Measuring of Firm Specific productivities: Evidence From Japanese Plant Level Panel Data

Ichimura, H [1], Y. Konishi [2] and Y. Nishiyama[3]

[1] Graduate School of Public Policy, University of Tokyo
[2] Institute of Economic Research, Hitotsubashi University, Tokyo
[3] Institute of Economic Research, Kyoto University, Kyoto
Email: konishi@ier.hit-u.ac.jp

Keywords: *segment level data, identification problem, observable and unobservable productivity shock*

EXTENDED ABSTRACT

In estimation of production function of firms, there are problems of endogeneity and self selection due to firm specific productivity shocks and entry/exit decisions. There are some methods proposed to handle the problems such as Olley and Pakes (1996) and Levinsohn and Petrin (1999, 2003). Here, endogeneity means input levels may not be independent of the "disturbances". The reason is that it is likely each firm determines the input levels depending on the firm-specific productivity, which is observable only for the firm, not econometricians, and thus the "disturbance" in the estimated equation involves the (unobserved) firm specific productivity shock, which should be highly correlated with the input levels, and other ordinary shocks. In the papers referred above, they consider that the endogeneity occurs only in the capital level, not in the labour input. It is a practical matter if this assumption is correct or not, but if it is incorrect, we will get inconsistent estimates.

In this paper, we suppose both capital and labour inputs are correlated with the productivity and propose an alternative semiparametric IV estimator. We adopt the lag variables of labour and capital as their instruments instead of investment or intermediate inputs unlike Olley and Pakes or Levinsohn and Petrin. Moreover, our econometric model automatically adapts to the effect of exit firm decision by each firm. We apply it to plant/segment level panel data of financial report of Japanese firms listed in Tokyo Stock Exchange. We found our estimator works well in empirical study in terms of sign and magnitude of technological parameters of inputs.

Using the estimation results, we decompose the so-called total factor productivity (TFP) or Solow residual into the firm specific productivity and other exogenous shocks based on the assumption that the exogenous shocks are uncorrelated with the inputs. Interestingly, we found that the firm specific productivity has not changed much in these five years. The fluctuation of TFP for each firm comes mostly from that of exogenous shocks, which we may think, the demand shock or other macroeconomic shocks. It is sometimes said that the productivity of Japanese economy has declined since the burst of babble economy, it may not be due to the productivity falls of Japanese firms, but due to a simple macroeconomic demand problem. Since we have investigated only some restricted number of industries, we need to extend it to other industries as well. Also, we can see the productivity changes only 2000~2005, it is not sufficient to make a strong statement. We will also need to extend it to cover 1990's as well.

1. INTRODUCTION

In Japan, after bursting babble economy since the middle of 1990's, the growth rate has not been increasing obviously, and it is said the productivity keeps declining. This period is sometime called "the lost decade". Recently, a number of researchers and the government try to get an answer what did occur in the period, and find an effective policy for raising industrial productivity and growth rate of GDP. Recently, we can use micro level data, for example, the plants and segments data as well as firms' data. It allows empirical researches to make more precise statistical analysis. Fukao and Kwon (2006) make the plant level data set in their project and use them to examine of productivity and they found the reasons of declining of productivities in last lost decade.

In the productivity analysis, the most common measure of productivity is Total Factor Productivity TFP, hereafter. Beginning with a pioneering work by Solow (1957), economists regard the constant term of Cobb-Douglas production function as the TFP. Production technology of a firm or an

economy is characterized by its production function (or cost function alternatively).

We briefly describe the production function. Cobb and Douglas (1928) proposes a production function with the following the form,

$$Y_{it} = AK_{it}^{\beta_k} L_{it}^{\beta_l} \tag{1}$$

where Y, K, L indicate the output level, capital and labour inputs respectively and A, β_k , β_l are parameters determining the production technology. We transform the Cobb-Douglas production function into a log-linear form.

$$y_{it} = \beta_0 + \beta_l l_{it} + \beta_k k_{it} + u_{it} \tag{2}$$

Equation (1) or equivalently (2) is called the Cobb-Douglas production function. Christensen, Jorgenson and Lau (1973) consider an extension of the Cobb-Douglas production function to the following more general and flexible functional form that including the polynomials of independent variables, that is called the Translog production function. These two functional forms are used widely in theoretical and empirical Economic research, and estimation of production function has been one of the main issues in empirical economics and econometrics. Especially, a lot of previous empirical works estimate the production function by least square method and treat the regression residuals as TFPs.

Though the regression residual is commonly used as an estimate of TFP, we should point out that the existence of an econometric problem an endogeneity problem. Endogeneity means here that after each firm observe their TFP (technology or productivity), they decide the levels of factor inputs. Then l_{it} and k_{it} and error terms must be correlated, which causes a bias in the OLS estimators. Obviously, the problem comes from that each firm can observe its own productivity but econometricians cannot.

There are some methods proposed to handle the problems such as Olley and Pakes (1996) and Levinsohn and Petrin (1999, 2003), hereafter we call O &P and L&P methods. They split out the error term into two parts as follows. ω_{it} represents the firm specific productivity or technological shock and η_{it} denotes the ordinary error term.

$$y_{it} = \beta_0 + \beta_l l_{it} + \beta_k k_{it} + \omega_{it} + \eta_{it} \tag{3}$$

They consider a correlation between ω_{it} and k_{it} explicitly, and it contributes to find an influence of the each firm productivity shock to their output growth. To the best of our knowledge, there are not many researches that apply these methods to Japanese plant level dataset. Fukao *et al.* (2007) is one of the important previous works, where they apply L & P method to estimate the production function of Japanese plant level. Their main interest is in Japanese wage function however, they use L & P estimation method to check the validity of the parameter of labor productivity supplementary.

In this paper, we apply L & P method to Japanese segments data of firms which belong to variety of industries, not only manufacturing industry but also service, commerce, whole sale trade, a real estate and car trucking industries. Because the endogeneity problem does not seem to be completely solved by O& P and L & P methods, we propose an alternative IV estimator. Applying our method to the same segment data set, and we also observe the firm specific productivities.

The following section shows a review of some papers that solve the endogeneity and the sample selection problems of the productivity analysis. Section 3 shows our alternative IV estimator to examine the firm specific productivity. While Section 4 gives results of the OLS, L & P and our method, we examine the firm specific productivity and decompose the productivity shock and the error term using the estimation results, in Section 5. Concluding remarks and future research are in Section 6.

2. REVIEW OF THE PREVIOUS WORKS

A number of previous researches are provided about measuring the TFP and macro productivity about economic growth. Here we ensure readers understand the meaning of terms used when discussing alternative methods.

Now we have a production function equation in equation (3). Suppose ω_{it} represents each firm's technology / productivity shocks that they are observable only for the firm, and each firm decides levels of factor inputs after observing the actual ω_{it} . Under this assumption, factor inputs and the productivity shock are correlated and it becomes a cause of the endogeneity problem in the estimation of equation (3).

Olley and Pakes (1996) and Levinsohn and Petrin (1999, 2003) show a solution to this problem using the firm's investment decisions as a proxy of ω_{it}

in (3). We can obtain accumulated K by standard perpetual inventories method as below, where K is the capital stock, I is the investment and δ is depreciation ratio.

$$K_{it+1} = (1-\delta)K_{it} + I_t \qquad (4)$$

Pakes (1996) proves that optimizing firms have investment functions that are strictly increasing in the unobservable productivity shock. We can write investment function as $i_t = i_t(\omega_t, k_t)$. The monotonicity allows investment function to be inverted to get $\omega_t = \omega_t(i_t, k_t)$. Including ω_{it} in the model, it gives a relation with k_{it} explicitly, and they could solve an endogeneity problem between k_{it} and ω_{it}. Inserting $\omega_t = \omega_t(i_t, k_t)$ in equation (3), we can write the model as a partially linear model, and obtain consistent semiparametric estimates of β_l and ϕ by Robinson (1987) as follows. Write

$$\begin{aligned} y_t &= \beta_0 + \beta_l l_t + \beta_k k_t + \omega_t(i_t, k_t) + \eta_t \\ &= \beta_l l_t + \phi_t(i_t, k_t) + \eta_t \text{ ,} \end{aligned} \qquad (5)$$

and subtract $E(y_t|i_t,k_t) = \beta_l E(l_t|i_t,k_t) + \phi_t(i_t,k_t)$ from equation (5). Then we obtain this equation,

$$y_t - E(y_t \mid i_t,k_t) = \beta_l\{l_t - E(l_t \mid i_t,k_t)\} + \eta_t \text{ .} \qquad (6)$$

Replacing the conditional expectations by nonparametric estimates as below, we can stimate equation (7) by least square method to obtain the consistent estimator of β_l.

$$y_t - \hat{E}(y_t \mid i_t,k_t) = \beta_l\{l_t - \hat{E}(l_t \mid i_t,k_t)\} + \eta_t \qquad (7)$$

In the second step, we identify β_k of the model. Assume ω_{it} follows a first order markov process, $\xi_t = \omega_t - E(\omega_t|\omega_{t-1})$ is uncorrelated with k_t, and put

$$\begin{aligned} \phi_t(i_t,k_t) &= \beta_0 + \beta_k k_t + \omega_t(i_t,k_t) \\ &= \beta_0 + \beta_k k_t + E(\omega_t \mid \omega_{t-1}) + \xi_t \end{aligned} \text{ .} \qquad (8)$$

Inserting equation (8) into (5), we have

$$\begin{aligned} y_t &= \beta_0 + \beta_l l_t + \beta_k k_t + \omega_t + \eta_t \\ &= \beta_0 + \beta_l l_t + \beta_k k_t + E(\omega_t \mid \omega_{t-1}) + \xi_t + \eta_t \end{aligned} \qquad (9)$$

where $\xi_t + \omega_t$ and k_t, l_t are uncorrelated. In the third step, we estimate β_0 and β_k, given β_0 and β_k, we can implement nonparametric estimation for $E(\omega_t|\omega_{t-1})$, and obtain $\hat{E}(\omega_t|\omega_{t-1})$, inserting $\hat{\beta}_l$ and $\hat{E}(\omega_t|\omega_{t-1})$ into (9) and we can estimate $y_t \approx \beta_0 + \hat{\beta}_l l_t + \beta_k k_t + \hat{E}_{(\beta_0,\beta_k)}(\omega_t \mid \omega_{t-1}) + \xi_t + \eta_t$ to get estimators of β_0 and β_k using non-linear least square or generalized method of moments. Levinsohn and Petrin (2003) show that the intermediate inputs can also be used to solve the endogeneity problem.

3. AN ALTERNATIVE ESTIAMTOR -IKN ESTIMATOR-

Olley and Pakes (1996) and Levinsohn and Petrin (1999, 2003) show how to use investment and intermediate inputs to control for correlation between capital inputs level and unobservable firm specific productivity shock. And they can identify the constant term and the parameters of inputs and surely they are consistent estimators.

However, the endogeneity problem of inputs level and unobservable firm specific productivity does not seem to be completely solved by these methods. They only consider the correlation between capital input level k_{it} and unobservable firm specific productivity shock ω_{it}. Because they adopt a estimation method by Robinson (1987), if l_{it} is also determined by firms depending ω_{it} like k_{it}, we can see $E(l_t|i_t,k_t) = E(l_t|\omega_t) = l_t$ and their procedure of getting $\hat{\beta}_l$ collapses. And if there was no the econometric technical problem as above, the assumption does not look reasonable in actual decision making of firms.

Thus we propose an alternative IV estimator and we names it Ichimura-Nishiyama-Konishi estimator, hereafter we call it IKN estimator. We suppose that the firm specific productivity influences labor input level as well as capital's one. We adopt the lag variables of labor and capital as their instruments instead of investment or intermediate goods like Olley and Pakes or Levinsohn and Petrin. Moreover, our model also includes the effect of entry-exit firm decision to confirm their productivity. We can rewrite equation (3) as

$$\begin{aligned} y_{it} &= \beta_0 + \beta_l l_{it} + \beta_k k_{it} + \omega_{it} + \varepsilon_{it} \\ &= \beta_0 + \beta_l l_{it} + \beta_k k_{it} + E(\omega_{it}|k_{it-1}, l_{it-1}) + \omega_{it} - E(\omega_{it}|k_{it-1}, l_{it-1}) + \varepsilon_{it} \\ &= \beta_0 + \beta_l l_{it} + \beta_k k_{it} + g(k_{it-1}, l_{it-1}) + \xi_{it} + \varepsilon_{it} \end{aligned} \qquad (10)$$

where,

$$g(k_{it-1}, l_{it-1}) \equiv E(\omega_{it}|k_{it-1}, l_{it-1}), \xi_{it} \equiv \omega_{it} - E(\omega_{it}|k_{it-1}, l_{it-1}) \text{ .}$$

From this equation, we could know immediately

$E(\xi_{it}|k_{it-1},l_{it-1})=0$, $E(\varepsilon_{it}|k_{it-1},l_{it-1})=0$, $E(k_{it}k_{it-1})\neq 0$ and $E(l_{it}l_{it-1})\neq 0$, then $f_k(k_{it-1})$ and $f_l(l_{it-1})$ for any functional forms of f_k and f_l are usable as instrumental variables for k_{it}, l_{it}. And also we can use a relationship of $E(\xi_{it}|k_{it-2},l_{it-2})=E[E(\xi_{it}|k_{it-1},l_{it-1})|k_{t-2},l_{t-2}]=0$, it means $f_k(k^s{}_{it-2})$ and $f_l(l^s{}_{it-2})$ are also usable as instrumental variables for k_{it}, l_{it}. Because k_{it}, l_{it} are endogenous variables, we apply instrumental variable method to estimate of our model. To implement estimation of equation (10), we adopt polynomial functions of $k_{it-1}, l_{it-1}, k_{it-2}, l_{it-2}$ as instrumental variables and we can approximate $g(k_{it-1}, l_{it-1})$ by trigonometric, splines and any other smoothed functions.

We describe IKN estimator's advantages and disadvantages briefly. We could allow for the correlation between ω_{it} and l_{it} as well as ω_{it} and k_{it}. We don't need to use investment as a proxy variable of ω_{it}, because usually it is hard to obtain segment level.' s investment data as Levinsohn and Petrin (2003) pointed it out. The demerit is that we use $k_{it-1}, l_{it-1}, k_{it-2}, l_{it-2}$ as instrument variables so that number of observation effectively used decreases.

4. ESTIMATION

We estimate the Cobb-Douglas production function (eq.2 & 4) by 3 methods which are OLS, L & P method and IKN method. Our data set is panel data of the Japanese segment level that covers the all kinds of industries and the period from 2000 to 2005.

4.1. Data

Our data set is from Nikkei NEEDS which is financial report, and our targets are listed companies that belong to the first section market at Tokyo Exchange Market. It is periods from 2000 to 2005 and segment level panel data. We should describe about a segment shortly. A segment belongs to a firm, it is sometime the smallest production unit, equal to the plant or constructed some sectors of the firm. Usually, each firm produces a large variety of goods; it is difficult to identify which technology is used to produce a good in firm-level analysis. Therefore, we sort the data and make groups by kinds of products and combine the segments if they produce same products by Japan Standard Industry Classification: JSIC 3 or 4-digits level in order to measure homogeneity technology in the group. We focus on 10 of the varieties industries in Table.1. We use their value added as their output variables that is dependent variable. They are composed by subtracting the cost of raw materials and sales administrative expense from the total amount of the sales. Independent variables are Capital (K) and Labour (L) are adopted fixed assets and work forces. For L&P estimation, an investment (I) variable is capital Expenditure.

4.2. Estimation results

We found different technology of capital and labour among these industries by IKN and obtained some reasonable results without medical products. OLS estimator results also look reasonable, but the estimators don't have consistency and tend to upper wards biases. In almost L & P results, we can not see the significance of the parameters of L. It suggests that O & P and L & P style's estimation can not identify $\hat{\beta}_l$ well. Contrastingly, our assumption of endogeneity problems seems to be valid. Moreover, some OLS and IKN estimator results are very similar. Though we discuss about them in Section 5, the phenomena imply the firm specific productivities are not existent or their fluctuating are sharply.

5. MEASURING THE FIRM SPECIFIC PRODUCTIVITY

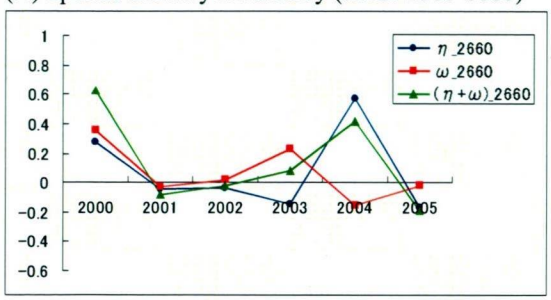
After bursting babble economy since the middle of 1990's, we had an economic stagnation for a while, and it is said that the bottom was 2001. Since then, we can see a very slightly economic recovery. Measuring the TFPs in the 2000-2005 periods, we might see the influences of the business fluctuations to the firm level productivities. Here, using the estimation results, we could decompose $\hat{\omega}_t + \hat{\eta}_t$ that is the firm specific productivity (pure TFP) and the error term, and we show results in Figure 1.We focus on observing the results of (B), (F) and (G). In previous productivity analysis, we usually discuss about the productivity by $\hat{\omega}_t + \hat{\eta}_t$. In these 3 results, $\hat{\omega}_t + \hat{\eta}_t$ seem to have upper wards trend, so we might conclude "the productivity increases in the period", but pure TFP $\hat{\omega}$ does not change actually. We should say the technological productivities keep stable in the period in the industries. We said the phenomena in previous section, (G)'s estimation results are very similar both of OLS and IKN. It means the correlation between 2 inputs and ω_{it} is not existent or quite small. In that situation, we can not find the

Table 1. Estimation Results of OLS, L &P and IKN. * and ** present 10% and 5% significant level.

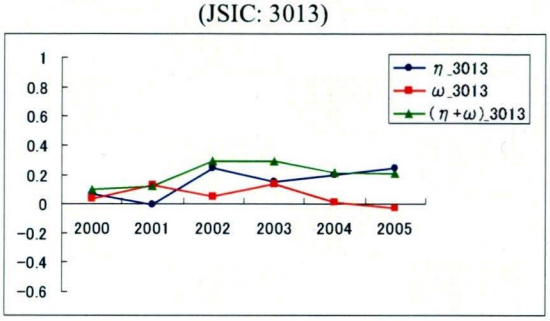
Drugs & Medicines (JSIC 1760)			
	OLS	L & P	IKN
lnK	0.927**	0.638**	0.819**
lnL	0.148	0.118	0.215
Obs.	351	349	179
Special Industry Machinery (JSIC 2660)			
	OLS	L & P	IKN
lnK	0.690**	1.145**	0.711**
lnL	0.310**	0.282	0.300**
Obs.	237	230	91
Motor Vehicles-Parts & Accessories (JSIC 3010)			
	OLS	L & P	IKN
lnK	0.684**	0.857**	0.767**
lnL	0.305**	0.205	0.237**
Obs.	510	501	267
Computer Programming Services (JSIC 3910)			
	OLS	L & P	IKN
lnK	0.547**	0.923**	0.608**
lnL	0.271**	0.291	0.344**
Obs.	514	487	226
Data Processing & Information Services (JSIC 3920)			
	OLS	L & P	IKN
lnK	0.606**	0.698**	0.631**
lnL	0.255**	0.312**	0.250**
Obs.	352	329	135
Common Motor Tracking (JSIC 4410)			
	OLS	L & P	IKN
lnK	0.683**	0.350**	0.676**
lnL	0.226**	0.292*	0.240**
Obs.	353	350	182
General Machinery & Equipment; Wholesale Trade (5310)			
	OLS	L & P	IKN
lnK	0.830**	1.267**	0.587**
lnL	0.227**	0.285	0.631**
Obs.	158	146	65
Electrical Machinery-Equipment & Supplies; Wholesale Trade (JSIC5330)			
	OLS	L & P	IKN
lnK	0.523**	0.694**	0.659**
lnL	0.322**	0.057	0.241**
Obs.	216	187	99
Department Stores & General Supermarkets (JSIC 5510)			
	OLS	L & P	IKN
lnK	0.546**	0.654*	0.526**
lnL	0.351**	0.304	0.365**
Obs.	246	187	135
Sales Agents of Buildings & Houses &Land (JSIC 6810)			
	OLS	L & P	IKN
lnK	0.695**	0.943**	0.726**
lnL	0.204**	0.156	0.192**
Obs.	822	677	340
Real Estate Lessors-Except House & Room Lessors(6910)			
	OLS	L & P	IKN
lnK	0.708**	0.462**	0.709**
lnL	0.105**	0.034	0.092**
Obs.	1188	1039	561

productivity changes of, $\hat{\omega}$ of (G) stays around "0" and don't change the level. We should note that there could be the industries which don't have correlation between inputs level and the firm-specific productivities.

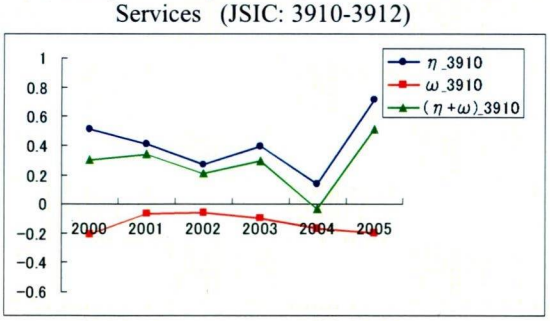
(A) Special Industry Machinery (JSIC: 2660-2668)



(B)Motor vehicles parts and accessories (JSIC: 3013)



(C) Computer Programming and other Software Services (JSIC: 3910-3912)



(D)Common Motor Tracking (JSIC: 4410-4412)

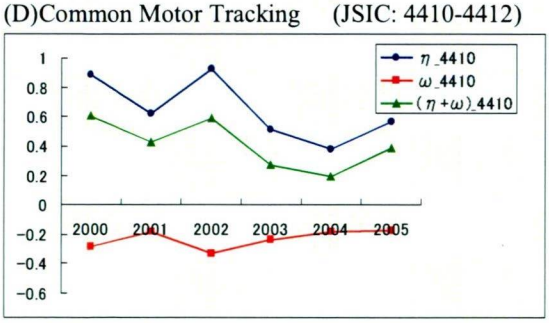
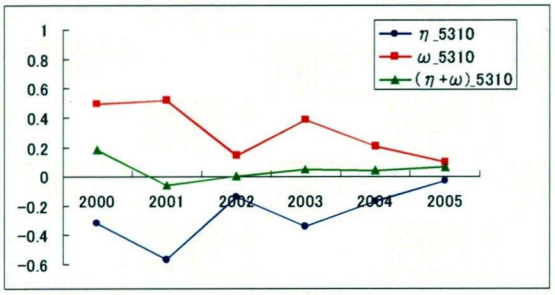
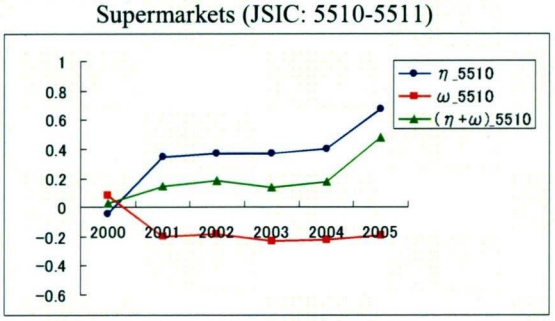


Figure 1. Decomposition of productivity shocks and other shocks.

(E) General Machinery & Equipment (Wholesale Trade) (JSIC: 5310-5314)



(F) Department Stores & General Merchandise Supermarkets (JSIC: 5510-5511)



(G) Sales Agents of Buildings & Houses & Land Subdividers & Developers (JSIC: 6810-6812)

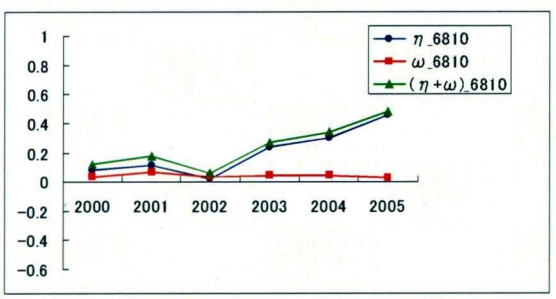


Figure 1. Decomposition of productivity shocks and other shocks. (Continued).

6. CONCLUSION AND FUTURE RESEARCH

The alternative estimator IKN presented in this paper provides a new measure for the segment level productivity. We found different technology of capital and labour among these industries by IKN. We also applied L&P estimation procedure to Japanese data of some financial reports. We proposed an alternative estimation method to O&P and L&P for production function under a stochastic firm- and time- specific technology which causes a nuisance endogeneity. This procedure allows that labour depends on the technology level unlike O&P or L&P, and exit decisions are endogenous automatically under certain conditions. We applied this method and obtained some reasonable results for machinery and equipments, car parts, trucking, department stores, estate agents and so on. We will apply this method to other industries. We measure firm- & time specific production skills in a similar manner as TFP. Using the above measure, we could decompose he firm specific productivity (pure TFP) and the error term.

Konishi and Nishiyama (2002) pointed out that Cobb-Douglas and Trans log production function are not adequate functions for measuring the productivity based on firm specific analysis, and they show the necessity to check the functional form statistically by Hong and White (1995) nonparametric functional form test. In this paper, we adopt the Cobb-Douglas production function basically, so we will construct the Hausman and Hong and White type test for our estimator. Moreover, we will compares the properties of these alternative estimators theoretically and numerically. Finally, using the measuring the productivities results, we will aggregate them into industry level as in L & P in order to observe the change of productivities in recent years.

7. REFFERENCES

Christensen, L. R., D. W. Jorgenson and L. J. Lau (1973), Transcendental logarithmic production frontiers, *Review of Economics and statistics,* 55, 28-45.

Cobb, C. W. and P. H. Douglas (1928), A theory of production, *American Economic Review,* 18, Supplement, 139-165.

Fukao, K. and H.U. Kwon (2006), Why did Japan's TFP growth slow down in the lost Decade? An empirical analysis based on firm-level data of manufacturing firms, *The Japanese Economic Review*, 57 (2), 195-227.

Fukao, K., R. Kambayashi, D. Kawaguchi, H.U. Kwon, Young Gak Kim and I, Yokoyama (2007), Deferred compensation: Evidence from employer-employee matched data from Japan, mimeo.

Hong, Y., and H. White. (1995), Consistent specification testing via nonparametric series regression. *Econometrica*, 63, 1133-1159.

Konishi, Y. and Y. Nishiyama (2002), Nonparametric Test for Translog Specification of Production Function in Japanese Manufacturing Industry, in A.E.

Rizzoli and A.J. Jakeman (eds.), *Integrated Assessment and Decision Support, Proceedings of the 1st Biennial Meeting of the International Environmental Modelling and Software Society,* 2002, 2(4), 597-602.

Levinsohn, J. and A. Petrin (1999) , When industries become more productive, Do firms? : investigating productivity dynamics, *NBER working paper series*; working paper 6893; Cambridge: National Bureau of Economic Research.

Levinsohn, J and A. Petrin (2003), Estimating production functions using inputs to control for unobservables, *Review of Economic Studies*, 70 Issue 2 , 317-411.

Olley, G., S. and A. Pakes (1996), The dynamics of productivity in the telecommunications equipment Industry, *Econometrica,* 64, 1263-1297.

Pakes, A (1996), Dynamic structural models problems and prospects: mixed continuous discrete control and market interaction, In Sims, C. (Ed.), *Advances in Econometrics*, Sixth World Congress, II, 171-259.

Robinson, P. M. (1988), Root-n consistent semiparametric regression, Econometrica, 55(4), 931-954.

Solow, R. M. (1957), Technical change and aggregate production function, *Review of Economics and Statistics*, 39, 312-320.

ランクサイズ回帰の検定について*

小西葉子† 西山慶彦‡

2008年2月

概 要

多くの実証研究では都市サイズ，企業の資産や売上高の規模などの研究対象がパレート性を持つことを，ランクサイズ回帰で観察してきた．具体的には，順位の対数値をその規模の対数値に回帰することにより，その係数が-1になるかを調べる．また，パレート性の有無には，二次項の係数が0であることも条件になるので，本稿では，二次項を加えたものを回帰モデルとする．パレート性の検証には，一次項，二次項それぞれのt検定と，一次項の係数が-1，二次項の係数が0という複合仮説が成立しているかをF検定で調べる方法がある．しかし，分析対象がパレート分布に従う時，データ数が大きくなると，t値は発散してしまうため通常のt検定を行えないことがわかっており，F検定でも同様の問題が観察された．そこで本稿では，F値の棄却域をシミュレーションによって構成し，ランクサイズ回帰の複合仮説を検証可能とし，パレート性の検定の新たな手法として提案した．

JEL Classification: C12, C16, R12,

*一橋大学経済研究所，定例研究会報告 (2008年2月27日) に対する討論者であった野口晴子氏のコメントに感謝します．また，同定例研究会の参加者からも多くの有意義なコメントを頂きました．記して感謝致します．

†独立行政法人 経済産業研究所

‡京都大学経済研究所

1 はじめに

都市・地域経済学でランクサイズ回帰のあてはまりがよく，古くから応用されている分野に都市人口分布の分析がある．まず一国の都市の人口を大きい順に並べ替え，1位，2位 ... と順位（ランク）をつける．ランクサイズ回帰とは，都市の人口規模の対数値を当該都市の順位(ランク)の対数値に回帰したものである．すると，多くの国において定数項がほぼサンプルサイズの対数に等しく，傾きはほぼ -1 に等しくなるという結果が得られる．つまり人口規模が1番大きい都市から順に2番目の都市は $1/2$ の人口，3番目は $1/3, \ldots$ と減少していく．

$S_i, i = 1, \ldots, n$ をある国の都市 i の人口とし，$S_{(i)}$ をそれを大きい順に並べ替えた順序統計量とする．つまり $S_{(1)} \geq S_{(2)} \geq \ldots \geq S_{(n)}$ である．

$$\log S_{(i)} \approx \alpha_0 + \alpha_1 \log i, i = 1, \ldots, n \tag{1}$$

$\alpha_0 > 0$, $\alpha_1 < 0$ のとき，(1) 式はランクサイズ回帰モデルと呼ばれ，$\alpha_1 = -1$ のとき Zipf's law(ジップの法則) が成り立っているといえる．またパレート分布のパラメータが1の場合も同様の関係が観察される．

このような現象は様々な分野やトピックで観察されており，経済学では所得や資産が上位層に集中することや，有能な少数の従業員が全体の生産に大きく貢献することなどが事例として挙げられている．これは，パレートの $80:20$ の法則とも呼ばれ，集中度や不平等度の指標の一つとして用いられてきたが，近年では，マーケティングや経営の分野でも広く用いられている．この種の方法は，自然科学や実験など大規模データが利用可能な分野で応用されてきたが，近年マイクロデータの利用可能性が高まることによって，都市経済学以外の分野でも分析対象がパレート性をもつか，あるいはある種のべき乗分布に従うか否かを検証する研究が行われはじめた．齊藤・渡辺 (2007) では，我が国の法人企業の $\frac{1}{3}$ をカバーする約82万社のデータを用いて企業間関係の分析を行っている．そこでは，企業間のネットワーク構造に着目して，当該企業と「仕入先」,「販売先」,「大株主」に関する企業間ネットワークに，パレート性やべ

き乗の関係があることが発見されている.

このようにランクサイズ回帰による分析が盛んな理由の一つは，最小二乗法中心の簡便な方法で分析ができることにある.

==図 1 挿入==

図 1 は各国の都市人口について縦軸に人口の対数値，横軸にその大きさに対応するランクの対数値をとってプロットしたグラフである．グラフは概ね右下がりの直線になっている．実際，図 1 のデータで (1) 式についての回帰を行った場合，推定値が -1 に非常に近い．都市経済学におけるランクサイズ回帰では，長い間通常の t 検定によって $\alpha_1 = -1$ という帰無仮説が調べられてきたが，通常 t 値が非常に大きく，α_1 の推定値は -1 にかなり近いにも関わらず帰無仮説が棄却されるという実証研究が非常に多かった．この点に関して，数多くはないものの，先行研究が蓄積されてきている．詳しいことは後述するが，近年の研究から，都市のサイズがパレート分布に従っている場合にはランクサイズ回帰の，t 統計量が漸近的に発散することがわかり，当然このような場合は，通常の t 検定の棄却域は使用できない.

他方，Rosen and Resnick(1980) 等では，被説明変数がパレート分布に従っているかの簡便的な検定として，(1) 式の一次の項の二乗項を説明変数として加え，その推定値が 0 であれば，パレート分布に従っているとする分析方法も提案されている．ここでも t 検定が用いられているが，この場合も上と同じ問題を含む．Nishiyama, Osada and Sato(2007) では，パレート性の検定のために，一次の項と二次の項をそれぞれの帰無仮説の下で t 検定できるようにシミュレーションによって棄却域を求めた．しかし，この定式化で，パレート性を調べるのならば，一次項の係数が -1，二次項の係数が 0 という複合仮説が成立しているかを F 検定により調べるのが自然である．ただし，t 検定の際に問題になったように，真の分布がパレート分布であっても，サンプルサイズが大きくなると共に F 値が発散し，F 検定を行うと帰無仮説を棄却しやすくなることが懸念される.

そこで，本稿では，同様に二乗の項を説明変数に考慮して推定を行い，F 値の棄却域をシミュレーションによって構成し，ランクサイズ回帰のパレート性の検定手法として新たに提案

する.

さらに企業の資産規模のデータを用いて，二次項も含めてランクサイズ回帰の実証分析を行う．また，本稿で得られたt値，F値の臨界値を用いてパレート性の検定も行う．その際先行研究で指摘されていない，二次項を含んだ場合の説明変数間の相関の高さに着目した．一次項と二次項の相関係数について，漸近的にどのような挙動をとるのかを調べ新たな事実がわかった．

次節では，先行研究のレビューを行う．3節ではランクサイズ回帰のシミュレーションを行い，t統計量とF統計量の棄却域を構成する．またそれに基づく実証分析を行う．4節では，結論と今後の課題，付録では実証結果で得られた知見より，ランクサイズ回帰の説明変数間の相関係数の挙動について調べている．

2 先行研究

経済学でランクサイズ回帰のあてはまりがよく，古くから応用されている分野に都市経済学がある．これは，都市の人口規模の対数値をその大きさのランクの対数値に回帰すると，定数項がほぼサンプルサイズの対数に等しく，傾きはほぼ -1 に等しくなるというものである．この関係は，都市規模がi.i.d.でパラメータの値が1のパレート分布に従っているときに成立することが知られている．この文脈で様々な国に関して先駆的かつ包括的な分析を行ったのはRosen and Resnick(1980)であり，現在もなおこの分野で必ず引用される文献である．Soo (2005)はそれを更新したデータについて調べている．これらの論文では，最小二乗推定(OLS)によりランクサイズ回帰を行って点推定値を得ている．それと同時に，1. ランクの対数の係数に関するt検定に基づいて，傾きが -1 であるという仮説の検定，および2. 回帰モデルにランクの対数とその2乗項を説明変数に含めて後者の係数がゼロかどうかをt検定で調べることによるパレート性の検定が行われている．しかし，被説明変数は順序統計量であり，その結果，定義上被説明変数は，分散不均一と自己相関をもつ．そのため，古典的な回帰理論を適

用することはできない．それらを考慮して推定量の性質を調べたのが Gabeix and Ioannides (2003)，Gabaix and Ibragimov (2005) などであり，一致性，漸近正規性が証明されている．また，Nishiyama and Osada (2005)，Nishiyama, Osada and Sato(2007) は，OLS よりも有効性のある推定方法として trimmed OLS と GLS 法による推定方法を提案している．

それらの研究において，推定に関しては実は統計的にはあまり大きな問題は生じないことが示されているが，検定に関しては，Nishiyama and Osada (2005)，Nishiyama, Osada and Sato(2007) らが分散不均一と自己相関のために，傾き -1 であるという帰無仮説を標準的な t 検定で調べることはできないことを示している．加えて，帰無仮説の下でも $s^2 = \sum$(回帰残差)$^2/(n -$ 説明変数の数) が漸近的にゼロに収束するという問題が指摘されている．そのため，真の分布がパレート分布であっても，サンプルサイズが大きくなると共に t 値が発散し，t 検定を行うと帰無仮説を棄却しやすくなる．この問題を回避するため，それらの論文では修正した t 検定が提案されている．そこでは，通常の t 値が検定統計量として用いられているが，棄却域はシミュレーションによって構成されている．

Nishiyama, Osada and Sato(2007) では，パレート性の検定のために，1. で述べたように説明変数に二乗の項も含み，一次の項と二次の項をそれぞれの帰無仮説の下で t 検定できるようにシミュレーションによって棄却域を求めた．しかし，この定式化で，パレート性を調べるのならば，一次項の係数が -1，二次項の係数が 0 という複合仮説が成立しているかを F 検定により調べるのが自然である．ただし，t 検定の際に問題になったように，ここでも帰無仮説の下でも $s^2 = \sum$(回帰残差)$^2/(n -$ 説明変数の数) が漸近的にゼロに収束し，真の分布がパレート分布であっても，サンプルサイズが大きくなると共に F 値が発散し，F 検定を行うと帰無仮説を棄却しやすくなることが懸念される．

そこで，本稿では，同様に二乗の項を説明変数に考慮して推定を行い，F 値の棄却域をシミュレーションによって構成し，パレート性の検定の手法として提案する．

これらの提案が，非常に初歩的なことからもわかるようにランクサイズ回帰は，汎用的で長い間多くの分野で利用されてきているが，その統計的性質が明らかになっていないのが現状

である.

3 モンテカルロシミュレーションと実証分析

前節で述べたように，Gabeix and Ioannides (2003)，Gabaix and Ibragimov (2005) で，(2) 式のタイプのランクサイズ回帰の推定値の一致性と漸近正規性が証明された．Nishiyama, Osada and Sato(2007) は，パレート性の検証のために，Rosen and Resnick(1982) や Soo (2005) タイプの (4) 式についてシミュレーションを行い，帰無仮説を一次の項が -1，二次の項が 0 としてそれぞれの t 値の臨界値を求めている．さらに Soo (2005) がアップデートしたデータを用いてパラメータ推定し検定を行っているが，その推定値の挙動や問題点については指摘していない．ここで (4) 式は通常定義されるランクサイズ回帰の逆回帰となっていることに注意されたい．ただし，逆回帰であっても，一次の項に関して (2) 式が有する統計的性質はおそらく変わらないと予想されるが，理論的な検証は今後の課題である．

本節ではまず，パレート性を調べるために，(3)，(4) 式についてモンテカルロシミュレーションを行い，各推定値，t 統計量，回帰残差の二乗和を計算し，シミュレーションベースの t 統計量の臨界値を得る．次に，F 検定によって一次項の係数が -1，二次項の係数が 0 という複合仮説が成立しているかを調べる．その際，F 値の棄却域をシミュレーションによって構成し，パレート性の検定の手法とする．また，実証分析では日経 NEEDS のデータを用い，上場企業の資産についてランクサイズルールが成立しているかを調べる．

3.1 モンテカルロシミュレーション

各シミュレーションでは，パラメータが 1 のパレート分布から $n = 50, 100, 200, 500, 1000, 3000$ のデータを発生させ，(3) 式，(4) 式について OLS を行った．繰り返し計算は 10000 回行っている．

$$\log S_{(i)} = c + \alpha_1 \log i, i = 1, \ldots, n \tag{2}$$

$$\log S_{(i)} = c + \alpha_1 \log i + \alpha_2 \log^2 i, i = 1, \ldots, n \tag{3}$$

$$\log i = c + \beta_1 \log S_{(i)} + \beta_2 \log^2 S_{(i)}, i = 1, \ldots, n \tag{4}$$

表 1 はサンプルサイズごとのシミュレーション結果で，記述統計量である．$\alpha_1, \alpha_2, \beta_1, \beta_2$ は推定値，$\alpha_1 t$ は帰無仮説が $\alpha_1 = -1$，$\beta_1 t$ は帰無仮説が $\beta_1 = -1$ の両側検定，$\alpha_2 t$ は帰無仮説が $\alpha_2 = 0$，$\beta_2 t$ は帰無仮説が $\beta_2 = 0$ の両側検定を行ったときの t 統計量，SSR は回帰の残差二乗和である．

α_1 については，サンプルサイズが 50，500，3000 と大きくなる程，平均値が -1 に近くなり，α_2 は同様にサンプルサイズが大きくなると平均値が 0 に近くなり，両係数とも一致性が成り立っているようにみえる．

図 2 は (3) 式について $n = 50, 500, 3000$ のとき回帰を行い，その推定値の密度関数を描いたものである．左の列は一次項 α_1，右の列は二次項 α_2 の密度関数である．サンプル数が大きくなるほど，一次項は -1，二次項は 0 の周りに推定値が集まっているのがわかる．(4) 式の一次項 β_1, 二次項β_2 も (3) 式の結果と同じ挙動になっている．一方，表 1 より (3) 式，(4) 式通じて，サンプル数が大きくなる程 t 値の平均値は絶対値で大きくなり，範囲（レンジ）は広がっている．これは先行研究で指摘されているように，サンプルサイズが大きくなると t 値が発散傾向にあることと矛盾がない．図 3 は，左の列が一次項 α_1 の帰無仮説 $\alpha_1 = -1$ の下での t 統計量の密度関数，右の列が二次項 α_2 の帰無仮説 $\alpha_2 = 0$ の下での t 統計量の密度関数である．サンプルサイズが大きくなるほど，分散が大きくなっている．Nishiyama, Osada and Sato(2007) で指摘されているように，t 統計量の分母を構成する s^2 が，n が大きくなる程 0 に近づいていることが関係していると考えられる．

=表 1 挿入=

=図 2 挿入=

=図 3 挿入=

このようにパレート分布から発生させたデータを用いても，推定値に関して t 検定を行うと，ランクサイズルールが成立していても $\alpha = -1$ や $\alpha = 0$ を棄却しやすくなってしまう．そのため，シミュレーションによって得られた棄却域を用いることが解決策の一つになる．

表 2 は，(3) 式，(4) 式のランクサイズ回帰において，一次項，二次項の係数の有意性を t 検定で検証するためのシミュレーションで得られた棄却域である．α_1 は -1，α_2 は 0 が帰無仮説であり，各サンプルサイズに対する両側 1%, 5%, 10% 水準の臨界値である．

サンプル数が多くなるほど，臨界値の絶対値は大きくなり，棄却域が狭くなっている．t 検定は，t 値が絶対値で 2 より大きければ，当該係数に関して帰無仮説が棄却される．シミュレーションで得られた臨界値は，2 よりかなり大きく，サンプル数の増加に伴い大きくなっており，通常の t 統計量の臨界値を用いるより帰無仮説を棄却しにくくなっている．以上より，順回帰 ((3) 式) であっても逆回帰 ((4) 式) であっても，t 検定を行う場合は通常の棄却域は用いることが適切でないことわかった．

=表 2 挿入=

前述したが，(3) 式と (4) 式で，パラメータ 1 のパレート性の成立を調べるのならば，一次項の係数が -1，二次項の係数が 0 という複合仮説が成立しているかを調べるのが自然である．

ただし，t 検定の際に問題になったように，$s^2 = \sum(\text{回帰残差})^2/(n - \text{説明変数の数})$ が漸近的にゼロに収束し，真の分布がパレート分布であっても，サンプルサイズが大きくなると共に F 値が発散し，F 検定を行うと帰無仮説を棄却しやすくなることが懸念される．

図 4 は (3) 式について，シミュレーションで発生させたデータを用いて帰無仮説 $\alpha_1 = -1$, $\alpha_2 = 0$，(4) 式について帰無仮説 $\beta_1 = -1, \beta_2 = 0$ の下で計算した F 統計量の密度関数である．n が大きくなるほど，F 値が大きくなっていることがわかる．表 3 は F 統計量の棄却域である．

通常の F 統計量は，制約の数が等しい場合，サンプルサイズが大きくなるほど，臨界値が

小さくなる．(3) 式と (4) 式で,，パラメータ 1 のパレート性の成立を調べる場合，通常の F 統計量の臨界値は $n = 50$ 以上だとおおよそ 3 である．しかし，表 3 の結果では，t 統計量と同じように，サンプル数が大きくなる程，臨界値が大きくなっており，通常の F 検定を行うと，帰無仮説 $\alpha_1 = -1, \alpha_2 = 0$，帰無仮説 $\beta_1 = -1, \beta_2 = 0$ を棄却しやすくなってしまう．よって，順回帰 ((3) 式) であっても逆回帰 ((4) 式) であっても，F 検定を行う場合は通常の棄却域は用いることが適切でないことわかった．以上より，本稿では，シミュレーションで得た F 値の棄却域をパレート性の検定の手法として提案する．

=表 3 挿入=

=図 4 挿入=

3.2 実証研究

ここでは，日経 NEEDS の 2006 年の上場企業の資産のデータを用いてパラメータ 1 のパレート性の検定を行う．サンプル数は 1736 社である．図 5 は Y 軸が企業の資産総額の対数値で，X 軸は当該企業の順位の対数値である．概ね線形で傾きも -1 に近く何らかのべき乗関数に従っているように見える．しかし，順位の低いところにデータが集中しそこが非線形になっている．そのため，パラメータ 1 のパレート分布に従っているかは，二次項まで含めたランクサイズ回帰を行い，その係数の推定値が 0 になるかを調べることが望ましい．表 4 の左側は (2)〜(4) 式の推定結果である．表の中央は通常の t 検定，F 検定を行った場合の検定結果を示している．通常の，t 検定，F 検定の臨界値を用いると各式において帰無仮説を棄却してしまう．一方右側は，本稿で得られたシミュレーションベースの棄却域による検定結果である．これを用いた場合，(2) 式，(3) 式では，パレート性を棄却しなかった．

=表 4 挿入=

=図 5 挿入=

ここで，(2) 式と (3) 式の推定値に着目したい．パラメータ 1 のパレート分布に従う場合，対数をとって回帰モデルの形にすると，一次項のパラメータが -1 の (2) 式の形になる．(2) 式の推定値は -1.05 で t 検定でも棄却されなかった．そこでさらに，二次項を加えその係数が 0 になるか否かを確認したのが (3) 式である．検定結果では，F 検定でパレート性は棄却されなかった．t 検定においても，一次項，二次項それぞれの帰無仮説も棄却しなかったが，α_1 は 0.4 で，符号が逆となり，値が大きく異なった．このような症状の代表的な要因として多重共線性が考えられる．特にランクサイズ回帰では，$\log$ (ランク) と $\log^2$ (ランク) を説明変数に含むため多重共線性が起きやすい．実際サンプル 1000 では，説明変数間の相関は 0.989 と非常に 1 に近い．

本稿の実証分析でも，二次の項を加えたことで推定値が大きく変わった．また指摘はされていないが，Nishiyama, Osada and Sato(2007) らの都市規模に関するランクサイズ回帰でも，多重共線性がおきているようにみえるものとそうでないものがある．

このことについて，本稿では，$\log$ (ランク) と $\log^2$ (ランク) の相関係数の挙動を調べた．詳しい導出は付録を参照されたい．ランクサイズ回帰においては，n が大きいところでは説明変数間の相関係数が 1 になることがわかった．このような状況下では，通常推定自体が困難となることが多い．しかし，前小節のシミュレーション結果においては，二次の項を含んだ回帰でも，各推定値は漸近的に一致性があるように見受けられる．このような特殊な状況下で推定に際して何が起こっているのかに関する理論的な検討は今後の課題とする．

4 おわりに

本稿では，古くから様々な分野で応用されてきたランクサイズ回帰のパレート性の検定手法を提案した．多くの実証研究では都市サイズ，企業の資産や売上高の規模などの研究対象がパレート性を持つことを，ランクサイズ回帰を行うことで観察してきた．具体的には，規模の対数値にその順位の対数値を回帰することによって，その係数が -1 になることを調べる．ま

た，パレート性の有無には，二次項の係数が 0 であることも条件になるため，本稿では，二次項を加えて推定を行った．パレート性の検証には，一次項，二次項それぞれの t 検定と，一次項の係数が -1，二次項の係数が 0 という複合仮説が成立しているかを F 検定で調べる方法がある．しかし，分析対象がパレート分布に従う時，サンプルサイズが大きくなると，t 値は発散してしまうため通常の t 検定を行うことはできないことがわかっており，F 検定でも同様の問題が観察された．

そのため本稿では，F 値の棄却域をシミュレーションによって構成し，一次項の係数が -1，二次項の係数が 0 という複合仮説を検証可能とし，パレート性の検定の新たな手法として提案した．これが本稿の主な貢献である．

実証研究では，本稿で得られた t 値，F 値の臨界値を用いて二次項も含めてランクサイズ回帰を行ったところ，検定によりパレート性があることが検証されたが，二次項を含んだ場合に多重共線性の症状がみられた．このことより，一次項と二次項の相関係数について調べ，漸近的に相関係数が 1 となることが明らかになり，このことが，実証結果で推定値の大きさに影響を与えている可能性があることがわかった．

今後の課題は，二次項と F 統計量に関する統計的性質を明らかにすること，一次項と二次項の相関係数が漸近的に 1 になり完全な多重共線性を起こすことが推定値や検定統計量にどのような影響を与えるのかを調べることである．

付録　説明変数間の相関について

以下では n が大きくなったときの相関係数挙動を調べる．証明には以下のレンマを用いる．各レンマの証明は，Nishiyama and Osada (2005)，Nishiyama, Osada and Sato(2007) で示されている．

- (a) $\sum \log i = n \log n - n + \frac{1}{2} \log n + O(1)$
- (b) $\sum \log^2 i = n \log^2 n - 2n \log n + 2n + \frac{1}{2} \log^2 n + O(\log n)$

- (c) $\sum \log i(\frac{1}{n}+\ldots+\frac{1}{i}) = n\log n - 2n + \frac{1}{4}\log^2 n + O(\log n)$
- (d) $\sum \frac{\log i}{i} = \frac{\log^2 n}{2} + o(\log^2 n)$
- (e) $\sum \frac{\log^2 i}{i} = \frac{\log^3 n}{3} + o(\log^3 n)$
- (f) $\sum \frac{\log^2 i}{i^2} = O(1)$

レンマ 1

$n \to \infty$ のとき $\log i$ と $\log^2 i$ の相関係数は 1 になる.

証明

$\log i = [\log 1, \log 2, \ldots, \log n]'$ と $\log^2 i = [\log^2 1, \log^2 2, \ldots, \log^2 n]'$ の相関係数は以下で定義される.

$$\frac{\frac{1}{n}\sum_{i=1}^{n}\log^3 i - \left(\frac{1}{n}\sum_{i=1}^{n}\log i\right)\left(\frac{1}{n}\sum_{i=1}^{n}\log^2 i\right)}{\sqrt{\frac{1}{n}\sum_{i=1}^{n}\log^2 i - \left(\frac{1}{n}\sum_{i=1}^{n}\log i\right)^2}\sqrt{\frac{1}{n}\sum_{i=1}^{n}\log^4 i - \left(\frac{1}{n}\sum_{i=1}^{n}\log^2 i\right)^2}}$$

まず分子の第 1 項から計算する. レンマ (a),(b),(e) 使用

$$\begin{aligned}
\sum_{i=1}^{n}\log^3 i &= n\log^3 n - \sum i\{\log^3(i+1) - \log^3 i\} \\
&= n\log^3 n - \sum i\left[\left(\log i + \log\left(1+\frac{1}{i}\right)\right)^3 - \log^3 i\right] \\
&= n\log^3 n - \sum i\left[3\log^2 i\left(\log\left(1+\frac{1}{i}\right)\right) + 3\log i\left(\log^2\left(1+\frac{1}{i}\right)\right) + \log^3\left(1+\frac{1}{i}\right)\right] \\
&= n\log^3 n - \left[3\sum\left(\log^2 i - \frac{\log^2 i}{2i} + \frac{\log^3 i}{3i^2}\right) + 3\sum\left(\frac{\log i}{i} - \frac{\log i}{i^2}\right) + \sum\left(\frac{1}{i^2} + \frac{3}{2i^3}\right)\right] \\
&= n\log^3 n - \left[3\left(n\log^2 n - 2n\log n + 2n\right) - \frac{3}{2}\left(\frac{\log^3 n}{3}\right) + 3\left(\frac{\log^2}{2}\right)\right] \\
&= n\log^3 n - 3n\log^2 n + 6n\log n - 6n
\end{aligned}$$

よって, 分子第 1 項は以下のようになる.

$$\frac{1}{n}\sum_{i=1}^{n}\log^3 i = \log^3 n - 3\log^2 n + 6\log n - 6$$

分子の第2項は，レンマ (a),(b) より

$$\left(\frac{1}{n}\sum\log i\right)\left(\frac{1}{n}\sum\log^2 i\right) \approx \frac{1}{n}\left(n\log n - n + \frac{1}{2}\log n\right)\frac{1}{n}\left(n\log^2 n - 2n\log n + 2n\right)$$

$$\approx (\log n - 1)\left(\log^2 n - 2\log n + 2\right) = \log^3 n - 3\log^2 n + 4\log n - 2$$

となる．よって分子は，$2\log n - 4$ である．レンマにより，分子は以下の近似が成立する．

$$\frac{1}{n}\sum_{i=1}^{n}\log^3 i - \left(\frac{1}{n}\sum_{i=1}^{n}\log i\right)\left(\frac{1}{n}\sum_{i=1}^{n}\log^2 i\right)$$

$$= \log^3 n - 3\log^2 n + 6\log n - 6 + o(1)$$

$$-\log n - 1 + o(1)\log^2 n - 2\log n + 2 + o(1)$$

$$= 4\log^2 n - 16\log n + 20 + o(1).$$

次に，分母 $\frac{1}{n}\sum_{i=1}^{n}\log^2 i - \left(\frac{1}{n}\sum_{i=1}^{n}\log i\right)^2$ と $\frac{1}{n}\sum_{i=1}^{n}\log^4 i - \left(\frac{1}{n}\sum_{i=1}^{n}\log^2 i\right)^2$ について考える．
レンマ (b) より，

$$\frac{1}{n}\sum_{i=1}^{n}\log^2 i - \left(\frac{1}{n}\sum_{i=1}^{n}\log i\right)^2$$

$$= \log^2 n - 2\log n + 2 + \tfrac{\log^2 n}{2n} - \tfrac{\log n}{n} + O(n^{-1}) - \left[\log n - 1 + \tfrac{\log n}{2n} + O(n^{-1})^2\right]$$

$$= 1 - \frac{\log^2 n}{2n} + O(n^{-1})$$

$\frac{1}{n}\sum_{i=1}^{n}\log^3 i$, の近似で用いたのと同様の変形により，$\sum_{i=1}^{n}\log^4 i = n\log^4 n - 4n\log^3 n - 3n\log^2 n + 6n\log n - 6n + O(\log n)$ であることがわかる．従って，レンマ (b) を用いると