The Validity of Stata at Microeconometrics:
The Case of Wage Regression of Japanese Long-term Care Workers

Abstract

This paper discusses the validity of Stata for microeconometrics. Stata is command-driven software that is often used often for econometrics; however, analytical methods for econometrics are limited. Thus, we use factor analysis to determine the wage regression of Japanese long-term care workers using, data from an established annual survey. Such a method is not often taught at educational institutions for use with microeconometrics. However, we apply the method by using *factor* commands. The results show that our model is more suitable than those without factors. Thus, we suggest that other valid methods can be employed more frequently with microeconometrics.

Keywords: microeconometrics; Stata; long-term care workers; wage regression; factor analysis; *factor*
JEL Classification: C87; I11; J31

1. Introduction

Recently in Japan, evidence-based policy has been given greater emphasis. This trend implies that the importance of quantitative policy analysis has been increasing. Thus, it seems that the importance of econometrics has been growing.

Economics is a field that substantiates economic theories. Hence, microeconometrics emphasizes "causal relationships". Thus, in recent microeconometric investigations, the difference-in-difference technique, which uses panel data, and propensity score matching estimation, which is the comparison of the effects of policy on actors who have the same characteristics, have become increasingly important.

However, econometrics needs software. In this regard, we have been able to use numerous software packages for analysis. The most popular of such software has probably been Fortran. However, for contemporary microeconometric studies, the most frequently used software is Stata[1]. Indeed, many countries and educational institutions employ it.

The purpose of this current study is to consider the Stata's validity for

[1] See Cox (2001) for a comparison of Fortran and Stata.

microeconometrics. Further, by using the command-driven nature of Stata, we consider whether, and how, we can improve an economic model's accuracy. In order to achieve this goal, we regress the wage equation of Japanese long-term care workers.

The main result is that by using factor analysis based on worker's motivations, we establish that the equation that includes factors is more accurate than the equation without factors. Thus, by using factor analysis, it seems that we can improve an economic model's accuracy.

The rest of this paper is as follows. Section 2 describes Stata. In section 3, we discuss the theory of Japanese long-term care workers and an identification strategy that empirically supports the theory. We also provide detail about the data. Section 4 presents the results, and section 5 is the conclusion.

2. Stata

Stata is command-driven software that is the most frequently used for microeconometrics. It was invented by the Stata Corporation in 1985. Since then, it has been frequently updated. The current version is Stata 15. We can obtain Stata by purchasing a license.

We can use Stata at many universities; indeed, lectures are held using Stata. Moreover, Stata is used not only for econometrics but also for medical science and social epidemiology. There are also academic publications *The Stata Journal* and *Stata Technical Bulletin*.

With regard to microeconometrics, Stata is used for the least squares, maximum likelihood, and instrument variable estimation methods. The least squares method has the command *reg*, the maximum likelihood method has the commands *probit* and *logit*, and the instrument variable estimation method has the command *ivreg*. These estimation methods are frequently used for microeconometrics[2].

However, factor analysis, cluster analysis, the analysis of variances, and Poisson regression are more frequently used than the aforementioned methods for medical science and social epidemiology. Factor analysis has the command *factor*, cluster analysis has the command *cluster*, the analysis of variances has the command *anova*, and Poisson regression has the command *poisson*.

The commands that are used for medical science and social epidemiology are not often used for microeconometrics. Even so, a few microeconometrics studies have used these

---

[2] See Cox et al. (2010) for an exampleof the methods' use for geography. Stata's graphics are also useful for many analytical techniques (Cox: 2004).

methods. Thus, it seems that the importance of such methods for microeconometrics will increase.

3. Long-term Care Workers in Japan

In Japan, the demand for long-term care is increasing. The reason the aging population. However, in Japan, the insufficient supply of long-term care is a serious problem. The cause is the reducing number of care workers. Such a reduction has many reasons. One is the workers' low wages[3].

A well-known study of the wage regression of long-term care workers is that of Zhou (2009)[4]. Based on this study, a great deal of research has analyzed wage regression. Moreover, in this current study we analyze wage regression. However, in addition to wage regression, we undertake factor analysis.

We use data about long-term care workers from the *Fact-Finding Survey on Long-term Care Work, 2013*. These data are collected every year by the Care Work Foundation for the Japanese Ministry of Health, Labor, and Welfare. The sample of offices used for the data is chosen randomly by the Care Work Foundation. The sample of workers is chosen by each office. The workers' answers are then directly returned to the Care Work Foundation and not through the offices.

We obtained the data from the Center for Social Research and Data Archives, The Institute of Social Science, Tokyo University. On December 22, 2016, we applied to The Institute of Social Science to use the data; we then downloaded the data that day. The application number of the data is 12656.

In this study, we use factor analysis based on workers' motivations to obtain jobs. In economics, the main incentive of workers is generally money. However, in behavioral economics, intrinsic and social motivations are also important incentives for workers. Thus, in this study, we analyze the detail of workers' motivations.

We regress wages with the following equation.

$$Ln(Wage\ of\ month)_i = \beta_K Motivations_i + \gamma Control\ Variables_i' + \varepsilon_i$$

---

[3] According to Hotta (2009), the reduction number of employees is caused by increasing stress. In addition, according to Owa (2010), to prevent the reduction in number, it is useful to improve employees' intrinsic motivation. Hanaoka (2009) pointed out that relatively low wage increases the number of those who leave the long-term care industy.

[4] See Kato (2017) for a survey such studies.

The dependent variable of this equation is the *log of the monthly wage of worker* i. The first item on the right-hand side of the equation is worker i's motivations. We determine these motivations by factor analysis. The second item on the right-hand side of the equation is the vector of control variables. The latter is composed of four education dummy variables; the number of years of job tenure; the number of years of experience; the squares of these years; a gender dummy, which has a value of 1 if a worker is female; two job rank dummy variables; a work-style dummy which has a value of 1 if a worker is part-time; and five dummy variables that provide the size of the offices based on the number of employees. The estimation is the least squares method. We use White robustness standard errors.

4. Results

The workers' motivations are follows.

(1) I feel that it is worth doing this job.
(2) This job will be needed in the future.
(3) I want to contribute to society.
(4) I want to participate in society
(5) I like the elderly.
(6) I have experienced family care.
(7) My skills are useful in this job.
(8) I want the knowledge and skills provided by this job.
(9) I want money.
(10) I can work as I wish.
(11) There are no other jobs that I want.
(12) Other reasons.
(13) I have no reason to work.

Figure 1. presents the correlation of each motivation of the workers using Stata analysis. The command is *correlate*.

Figure1. Correlation of Motivations using Stata Analysis

```
. correlate Motivation1 Motivation2 Motivation3 Motivation4 Motivation5 Motivation6 Motivation7 Motivation8 Motivation9 Motivation10
>  Motivation11 Motivation12 Motivation13
(obs=18,881)
```

	Motiv~n1	Motiv~n2	Motiv~n3	Motiva~4	Motiva~5	Motiva~6	Motiva~7	Motiva~8	Motiva~9	Motiv~10	Motiv~11	Motiv~12	Motiv~13
Motivation1	1.0000												
Motivation2	0.1712	1.0000											
Motivation3	0.2898	0.1811	1.0000										
Motivation4	0.2327	0.1303	0.3218	1.0000									
Motivation5	0.2190	0.0659	0.2103	0.1726	1.0000								
Motivation6	0.0125	0.0079	0.0588	0.0635	0.0892	1.0000							
Motivation7	0.0962	0.1105	0.0670	0.1189	0.0303	-0.0283	1.0000						
Motivation8	0.1414	0.1599	0.1387	0.1480	0.1263	0.1171	0.1130	1.0000					
Motivation9	0.0384	0.0469	0.0254	0.0566	0.0077	0.0025	0.1081	0.0552	1.0000				
Motivation10	-0.0237	-0.0166	-0.0147	0.0712	-0.0268	0.0162	0.0982	0.0574	0.1018	1.0000			
Motivation11	-0.1861	0.0038	-0.0989	-0.0534	-0.0934	-0.0377	-0.0398	-0.0571	0.0199	-0.0153	1.0000		
Motivation12	-0.1434	-0.0879	-0.0728	-0.0525	-0.0633	-0.0547	-0.0741	-0.0586	-0.0147	-0.0422	-0.0009	1.0000	
Motivation13	-0.1940	-0.1369	-0.1273	-0.0794	-0.1043	-0.0816	-0.1368	-0.1035	-0.0365	-0.0773	-0.0626	-0.0391	1.0000

The motivations numbered 11, 12, and 13 correlate negatively. The largest absolute numbers of covariance are those of Motivation3 and Motivation4. This relationship implies prosocial motivation and intrinsic motivation[5].

[5] See Besley and Ghatak (2005) regarding "Motivated Agent." This suggests that the compensation of an intrinsically motivated agent is lower than that of a non-intrinsically motivated agent. This hypothesis is based on Perry and Wise (1990), and Benabou and Tirole (2003). With regard to social motivation, see Benabou and Tirole (2006).

We then create factors. The command is *factor*. As we suggested from the results shown in Figure 1, workers' motivations seem to correlate. Thus, we undertake principal component analysis. Further, we assume that there are five factors. Figure 2 presents the results of undertaking factor analysis. The command is *factor*, the option command for principal component analysis is *pcf*, and the factor number is *factors (5)*.

Figure2. Factor Analysiswith the Command *factor* , *pcf* , and *factors (5)*

```
. factor Motivation1 Motivation2 Motivation3 Motivation4 Motivation5 Motivation6 Mot
> ivation7 Motivation8 Motivation9 Motivation10 Motivation11 Motivation12 Motivation
> 13, pcf factors (5)
(obs=18,881)

Factor analysis/correlation                        Number of obs    =     18,881
    Method: principal-component factors            Retained factors =          5
    Rotation: (unrotated)                          Number of params =         55
```

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor1	2.17842	0.93227	0.1676	0.1676
Factor2	1.24616	0.17391	0.0959	0.2634
Factor3	1.07225	0.00881	0.0825	0.3459
Factor4	1.06344	0.04271	0.0818	0.4277
Factor5	1.02073	0.07661	0.0785	0.5062
Factor6	0.94412	0.02336	0.0726	0.5789
Factor7	0.92076	0.02661	0.0708	0.6497
Factor8	0.89415	0.05581	0.0688	0.7185
Factor9	0.83834	0.01391	0.0645	0.7830
Factor10	0.82444	0.07346	0.0634	0.8464
Factor11	0.75098	0.09734	0.0578	0.9041
Factor12	0.65364	0.06105	0.0503	0.9544
Factor13	0.59258	.	0.0456	1.0000

```
LR test: independent vs. saturated:  chi2(78) = 1.3e+04 Prob>chi2 = 0.0000
```

Figure 3 presents the unique variances of the factors.

Figure3. Uniqe Variances of the Factors

```
Factor loadings (pattern matrix) and unique variances
```

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Uniqueness
Motivation1	0.6339	-0.2065	-0.0736	-0.2091	0.0442	0.5044
Motivation2	0.4379	0.1350	0.3034	-0.2978	-0.2952	0.5221
Motivation3	0.6238	-0.2235	0.0539	-0.1192	0.0825	0.5370
Motivation4	0.5739	-0.0338	-0.0721	-0.0297	0.1265	0.6474
Motivation5	0.4733	-0.3084	0.0293	0.1654	0.0977	0.6431
Motivation6	0.1916	-0.0806	0.1289	0.8191	-0.1669	0.2414
Motivation7	0.3272	0.4917	-0.1631	-0.2167	0.0323	0.5765
Motivation8	0.4513	0.1400	0.0472	0.2898	-0.1190	0.6763
Motivation9	0.1534	0.5018	-0.2280	-0.0141	0.0970	0.6631
Motivation10	0.0993	0.5529	-0.3644	0.2797	0.1598	0.4479
Motivation11	-0.2317	0.3546	0.5986	-0.0471	-0.3124	0.3625
Motivation12	-0.2438	-0.0176	0.3177	0.0242	0.8098	0.1830
Motivation13	-0.3849	-0.2960	-0.5258	-0.0742	-0.2607	0.4143

.

Figure 4 presents the rotation method. The command for rotation *rotate*. Further, we use the promax method, for which the option command is *promax*.

Figure4. Promax Rotation with the Commands *rotate* and *promax*

```
. rotate, promax

Factor analysis/correlation                    Number of obs    =     18,881
    Method: principal-component factors        Retained factors =          5
    Rotation: oblique promax (Kaiser off)      Number of params =         55
```

Factor	Variance	Proportion	Rotated factors are correlated
Factor1	2.12131	0.1632	
Factor2	1.34736	0.1036	
Factor3	1.16523	0.0896	
Factor4	1.14014	0.0877	
Factor5	1.06882	0.0822	

```
LR test: independent vs. saturated:  chi2(78) = 1.3e+04 Prob>chi2 = 0.0000
```

Figure 5 presents the rotated factor loadings and unique variances.

Figure5. Ratated Factor Loadings and Unique Variences

Rotated factor loadings (pattern matrix) and unique variances

Variable	Factor1	Factor2	Factor3	Factor4	Factor5	Uniqueness
Motivation1	0.7245	-0.0607	-0.1427	-0.1225	0.0707	0.5044
Motivation2	0.3792	-0.0903	0.4546	-0.1351	0.2289	0.5221
Motivation3	0.7088	-0.1174	-0.0611	-0.0229	-0.0164	0.5370
Motivation4	0.5665	0.1249	-0.1129	0.0138	-0.0203	0.6474
Motivation5	0.5245	-0.1466	-0.1556	0.2340	-0.0511	0.6431
Motivation6	-0.0670	0.0055	0.0464	0.8834	0.0996	0.2414
Motivation7	0.1658	0.5198	0.0856	-0.2388	0.0790	0.5765
Motivation8	0.2440	0.1919	0.1129	0.3627	0.1376	0.6763
Motivation9	-0.0440	0.5841	-0.0104	-0.0888	0.0136	0.6631
Motivation10	-0.1917	0.7548	-0.1594	0.1518	-0.0128	0.4479
Motivation11	-0.3805	-0.1246	0.7828	0.0509	0.0457	0.3625
Motivation12	-0.0338	-0.0308	-0.0008	-0.1223	-0.9015	0.1830
Motivation13	-0.2864	-0.1140	-0.4958	-0.1288	0.3910	0.4143

Figure 6 presents the factor rotation matrix.

Figure6. Factor Rotation Matrix

Factor rotation matrix

	Factor1	Factor2	Factor3	Factor4	Factor5
Factor1	0.9727	0.3743	0.2088	0.2564	0.1862
Factor2	-0.1351	0.8374	0.5287	-0.1359	0.0139
Factor3	0.0822	-0.3385	0.7620	0.1555	-0.3657
Factor4	-0.1371	0.1095	-0.1618	0.9419	-0.1532
Factor5	0.1003	0.1796	-0.2648	-0.0672	-0.8988

Figure 7 presents the prediction of factors. The command for the prediction of factors is *predict* and Factor1, Factor2, et cetera.

Figure7. The Prediction of Factors with the Command *predict*

```
. predict Factor1 Factor2 Factor3 Factor4 Factor5
(regression scoring assumed)

Scoring coefficients (method = regression; based on promax(3) rotated factors)
```

Variable	Factor1	Factor2	Factor3	Factor4	Factor5
Motivation1	0.33111	-0.02036	-0.05882	-0.10161	0.06813
Motivation2	0.21354	-0.01237	0.43675	-0.16349	0.23826
Motivation3	0.33039	-0.05778	-0.00002	-0.00541	-0.02299
Motivation4	0.27067	0.11784	-0.03891	0.02619	-0.03384
Motivation5	0.23528	-0.10096	-0.11521	0.23367	-0.08285
Motivation6	-0.01784	-0.00696	-0.00554	0.78647	0.00054
Motivation7	0.11142	0.42148	0.14866	-0.23284	0.09185
Motivation8	0.14090	0.16561	0.12301	0.30922	0.08713
Motivation9	0.00795	0.45112	0.04257	-0.08858	0.01310
Motivation10	-0.06388	0.56049	-0.09886	0.13572	-0.04200
Motivation11	-0.12063	-0.05025	0.64181	-0.00036	0.06185
Motivation12	-0.00613	-0.00902	-0.01882	-0.01256	-0.84597
Motivation13	-0.19617	-0.15255	-0.45715	-0.13782	0.38338

Factor1 is negatively correlated with Motivation6 and 10 to 13. This factor seems to be a positive action among workers. Factor2 is negatively correlated with Motivation1, 2, 3, 5, 6, 11, 12, and 13. Factor3 is negatively correlated with Motivation1, 3, 4, 5, 6, 10, 12, and 13. Factor4 is negatively correlated with Motivation 1, 2, 3, 7, 9, 11, 12, and 13. Factor5 is negatively correlated with Motivation3, 4, 5, 10, and 12.

We then regress workers' monthly wages. We define the control variables as *experience years, tenure years,* and dummy variables based on the workers' level of education. Each variable relates to the workers' human capital. We define each variable on Stata as follows. *Experience years* is "year_of_experience," *tenure years* is "year_of_tenure," and the six school dummy variables are "care_highschool," "other_highschool," "care_professional," "other_professional," "care_university," and "other_university." We then add five dummy variables based on the number of employees at the offices. We define these variables as "number_of_employee" together with a number from 2 to 6. The number given relates to the number of employees working at the offices. We also add the age of each worker, which we call *age*. In addition, the work-style dummy has a value of 1 if a worker is part-time. We name this dummy variable *non_regular_job*. Then, we add a gender dummy variable named *female*, which has a value of 1 if a worker is female. Further, we add two job rank variables, *manage* and *middle*. Each variable has a value of 1 if a worker is a manager or in middle management. We also add White robustness standard errors, which have an option command of *robust*.

Figure 8 presents the results of wage regression and the command *reg* without factors.

The least squares method has the command *reg*.

Figure8. Wage Regression with Command *reg* and without Factors

```
. reg log_of_wage year_of_experience square_of_experience year_of_tenure square_of_t
> enure care_highschool other_highschool care_professional other_professional care_u
> niversity other_university manage middle number_of_employee_2 number_of_employee_3
>  number_of_employee_4 number_of_employee_5 number_of_employee_6 non_regular_job fe
> male age, robust

Linear regression                               Number of obs     =     16,468
                                                F(20, 16447)      =     347.36
                                                Prob > F          =     0.0000
                                                R-squared         =     0.3566
                                                Root MSE          =     .40352
```

log_of_wage	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
year_of_experie~e	.2098596	.0424618	4.94	0.000	.1266298	.2930894
square_of_exper~e	-.2461163	.0512446	-4.80	0.000	-.3465613	-.1456714
year_of_tenure	.0947581	.0439553	2.16	0.031	.0086009	.1809153
square_of_tenure	-.044643	.0503084	-0.89	0.375	-.1432529	.053967
care_highschool	.0004887	.0253104	0.02	0.985	-.0491224	.0500997
other_highschool	.0558656	.0159361	3.51	0.000	.0246291	.0871022
care_professional	.0653433	.0188591	3.46	0.001	.0283774	.1023093
other_professio~l	.0738088	.0178518	4.13	0.000	.0388172	.1088003
care_university	.0860759	.019492	4.42	0.000	.0478695	.1242824
other_university	.0765294	.0179831	4.26	0.000	.0412806	.1117783
manage	.2423908	.0090538	26.77	0.000	.2246443	.2601373
middle	.1279574	.00691	18.52	0.000	.1144131	.1415017
number_of_emplo~2	-.0227193	.0136244	-1.67	0.095	-.0494245	.0039859
number_of_emplo~3	.0161978	.0126269	1.28	0.200	-.0085523	.0409479
number_of_emplo~4	.0545255	.0128959	4.23	0.000	.0292481	.079803
number_of_emplo~5	.1160911	.0136664	8.49	0.000	.0893034	.1428788
number_of_emplo~6	.1390842	.0146412	9.50	0.000	.110386	.1677825
non_regular_job	-.5182016	.0091257	-56.78	0.000	-.536089	-.5003142
female	-.0762113	.0074924	-10.17	0.000	-.0908973	-.0615253
age	-.0002724	.0003363	-0.81	0.418	-.0009315	.0003867
_cons	12.0838	.0245153	492.91	0.000	12.03575	12.13186

"Year of experience" has positive and statistically significant correlations. The "square of experience year" has negative and statistically significant correlations. These findings mean that the general human capital of workers is diminishing with years of experience. However, "year of tenure" has positive and statistically significant correlations. Further, the "square of tenure" has no statistical significance. These findings mean that the relationship between wages and specialist human capital is linear and that specialist human capital is not diminishing. Thus, in the Japanese long-term care industry, over a long period, specialist human capital is needed more than general human capital. This finding also suggests that the relationship between workers and users is important in the Japanese long-term care industry.

The other variables mostly have statistically significant correlations. For example, all the school dummy variables have positive and statistically significant correlations; further, the largest coefficient is that of "care university." This finding means that universities that run courses on long-term care are providing the necessary practical skills and knowledge.

Figure 9 presents the results of the regression of wages alongside the factors that we determined.

Figure9. Wage Regression with the Command *reg* and with Factors

```
. reg log_of_wage year_of_experience square_of_experience year_of_tenure square_of_t
> enure care_highschool other_highschool care_professional other_professional care_u
> niversity other_university manage middle number_of_employee_2 number_of_employee_3
>  number_of_employee_4 number_of_employee_5 number_of_employee_6 non_regular_job fe
> male age Factor1 Factor2 Factor3 Factor4 Factor5 , robust

Linear regression                               Number of obs     =     16,468
                                                F(25, 16442)      =     292.60
                                                Prob > F          =     0.0000
                                                R-squared         =     0.3687
                                                Root MSE          =     .39977
```

log_of_wage	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
year_of_experie~e	.2108536	.0420034	5.02	0.000	.1285223	.2931849
square_of_exper~e	-.243087	.0507698	-4.79	0.000	-.3426013	-.1435728
year_of_tenure	.0903291	.0435263	2.08	0.038	.0050129	.1756454
square_of_tenure	-.0442036	.0498456	-0.89	0.375	-.1419064	.0534993
care_highschool	-.0044715	.0250686	-0.18	0.858	-.0536086	.0446656
other_highschool	.0498407	.0158366	3.15	0.002	.0187993	.0808822
care_professional	.0547807	.0188088	2.91	0.004	.0179134	.091648
other_professio~l	.0631432	.0177515	3.56	0.000	.0283483	.097938
care_university	.0715003	.0194787	3.67	0.000	.03332	.1096806
other_university	.0636211	.0178854	3.56	0.000	.0285638	.0986785
manage	.2322052	.0090166	25.75	0.000	.2145318	.2498786
middle	.125468	.006868	18.27	0.000	.1120061	.13893
number_of_emplo~2	-.0204431	.0134734	-1.52	0.129	-.0468524	.0059662
number_of_emplo~3	.0182626	.0125118	1.46	0.144	-.0062619	.0427872
number_of_emplo~4	.0568615	.0127721	4.45	0.000	.0318268	.0818963
number_of_emplo~5	.1167979	.0136157	8.58	0.000	.0901097	.1434861
number_of_emplo~6	.1366646	.0145474	9.39	0.000	.1081501	.1651791
non_regular_job	-.4987747	.0091995	-54.22	0.000	-.5168068	-.4807427
female	-.0644648	.0075073	-8.59	0.000	-.0791799	-.0497498
age	4.92e-06	.000335	0.01	0.988	-.0006517	.0006616
Factor1	.0228695	.003217	7.11	0.000	.0165638	.0291751
Factor2	-.0265753	.0037905	-7.01	0.000	-.0340051	-.0191454
Factor3	.0131117	.0030471	4.30	0.000	.007139	.0190843
Factor4	-.0485344	.0033041	-14.69	0.000	-.0550108	-.0420581
Factor5	.0025509	.0033979	0.75	0.453	-.0041094	.0092111
_cons	12.0653	.0244733	493.00	0.000	12.01733	12.11327

Except for Factor5, the factors have statistically significant correlations. However, the sign of these are not homogeneous. Factor1 and Factor3 are positively correlated. However, Factor2 and Factor4 are negatively correlated. These findings suggest that Factor1 and Factor3 increase workers' productivity, while Factor2 and Factor4 decrease productivity.

Further, in Figure 9 the F statistics and coefficients of determination are larger than those of the results without factors. However, the root-mean-square-error is smaller than that of the results without factors. These findings mean that the model is more precise with factors than without factors. Thus, it seems plausible to use factor analysis.

5. Conclusion

The importance of econometrics has been increasing, which suggests that the generation of econometrics is necessary. In econometrics, instrumental variable methods have been emphasized. However, the conditions for using such methods are not realistic enough for analysis[6]. Thus, more useful analytical methods are needed for econometrics.

In this study, we discuss the validity of Stata. We find that the use of factor analysis makes an equation model more suitable than the applying Stata without it. Factor analysis is not often used for microeconometrics; however, we show that we can employ this method with Stata. Moreover, in Stata 15, we use a greater number of variate methods[7]. Stata has also been frequently updated. Thus, the validity of Stata is increasing.

However, we have a number of problems related to this study. The first is the method of determining the factors. We assume that there are five factors. However, this assumption has little basis. Thus, a more plausible assumption is needed. The second problem is that the data may have selection bias. In this regard, the workers who complete the survey questionnaires are chosen by their offices. We need to conduct our analysis with different data. Lastly, the Mincer equation has a difficulty. We regress a simple Mincer equation; however, wages are determined by many factors that we do not describe. Consequently, analysis in greater detail is required.

6. References

Bènabou, R. and Tirole, J. (2003) "Intrinsic and Extrinsic Motivation", *Review of Economic Studies*, Vol. 70, No. 3, pp. 489-520.

Bènabou, R. and Tirole, J. (2006) "Incentive and Prosocial Behavior", *American Economic Review*, Vol. 96, No. 5, pp. 1652-1678.

Besley, T. and Ghatak, M. (2005) "Competition and Incentives with Motivated Agents", *American Economic Review*, Vol. 95, No. 3, pp. 616-636.

Cox, N. J. (2001) "Speaking Stata: How to Repeat Yourself without Going Mad", *Stata*

6 See Heckman (1997).

7 According to the Stata Home Page (2017), we can use latent class analysis and a finite mixture model with Stata 15.

Journal, Vol. 1, No. 1, pp. 86-97.

Cox, N. J. (2004) "Speaking Stata: Graphing Model Diagnostics", *Stata Journal*, Vol. 4, No. 4, pp. 449-475.

Cox, N. J., Mindrescu, M. and Evans, I. S. (2010) "Climatic Implications of Cirque Distribution in the Romanian Carpathians: Palaeowind Directions during Glacial Periods", *Journal of Quaternary Science*, Vol. 25, No. 6, pp. 875-888.

Hanaoka, C. (2009) "Relative Wages and Direct Care Worker Turnover in Public Long-term Care Insurance System in Japan", *Quarterly of Social Security Research*, Vol. 45, No. 3, pp. 269-285. (in Japanese)

Heckman, J. (1997) "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations", *Journal of Human Resources*, Vol. 32, No. 3, pp. 441-462.

Hotta, S. (2009) "Human Resource Management to Reduce Job Stress among Caregivers in Japan's Long-Term Care Insurance Facilities", *Quarterly of Social Security Research*, Vol. 46, No. 2, pp. 150-163. (in Japanese)

Kato, Y. (2017) "The Perspective of Analysis about Long-term Care Workers: Organizations and Intrinsic Motivation, Social Capital", *Econoinformatics*, Vol. 95, pp. 1-12. (in Japanese)

Owa, M. (2010) "Effect of Job Satisfaction on Intention to Stay at the Workplace among Care Workers", *Research Journal of Care and Welfare*, Vol. 17, No. 1, pp. 16-23. (in Japanese)

Perry, J. L. and Wise, L. R. (1990) "The Motivation of Public Service", *Public Administration Review*, Vol. 50, No. 3, pp. 367-373.

Stata Home Page (2017) https://www.lightstone.co.jp/stata/stata15_new.html, 2017/25/12 Access.

Zhou, Y. (2009) "Shortage of Long-Term Care Workers in Japan", *Academic Studies Journal of Health Care and Society*, Vol. 19, No. 2, pp. 151-168. (in Japanese)

Acknowledgements

The author is very grateful to Associate Prof. Suzuki Jun (Kobe University) and Prof. Takayuki Nago (Kobe University). The author is indebted to the Tokyo University Social Science Institute Data Archive Center, through which the data from the 2013 Fact-Finding Survey on Long-term Care Work were obtained. Any remaining errors are the author's responsibility.

This study is supported by the Japan Society for the Promotion of Science (JSPS) (Number 17H02505).