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

Fu Zixuan

Supervised by Prof.Thierry Magnac

June 21, 2024

Questions

- Out of the top 20% hospitals in France[1], 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?

[1] in terms of labor (nurses) employment efficiency

Roadmap

1. Estimate the efficiency.
   - Y: Labor input (number of full time equivalent nurses).
   - X: Hospital output (e.g., inpatient/outpatient stays, medical sessions).
2. **Select the 20% most efficient hospitals.**

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), the *conditional input demand function*

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

  The smaller the θ_i, the less input is needed to produce the same amount of output, the more efficient the hospital is.

Literature: Invidious decision

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

Hospital Types

Year	Teaching	Normal Public	Private FP	Private NP	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

Output

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

The hospitals differ not only in efficiency but also in the mix of services they provide.

[2] Short term acute care

Fixed effect estimation

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:		log(ETP_INF)	
	Within Group	First Difference	*System GMM*
Model:	(1)	(2)	(3)
Variables			
log(SEJHC_MCO)	0.10^{***}	0.07^{***}	0.70^{***}
	(0.00)	(0.01)	(0.05)
log(SEJHP_MCO)	0.02^{***}	0.01^{***}	-0.05
	(0.00)	(0.00)	(0.04)
...	...	...	...
log(SEANCES_MED)	0.02^{***}	0.02^{***}	0.07^{***}
	(0.00)	(0.00)	(0.03)
log(SEJ_PSY)	0.00	0.00	0.07^{***}
	(0.00)	(0.00)	(0.01)
Fit statistics			
Observations	15335	13502	11536
n	1833	1833	1833
T	9	9	9

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

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):

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

The Selection Task

- Select the bottom 20% [3] out of the population of θ_i, those i whose $\theta_i < G^{-1}(0.2)$
- Control the overall false discovery rate at 20%,

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

[3] The most efficient 20%.

Problem Formulation

The **loss** function is $(h_i = 1\{\theta_i < \theta_\alpha = G^-1(\alpha)\})$

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

Therefore, the problem is to find δ such that

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

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

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)

Prior Distribution G

Observe[4]

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

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

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

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

▸ Appendix

[4] $Y_{it} = \theta_i + \varepsilon_{it} + x_{it}(\beta - \hat{\beta})$

Tail probability

Given the two sufficient statistics, the posterior tail probability is

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

We want to find a cutoff λ such that both constraints are satisfied [5]:

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

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

Estimate G

Following (Koenker and Mizera, 2014; Andersen and Andersen, 2010) The primal problem:

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

where $T(f_i) = \int p(y|\alpha) f_i d\alpha$ and $K(f_i) = \int f_i d\alpha$.
Discretize the support:

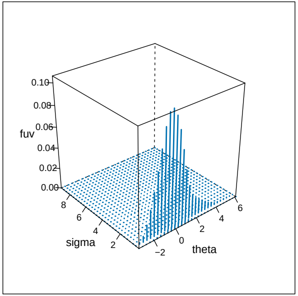
$$\min_{f=dG} \left\{ -\sum_i \log g(y_i) \middle| g = Af,\ 1^T f = 1 \right\}$$

where $A_{ij} = p(y_i|\alpha_j)$ and $f = (f(\alpha_1), f(\alpha_2), \ldots, f(\alpha_m))$.
The dual problem:

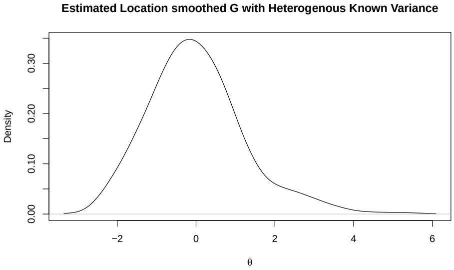
$$\max_{\lambda,\mu} \left\{ \sum_i \log \lambda_1(i) \middle| A^T \lambda_1 < \lambda_2 1,\ (\lambda_1 > 0) \right\}$$

The estimated $\hat{G}$

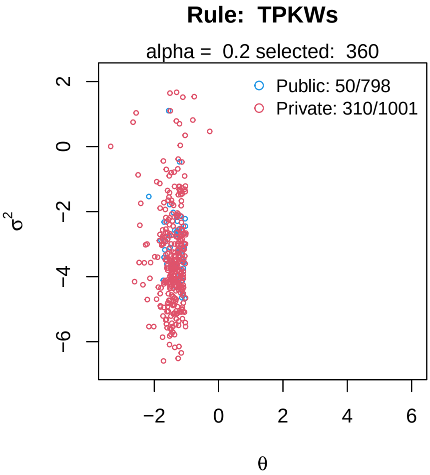
Case 1: σ_i unknown, only S_i observed.

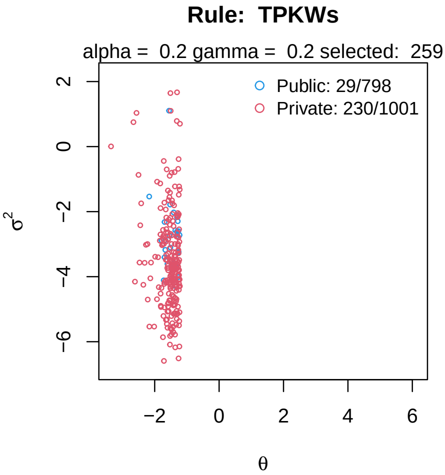


Case 2: σ_i known.

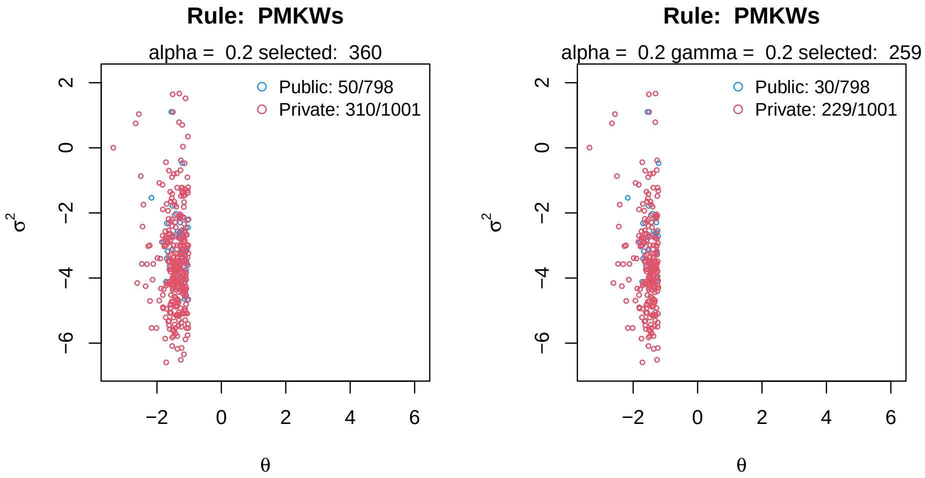


Unknown σ_i: Posterior Tail probability

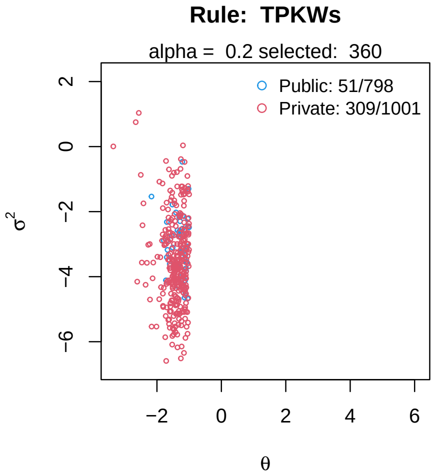


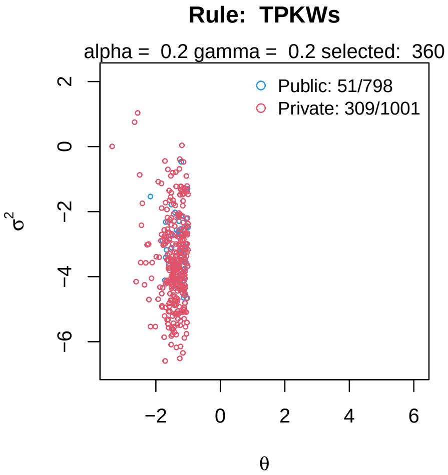


Unknown σ_i: Posterior Mean

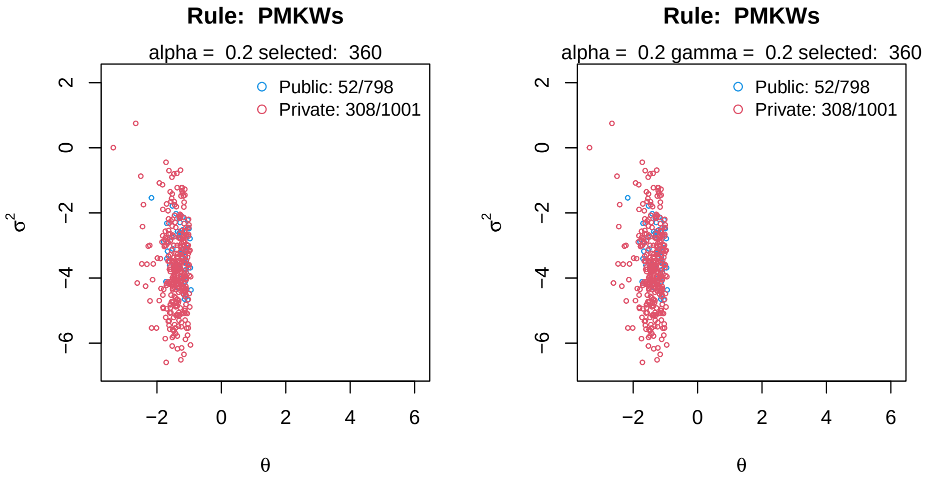


Known σ_i: Posterior Tail probability

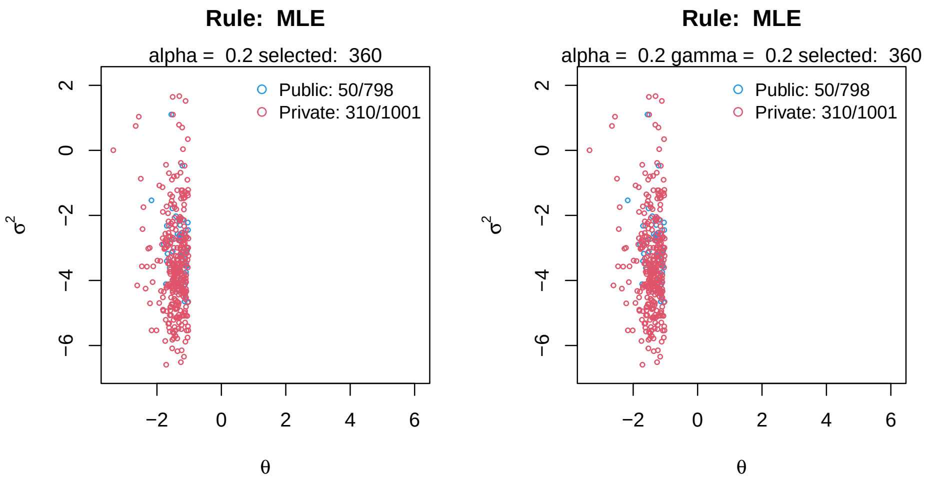




Known σ_i: Posterior Mean



"Face value"


Rule: MLE
alpha = 0.2 selected: 360
Public: 50/798
Private: 310/1001
σ2
θ
Rule: MLE
alpha = 0.2 gamma = 0.2 selected: 360
Public: 50/798
Private: 310/1001
σ2
θ

Conclusion

- Difference in whether to assume known σ_i.
- Control for the False Discovery Rate.
- Private (FP and NP) hospitals are indeed more "efficient".

Limitation

- Interpretation of the θ_i.
- Specification, endogeneity *etc.*
- Normality assumption on ε_{it}. Next

References I

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.

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

References II

Blundell, R. and Bond, S. (1998b). Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics*, page 115–143.

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.

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.

References III

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.

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

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

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

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	
	OLS	Lagged IV
Model:	(1)	(2)
Variables		
Constant	1.59***	1.58***
	(0.067)	(0.069)
STAC inpatient	0.278***	0.277***
	(0.012)	(0.013)
...	...	...
Private Forprofit	-0.303***	-0.280***
	(0.061)	(0.065)
Private Nonprofit	-0.215***	-0.188***
	(0.056)	(0.055)
Teaching	0.717***	0.709***
	(0.056)	(0.056)
Fit statistics		
Observations	15,335	13,402
R^2	0.835	0.837

Clustered (FI) 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, 1998a).

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

Issues: Weak instruments (Blundell and Bond, 1998b) and the proliferation of instruments (Roodman, 2007).

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

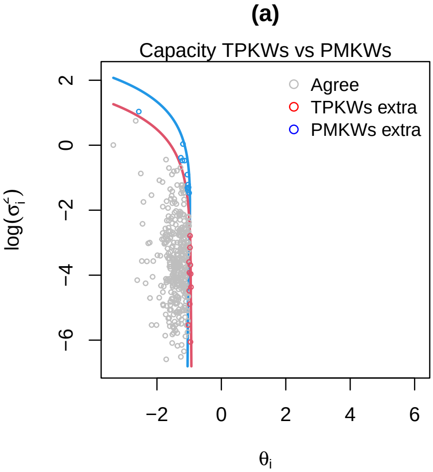
Normality assumption on ε_{it}

Estimate the fixed effect θ_i by

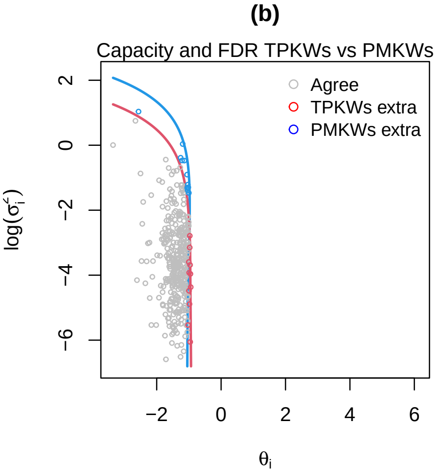
$$\begin{aligned}\hat{\theta}_i =& \frac{1}{T}\sum(\theta_i + \varepsilon_{it} + x_{it}(\beta - \hat{\beta})) \\ \xrightarrow{N\to\infty}& \theta_i + \frac{1}{T}\sum_t \varepsilon_{it}\end{aligned}$$

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)

TP vs PM

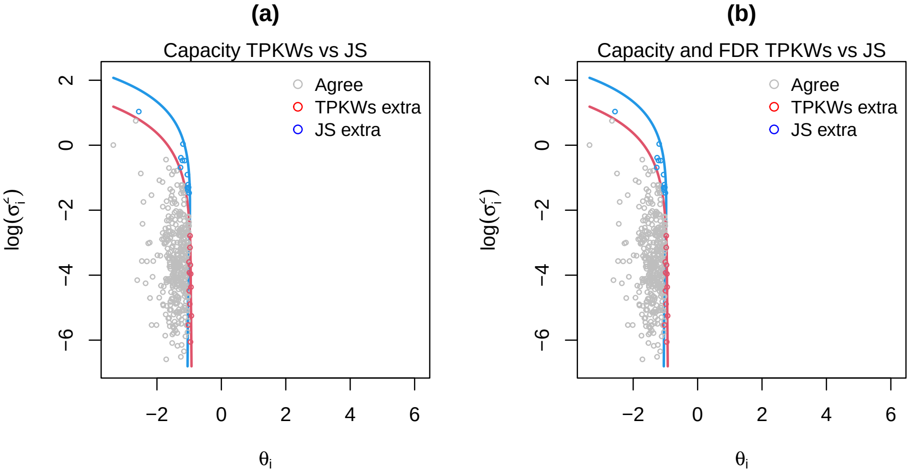


(a)
Capacity TPKWs vs PMKWs
Agree
TPKWs extra
PMKWs extra
log(σ²ᵢ)
2
0
−2
−4
−6
−2
0
2
4
6
θᵢ

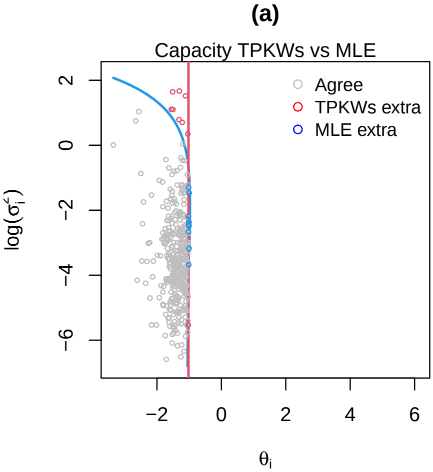


(b)
Capacity and FDR TPKWs vs PMKWs
Agree
TPKWs extra
PMKWs extra
log(σ²ᵢ)
2
0
−2
−4
−6
−2
0
2
4
6
θᵢ

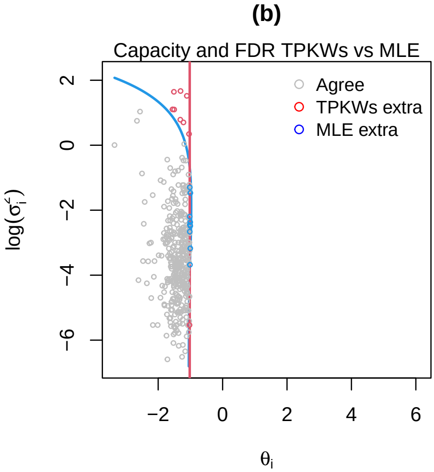
TP vs JS


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

TP vs MLE



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



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