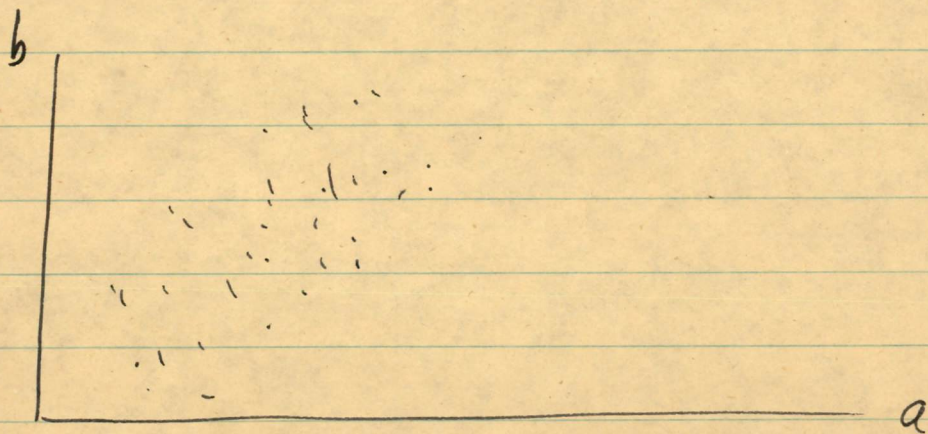


Regression Fallacy & Stefaniak data

Still, if Marshall's index has bias, Census may feelably
show regressive assmt. If rank by sales value &
relate to assmt, regn. fallacy will show regressive assmt.

Which way is his index biased?

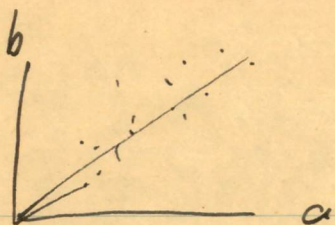
He ~~rank by~~ doesn't rank, but the choice of
using a/b or b/a probably has same effect.



* Let $\sum a = \sum b$, so $\sum a / \sum b = 1$

and $\sum b / \sum a = 1$

There is no obvious reason for bias. I think
Marshall's is probably much better than other methods.



$$\frac{\frac{\sum(a_i/b_i)}{n}}{\frac{\sum a_i}{\sum b_i}} \stackrel{?}{=} \frac{\frac{\sum b}{\sum a}}{\sum(\frac{b_i}{a_i})/n}$$

$$\frac{\sum(\frac{a}{b})}{n} \cdot \frac{\sum(\frac{b}{a})}{n} \stackrel{?}{=} \frac{\sum a \cdot \sum b}{\sum a \cdot \sum b} = 1$$

$$\text{Let } \frac{a_1}{b_1} = \frac{1}{2} \quad \text{and} \quad \frac{a_2}{b_2} = \frac{1}{3}$$

$$\frac{\frac{1}{2} + \frac{1}{3}}{2} \cdot \frac{2 + 3}{2} = \frac{\frac{5}{6} \cdot 5}{4} = \frac{25}{24}$$

So the index is not symmetrical. This means probably

it does not avoid regression bias.

But it may be much less biased than some other common methods: least squares, e.g.

$$\frac{a}{b}: \quad \frac{1}{1}, \quad \frac{2}{3}, \quad \frac{4}{7}$$

$$\begin{array}{r} 63 \\ 12 \\ \hline 126 \\ 63 \\ \hline 756 \end{array}$$

$$\begin{array}{r} 47 \\ 17 \\ \hline 329 \\ 47 \\ \hline 799 \end{array}$$

$$\frac{21 + 14 + 12}{3 \cdot 21} \stackrel{?}{=} \frac{3}{1 + \frac{3}{2} + \frac{7}{4}} = \frac{3 \cdot 4}{4 + 6 + 7}$$

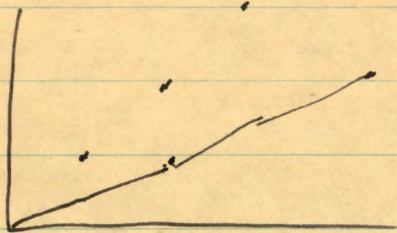
$$\frac{47}{63} \stackrel{?}{=} \frac{12}{17}$$

$$756 \neq 799$$

$$\frac{\frac{a_1}{b_1} + \frac{a_2}{b_2}}{n} \stackrel{?}{=} \frac{n}{\frac{b_1}{a_1} + \frac{b_2}{a_2}}$$

$$\cancel{1} + \frac{b_1 a_2}{a_1 b_2} + \frac{b_2 a_1}{a_2 b_1} + \cancel{1} \stackrel{?}{=} \cancel{4} \quad 2$$

$$a_2^2 b_1^2 + b_2^2 a_1^2 \stackrel{?}{=} 2 a_1 a_2 b_1 b_2$$



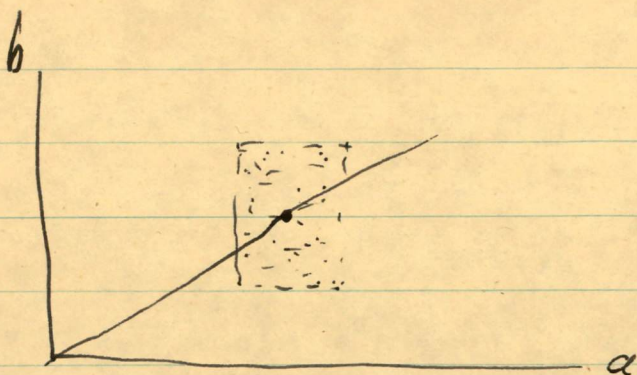
$$\frac{\frac{1}{2} + \frac{1}{2}}{2} \stackrel{?}{=} \frac{2}{2+2} \quad \text{yes}$$

$$\frac{\frac{a}{b} + \frac{c}{d}}{2} \stackrel{?}{=} \frac{2}{\frac{b}{a} + \frac{d}{c}}$$

$$1 + \frac{bc}{da} + \frac{da}{bc} + 1 \stackrel{?}{=} 4$$

$$\text{If } \frac{a}{b} = \frac{c}{d} = 1, \text{ then } a+c = b+d$$

So ?

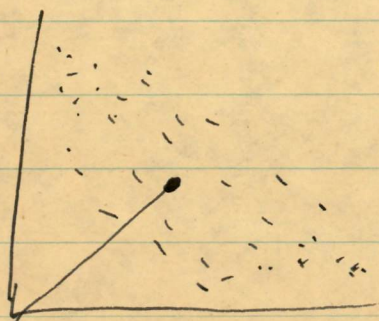


$$\sum a = \sum b$$

$$\frac{\sum \frac{a}{b}}{n} \approx 1$$

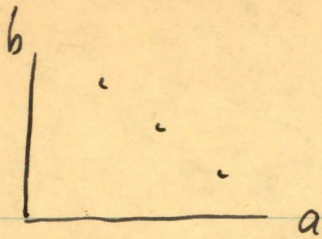
So Manvel's index = 1, meaning relationship not affected by size.

Regression lines on other hand wd. be vert. & horiz.
So Manvel's index is much freer of bias.



Manvel's index not affected by whether relationship is + or -.

This is bad! A neg. relationship is the most extreme asmt. bias imaginable.



$$\frac{a_i}{b_i} : \frac{1}{3}, \frac{2}{2}, \frac{3}{1}$$

$$\sum a = 6$$

$$\sum b = 6$$

$$\sum \frac{a}{b} = \frac{1 + 3 + 9}{3} = \frac{13}{3}$$

$$\frac{\frac{\sum (\frac{a}{b})}{n}}{\frac{\sum a}{\sum b}} = \frac{\frac{13}{3}}{1} = \frac{13}{3}$$

$$\frac{a_i}{b_i} : \frac{1}{1}, \frac{2}{2}, \frac{3}{3}$$

$$\sum \frac{a}{b} / n = 1$$

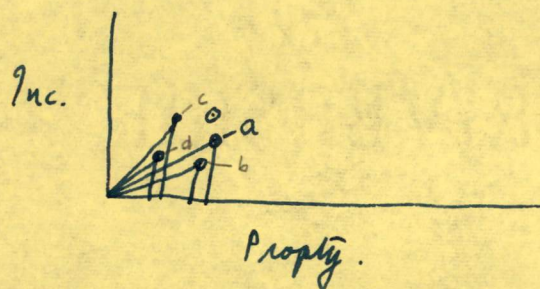
$$r_{index} = 1$$

How to escape regression fallacy? (partly)

Use ratios, (which are indep. of absolute size), as indep. vble. by which ordering is made.

$\frac{\text{Labor}}{\text{acre}}$, e.g. Then see trend of acres as labor ÷ acres rises.

$\frac{\text{Inc.}}{\text{Propty.}}$



3 different orderings possible:

- 1) By Inc. ~~acdb~~ bdac
- 2) By Propty. ~~abcd~~ dcba
- 3) By $\frac{\text{Inc.}}{\text{Propty.}}$ (or the reverse) ~~cdab~~ badc

Doesn't entirely escape reg. fall., but partly. Then order by deciles and find % of inc. & propty. in each — tells you more than a linear regn. coeff. etc., but get those, too. Order by deciles, of. Lor. curves for propty. & inc.

Make this from. 1. Order by $\frac{\text{Inc.}}{\text{Propty.}}$, get Lor. data. Top 10% will

have no propty., and probly < 10% of Inc.

2. Order by $\frac{\text{Propty.}}{\text{Inc.}}$, get Lor. data. Top 10% will

have some inc., probably > 10%, and much more than 10% of the propty.

Should be quite striking.

Try to get some such into the study. Show that land not a regression tax base.

Do it w. labor/acre, e.g., for indy., & plant/acre, & deprec. plant / Σ acres.

See my
Indy.
study
for data.

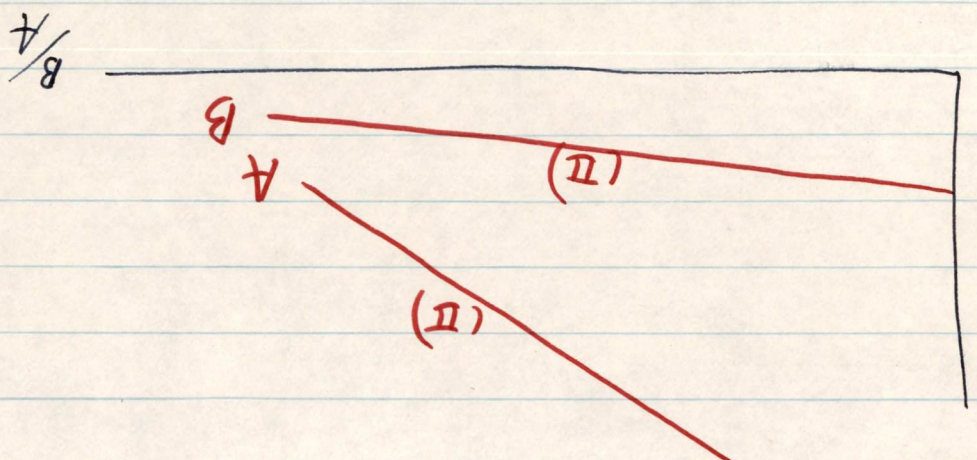
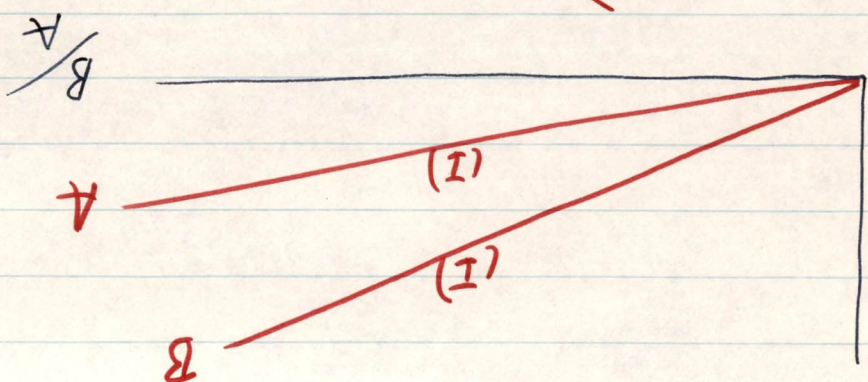
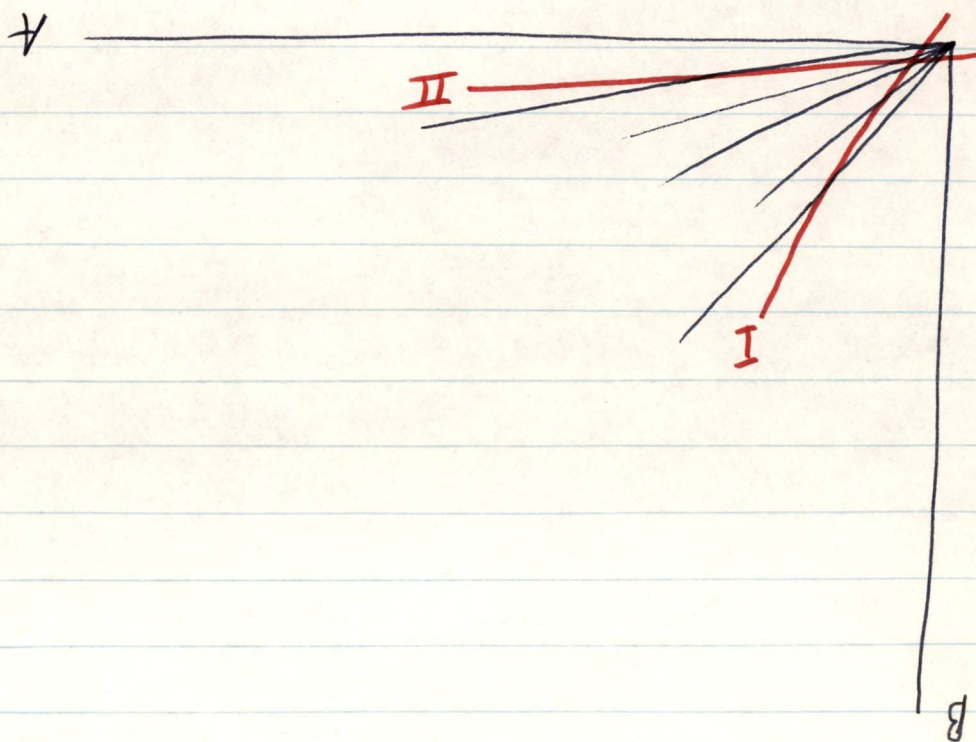
Reqm. fallacy.

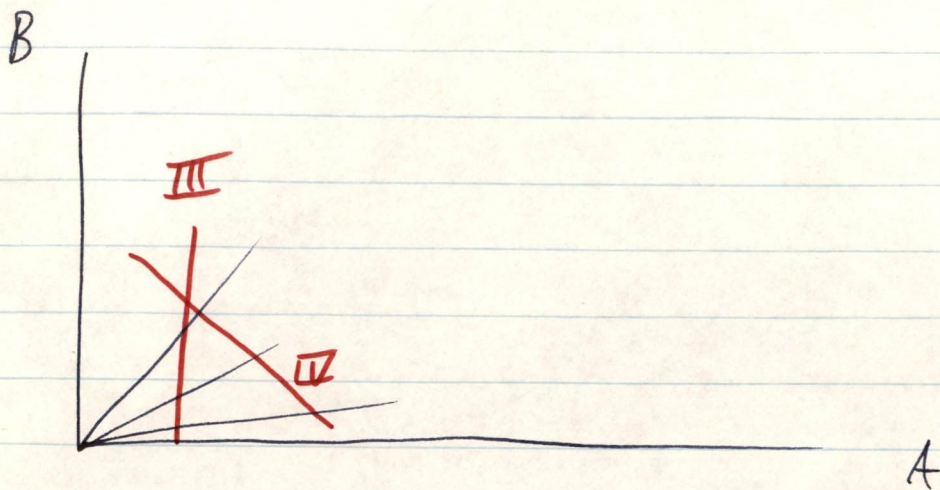
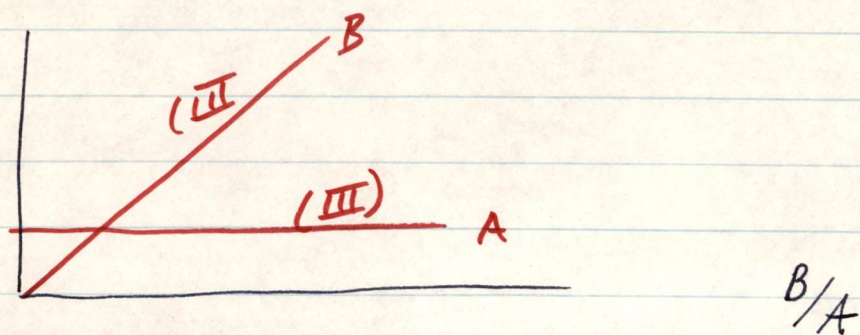
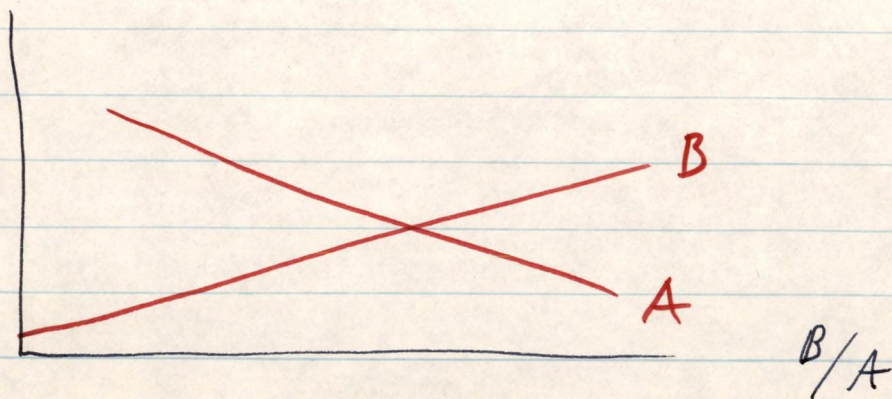
2

Try for some resid. data on bldg value
~~l.v.~~

✓ Another escape from reqm. fall. is when 2 vbles can be added, e.g. l.v. & bldg.v., or plant area & land area.

Then take reqm. of $(\frac{a}{b})$ or $(a + b)$, or v.v.





In only one case, Case II, will B decline as B/A increases. ~~And~~ That is ^{one} ~~the~~ case I seek.

Case I is the same, though. In (I), A declines as A/B rises.

I should always get one or the other.

Coeff. of dispersion

1. Md. AR

a. $Las/Laav$, rank by. (Define as AR_i)

b. Report Md. AR

2. Define $|AR_i - MdAR| \equiv$ ~~dispersion~~ deviation, d .

$\sum d/n \equiv$ ave. deviation,

$$\frac{\sum d/n}{MdAR} \equiv \text{coeff. of dispersion} = \frac{\sum d}{n \cdot MdAR}$$

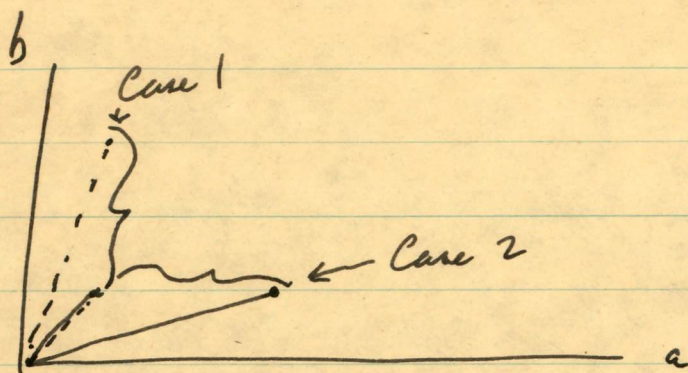
"Price-related differential of assessment ratios" (Size-related differential of ~~any~~ ratios!)

1. $\sum_0^n AR_i / n$ (unweighted sum of ratios)

2. $\frac{\sum Las}{\sum Laav}$

(Related to Pearson's r ?)

$$3. \frac{\sum_0^n AR_i / n}{\sum Las / \sum Laav} = \frac{\sum Laav \cdot \sum_0^n AR_i}{n \cdot \sum Las}$$



Case 1. ~~$\frac{a_1}{b_1}$~~ $\frac{b_1}{a_1} = \frac{1}{1}$

$$\frac{b_2}{a_2} = \frac{3}{1}$$

$$\frac{\frac{b_1}{a_1} + \frac{b_2}{a_2}}{n \cdot \frac{\sum b}{\sum a}} = \frac{1 + 3}{2 \cdot \frac{4}{2}} = \frac{4}{4} = 1$$

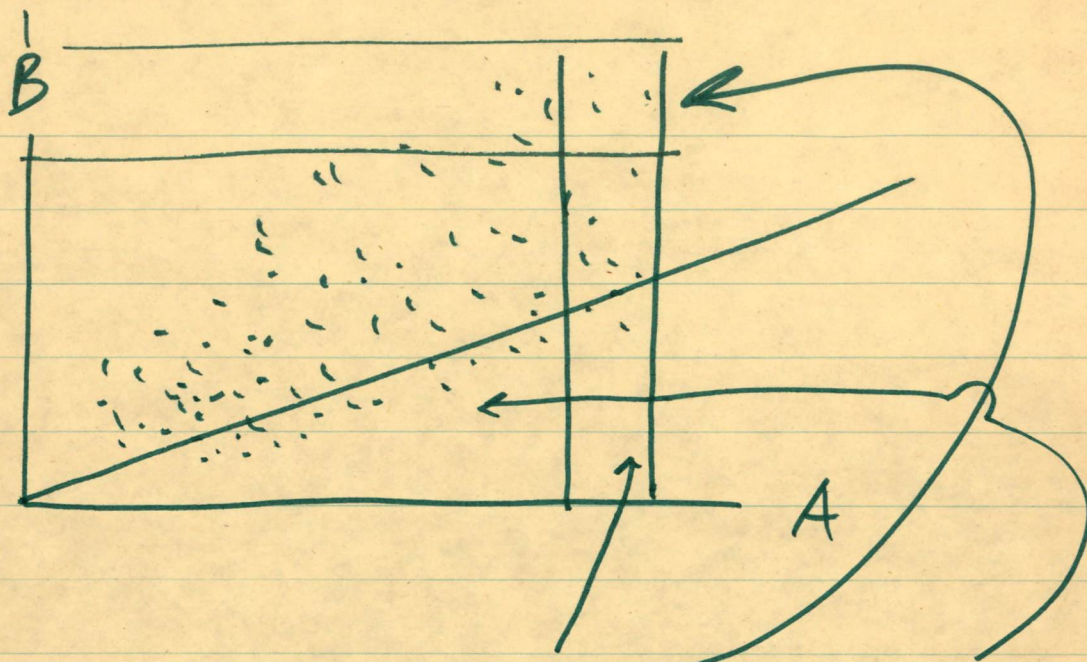
Case 2 $\frac{b_1}{a_1} = \frac{1}{1}$ $\frac{b_2}{a_2} = \frac{1}{3}$

$$\frac{\frac{b_1}{a_1} + \frac{b_2}{a_2}}{n \cdot \frac{\sum b}{\sum a}} = \frac{1 + \frac{1}{3}}{2 \cdot \frac{2}{4}} = \frac{\frac{4}{3}}{1} = \frac{4}{3}$$

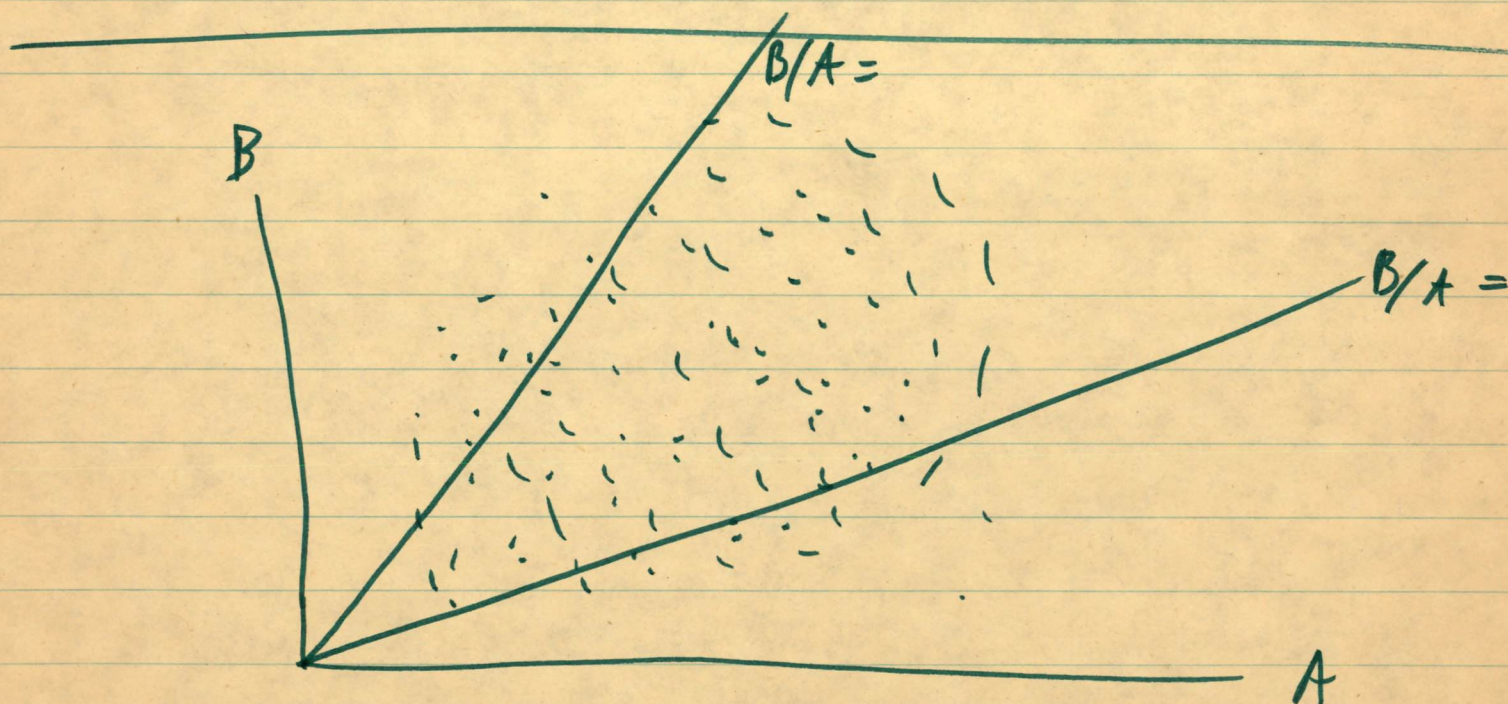
$$\sum_0^n \frac{a}{b}$$

$$\frac{\sum a}{\sum b} \stackrel{?}{=} \frac{a_1}{b_1} \cdot b_1 + \frac{a_2}{b_2} \cdot a_2 = \sum a$$

$$\frac{\left[\frac{a_1}{b_1} \cdot b_1 + \frac{a_2}{b_2} \cdot b_2 + \dots + \frac{a_n}{b_n} \cdot b_n \right] \cancel{\frac{b_i}{a_i}}}{\cancel{\frac{a_i}{b_i}} \cdot \frac{b_1}{a_1} \cdot a_1 + \dots + \frac{b_n}{a_n} \cdot a_n} = \frac{\sum a}{\sum b}$$



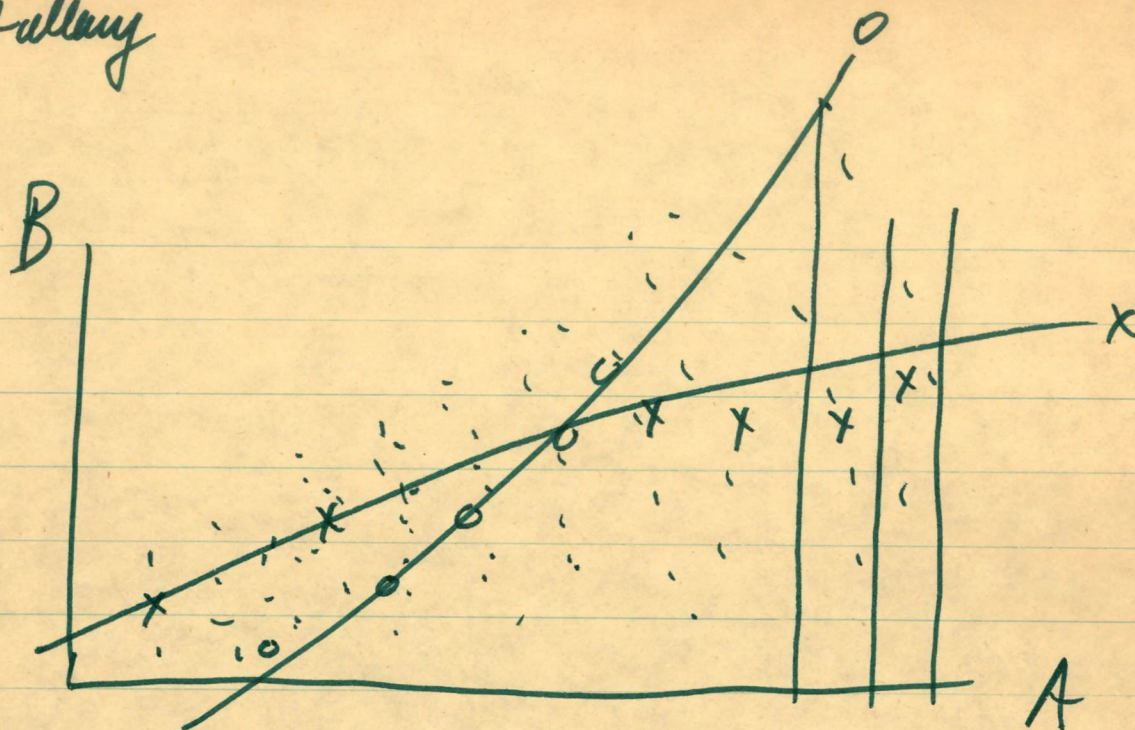
Top deciles ranked by A, B, and A/B



The A's are larger in decile 1; the B's in 10
To neutralize, take $B+A$ in ea. decile.

The superiority of ratio ranker is not in neutrality, then,
but can achieve neutrality elsewhere by adding $B+A$. The
superiority is in that ~~each~~ deciles contain more equal units.

Regm. Fallang

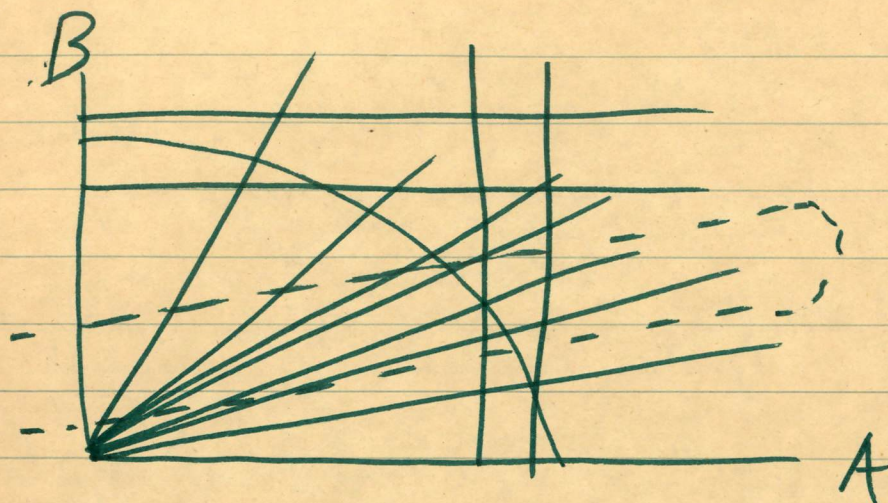


x Rank by A, report on $\bar{B}$
 o " " B', " " $\bar{A}$

If you fit a curve that minimizes vert. deviations,
 that is similar to ranking by A, reports $\frac{1}{n} \sum B$.
 And v.v. for min. horiz. devns.

A neutral msre. wd. be to average the two
 lines.

From my devile data I can simulate this
 graphically.



1. If points are inside dashed area, as B/A rises,
 A falls rapidly
 B falls slowly

1st 2 deciles may be ignored as eccentric?

Perhaps no. But some deciles may be eccentric.

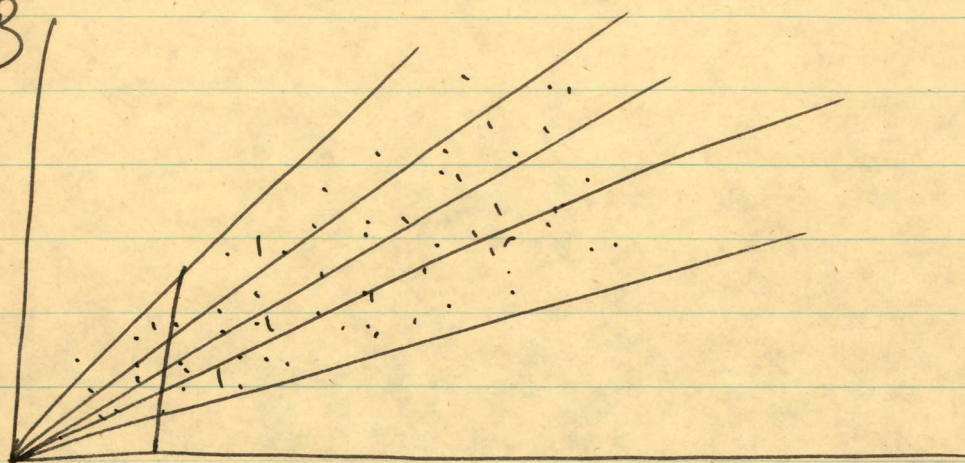
To see major trend, can simply cut in half.

✓ The $\frac{1}{2}$ of larger B/A have only 35% of the B , & 28% of the A . They are smaller.

~~1492~~

TLA

B

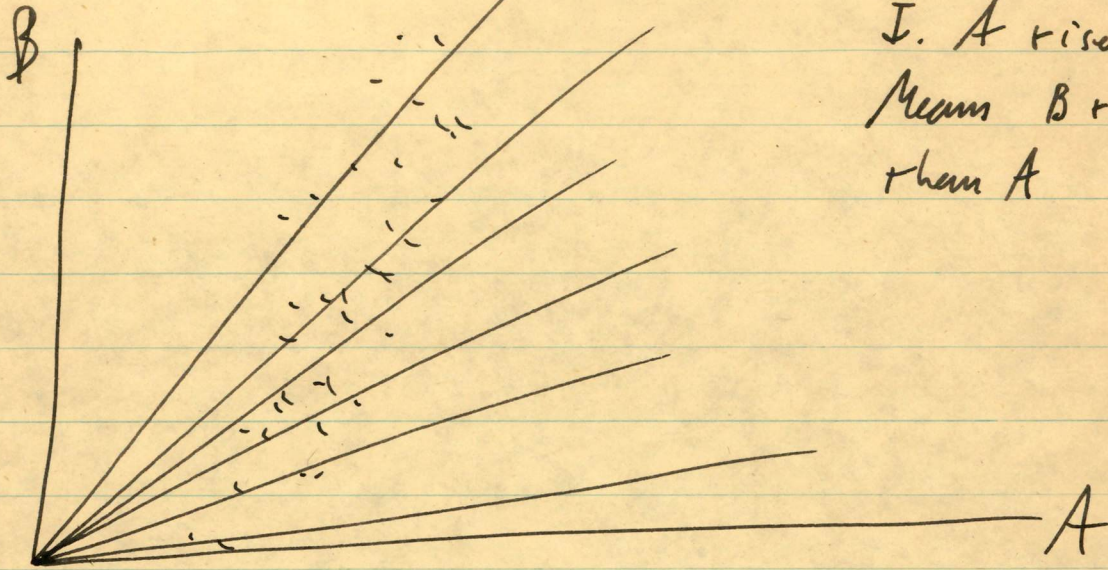


A DPa

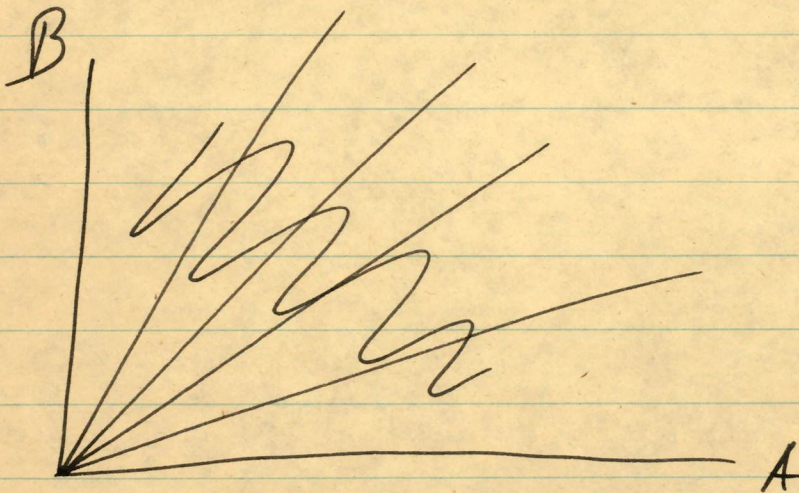
Advantages of ranking by ratios.

1. Deciles are nearer same size, so needn't use d.g. method.
2. May escape regn. fallacy altogether? If rank by B/A , get same deciles backwards.
3. Ranking by A/B is same as by $A/(A+B)$
4. When rank by A/B , report in $A+B$
5. If rank by A/B and B increases or A decreases, have unambiguous result.

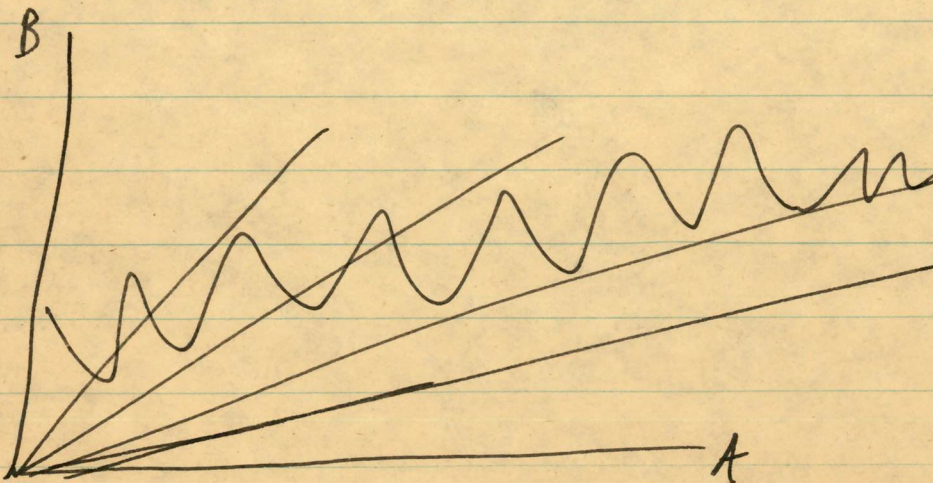
If rank by B/A , can identify relationship, as move to top deciles.



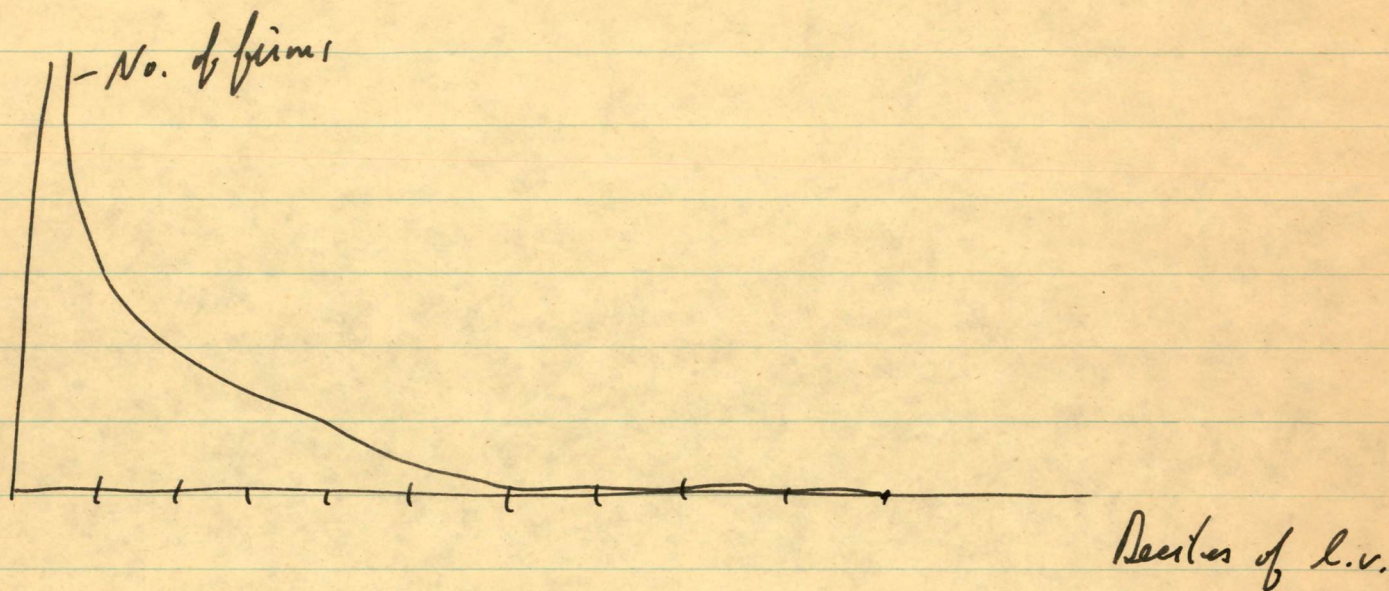
I. A rises & B rises
Means B rises faster
than A



II. B rises, A falls
Means neg. relation.



III.
If you have this,
rank by B/A and
B values fall, A falls
Means B rises slower
than A



New ~~program~~ meaning for decile.

Rank firms by A

Now a decile is the no. of firms that have 10% A.
Call them "deciles" deciguants.

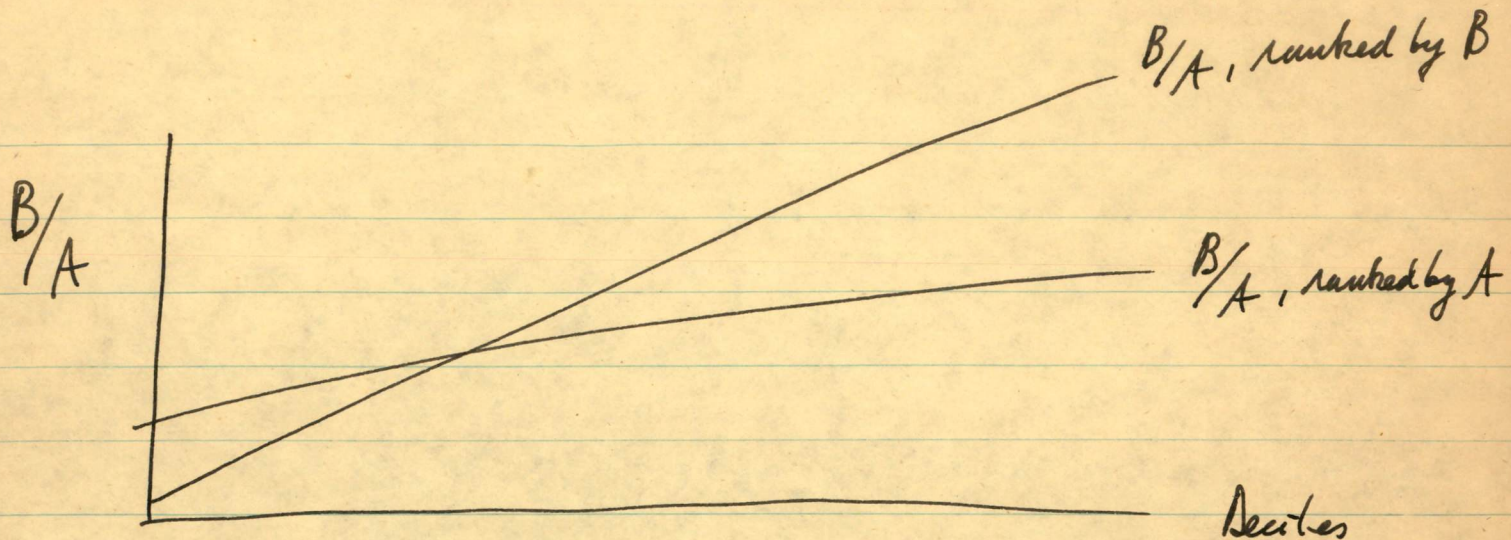
Report by deciguants: the no. of firms, & %

The quant. of B, & %.

Since the dg has 10% of A, the % of B
is a ratio automatically.

E.g. top 9 dgs ranked by ETA have (10% of firms &) 74%
of TVLaav.

Cf. top 9 dgs ~~of~~ ranked by TVLaav, have (20% of firms
and) 55% of ETA. This shows that TVLaav
grows w. ETA more than v.v., a fact not so visible from
ration by deciles. These latter give too much weight to bottom dg.
which is broken into 9 parts.



Here is how regression fallacy will appear when ~~ranked by~~ report by deciles & report on ratio.

$$\frac{\frac{d \Sigma A}{g \Sigma A}}{\frac{d \Sigma B}{g \Sigma B}} = \frac{d \Sigma A}{d \Sigma B} \times \frac{g \Sigma B}{g \Sigma A}$$

Deelte	Ronde 1		ETA/TVLAAV	Ronde 2		ETA/TVLAAV	TVLAAV/ETA
	% ETA	% TVLAAV		% ETA	% TVLAAV		
1	.0	.009	0	.0	.222	0	
2	.0	.027	0	.001	.004	4	
3	.002	.015	.13	.003	.007	2.3	
4	.004	.015	.26	.006	.007	1.2	
5	.006	.024	.25	.009	.010	1.1	
6	.008	.024	.33	.016	.177	11.0	
7	.013	.029	.45	.024	.012	.50	
8	.019	.038	.50	.045	.019	.42	
9	.039	.078	.50	.093	.046	.50	
10	.909	.741	1.23	.803	.496	.62	

$$\begin{array}{r} 29 \overline{) 13.0} \\ 116 \\ \hline 140 \end{array}$$

$$\begin{array}{r} 38 \overline{) 19.0} \\ 0 \end{array}$$

$$\begin{array}{r} 74 \overline{) 91.} \\ 74 \\ \hline 170 \\ 148 \\ \hline 220 \end{array}$$

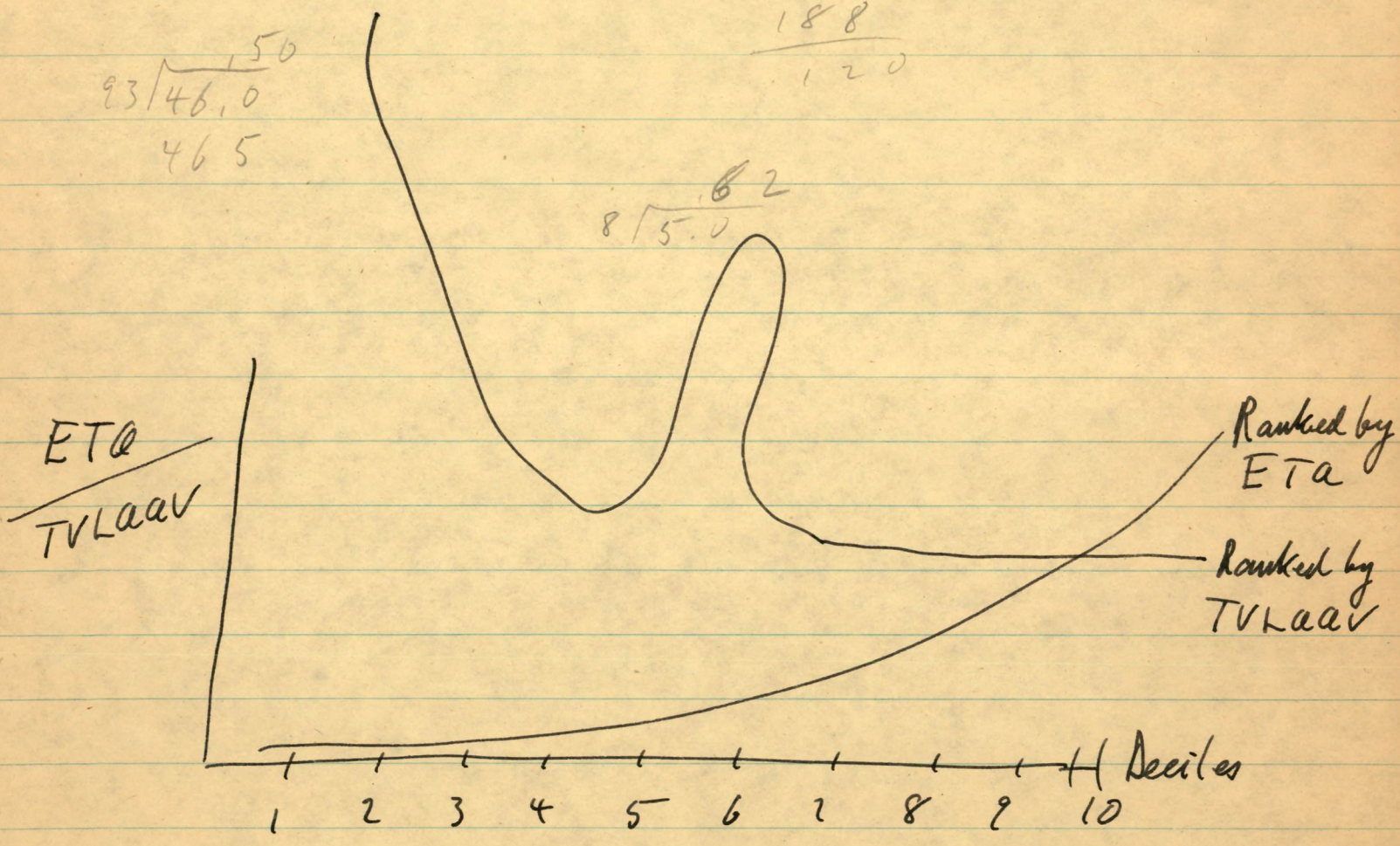
$$\begin{array}{r} 15 \overline{) 2.0} \\ 15 \\ \hline 50 \end{array}$$

$$16 \overline{) 177.}$$

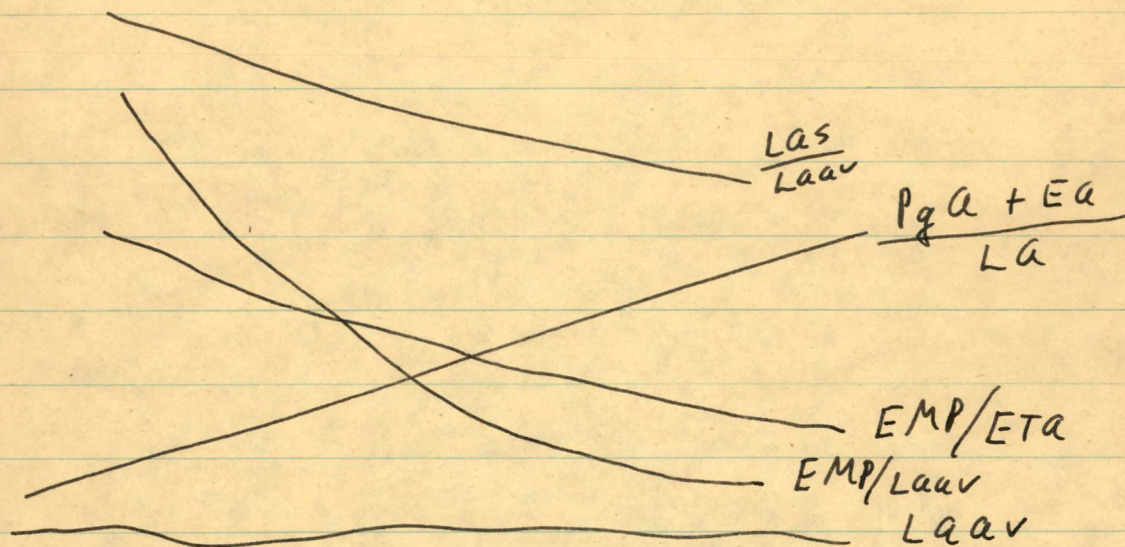
$$\begin{array}{r} 93 \overline{) 46.0} \\ 465 \end{array}$$

$$\begin{array}{r} 49 \overline{) 20.0} \\ 188 \\ \hline 120 \end{array}$$

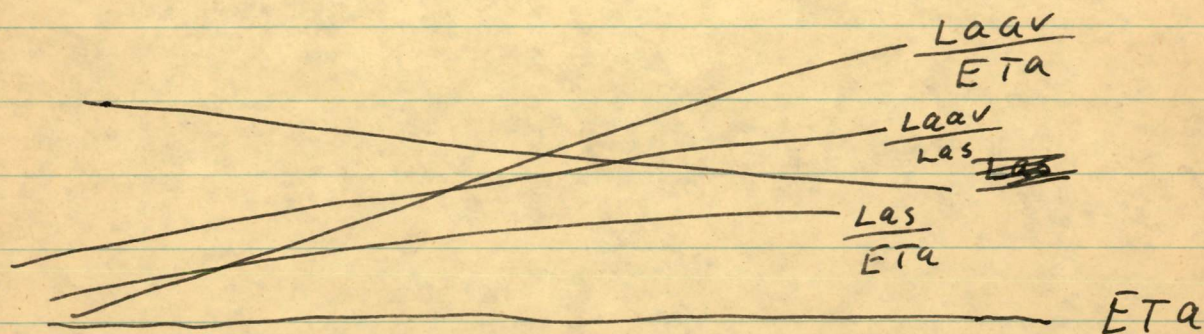
$$8 \overline{) 5.0} \quad .62$$



To present results



So for each indep. variable, I need ratios of deviate results

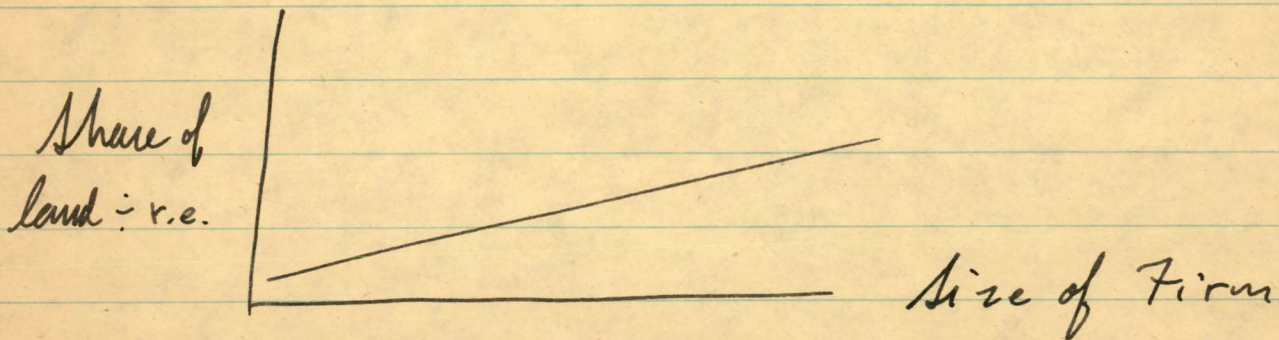


Methodology

Decile method shortcomings.

When most of study vbles. are in top decile, and you change ranking vble, you may reorder firms greatly but all that shows up in results is those that shift ^{across} ~~below~~ 90th percentile.

Percentile analysis of top decile may solve this.



Reporting results as ratios maintaining impact of smaller firms.

