QUANT MACRO: PROBLEM SET 5

Mario Serrano

QUESTION 1. FACTOR INPUT MISSALOCATION

1.1 Firm-specific output, capital and productivity are, respectively, y_i; k_i and z_i. Assume that ln z_i and ln k_i follow a joint normal distribution. Assume that the correlation between ln z_i and ln k_i is zero, the variance of ln z_i is equal to 1.0, the variance of ln k_i is equal to 1.0, and that average s and k is equal to one. Then simulate 10,000,000 observations and plot the joint density in logs and in levels. We are going to assume that these 10,000,000 observations are your complete (or administrative) data that captures the entire population/universe of firms in a given country.

From the statement, we know expectations from the level variables & they follow a normal distribution. With them, we can calculate the corresponding for the los variables by knowing that the exponential of a variable which follows a normal distribution, have a log-normal distribution. The density function is for those variables is:

$$f(x;\mu,\sigma) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-(\ln(x)-\mu)^2/2\sigma^2}$$

Integrating this along all domain for each of our variables, we can obtain obtain an array for the means of the log variables according to:

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) dx$$

For the case of log z, we have to rearrange the function since we have the expectation of s. We know:

$$s = z^{\frac{1}{1-\gamma}}$$

Therefore:

$$E(s) = \frac{1-\gamma}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\ln(s^{1-\gamma})-\mu_z}{\sigma}\right)^2} = 1$$

We obtain:

E(ln k) = -0.5
E(ln z) = -1.25

Since they have no correlation, the variance covariance matrix is:

$$\text{Cov(lnzi, lnki)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Now, we can draw a sample from the pseudo-random variable generated with a multivariate normal distribution for these the level and the log variables. Plotting results, we obtain:

Figure 1: Level variables distribution

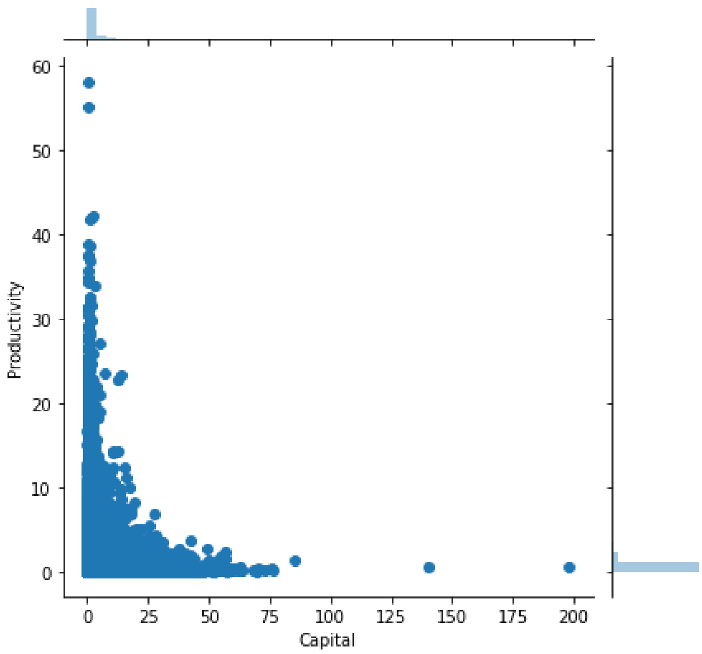
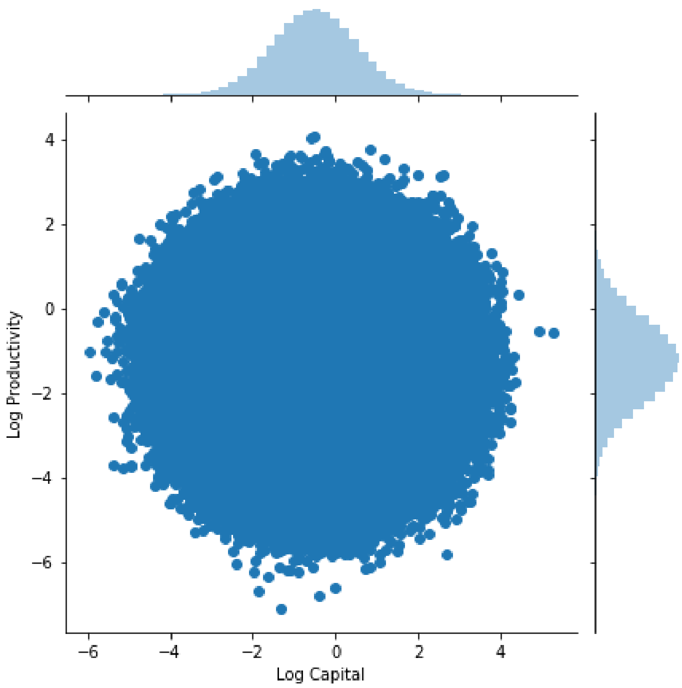


Figure 2: Log variables distribution



3. Solve the following maximization problem:

$$Y^e = \max_{k_i} \sum_i s_i^{1-\gamma} k_i^\gamma$$

subject to K = sum(k_i) where K is a parameter equal to the aggregate capital computed by adding up your complete data. To solve this problem, use s_i from the actual data that you created in item 1.

The typical social planner problem is

$$\max_{\{k_1,..,k_I\}} \sum_{i=1}^{I} s_i k_i^\gamma$$

Subject $\sum_{i=1}^{I} k_i = K$

We can rewrite it taking out the first firm as

$$\max_{\{k_1,\ldots,k_I\}} s_1 k_1^\gamma + \sum_{i \neq 1}^{I} s_i k_i^\gamma$$

And plugging into it the resource constraint

$$\max_{\{k_2,\ldots,k_I\}} s_1 \left(K - \sum_{i \neq 1}^{I} k_i \right)^\gamma + \sum_{i \neq 1}^{I} s_i k_i^\gamma$$

If we compute the FOC, we obtain the optimal condition for maximizing aggregate production

$$s_1 k_1^{\gamma - 1} = s_i k_i^{\gamma - 1} \quad \forall i$$

Since in this case output is $y_i = s_i^{1-\gamma} k_i^\gamma$ we have:

$$s_1^{1-\Upsilon} k_1^{\Upsilon-1} = s_i^{1-\Upsilon} k_i^{\Upsilon-1}$$

Taking k1 to one side and summing across i we obtain:

$$k_1 = s_i/S*K$$

With it, we can compute optimal allocations for each firm.

4. Compare the optimal allocations k_e against the data.

Plotting actual & optimal allocations against productivity we obtain

Figure 3: Actual allocations

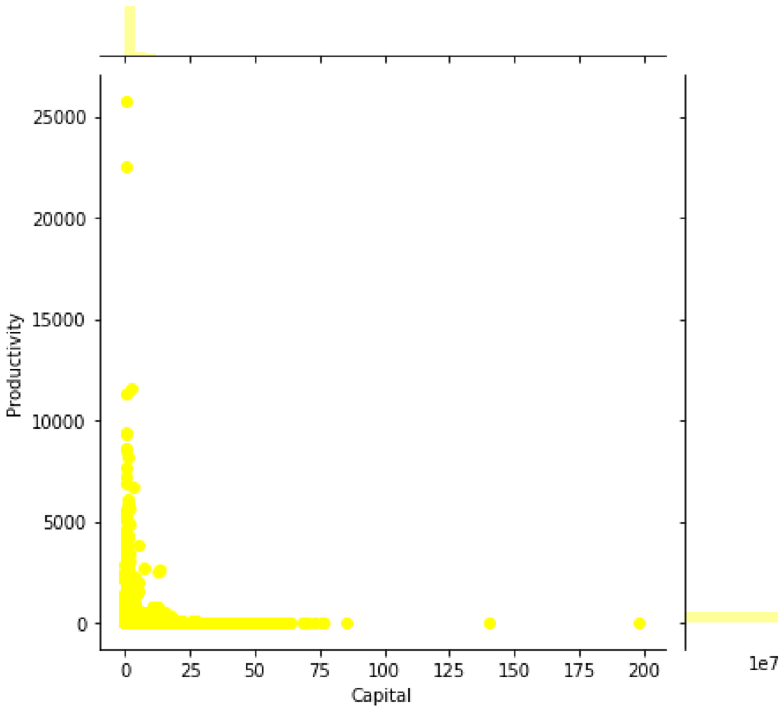
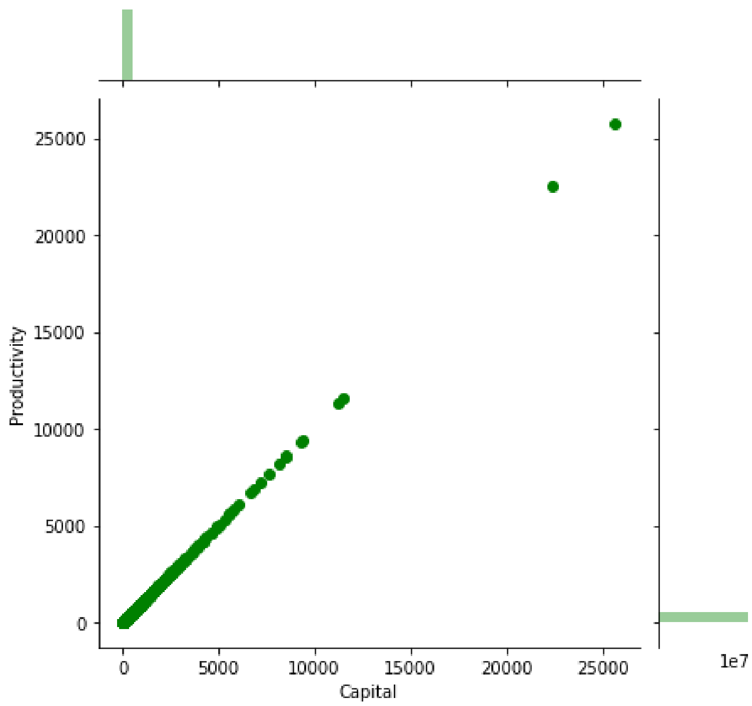


Figure 4: Optimal allocations



As we see, in the optimal case there is a perfect match between productivity level and capital for each firm, ensuring the aggregate output maximization. In the actual case, capital is spread and there are some high productivity firms that do own a low level of capital and vice versa. This will be traduced in a drastic reduction in aggregate output.

5. Compute the ouptut gains from reallocation

Applying

$$\left(\frac{Y^e}{Y^a}-1\right)*100$$

We obtain an output gain of 38,86%

6. Redo items (2)-(5) assuming that the correlation between ln z_i and ln k_i is 0.50. Redo with correlation -0.50.

<u>Case I: Correlation 0.5</u>

Figure 5: Level/Log variables distribution

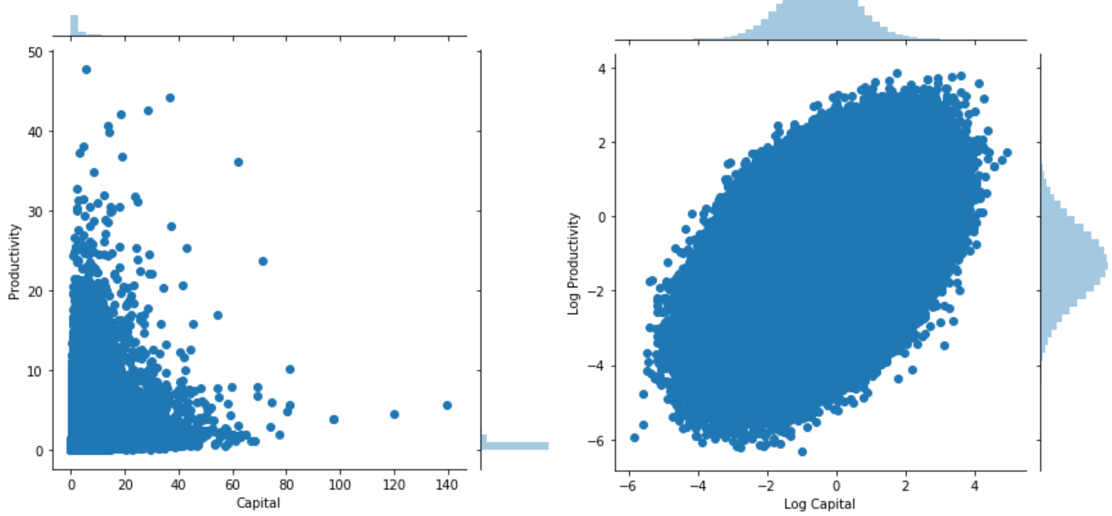
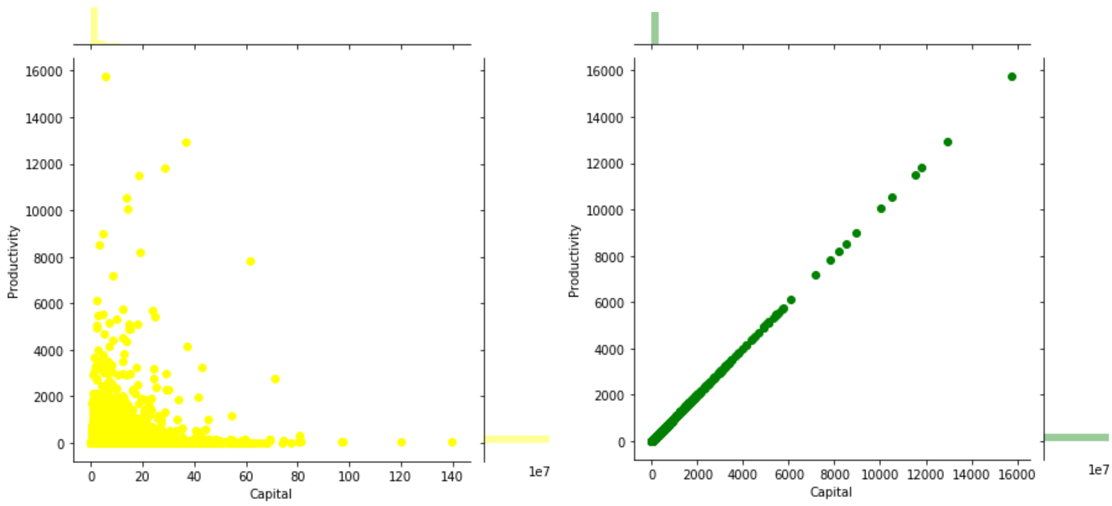


Figure 6:Actual/Optimal allocations



As we see, now there is a more efficient relation between firms' productivity and capital owned, as we expected after introduce a positive correlation.

The output gain now is 76%

Case I: Correlation -0.5

Figure 7:Level/Log variables distribution

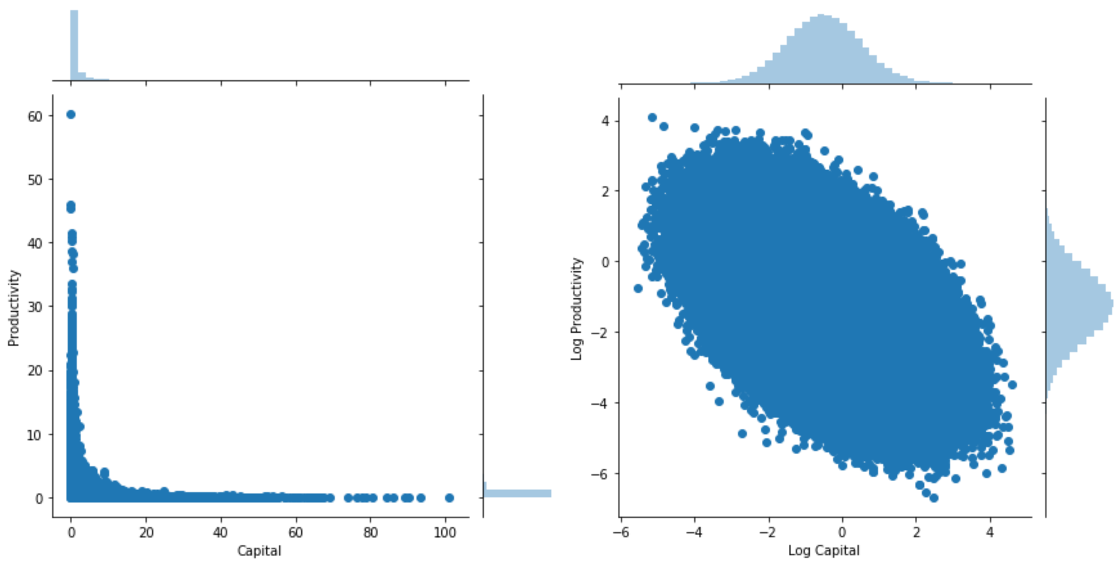
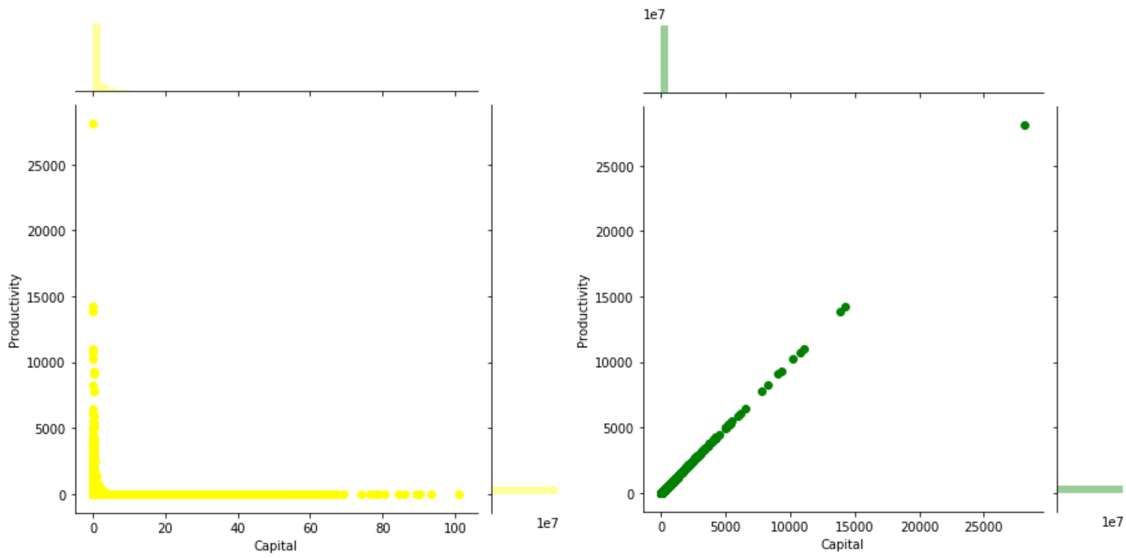


Figure 8: Actual/Optimal allocations



Here we face the opposite situation: since now we have introduced a negative correlation, distribution is even worse than in the no correlation case. Aggregation output will be the least of the three cases.

QUESTION 2: HIGHER SPAN OF CONTROL

1. Redo the previous Question 1 for gamma = 0,8. Discuss your results.

Now means are

E(ln k) = -0.5
E(ln z) = -2.5

Plotting the new distribution for the level & log variables

Figure 9: Level variables distribution

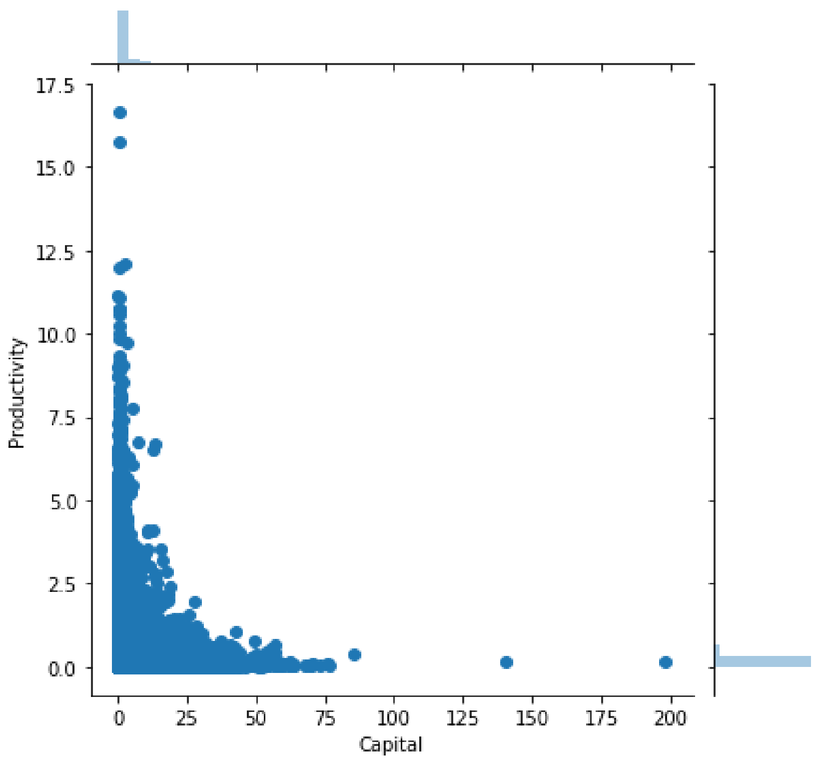
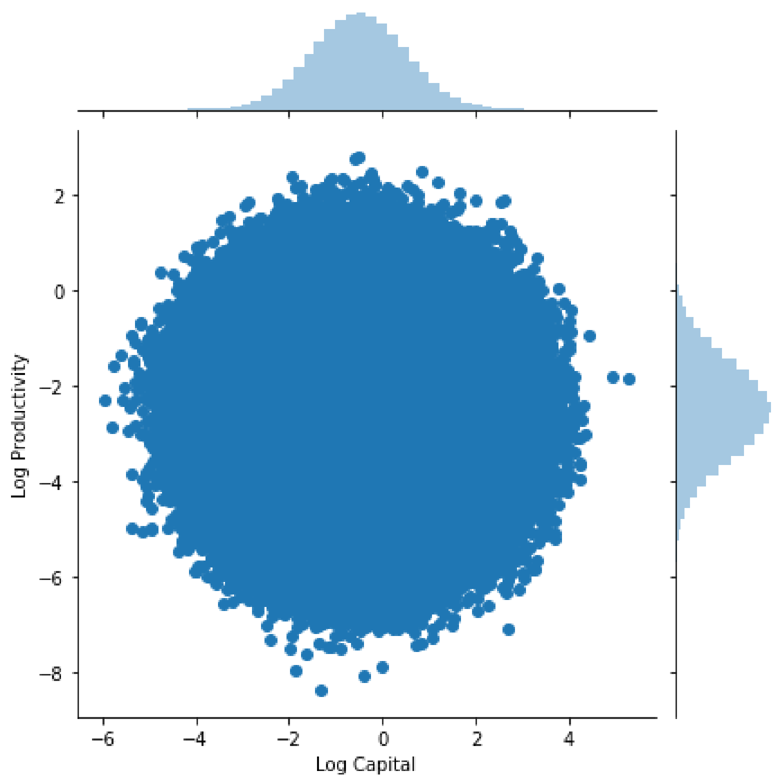


Figure 10: Log variables distribution



Which are very similar than in the previous case.

For the optimal allocations against the data we obtain

Figure 11: Actual allocations

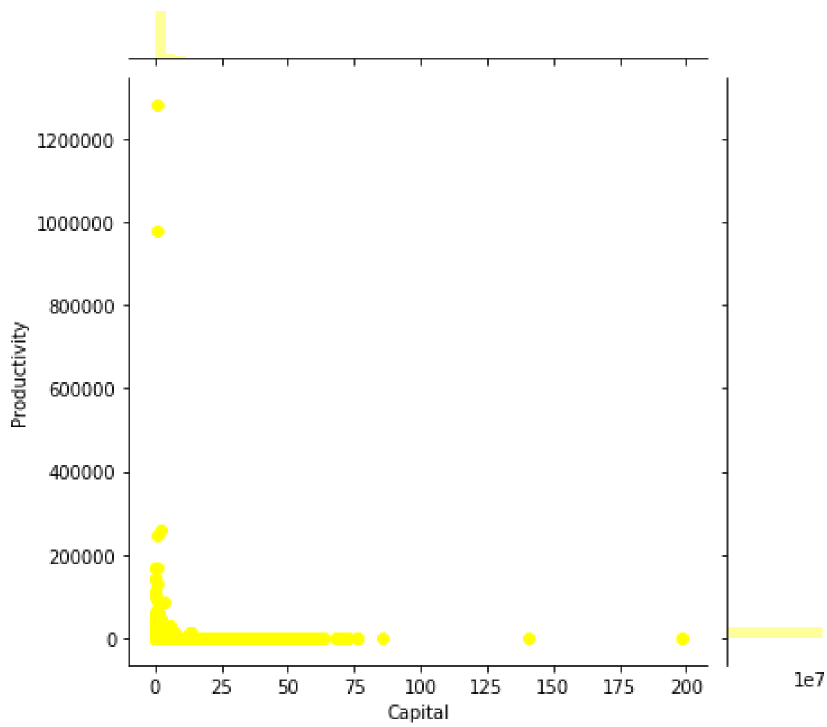
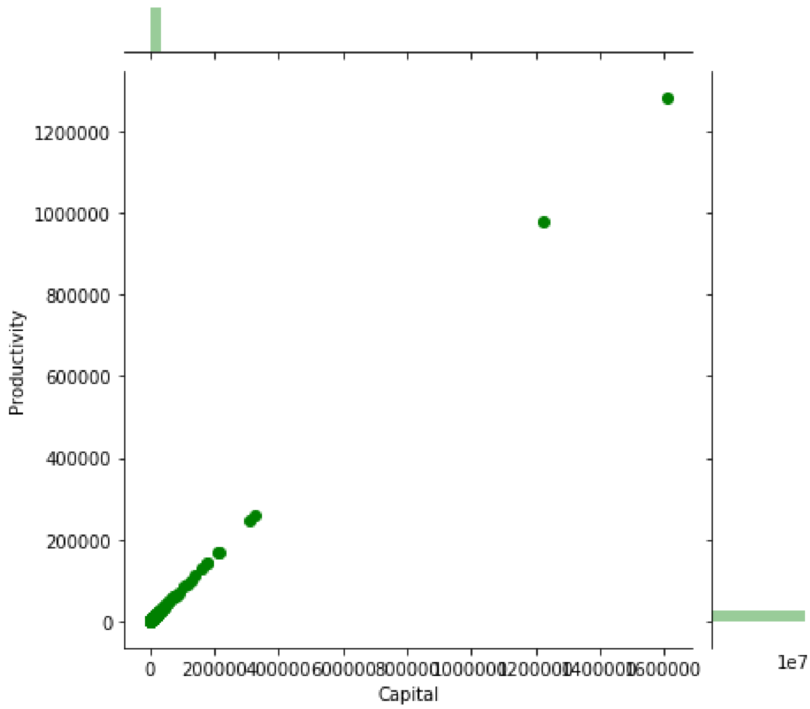


Figure 12: Optimal allocations



In the optimal allocation graph the situation is the same as before: each firm gets exactly what is optimal depending on its productivity. The actual distribution has now some differences. There are more extreme cases than in the previous case, meaning even more gap between productivity and owned capital for some firms.

Output gains from reallocation are now 664%

QUESTION 3: FROM COMPLETE DISTRIBUTIONS TO RANDOM SAMPLES

3.1 Please, random sample (without replacement) 10,000 observations. That is, your data sample implies a sample-to-population ratio of 1/1,000. What is the variance of ln z_i and ln k_i in your random sample? How do they compare compare to the complete data? How about the correlation between ln z_i and ln k_i?

We generate a random sample from the data & compute sample and population variance

Population Variance

----

var(log_k) = 0.9997189609301986
var(log_z) = 1.000336092984781

Sample Variance

----

var(log_k) = 0.9956283136873517
var(log_z) = 1.0254930570675518

To calculate correlation, we define a function to compute the Pearson coefficient

$$\rho_{X,Y} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

And we obtained

Population Correlation

----

corr(logk,logz) = 0.00024640542474171613

Sample Correlation

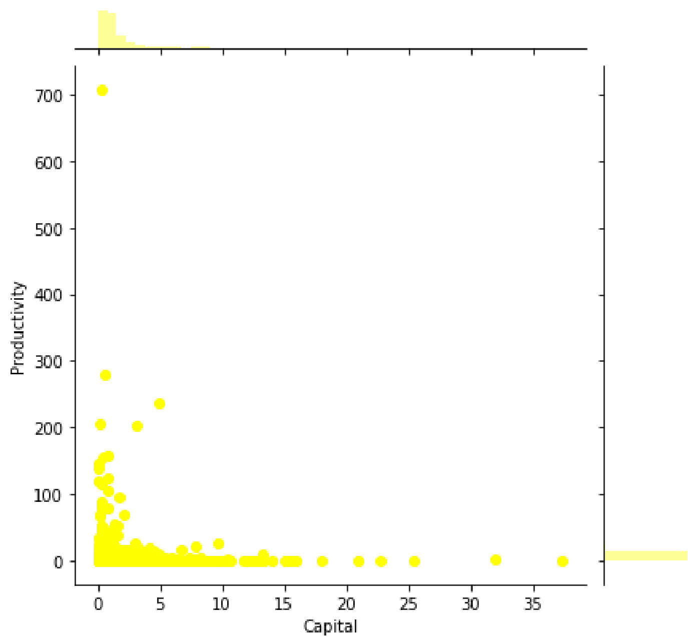
----

corr(logk,logz) = -0.004946059112903485

3.2 Redo items (3) to (5) in Question 1 for your random sample of 10,000 firms. Compare your results for misallocation using your random sample to the results obtained using the complete distribution.

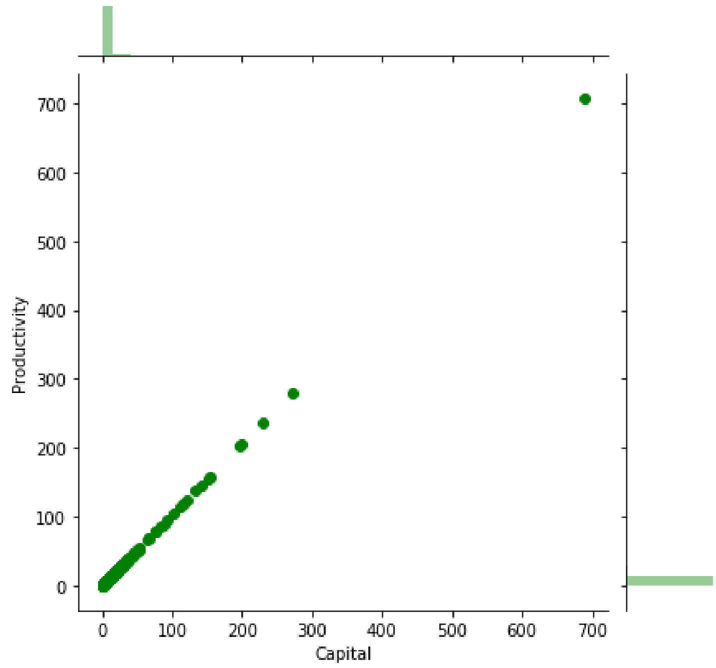
We redo exercise 1 with the new sample, and plot again the results of actual & optimal allocation against productivity obtaining

Figure 13: Actual allocations



As we see, the distribution has similar side face, but now there are more dispersion. This makes sense since now we are analysing a smaller sample from the entire population, so enlargering data will approach the graphic to the true distribution.

Figure 14: Optimal allocations

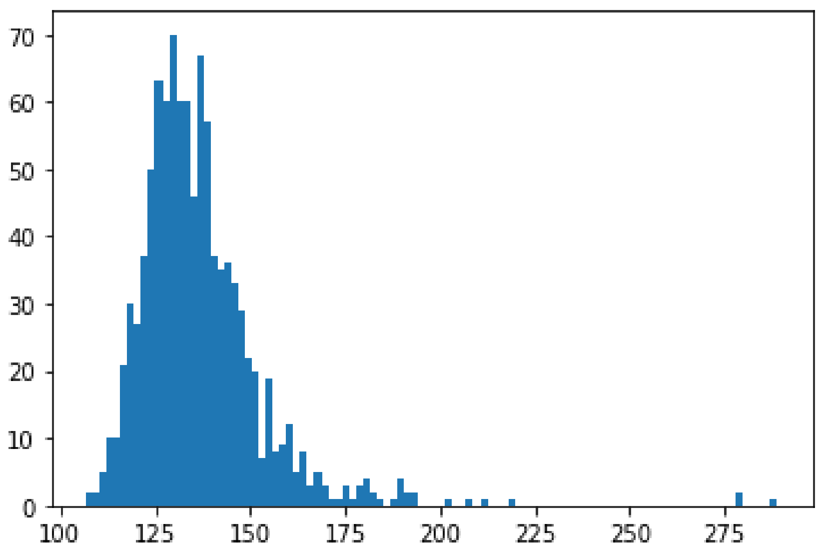


The gains from reallocation are now 39,30%

3.3 Do the previous two items 1,000 times. Notice that each random sample is drawn from the entire population. This implies that you will compute 1,000 measures of misallocation. Show the histogram of the output gains, and provide some statistics of that distribution of these output gains, in particular, the median. Discuss your results.

We iterate with the same process 1000 times and obtain

Figure 15: Output gains frequency



In which we can see that the most probable gain from reallocation depending on the sample draw is between 25 – 50 %. The median and the mean confirm this conclusion being both of them around 135.

Mean: 134.42
Median: 131.96

3.4. What is the probability that a random sample delivers the misallocation gains within an interval of 10% with respect to the actual misallocation gains obtained from complete data?

From our iteration, we compute the percentage of times for which the output gains is between these limits. We obtained a percentage of 69,9%

3.5. Redo items (1)-(4) for three different sample-to-population ratios. In particular, use do the cases in which your random sample extracts 100 observations, 1,000 observations and 100,000 observations. That is, your sample implies a sample-to-population ratio of, respectively, 1/100,000, 1/10,000, and 1/100. Compare your results in items (1)-(4) to those obtained with the previous random sample size and the complete data.

<u>**Case I: 100 000 observations**</u>

Population Variance

----

var(log_k) = 0.9997189609301986
var(log_z) = 1.000336092984781

Sample Variance

----

var(log_k) = 1.0009784335787149
var(log_z) = 1.0003251029471232

Population Correlation

----

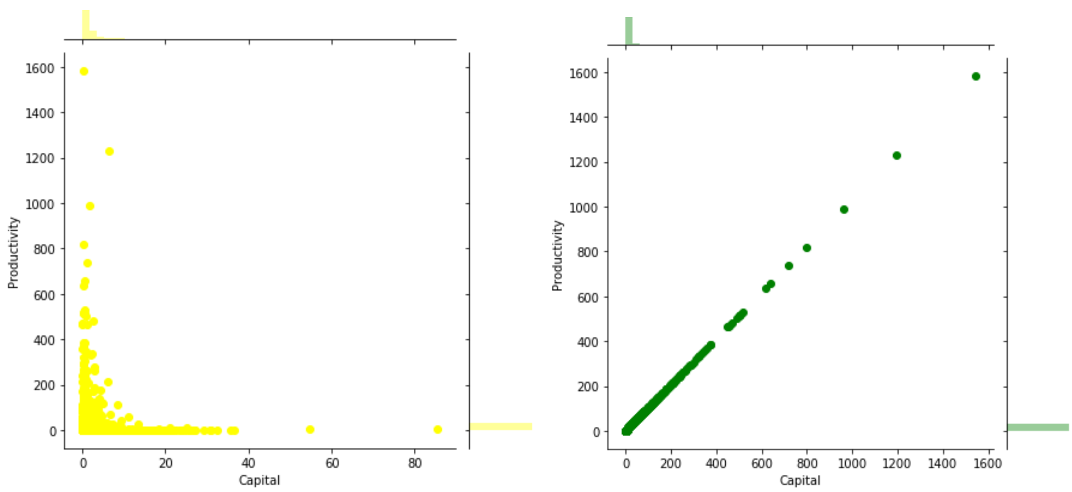
corr(logk,logz) = 0.00024640542474171613

Sample Correlation

----

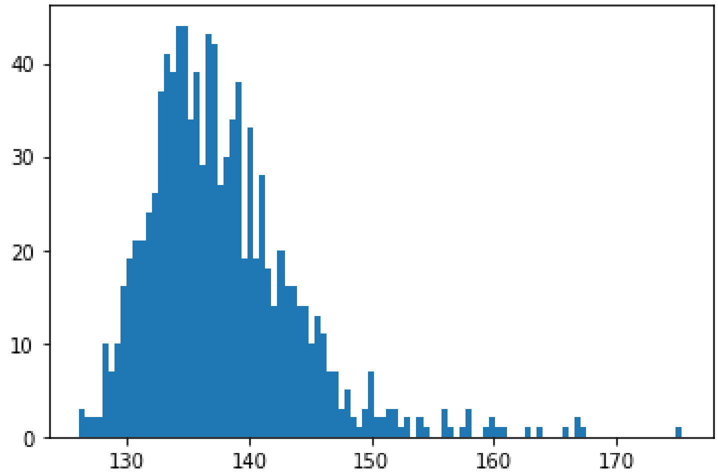
corr(logk,logz) = -0.001691799773064142

Figure 16:Actual/Optimal allocaations



Median: 136.74
Mean: 137.72
Output gains: 40.157 %

Figure 17: Output gains frequency



Percentage of gains inside bonds: 97.0

<u>Case II: 1000 observations</u>

Population Variance

----

var(log_k) = 0.9997189609301986
var(log_z) = 1.000336092984781

Sample Variance

----

var(log_k) = 0.9370048285141288
var(log_z) = 0.9784286494022565

Population Correlation

----

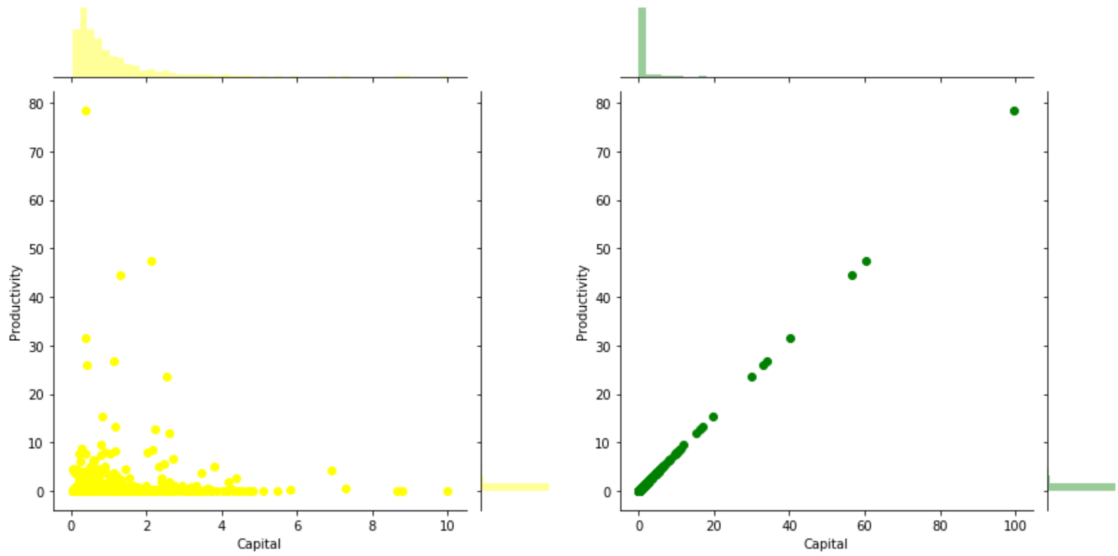
corr(logk,logz) = 0.00024640542474171613

Sample Correlation

----

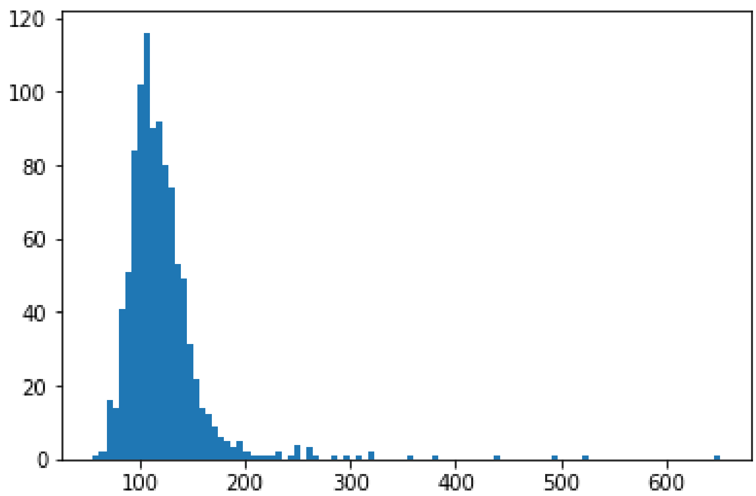
corr(logk,logz) = 0.050198093227372606

Figure 18: Actual/Optimal allocaations



Mean: 121.34454974087217
Median: 114.15821279770115
Output gains: 99.76473403334451%

Figure 19: Output gains frequency



Percentage of gains inside bonds: 24.4%

<u>Case III: 100 observations</u>

Population Variance

----

var(log_k) = 0.9997189609301986
var(log_z) = 1.000336092984781

Sample Variance

----

var(log_k) = 1.0901824358664627
var(log_z) = 1.063987516470223

Population Correlation

----

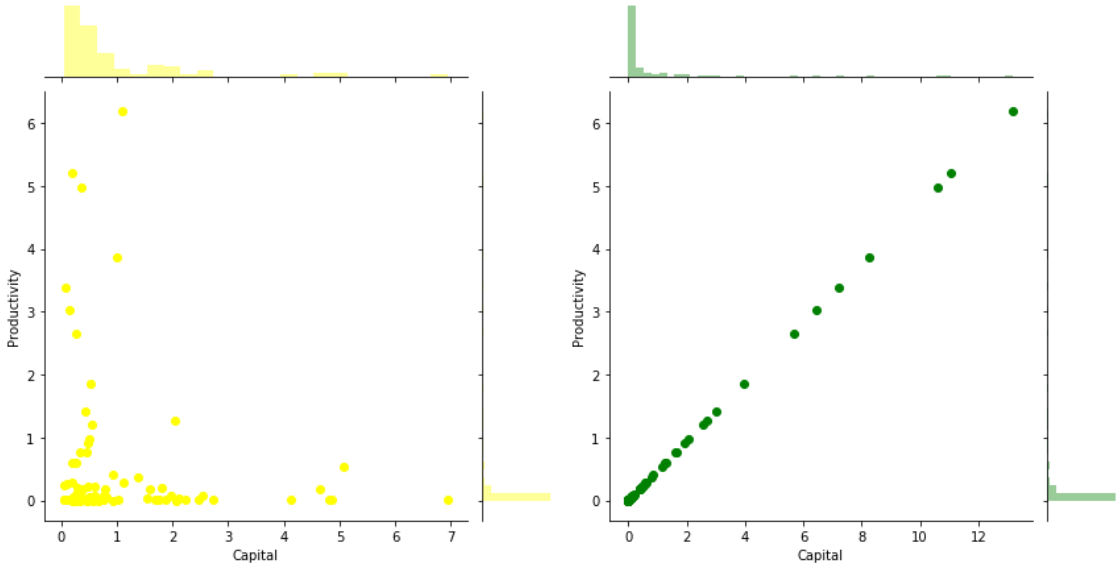
corr(logk,logz) = 0.00024640542474171613

Sample Correlation

----

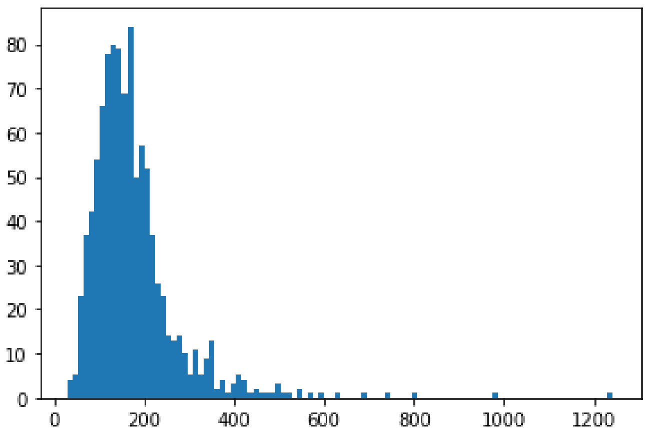
corr(logk,logz) = -0.04085686317748471

Figure 20: Actual/Optimal allocaations



Mean: 174.9684208012289
Median: 156.4882050521470
Output Gains: 113.63170498027087%

Figure 21: Output gains frequency



Percentage of gains inside bonds: 18.2%