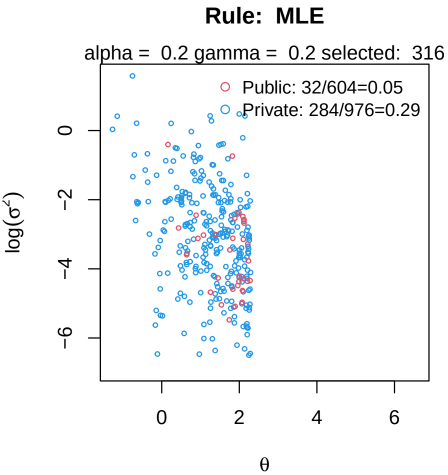


$G(\theta)$: Posterior Tail probability (0.2,0.1)



"Face value"


Rule: MLE
alpha = 0.2 selected: 316
Public: 32/604=0.05
Private: 284/976=0.29
log(σ²)
0
−2
−4
−6
0
2
4
6
θ


Rule: MLE
alpha = 0.2 gamma = 0.2 selected: 316
Public: 32/604=0.05
Private: 284/976=0.29
log(σ²)
0
−2
−4
−6
0
2
4
6
θ

Fixed effect estimation

Assume that $\mathbb{E}[\varepsilon_{it}|\theta_i, x_{i1}, \ldots, x_{i,t-1}] = 0$.
First Difference GMM: use lagged level as instruments for current difference

$$\mathbb{E}[x_{i,t-2}(\Delta y_{it} - \beta \Delta x_{it})]$$

System GMM: use lagged difference as instruments for current levels

$$\mathbb{E}[\Delta x_{i,t-1}(y_{it} - \beta x_{it})] \quad \text{if} \quad \mathbb{E}\left[\Delta x_{i,t-1}(\theta_i + \varepsilon_{i,t})\right] = 0$$

Results

Dependent Variable:	Nurses			
	Within Group	First Difference	FD GMM	SYS GMM
Model:	(1)	(2)	(3)	(4)
Variables				
STAC inpatient	0.10^{***}	0.07^{***}	0.13^{***}	0.48^{***}
	(0.00)	(0.01)	(0.03)	(0.02)
STAC outpatient	0.02^{***}	0.01^{***}	0.02	0.05^{*}
	(0.00)	(0.00)	(0.01)	(0.02)
Medical sessions	0.02^{***}	0.02^{***}	0.02^{***}	0.06^{***}
	(0.00)	(0.00)	(0.00)	(0.01)
External consultations	0.00	0.00	0.01	0.07^{***}
	(0.00)	(0.00)	(0.01)	(0.01)
Emergency	0.01^{***}	0.01	0.05	-0.04^{**}
	(0.00)	(0.00)	(0.04)	(0.01)
Long-term & follow-up	0.01^{***}	0.01^{***}	-0.06^{**}	0.07^{***}
	(0.00)	(0.00)	(0.02)	(0.01)
Home care	0.01^{***}	0.02^{**}	-0.00	0.01
	(0.00)	(0.01)	(0.02)	(0.02)
Psychiatric care	0.02^{***}	0.01	0.01	0.07^{**}
	(0.00)	(0.01)	(0.02)	(0.02)
Fit statistics				
n	1690	1690	1690	1690
T	9	9	9	9
Sargan Test: chisq			128.32	360.97
Sargan Test: df			48.00	104.00

Heteroskedasticity-robust standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Conclusion

- Whether to control for **False discovery rate**$\rightarrow$Control for FDR shrinks the selection set.
- Whether to assume known σ_i makes a difference$\rightarrow$ Assume unknown σ_i makes the FDR constraints bind, thus less selected than assuming σ_i known.
- Private (FP and NP) hospitals are indeed more "efficient" $\rightarrow$ Caution.

Limitation

- Interpretation of the θ_i: The Schmidt and Sickles/Pitt and Lee models treat all time invariant effects as inefficiency. Greene (2005) treats time invariant components as only unobserved heterogeneity.
- Specification, endogeneity, normality assumption on ε_{it}.*etc.* Next

References I

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

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, M. and Bover, O. (1995). Another look at the instrumental variable estimation of error-components models. *Journal of econometrics*, 68(1):29–51.

Basu, P., Cai, T. T., Das, K., and Sun, W. (2018). Weighted false discovery rate control in large-scale multiple testing. *Journal of the American Statistical Association*, 113(523):1172–1183.

Blundell, R. and Bond, S. (1998). Initial conditions and moment restrictions in dynamic panel data models. *Journal of econometrics*, 87(1):115–143.

References II

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.

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

Greene, W. (2005). Fixed and random effects in stochastic frontier models. *Journal of productivity analysis*, 23:7–32.

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.

References III

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 Mizera, I. (2014). Convex optimization, shape constraints, compound decisions, and empirical bayes rules. *Journal of the American Statistical Association*, 109(506):674–685.

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.

Roodman, D. (2007). A short note on the theme of too many instruments. *Center for Global Development Working Paper*, 125(10.2139).

Conditional Input Demand Function

In standard microeconomics, the profit maximization problem is

$$\max_{\vec{y}} \sum k_i y_i - \sum p_i x_i \quad \text{subject to} \quad f_i(x_1, x_2, \ldots, x_n) = y_i$$

where p_i is the price of input i and f is the cost function.
The cost minimization problem is thus

$$\min_{\vec{x}} \sum p_i x_i \quad \text{subject to} \quad f_i(x_1, x_2, \ldots, x_n) = y_i$$

Thus, the factor demand function/correspondence is

$$x_i = x_i(p_1, p_2, \ldots, p_n, y_1, y_2, \ldots, y_m)$$

Back

Input demand function vs Production function

- We can remain agnostic as to the nature of the appropriate formula for the aggregation of outputs and use as many different products as desired.
- When input prices have low variability. Conditional factor demand can be estimated without information on input prices. Even if we add prices, due a lack of variability, the price parameters will be poorly estimated.
- From $x_i = x_i(p_1, p_2, \ldots, p_n, y_1, y_2, \ldots, y_m)$, we do not need to observe a complete list of inputs. But we do need to observe all input prices (can be ignored if almost no variability) and all outputs. While in the production function, it is the other way around (need to observe all inputs). Since, in our case, output is more *observable* than input (because capital is not easily observed), this approach is preferred.

First glance

Dependent Variable:	Nurses	
	Pool	Pool exclude
Model:	(1)	(2)
Variables		
Constant	1.39***	1.36***
	(0.025)	(0.024)
STAC inpatient	0.279***	0.293***
	(0.005)	(0.005)
STAC outpatient	0.050***	0.034***
	(0.003)	(0.003)
Medical sessions	0.061***	0.063***
	(0.002)	(0.002)
External consultations	0.057***	0.040***
	(0.002)	(0.001)
Emergency	0.016***	0.023***
	(0.001)	(0.001)
Long-term & follow-up	0.076***	0.072***
	(0.002)	(0.002)
Home care	0.016***	0.028***
	(0.003)	(0.003)
Psychiatric care	0.073***	0.075***
	(0.003)	(0.004)
Fit statistics		
Observations	13,402	12,279
R^2	0.821	0.819

Heteroskedasticity-robust standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Dependent Variable:	Nurses	
	Dummy	Dummy exclude
Model:	(1)	(2)
Variables		
Constant	1.58***	1.50***
	(0.027)	(0.028)
STAC inpatient	0.276***	0.290***
	(0.005)	(0.005)
STAC outpatient	0.057***	0.048***
	(0.004)	(0.004)
Medical sessions	0.063***	0.068***
	(0.002)	(0.002)
External consultations	0.027***	0.028***
	(0.002)	(0.002)
Emergency	0.021***	0.018***
	(0.001)	(0.001)
Long-term & follow-up	0.069***	0.067***
	(0.002)	(0.002)
Home care	0.026***	0.025***
	(0.003)	(0.003)
Psychiatric care	0.063***	0.071***
	(0.003)	(0.004)
Private Forprofit	-0.270***	-0.245***
	(0.027)	(0.027)
Private Nonprofit	-0.180***	-0.160***
	(0.021)	(0.022)
Teaching	0.707***	
	(0.021)	
Fit statistics		
Observations	13,402	12,279
R^2	0.838	0.821

Heteroskedasticity-robust standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Second glance

Dependent Variable:	Nurses			
	Teaching	Public	Forprofit	Nonprofit
Model:	(1)	(2)	(3)	(4)
Variables				
Constant	3.28***	1.40***	1.41***	1.00***
	(0.123)	(0.099)	(0.041)	(0.059)
STAC inpatient	0.108***	0.328***	0.261***	0.343***
	(0.016)	(0.018)	(0.007)	(0.012)
STAC outpatient	0.131***	0.078***	0.048***	0.046***
	(0.013)	(0.005)	(0.005)	(0.010)
Medical sessions	0.058***	0.049***	0.075***	0.093***
	(0.008)	(0.003)	(0.002)	(0.006)
External consultations	0.018**	0.025***	-0.003	0.002
	(0.009)	(0.004)	(0.006)	(0.005)
Emergency	0.049***	-0.008**	0.034***	0.024***
	(0.004)	(0.003)	(0.002)	(0.004)
Long-term & follow-up	0.058***	0.051***	0.057***	0.117***
	(0.005)	(0.003)	(0.003)	(0.007)
Home care	0.020**	0.029***	0.049***	-0.012
	(0.010)	(0.003)	(0.007)	(0.009)
Psychiatric care	0.029***	0.071***	0.076***	0.049***
	(0.005)	(0.004)	(0.007)	(0.017)
Fit statistics				
Observations	1,123	5,260	4,415	2,604
R^2	0.780	0.863	0.742	0.754

Heteroskedasticity-robust standard-errors in parentheses
*Signif. Codes: ***: 0.01, **: 0.05, *: 0.1*

Panel data Estimator

- Strict exogeneity: Within Group/First Diffrence

$$E[\epsilon_{it}|x_{i1},\ldots,x_{iT},\theta_i] = 0$$

- Relaxed: First Difference GMM (Arellano and Bond, 1991), System GMM (Arellano and Bover, 1995; Blundell and Bond, 1998).

$$E[\epsilon_{it}|x_{i1},\ldots,x_{it-p},\theta_i] = 0$$

Issues: Weak instruments (**?**) and the proliferation of instruments (Roodman, 2007).

$$\mathbb{E}[\Delta x_{i,t-1}(y_{it} - \beta x_{it})] \quad \text{if} \quad \mathbb{E}\left[\Delta x_{i,t-1}(\theta_i + \varepsilon_{i,t})\right] = 0$$

NPMLE Computation Methods

The primal problem:

$$\min_{f=dG}\left\{-\sum_i \log g(y_i)\middle| g(y_i) = T(f),\ K(f) = 1,\ \forall i\right\}$$

where $T(f) = \int p(y_i|\theta) f d\theta$ and $K(f) = \int f d\theta$.
Discretize the support:

$$\min_{f=dG}\left\{-\sum_i \log g(y_i)\middle| g = Af,\ 1^T f = 1\right\}$$

where $A_{ij} = p(y_i|\theta_j)$ and $f = (f(\theta_1), f(\theta_2), \ldots, f(\theta_m))$.
The dual problem:

$$\max_{\lambda,\mu}\left\{\sum_i \log \lambda_1(i)\middle| A^T\lambda_1 < \lambda_2 1,\ (\lambda_1 > 0)\right\}$$

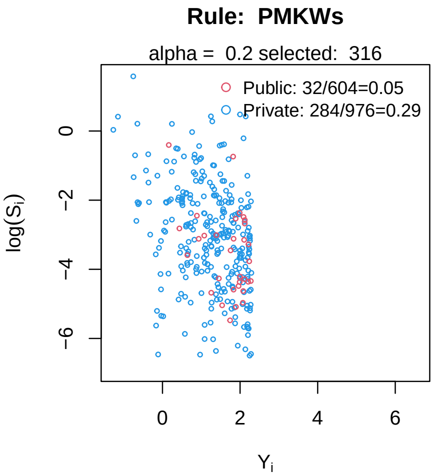
Normality assumption on ε_{it}

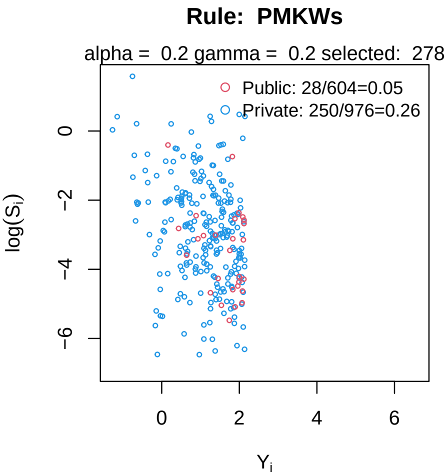
Estimate the fixed effect θ_i by

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

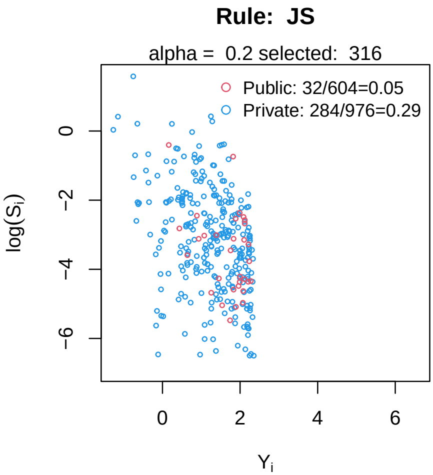
When T is relatively small (or even fixed), can't use central limit theorem to claim that $\hat{\theta}_i \xrightarrow{d} \mathcal{N}(\theta_i, \frac{\sigma_i^2}{T})$. $\longrightarrow$ Assume that $\varepsilon_{it} \sim \mathcal{N}(0, \sigma_i^2)$. Back (main) Back (end)

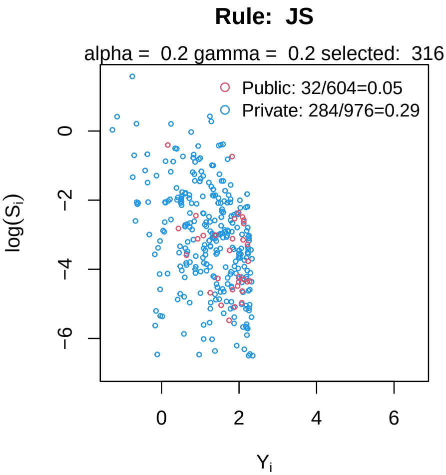
$G(\theta, \sigma)$: Posterior Mean



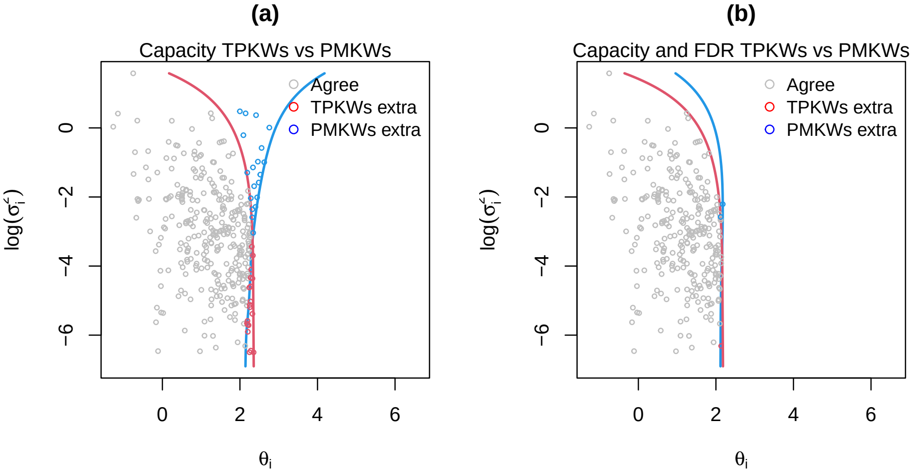


$G(\theta, \sigma)$: James-Stein Shrinkage





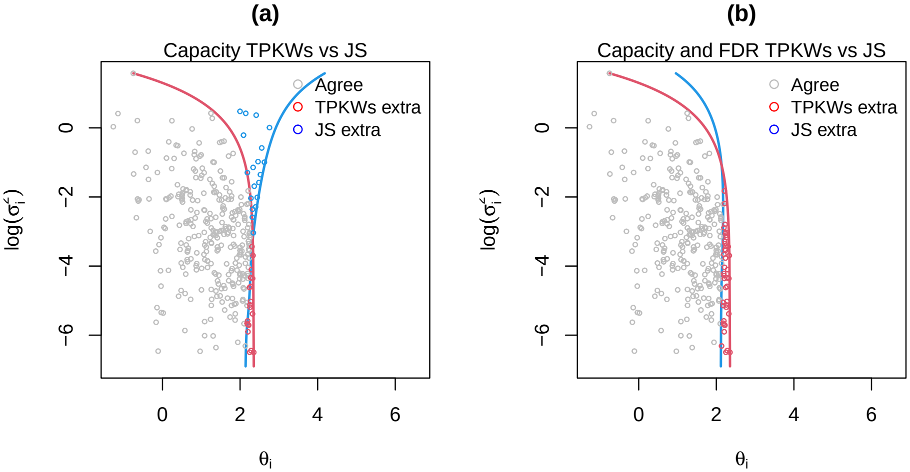
TP vs PM



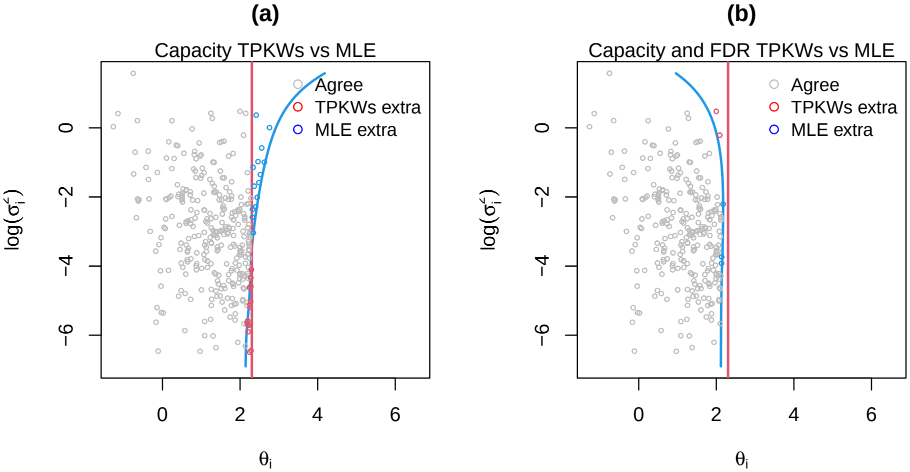
(a)
Capacity TPKWs vs PMKWs
Agree
TPKWs extra
PMKWs extra
(b)
Capacity and FDR TPKWs vs PMKWs
Agree
TPKWs extra
PMKWs extra
$\log(\sigma_i^2)$
θ_i

Back

TP vs JS


(a)
Capacity TPKWs vs JS
(b)
Capacity and FDR TPKWs vs JS
Agree
TPKWs extra
JS extra
$\log(\sigma_i^2)$
θ_i
0
−2
−4
−6
0
2
4
6

TP vs MLE



(a)
Capacity TPKWs vs MLE
(b)
Capacity and FDR TPKWs vs MLE
Agree
TPKWs extra
MLE extra
$\log(\sigma_i^2)$
θ_i
0
−2
−4
−6
0
2
4
6