보건사회연구 40(2), 2020, 416-445
Health and Social Welfare Review
http://dx.doi.org/10.15709/hswr.2020.40.2.416

Temporary-Permanent Workers' Wage Gap across the Wage Distribution:

A Simple Comment on the Use of a Linear Probability Model Instead of a Binary Probability Model when using Unconditional Quantile Regression with Individual Fixed-Effects

Choi, Yohan
(Korea Institute for Health and Social Affairs)

In order to estimate the temporary-permanent workers' wage gap which is caused by the difference in contract types across the marginal wage distribution with controlling for individual unobserved heterogeneity, recent studies applied UQR (unconditional quantile regression) with individual FE (fixed-effects). Since, as a dependent variable, UQR uses the RIF (recentered influence function) which has a binary outcome, UQR is a BPM (binary probability model) that has a theoretical interest in the effects on a latent dependent variable which is continuous. Based on the empirical results of Firpo et al. (2009) and the widely held belief that a LPM (linear probability model) well approximates a BPM, subsequent studies have generally used a LPM instead of a BPM. In the case of controlling for individual FE, linear FE regression and logistic FE regression can be used for a LPM and a BPM, respectively. However, these two regressions have a critical difference that individuals having no longitudinal variation in the RIF are excluded in logistic FE regression and not in linear FE regression. In a strict sense, however, individuals having no longitudinal variation in the RIF have to be excluded from the sample because the partial effects of explanatory variables on a latent continuous dependent variable are not non-existent but just not identified in these individuals. By analyzing panel data of South Korea, I find that the inclusion of individuals having no longitudinal variation in the RIF substantially underestimates both the temporary-permanent wage gap and the confidence interval of that, especially at the extreme quantiles. Even so, if the aim of the study is to just use the within-variance and not to strictly control for individual FE based on the binary choice model, it could be possible to include individuals having no longitudinal variation in a dependent variable in the analysis.

Keywords: Temporary Contracts, Wage Gap, Unconditional Quantile Regression, Fixed-Effects, Linear Probability Model

■ 투고일: 2020. 1. 14. ■ 수정일: 2020. 5. 14. ■ 게재확정일: 2020. 5. 19.

Ⅰ. Introduction

In many countries nowadays, temporary employment constitutes a large portion of total wage employment. Since job security is clearly different between temporary and permanent employment largely due to the employment protection legislation, researchers have shown interest in the different economic outcomes of temporary and permanent contracts. Especially, the wage gap between temporary and permanent contracts has been investigated through many studies. Since the temporary-permanent wage gap could reflect various characteristics of labor market, the investigation of this gap could be helpful to understand the structure of labor market segmentation.

However, because theories also suggest that low-skilled temporary workers can experience more severe wage penalty than highly skilled temporary workers, there were worries that the temporary-permanent difference in the conditional mean of wage could hide the genuine figure of the temporary-permanent wage gap. Thus, several studies (Bosio, 2014; Cochrane et al., 2017; Kim and Kim, 2018) tried to examine the temporary-permanent wage gap across the marginal wage distribution by relying on unconditional quantile regression (UQR) developed by the seminal work of Firpo et al. (2009). Nonetheless, these studies have an important limitation in an empirical sense because they do not control for individual unobserved heterogeneity. Prior studies consistently show that the temporary-permanent wage gap could be severely biased when individual fixed-effects are not controlled for (Booth et al., 2002b; Mertens et al., 2007; Lee and Kim, 2009; Lee, 2011).

Therefore, recent studies tried to use UQR with individual fixed-effects in investigating the temporary-permanent wage gap across the wage distribution (Lass and Wooden, 2019; Choi, 2020). However, when using UQR with individual fixed-effects, there arises one empirical issue which has to be considered.

The previous studies which are mentioned above all use linear regression. However, UQR is a binary probability model that has a theoretical interest in the effects on latent dependent variable which is continuous because as a dependent variable UQR uses the recentered influence function (RIF) which has a binary outcome. The reasons that previous studies use a linear probability model instead of a binary probability model at applying UQR are (a) that Firpo et al. (2009, p.965) showed that the results are almost same between linear UQR and logistic UQR and (b) that there exists a widely held belief that a linear probability model well approximates a binary probability model.

And in the case of controlling for individual fixed-effects, linear fixed-effects regression and logistic fixed-effects regression (Andersen, 1970; Chamberlain, 1980) can be used for a linear probability model and a binary probability model, respectively. However, these two regressions have a critical difference that individuals having no longitudinal variation in the RIF are excluded in logistic fixed-effects regression and not in linear fixed-effects regression.

This difference is due to that whereas linear regression assumes a continuous dependent variable, logistic regression assumes that a binary outcome is determined by a latent continuous variable. It is evidently not a new finding but an elementary knowledge of statistics. For this reason, individuals having no longitudinal variation in the RIF have to be excluded from the sample since UQR is a binary probability model and then the partial effects of explanatory variables are not identified in individuals having no longitudinal variation in the RIF. On the other hand, the inclusion of these individuals in the sample means that in these individuals explanatory variables have no effect on a latent continuous variable which determines the RIF. This statement is obviously wrong since the RIF is a function of a continuous dependent variable which is an hourly wage in this study. Intuitively, the inclusion of individuals having no longitudinal variation in the RIF will underestimate the temporary-permanent wage gap, which means that the estimator of temporary-permanent wage gap will shrink toward zero. That is, the inclusion of these individuals in the sample is like assuming that in these individuals there are no difference in the temporary-permanent wage gap.

However, Lass and Wooden (2019) and Choi (2020) did not exclude individuals having no longitudinal variation in the RIF from the sample when using UQR with individual fixed-effects. Also, I find that studies in other fields also did not exclude these individuals from the sample when using UQR with individual fixed-effects as will be discussed in Section Ⅵ.

In order to get empirical evidence about this argument, I estimate the hourly wage gap between temporary and permanent contracts using panel data of South Korea and UQR with individual fixed-effects. Because South Korea has a very large portion of temporary employment among total wage employment compared to other OECD countries, the issues about temporary workers could have important policy implications.

The remainder of this study is constructed as follows. Section Ⅱ and Section Ⅲ provide theoretical and statistical considerations, respectively. Section Ⅳ explains data and Section Ⅴ presents regression results. Finally, I discuss the conclusion of this study in Section Ⅵ.

Ⅱ. Theoretical consideration

There are several explanations that the difference in contract types can result in the difference in wage rates between temporary and permanent contracts. While the compensating wage differential theory (Rosen, 1986) suggests that temporary contracts can have a wage premium than permanent contracts, the wage penalty of temporary contracts are suggested by the buffer stock model (Booth et al., 2002a), the insider-outsider theory (Lindbeck and Snower, 1989, 2001), and the efficiency wage theory (Guell, 2003).

First, the compensating wage differential theory (Rosen, 1986) says that since in competitive labor market employers have to compensate for job insecurity of temporary contracts by higher wages, temporary contracts will have a wage premium than permanent contracts. Second, the buffer stock model (Booth et al., 2002a) argues that, compared to permanent workers, temporary workers are more likely to be used as a buffer stock which can be easily laid off when employers confront with financial difficulties and then temporary workers have a higher possibility to take the residual tasks of firms than permanent workers. This difference in the tasks can eventually result in the difference in wage rates between temporary and permanent workers. Third, the insider-outsider theory (Lindbeck and Snower, 1989, 2001) focuses on the difference in a bargaining power to negotiate wages with employers between temporary and permanent contracts. Compared to temporary workers, permanent workers have a higher bargaining power due to their higher firing costs. So, although permanent workers require to increase their wages above the market-clearing level, employers are hard to replace them with other workers. And the fourth explanation is based on the efficiency wage theory. The efficiency wage theory says that higher wages are helpful to enhance the efforts of workers. However, in the case of temporary workers, employers can insure the efforts of them by using a contract renewal and then temporary workers who want a contract renewal would work hard without a higher wage rate (Guell, 2003).

Empirical results estimated by linear fixed-effects regression show that while the conditional mean of hourly wage is lower in temporary contracts than in permanent contracts in European countries (Booth et al., 2002b; Mertens et al., 2007) and in South Korea (Lee and Kim, 2009; Lee, 2011), that is higher in permanent contracts than in temporary contracts in Australia (Lass and Wooden, 2019).

However, the above theories also suggest that the temporary-permanent wage gap can be different across the marginal wage distribution. First, since the competitive wage theory assumes that temporary workers choose their jobs alternative to permanent jobs not unemployment, this theory could have more explanatory power for high-skilled workers than for low-skilled workers. Also, the other explanations assume that, despite the inferior position, workers having temporary jobs do not easily translate to permanent position and then it leads to new equilibrium. So, contrary to the compensating wage differential theory, these explanations are more appropriate to explain for low-skilled workers than for high-skilled workers.

Several studies investigate the temporary-permanent hourly wage gap across the marginal wage distribution using UQR (Bosio, 2014; Cochrane et al., 2017; Kim and Kim, 2018) showing that the wage penalty of temporary contracts is concentrated largely at the lower wage distribution and this is reduced or converted toward the upper wage distribution. But these studies do not control for individual unobserved heterogeneity although the needs of controlling for individual fixed-effects are sufficiently recognized in the literature. It might be partly due to that scholars were not familiar with how to combine UQR with individual fixed-effects.

More recently, Lass and Wooden (2019) and Choi (2020) estimate the temporary-permanent wage gap across the marginal wage distribution by using UQR with individual fixed-effects. Lass and Wooden analyze for Australia and Choi analyzed for Korea. Choi finds that temporary contracts generate a wage penalty than permanent contracts and this penalty is largely concentrated on low-wage workers.

However, as discussed in the introduction and will be discussed in next section in more detail, their estimates could be substantially underestimated in absolute values because their samples include individuals having no longitudinal variation in the RIF despite that the partial effects of explanatory variables are not identified in these individuals.

Ⅲ. Statistical consideration

The temporary-permanent contracts' wage gap across the marginal wage distribution can be estimated using UQR suggested by Firpo et al. (2009). Firpo et al. (2009) found that through using the RIF of an original dependent variable as a new dependent variable the partial effects of explanatory variables on a functional of the marginal distribution of an original dependent variable can be consistently estimated (see Corollary 1 in Firpo et al., 2009). It is clearly very powerful method because existing regression models can easily be applied by using just the RIF of an original dependent variable as a new dependent variable.

And in the case of a quantile, the RIF will have two values as in equation (1) (Firpo et al., 2009, p.958).

$$(1)\quad RIF(y;q_\tau) = q_\tau + \frac{\tau - 1\{y \le q_\tau\}}{f_Y(q_\tau)},$$

where y is a dependent variable, τ is a specific quantile, q_τ is a value of a τth quantile, $1\{\bullet\}$ is an indicator function having a value of zero or one and $f_Y(\bullet)$ is a density function of a dependent variable. Due to the indicator function, the RIF of the specific quantile has only two values and thus the UQR which uses the RIF as a dependent variable becomes a binary probability model. However, based on the finding of Firpo at el. (2009, p.965) that the results of ordinary least squares are almost same with the results of logistic regression and the widely held belief that a linear probability model well approximates a binary probability model, subsequent studies have used a linear probability model when applying UQR.

However, unobserved factors such as ability and motivation can be highly correlated with contract types. Previous studies consistently showed that the estimators of the temporary-permanent wage gap are highly different between ordinary least squares and linear fixed-effects regression (Booth et al., 2002b; Mertens et al., 2007; Lee and Kim, 2009; Lee, 2011). Therefore, Lass and Wooden (2019) and Choi (2020) analyzed the temporary-permanent wage gap across the wage distribution with controlling for individual fixed-effects. Because UQR is a method just using the RIF of an original dependent variable as a new dependent variable, linear fixed-effects regression or logistic fixed-effects regression can be used. Borgen

(2016) explains practical procedures to apply linear UQR with individual fixed-effects when using Stata and provides a user-written command (xtrifreg).

In the case of fixed-effects regression, however, there exists a critical difference between linear fixed-effects regression and logistic fixed-effects regression. It is that individuals who have no longitudinal variation in the RIF are excluded in logistic fixed-effects regression but not in linear fixed-effects regression. It is because in logistic fixed-effects regression (Andersen, 1970; Chamberlain, 1980) the observations of these individuals do not contribute to the log-likelihood function. On the other hand, linear fixed-effects regression does not exclude individuals who have no longitudinal variation in a dependent variable. Intuitively, this difference can result in substantial difference in estimators.

Above all, this difference is theoretically important because the view that a latent continuous variable determines a binary outcome which is an essential assumption of a probability model is reflected only in logistic fixed-effects regression and not in linear fixed-effects regression. In logistic regression, a binary outcome is generally assumed to be determined by a latent variable as in equation (2).

$$(2)\ \ y_i = \begin{cases} 1 & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* \leq 0 \end{cases}, \qquad y_i^* = x_i\beta + e_i,$$

where y_i^* is a latent variable which is continuous, y_i is a binary outcome, $x_i\beta$ is the a conditional mean of a latent variable and e_i is an error term which is assumed to have the logistic distribution. This is a logistic regression model. If an error term follows the normal distribution, it becomes a probit regression model and the argument of this study is equally applied. Logistic fixed-effects regression (Andersen, 1970; Chamberlain, 1980) uses the number of positive (or negative) responses in each individual as a sufficient statistic to remove individual fixed-effects. And in each individual having no longitudinal variation in a binary outcome, the log-likelihood function conditioned on the sufficient statistic appears to have constant values and they are excluded from the sample. I provide these procedures in Appendix. In equation (2), the most notable thing is that in a binary probability model the theoretical interest is in the effects of explanatory variables on a latent continuous dependent variable that is expressed as β. The simple equation is the theoretical basis of the binary choice model, meaning that the higher the value of a latent variable y_i*, the more probable the positive response of a binary outcome y_i.

Therefore, the theoretical reason why individuals having no longitudinal variation in a binary outcome are excluded from the sample in logistic fixed-effects regression is that the causal effects of explanatory variables on a latent variable are not identified in these individuals. In here, I describe this argument in a very intuitive level. Figure 1 shows the values of a latent variable which determines the binary outcome over time for three imaginary persons. I assume that dependent variable has the value of one (positive response) if the values of a latent variable is over zero and has the value of zero (negative response) if the value of a latent variable is not over zero.

We can see that in all three persons explanatory variables have same variations in a latent variable except for the baseline values mean individual fixed-effects are different across them. I assume that there is only one explanatory variable, time and then, we could infer that time has same effects on the latent variables of three persons. However, differently with person B, we cannot find the longitudinal variation in a binary outcome in person A and C since in each person all latent variables have the same signs. Therefore, while in Person B the binary outcomes of zero and one are both observed and we can identify the causal effects of time on a dependent variable, it is impossible to identify the time effect on a dependent variable in persons A and C. That is, A and B have to be excluded from the sample.

However, linear regression does not rely on a latent variable approach. It is just like saying that a binary outcome itself is a latent variable as in Figure 2. Therefore, in linear regression, person A and C are assumed that their binary outcomes are not affected by time. Regarding the temporary-permanent wage gap, this means that in Person A and C explanatory variables do not have no effects on wage rates, which is evidently wrong.

Figure 1. Values of a latent variable of three imaginary persons: The case of logistic regression

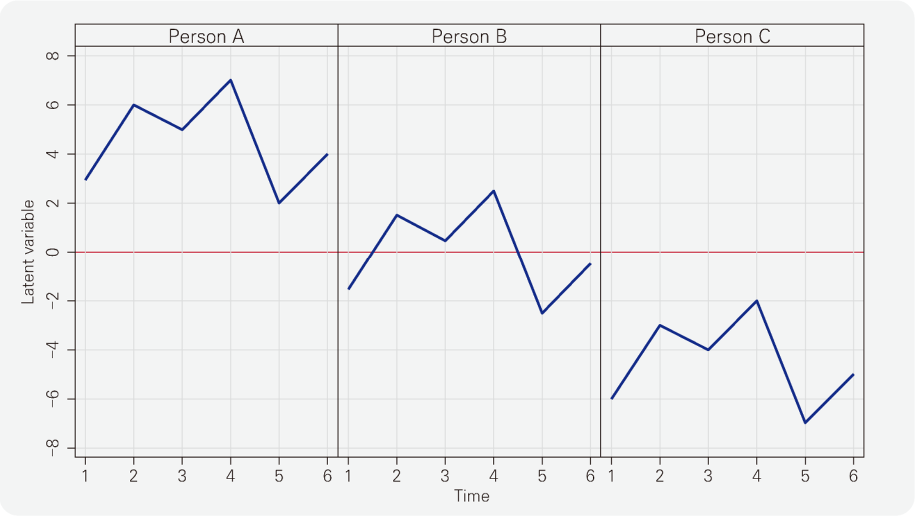
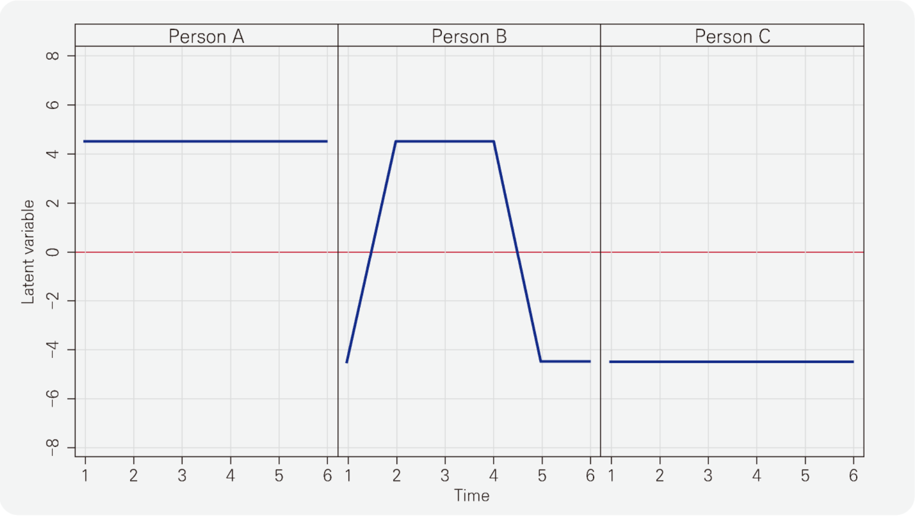


Figure 2. Values of a latent variable of three imaginary persons: The case of linear regression



Therefore, linear fixed-effects regression could never be a good approximation of logistic fixed-effects regression even when ordinary least squares well approximate logistic regression. However, it could be somewhat complex to apply logistic fixed-effects regression in applying UQR because partial effects need to be additionally calculated from the estimated coefficients and because these partial effects rely on unobserved factors which was not estimated and just removed. Calculation of partial effects in non-linear models with individual fixed-effects is possible but difficult. A much simpler way is to just exclude individuals having no longitudinal variation in the RIF from the sample and to analyze this reduced sub-sample by linear fixed-effects regression under the general belief that linear regression approximates logistic regression well at least when using same sample.

Based on the above discussions, we can naturally expect that the inclusion of individuals having no variation in the RIF would result in the underestimation of both the temporary-permanent wage gap. That is, a linear probability model assumes that among individuals having no longitudinal variation in the RIF there is no effect of explanatory variables on a latent dependent variable and this assumption generates the attenuation bias in estimators. Also, due to the reduced sample size, the confidence interval of that would be also underestimated. Especially, these underestimations may be more severe in extreme quantiles since the more extreme quantiles, the more individuals having no longitudinal variation in the RIF.

In order to examine this prediction empirically, I estimate the temporary-permanent wage gap using linear UQR with individual fixed-effects using both the whole sample and the sub-sample excluding individuals who have no longitudinal variation in the RIF and compare these results.

Ⅳ. Data

As a data source, I use the 4th to 20th waves (2001-2017) of the Korean Labor and Income Panel Study (KLIPS). The KLIPS is a nationally representative annual panel survey of households and their members in South Korea. The 1st wave collected 5,000 households and the 12th wave additionally collected 1,415 households. In 20th wave, 67.1 percent of households at the 1st wave and 84.4 percent of households at the 12th wave were surveyed. The KLIPS includes sufficient information about labor market behavior and individual characteristics. Also, due to sufficiently long waves, it is appropriate to apply a fixed-effects approach using the KLIPS. The sample includes men and women aged 20-64 who are not in regular education. And I analyze men and women separately.

Employment contracts are classified into temporary and permanent contracts. Like the major surveys in Korea, the KLIPS also basically classifies wage employment into three categories: (a) permanent contracts or temporary contracts of one year or more, (b) temporary contracts of less than one year and one month or more, and (c) temporary contracts of less than one month and casual contracts. Additionally, the KLIPS surveys the self-reported status of regular or irregular worker. I classify wage workers who respond as being in category (a) and self-reported regular status into permanent workers and the other wage workers into temporary workers.

The value of hourly wage is used as a dependent variable. The hourly wage is calculated through dividing monthly wages by the total weekly working hours and 4.3 and adjusted by the consumer price index from the Bank of Korea. And in line with previous studies, control variables include age and its square term, final educational attainment, marital status, living with children aged 0-18, residential area, industry, occupation, firm sizes, tenure years and its square term, whether a firm having a labor union, labor union membership and year dummies. For readability, I present summary statistics in Appendix.

I will compare the results from using the whole sample and the results from using sub-sample which excludes individuals who have no longitudinal variation in the RIF. Table 1 presents the number of observations in each sample at each quantile. The number of whole sample is, of course, same across all quantiles. On the other hand, the number of the sub-sample is considerably smaller than the number of the whole sample and, as expected, the number of the sub-sample decreases toward extreme quantiles. Large portion of individuals who have no longitudinal variation in the RIF means that the regression results using the whole sample could be

Table 1. The number of observations in each sample at each quantile

	Men					
	Whole sample			Sub-sample		
	Total	Temp	Perm	Total	Temp	Perm
10th quantile	46,051	11,375	34,676	14,412	5,413	8,999
20th	46,051	11,375	34,676	22,975	7,583	15,392
30th	46,051	11,375	34,676	27,103	8,104	18,999
40th	46,051	11,375	34,676	28.833	7,759	21,074
50th	46,051	11,375	34,676	28,525	6,805	21,720
60th	46,051	11,375	34,676	27,303	5,552	21,751
70th	46,051	11,375	34,676	24,029	4,077	19,952
80th	46,051	11,375	34,676	19,208	2,499	16,709
90th	46,051	11,375	34,676	12,187	1,266	10,921
	Women					
	Whole sample			Sub-sample		
	Total	Temp	Perm	Total	Temp	Perm
10th quantile	30,843	12,302	18,541	10,551	5,598	4,953
20th	30,843	12,302	18,541	14,944	7,690	7,254
30th	30,843	12,302	18,541	17,164	8,419	8,745
40th	30,843	12,302	18,541	18,140	8,268	9,872
50th	30,843	12,302	18,541	17,620	7,359	10,261
60th	30,843	12,302	18,541	15,951	5,947	10,004
70th	30,843	12,302	18,541	13,117	4,147	8,970
80th	30,843	12,302	18,541	10,393	2,672	7,721
90th	30,843	12,302	18,541	7,124	1,590	5,534

Data: The 4th to 20th waves of the KLIPS severely underestimated. As discussed in section Ⅲ, because, for these individuals, the partial effects of explanatory variables on the RIF are not existent but just not identified, the inclusion of individuals having no longitudinal variation in the RIF is undesirable and may result in the underestimation of the temporary-permanent wage gap.

V. Results

Table 2 presents the results of UQR with individual fixed-effects using both the whole sample and the sub-sample. And Figure 3 and 4 summarize these results. Through these figures, we can clearly see that, as expected, both the temporary-permanent wage gap and the confidence intervals of that tend to shrink toward zero when not excluding individuals having no longitudinal variation in the RIF compared to when excluding these individuals from the sample. All absolute values of coefficients and standard errors are bigger in the results of the sub-sample than in the results of the whole sample across all quantiles.

And this underestimation appears to be more severe in extreme quantiles which is also expected. Especially, the absolute values of the coefficients of temporary contracts in the sub-sample are over 1.5 times of that in the whole sample at 10, 60, 70 and 90 quantiles in men and at 10, 70 and 90 quantiles in women. Especially in 10th quantile, while the wage penalty of temporary workers is 14.7 percent among men and 9.8 percent among women in the results of the whole sample, that appears 27.9 percent among men and 17.6 percent among women in the results of the sub-sample. Also, the confidence interval of the temporary-permanent wage gap becomes massive at the upper wage distribution when excluding individuals having no longitudinal variation in the RIF. In 90th quantile, the standard errors in the results of the sub-sample are nearly six times larger than that in the results of the whole sample.

Except the statistical discussion of this study, there are several findings which are worthy of notice in results from the sub-sample. First, I cannot find evidence about the protective effects of minimum wage on the severe wage penalty of low-wage temporary workers. These findings are in conflict with the results of Cochrane et al. (2017) and the results of Lass and Wooden (2019) in Australia. Cochrane et al. (2017) and Lass and Wooden (2019) find that the temporary-permanent wage gap tends to be alleviated in the extreme lower wage distribution and they attribute that to the high level of minimum wage.

Second, while women show that the wage penalty of temporary workers seem to decrease consistently toward the upper wage distribution, men show that the wage penalty seems to increase from the median to the upper wage distribution. Also, the results of men are contrasting with the compensating wage differential theory which predicts the wage premium of high-skilled temporary workers. The reverse trend of men was not founded even in the studies for other countries. It seems that permanent contracts predominantly take the essential tasks of firms and fulfill the needs of high skill in male labor market of Korea. Also, because it could mean that there is substantial wage difference between male and female permanent workers, it needs to carefully interpret the low temporary-permanent wage gap in women at the upper wage distribution.

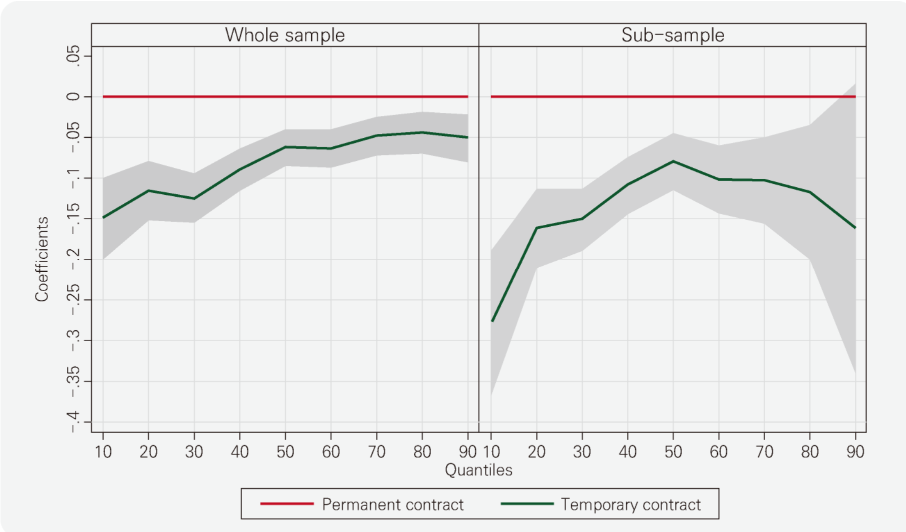
I think that these short reviews are sufficient to achieve the goals of this study and the more detailed discussions of the empirical results are somewhat above the range of this study. In Appendix, to examine whether the above results are due to too many control variables, I additionally present the results by using only age and year dummies as control variables. These results appear to be essentially same.

Table 2. Temporary-permanent hourly wage gap using UQR with individual fixed-effects

	Men					
	Whole sample			Sub-sample		
	Coef.	S.E.	Within R^2	Coef.	S.E.	Within R^2
10th quantile	-0.147***	0.026	0.09	-0.279***	0.046	0.19
2oth	-0.115***	0.019	0.13	-0.161***	0.025	0.19
3oth	-0.125***	0.016	0.15	-0.151***	0.020	0.20
4oth	-0.090***	0.014	0.16	-0.109***	0.018	0.21
5oth	-0.061***	0.012	0.16	-0.079***	0.018	0.21
6oth	-0.063***	0.012	0.16	-0.101***	0.021	0.22
7oth	-0.048***	0.012	0.15	-0.103***	0.027	0.22
8oth	-0.044**	0.013	0.12	-0.116**	0.041	0.20
9oth	-0.050**	0.015	0.08	-0.163+	0.092	0.08
	Women					
	Whole sample			Sub-sample		
	Coef.	S.E.	Within R^2	Coef.	S.E.	Within R^2
10th quantile	-0.098***	0.017	0.07	-0.176***	0.036	0.14
2oth	-0.087***	0.014	0.11	-0.127***	0.020	0.17
3oth	-0.088***	0.012	0.15	-0.116***	0.016	0.21
4oth	-0.084***	0.012	0.14	-0.106***	0.015	0.20
5oth	-0.071***	0.012	0.14	-0.090***	0.017	0.19
6oth	-0.051***	0.013	0.13	-0.074***	0.020	0.19
7oth	-0.036*	0.014	0.11	-0.061*	0.030	0.18
8oth	-0.003	0.014	0.10	0.003	0.045	0.17
9oth	0.015	0.018	0.09	0.077	0.100	0.18

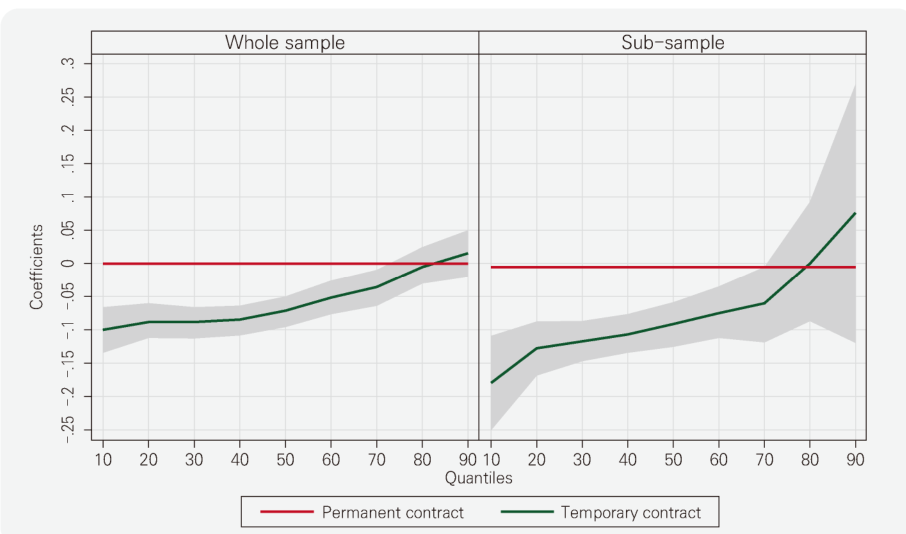
Note: Presented coefficients are the coefficients of contract type (permanent = 0, temporary = 1). Control variables include age and its square term, final educational attainment, marital status, living with children aged 0-18, residential area, industry, occupation, firm sizes, tenure years and its square term, whether a firm having a labor union, labor union membership and year dummies. +p<0.1, *p<0.05, **p<0.01, ***p<0.001.

Figure 3. Temporary-permanent wage gap across the wage distribution for men



Note: Coefficients and the 95% confidence intervals are figured. Individuals having no longitudinal variation in the RIF are excluded in the sub-sample.

Figure 4. Temporary-permanent wage gap across the wage distribution for women



Note: Coefficients and the 95% confidence intervals are figured. Individuals having no longitudinal variation in the RIF are excluded in the sub-sample.

VI. Concluding remarks

Recent studies have widely used UQR to examine the wage difference between temporary and permanent contracts across the marginal wage distribution. Also, more recent studies tried to combine UQR with a fixed-effects approach to control for individual unobserved heterogeneity. Although UQR is a binary probability model which uses the binary RIF as a dependent variable, these studies all have used a linear probability model under the widely held belief that a linear probability model well approximates a binary probability model.

In the case of controlling for individual fixed-effects, however, individuals having no longitudinal variation in the RIF have to be removed from the sample because the partial effects of explanatory variables are not identified in these individuals. It is due to that in a binary probability model the theoretical interest is in the effects of explanatory variable on a latent dependent variable which is continuous. As shown in this study, the inclusion of these individuals can result in the underestimation of the partial effects of explanatory variables. It is obvious because the inclusion of individual having no longitudinal variation in the RIF is equal to assume that there is no temporary-permanent wage gap in these individuals, which is wrong. Therefore, to use linear fixed-effects regression instead of logistic fixed-effects regression when using UQR, we at least have to remove individuals having no longitudinal variation in the RIF. And it is the conclusive recommendation of this study.

This is evidently not a sophisticated argument but a very elementary knowledge of statistics. Nonetheless, Lass and Wooden (2019) and Choi (2020) did not exclude individuals having no longitudinal variation in the RIF and so their estimates may be underestimated. However, it seems to be not a problem of these two studies. Borgen (2016) also do not exclude these individuals. Their whole sample size (p.407) and the analyzed sample size are same (pp.412-413). He discussed the way to combine UQR with a fixed-effects approach and provided the user-written Stata command to use UQR with individual fixed-effects.

Additionally, I find that the studies also in other fields do not exclude clusters which are classified by each fixed effects unit and have no variation in the RIF. It can be easily confirmed by checking that the observations of the analyzed samples are same across quantiles (Apel and Powell, 2019, p.208; Wang and Lien, 2018, p.105; Campbell and Tavani, 2019, p.671; Song, 2018, pp.339-340; Freire and Rudkin, 2019, pp.132-133; Maroto, 2018, p.2276; Ma et al., 2019 p.4774, p.4776; Le, 2019, p.212; Rudkin and Sharma, 2019, pp.13-14). And I found no study excluding observations which have no variation in the RIF in clusters which are classified by each fixed effects unit.

Above all, the argument of this study is applied to all cases in which a linear probability model is used instead of a discrete probability model. However, it seems that there exists a large ignorance of this issue. I find only one study of Beck (2020) which discusses the different sample sizes between linear fixed-effects regression and logistic fixed-effects regression, but he concludes that it is difficult to say which one is right and which one is wrong and recommends the presentation of both results using the whole sample and the sub-sample.[1] On the other hand, this study argued that in each fixed-effects unit observations having no variation in a dependent variable is the source of the underestimation bias and has to be excluded from the sample. Since these two approaches can lead to very different conclusions, the much formal discussions would need to be carried out about this issue which is elementary but important evidently.

However, there remains one thing worthy of notice. In Table 1, we can see that sample sizes are highly different across quantiles, meaning that in the sum-samples the distribution of covariates would be highly different across quantiles. Compared to the sum-samples at high quantiles, the sub-samples at low quantiles are more likely to include those who have low educational level, small firm size, non-union

1) Beck (2020) shows same results with my results which is that the absolute values of coefficients and the standard errors are smaller when using the whole sample than using the sub-sample which excludes clusters not having longitudinal variation in a dependent variable.

membership, and so on (see Table A1-A2). Therefore, if the temporary-permanent wage gap is severely moderated by these factors, it is hard to interpret the estimated distributional figures of the temporary-permanent wage gap as the results of contract types solely. That is, the difference in the temporary-permanent wage gap across the wage distribution could be resulted from the difference in the distribution of covariates and not from the different effects of contract types on wage rates across the wage distribution.

But the aim of investigating the temporary-permanent wage gap across the marginal wage distribution is, precisely, to examine these differences in the moderating effects on the effects of contract types on wage rates. In terms of a latent variable, it can be said that the effects of treatment variables on a latent variable are moderated by the other covariates which also have effects on a latent dependent variable. There are two ways to consider this possibility. One is to model interaction effects, and the other is to use different sub-samples which have different distribution of covariates. Evidently, the use of different sub-samples at different quantiles is exactly same with the second way. Therefore, the use of same sample at different quantiles cannot reflect the theoretical reason to use UQR at all and, therefore, when applying UQR with individual fixed-effects, researchers also have not to use the whole sample which does not exclude individuals having no longitudinal variation in the RIF.

Thus far, I have discussed the issue in a strict sense. But one can want to reflect no effect of explanatory variables on a binary outcome among individuals having no longitudinal variation in a binary outcome. That is, the aim of researchers is to just use the within-variance. In this case, it could be possible to include individuals having no longitudinal variation in a dependent variable. However, because this approach is not based on a probability model having a theoretical interest in the effects on a latent dependent variable and hard to say that it controls for individual fixed-effects based on the binary choice model, researchers should be careful to consider the inclusion of clusters having no variation in a discrete outcome in the analysis when using a fixed-effects model.

최요한은 서울대학교에서 사회복지학 박사과정을 수료하였으며, 현재 한국보건사회연구원에서 전문연구원으로 재직 중이다. 현재는 생애주기에 따른 소득동태를 연구 중이다.
(E-mail: yohann.choi@daum.net)

Appendix

Appendix 1. Logistic fixed-effects regression

Logistic fixed-effects regression (Andersen, 1970; Chamberlain, 1980) uses the number of positive (or negative) responses in each individual as a sufficient statistic to remove individual fixed-effects. When conditioned on the number of positive responses, individual log-likelihood function can be expressed as follows.

$$\text{(A1)}\quad p(Y_i|X_i,c_i,y_{i+}) = \frac{p(Y_i|X_i,c_i)}{p(y_{i+}|X_i,c_i)} = \frac{p(Y_i|X_i,c_i)}{\Sigma_{z:z_+=y_+}p(y_{i+}|X_i,c_i)}$$

$$= \frac{\Pi_{t=1}^{T}p(y_{it}|x_{it},c_i)}{\Sigma_{z:z_+=y_+}\left\{\Pi_{t=1}^{T}p(z|x_{it},c_i)\right\}} = \frac{\Pi_{t=1}^{T}\dfrac{\exp(x_{it}\beta+c_i)^{y_{it}}}{1+\exp(x_{it}\beta+c_i)}}{\Sigma_{z:z_+=y_+}\left\{\dfrac{\exp(x_{it}\beta+c_i)^{z_t}}{1+\exp(x_{it}\beta+c_i)}\right\}}$$

$$= \frac{\Pi_{t=1}^{T}\exp(x_{it}\beta+c_i)^{y_{it}}}{\Sigma_{z:z_+=y_+}\left\{\exp(x_{it}\beta+c_i)^{z_t}\right\}}$$

$$= \frac{\exp(\Sigma_t y_{it}x_{it}\beta)\exp(\Sigma_t y_{it}c_i)}{\Sigma_{z:z_+=y_+}\left\{\exp(\Sigma_t z_{it}x_{it}\beta)\exp(\Sigma_t z_{it}c_i)\right\}}$$

$$= \frac{\exp(\Sigma_t y_{it}x_{it}\beta)\exp(\Sigma_t y_{it}c_i)}{\Sigma_{z:z_+=y_+}\left\{\exp(\Sigma_t z_{it}x_{it}\beta)\exp(\Sigma_t z_{it}c_i)\right\}} = \frac{\exp(\Sigma_t y_{it}x_{it}\beta)}{\Sigma_{z:z_+=y_+}\exp(\Sigma_t z_{it}x_{it}\beta)}$$

$$= p^*(Y_i|X_i,y_{i+}),\quad Y_i=(y_{i1},\dots,y_{iT}),\ X_i=(x_{i1},\dots,x_{iT}),\ y_{i+}=\Sigma_{t=1}^{T}y_{it},$$

$$z=\left\{(z_1,\dots,z_T): z_t\in(0,1),\ z_+=\Sigma_{t=1}^{T}z_t=y_{i+}\right\}.$$

So, individual fixed-effects appear to be removed conditioned on the number of positive responses. And the log-likelihood function of individuals having no variation in a binary outcome has the constant value of one and their observations do not contribute to the total log-likelihood function.

Appendix 2. Sample statistics

Table A1. Summary statistics for men

	Whole sample	Sub-sample								
		10th	20th	30th	40th	50th	60th	70th	80th	90th
Temporary contract	24.7%	37.8%	33.3%	30.3%	27.4%	24.4%	21.0%	17.8%	14.3%	12.5%
Hourly wage	13778 (10370)	9180 (6139)	10317 (6287)	11264 (7967)	12089 (8049)	13123 (8640)	14313 (9181)	15857 (10200)	17864 (12052)	21031 (14650)
Age	41.1 (10.3)	40.6 (11.5)	40.6 (10.9)	40.6 (10.5)	40.5 (10.1)	40.5 (9.8)	40.5 (9.5)	40.8 (9.3)	41.1 (9.1)	42.3 (9.2)
High school	52.7%	70.4%	66.2%	62.4%	58.1%	54.1%	49.3%	44.3%	38.5%	33.6%
College	15.2%	14.2%	15.6%	16.6%	17.2%	17.3%	17.1%	16.8%	16.1%	13.4%
University	32.1%	15.4%	18.2%	20.9%	24.8%	28.6%	33.6%	38.8%	45.5%	53.0%
Married	72.3%	59.0%	63.3%	67.1%	70.2%	73.3%	76.4%	79.8%	82.5%	86.1%
Children in a household	50.1%	38.1%	42.0%	45.9%	49.3%	52.3%	55.6%	59.2%	62.0%	62.5%
Metropolitan areas	44.6%	41.4%	40.9%	42.0%	42.5%	42.6%	44.3%	45.3%	48.1%	49.6%
Major cities	29.4%	32.7%	32.0%	31.5%	31.0%	30.8%	29.4%	28.3%	26.7%	26.9%
Other cities	26.0%	25.9%	27.1%	26.5%	26.5%	26.6%	26.3%	26.4%	25.2%	23.5%
Tenure years	7.2 (7.7)	4.9 (6.1)	5.5 (6.5)	6.0 (6.8)	6.5 (7.1)	7.1 (7.4)	7.6 (7.5)	8.5 (8.0)	9.6 (8.5)	11.1 (9.1)
Labor union in a firm	22.4%	12.9%	14.3%	16.2%	18.2%	20.9%	24.0%	28.4%	33.2%	40.6%
Labor union membership	13.4%	8.7%	9.6%	10.8%	11.7%	13.4%	15.2%	17.6%	20.1%	23.5%
Observations	46,051	14,412	22,975	27,103	28.833	28,525	27,303	24,029	19,208	12,187

Note: Mean (standard deviation) or percentage is presented. Summary statistics of industry, occupation and firm sizes are not presented.

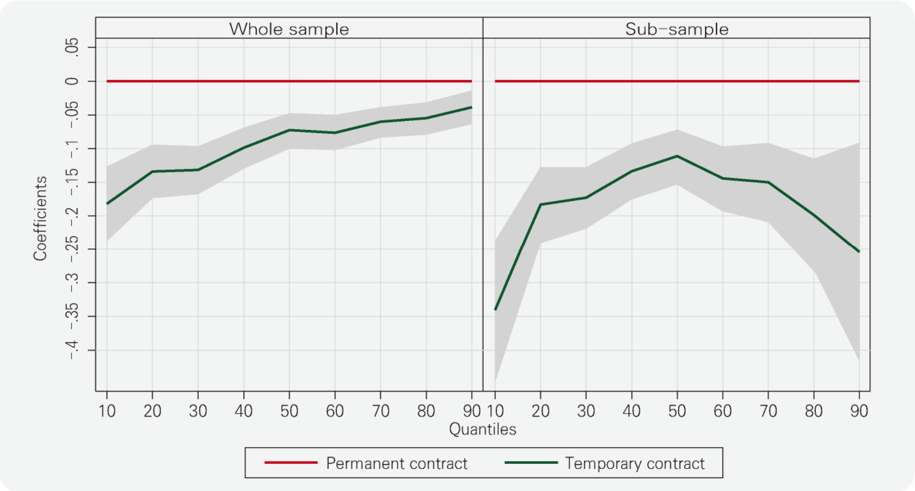
Table A2. Summary statistics for women

	Whole sample	Sub-sample								
		10th	20th	30th	40th	50th	60th	70th	80th	90th
Temporary contract	39.9%	53.0%	51.5%	49.2%	45.9%	42.3%	38.1%	33.0%	27.9%	25.8%
Hourly wage	8988 (6718)	6187 (3873)	6563 (4087)	6943 (4344)	7428 (4577)	8078 (4907)	8929 (5456)	10135 (6237)	11897 (8125)	14121 (9848)
Age	39.8 (11.2)	43.2 (11.5)	42.2 (11.4)	41.2 (11.3)	40.4 (11.2)	39.3 (11.0)	38.2 (10.7)	37.1 (10.4)	36.8 (10.0)	37.9 (9.8)
High school	58.7%	80.4%	77.7%	73.8%	68.9%	62.5%	54.8%	46.0%	37.5%	33.2%
College	17.4%	9.6%	11.8%	14.2%	16.8%	18.6%	21.1%	23.2%	22.1%	19.2%
University	23.9%	10.0%	10.5%	12.1%	14.3%	18.9%	24.1%	30.8%	40.4%	47.6%
Married	62.7%	65.8%	64.2%	62.8%	62.0%	60.1%	58.5%	58.3%	60.4%	67.4%
Children in a household	42.4%	38.3%	39.4%	40.3%	41.4%	41.7%	42.6%	44.2%	47.4%	51.8%
Metropolitan areas	45.2%	38.6%	41.2%	43.3%	43.9%	45.3%	47.0%	46.9%	47.9%	48.2%
Major cities	29.5%	33.7%	32.0%	30.7%	30.6%	30.3%	29.0%	28.5%	28.2%	27.7%
Other cities	25.3%	27.7%	26.8%	26.0%	25.5%	24.4%	24.1%	24.6%	24.0%	24.2%
Tenure years	4.5 (5.5)	3.5 (5.2)	3.5 (4.1)	3.5 (4.1)	3.7 (4.3)	3.8 (4.5)	4.1 (4.7)	4.5 (5.1)	5.4 (5.9)	7.0 (7.4)
Labor union in a firm	14.1%	6.3%	7.1%	7.6%	8.6%	10.3%	12.7%	16.2%	23.5%	30.9%
Labor union membership	6.8%	2.7%	3.3%	3.4%	4.0%	5.0%	6.6%	8.7%	12.7%	16.6%
Observations	30,843	10,551	14,944	17,164	18,140	17,620	15,951	13,117	10,393	7,124

Note: Mean (standard deviation) or percentage is presented. Summary statistics of industry, occupation and firm sizes are not presented.

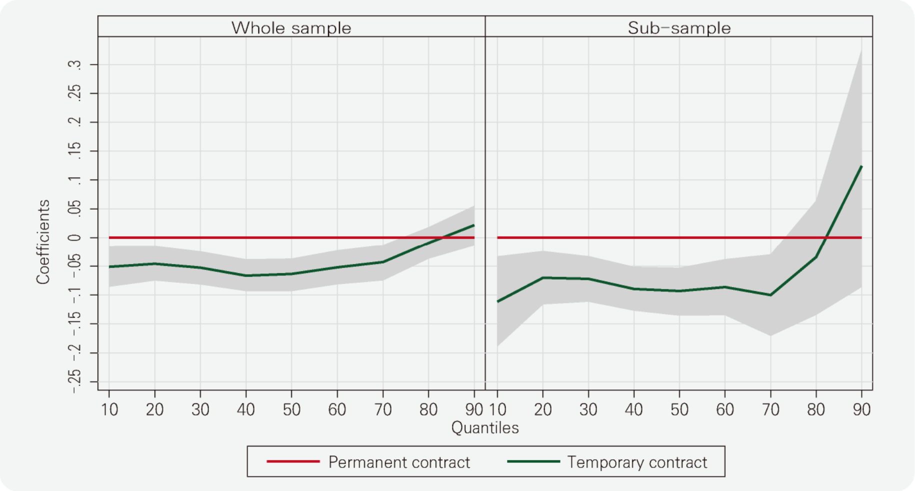
Appendix 3. Additional results

Figure A1. Temporary-permanent wage gap across the wage distribution: Additional results for men



Note: Control variables include only age and its square term and year dummies.

Figure A2. Temporary-permanent wage gap across the wage distribution: Additional results for women



Note: Control variables include only age and its square term and year dummies.

References

Andersen, E. B. (1970). Asymptotic properties of conditional maximum-likelihood estimators. *Journal of the Royal Statistical Society: Series B (Methodological), 32*(2), pp.283-301.

Apel, R., & Powell, K. (2019). Level of Criminal Justice Contact and Early Adult Wage Inequality. *RSF: The Russell Sage Foundation Journal of the Social Sciences, 5*(1), pp.198-222.

Beck, N. (2020). Estimating Grouped Data Models with a Binary-Dependent Variable and Fixed Effects via a Logit versus a Linear Probability Model: The Impact of Dropped Units. *Political Analysis, 28*(1), pp.139-145.

Booth, A. L., Francesconi, M., & Frank, J. (2002a). *Labour as a buffer: do temporary workers suffer? IZA Discussion Paper 673*. Bonn: Institute for the Study of Labor.

Booth, A. L., Francesconi, M., & Frank, J. (2002b). Temporary jobs: stepping stones or dead ends?. *The economic journal, 112*(480), F189-F213.

Borgen, N. T. (2016). Fixed-effects in unconditional quantile regression. *The Stata Journal, 16*(2), pp.403-415.

Bosio, G. (2014). The implications of temporary jobs on the distribution of wages in Italy: An unconditional IVQTE approach. *Labour, 28*(1), pp.64-86.

Campbell, T., & Tavani, D. (2019). Marx-biased technical change and income distribution: A panel data analysis. *Metroeconomica, 70*(4), pp.655-687.

Chamberlain, G. (1980). Analysis of Covariance with Qualitative Data. *The Review of Economic Studies, 47*(1), pp.225-238.

Choi, Y. (2020). Temporary-permanent contracts' wage gap across the wage distribution in South Korea: Using unconditional quantile regression with individual fixed-effects. *Quarterly Journal of Labor Policy, 20*(1), pp.1-28.

Cochrane, B., Pacheco, G., & Chao, L. (2017). Temporary-permanent wage gap: Does type of work and location in distribution matter?. *Australian Journal of*

Labour Economics, 20(2), 125.

Firpo, S., Fortin, N. M.. & Lemieux, T. (2009). Unconditional quantile regressions. *Econometrica, 77*(3), pp.953-973.

Freire, T., & Rudkin, S. (2019). Healthy food diversity and supermarket interventions: Evidence from the Seacroft Intervention Study. *Food Policy, 83*, pp.125-138.

Guell, M. (2003). Fixed-term Contracts and Unemployment: An Efficiency Wage Analysis. Working Papers 812, Princeton University, Department of Economics, Industrial Relations Section.

Kim, Y., & Kim, G. (2018). Wage gap between regular and non-regular workers acorss the wage distribution. *Journal of Vocational Education & Training, 21*(3), pp.167-190.

Lass, I., & Wooden, M. (2019). The structure of the wage gap for temporary workers: Evidence from Australian panel data. *British Journal of Industrial Relations, 57*(3), pp.1-26.

Le, D. (2019). Household wealth and gender gap widening in height: Evidence from Adolescents in Ethiopia, India, Peru, and Vietnam. Economics & Human Biology.

Lee, I. J. (2011). Wage differentials between standard and non-standard workers: Evidence from and establishment-worker matched data. *Korean Journal of Labor Economics, 34*(3), pp.119-139. (Written in Korean)

Lee, I. J., & Kim, T. G. (2009). Wage differentials between standard and non-standard workers: Assessing the effects of labour unions and firm size. *Korean Journal of Labor Economics, 32*(3), pp.1-26. (Written in Korean)

Lindbeck, A., & Snower, D. J. (1989). The insider-outsider theory of employment and unemployment. Cambridge, MA: MIT Press.

Lindbeck, A., & Snower, D. J. (2001). Insiders versus outsiders. *Journal of Economic Perspectives, 15*(1), pp.165-188.

Ma, W., Renwick, A., & Greig, B. (2019). Modelling the heterogeneous effects of stocking rate on dairy production: an application of unconditional quantile regression with fixed effects. Applied Economics, 1-12.

Maroto, M. (2018). Saving, Sharing, or Spending? The Wealth Consequences of Raising Children. *Demography, 55*(6), pp.2257-2282.

Mertens, A., Gash, V., & McGinnity, F. (2007). The cost of flexibility at the margin. Comparing the wage penalty for fixed-term contracts in Germany and Spain using quantile regression. *Labour, 21*(4-5), pp.637-666.

Rosen, S. (1986). The theory of equalizing differences. In O. Ashenfelter and R. Layard (eds.), Handbook of Labor Economics: Volume 1. Amsterdam: North Holland, pp.64-92.

Rudkin, S., & Sharma, A. (2019). Live football and tourism expenditure: match attendance effects in the UK. European Sport Management Quarterly, pp.1-24.

Song, S. (2018). Spending patterns of Chinese parents on children's backpacks support the Trivers-Willard hypothesis: Results based on transaction data from China's largest online retailer. *Evolution and Human Behavior, 39*(3), pp.336-342.

Wang, W., & Lien, D. (2018). Union membership, union coverage and wage dispersion of rural migrants: Evidence from Suzhou industrial sector. *China Economic Review, 49*, pp.96-113.

임금분포에 따른 유기계약근로자와 무기계약근로자 간의 임금격차:

개인의 고정효과를 통제한 무조건부분위회귀모델의 사용 시 이항확률모델 대신 선형확률모델을 사용하는 것에 대하여

최 요 한
(한국보건사회연구원)

한계임금분포에 따른 유기계약근로자(temporary workers)와 무기계약근로자(permanent workers) 간의 고용계약형태의 차이로 인하여 발생하는 임금격차를 개인의 미관측 이질성을 통제하여 추정하기 위하여, 최근의 연구들은 개인고정효과를 통제한 무조건부분위회귀모델(unconditional quantile regression, UQR)을 적용하였다. UQR은 종속변수로서 이항변수인 재중심화영향함수(recentered influence function, RIF)를 사용하므로, UQR은 이론적 관심이 연속변수인 잠재종속변수에 미치는 영향에 있는 이항확률모델이다. Firpo 외(2009)의 실증결과와 선형확률모델이 이항확률모델을 잘 근사한다는 일반적인 믿음에 기초하여, 후속연구들은 이항확률모델 대신 선형확률모델을 사용하여 왔다. 개인의 고정효과를 통제하는 경우에, 선형고정효과회귀모델과 로지스틱고정효과회귀모델이 각각 선형확률모델과 이항확률모델에 대하여 사용될 수 있다. 그러나 이 두 회귀모델은 중요한 차이를 가지는데, 그것은 RIF의 종단적 변량이 없는 개인들이 로지스틱고정효과회귀모델에서는 제외되지만 선형고정효과회귀모델에서는 그렇지 않다는 것이다. 그러나 엄밀하게는, RIF의 종단적 변량이 없는 개인들은 표본에서 제외되어야만 한다. 이는 이 개인들에서는, 설명변수가 연속변수인 잠재종속변수에 미치는 부분효과가 존재하지 않는 것이 아니라 단지 식별되지 않기 때문이다. 한국의 패널자료를 분석함으로써, 본 연구는 RIF의 종단적 변량이 없는 개인들의 포함이 유기계약과 무기계약 간의 임금격차와 이의 신뢰구간을 과소추정하며, 이는 특히 극단적 분위들에서 그러함을 발견하였다. 그렇지만, 만약 연구의 목적이 이항선택모델에 기초하여 개인의 고정효과를 엄밀하게 통제하는 것이 아닌 단지 개인내변량을 추출하는 것에 있다고 한다면, 종속변수에 변량이 없는 개인들을 분석에 포함하는 것은 가능한 선택일 수도 있다.

주요 용어: 유기계약, 임금격차, 무조건부분위회귀모델, 고정효과, 선형확률모델