Journal of Economic Theory and Econometrics, Vol. 28, No. 3, Sep. 2017, 22–35

Estimation of the impact of the statutory labor hours cut on labor earnings in Korea

Hosin Song*

Abstract In this paper we estimate the impact of the statutory labor hours cut in Korea on monthly labor earnings by the regression discontinuity design (RDD) method. The implementation of the statutory labor hours cut policy was sequentially extended based on the number of corporation's employees. The estimation results show that the statutory labor hours cut did not make the workers receiving the treatment better off on average throughout the entire period of 2004-2008 in the sense that it raised monthly labor earnings. However, the policy intervention is found to substantially improve the welfare of workers in the treatment group in 2007 and 2008.

Keywords: cross-validation, regression discontinuity, statutory labor hours cut

JEL codes : C14, C54

*Associate professor, Department of Economics, Ewha Womans University, Seoul, Korea. Email: hsong@ewha.ac.kr. I am grateful to two anonymous referees for their constructive comments. This paper is motivated by a research project with the Bank of Korea in 2014. I am grateful to Kang-hyun Oh at Bank of Korea for helpful discussion about the labor hours cut policy in Korea. The previous version of this paper was presented at the KAPF conference in the spring of 2016, and I am grateful to Seokjin Woo, Jinyoung Kim, Woojin Lee and other participants of the session. I am also grateful to Jieun Lee for her research assistance.

Received May 15, 2017, Revised August 30, 2017, Accepted August 30, 2017

1. INTRODUCTION

Statutory labor hours in Korea were reduced from 44 hours a week to 40 hours a week, starting in 2004.[1] This policy was sequentially implemented. The law required that the policy be implemented first at the biggest corporations, and later at successively smaller corporations. Specifically, in 2004, the statutory labor hours cut policy was imposed on all corporations having at least 1000 employees. In 2005, the policy was extended to all corporations with at least 300 employees. In 2006, it was extended again to all corporations with at least 100 employees. The threshold was reduced to 50 employees in 2007, and to 20 employees in 2008. Finally, it was extended to all corporations in 2011.[2]

The policy intervention decision was thus made by an observable "number of employees". Therefore, the impact of the policy intervention on outcome variables can be estimated by a quasi-experimental regression discontinuity design (RDD) approach. The purpose of this paper is to estimate the impact of the statutory labor hours cut on real wage income via RDD for each year of the period 2004-2008. Under the context of RDD, the treatment is the policy intervention "the statutory labor hours cut", and it is determined by the observable "number of employees of the corporation to which the worker belongs".

RDD has some advantages over the difference-in-difference (DID) model. To apply the DID model, the researcher needs to construct a comparison (or control) group against the treatment group. For the internal validity, the two groups should be similar in terms of unobservables as well as observables. But, it is not easy to find such a comparison group when using the observational data which is mostly relevant in social science study.[3] In contrast to DID, RDD can be considered when the treatment process is well known. Suppose that z_0 is a known threshold of an observable z. If the observable z is as large as z_0, the policy intervention happens, but it does not happen otherwise. The observations around the threshold z_0 are assumed to be similar in terms of unobservables as well as observables.[4] Hence, RDD does not require the construction of a comparison

[1] Strictly speaking, the statutory labor hours cut is restricted to regular labor hours only. It is not related to overtime labor hours.

[2] See Table 1 for details.

[3] One can think of matching ideas to construct the comparison group, but it may not be easy when one is not sure about the determination process of the treatment.

[4] This assumption is natural unless some entities can change their values of z strategically. In this case, such strategic behavior is almost impossible since moving to larger or smaller corporations is not easy for the worker.

group.

There are only a few empirical studies on the effects of a reduction in statutory working hours. Crepon and Kramarz (2002) focus on the statutory labor hours cut in France, while Hunt (1999) focuses on the case in Germany. However, both employ linear models using panel data.

As to the labor hours cut policy in Korea, there are two recent studies about working hours using micro data. Yoo and Lee (2014) constructed panel data using the Labor Force Survey at Establishments. They analyzed the impact of statutory labor hours cut on actual hours worked by allowing for time trend and fixed effect at the level of industry. Kim and Lee (2012) used Surveys on Labor Conditions by Establishment Type. They modified Hunt (1999) since relevant Korean panel data was not available for the direct application. Their model depends on policy intervention dummy and other dummy variables of size, industry and time as well as their interactions.[5] Both depend on a linear regression model. As far as the author knows, this paper is the first to employ a quasi-experimental RDD approach to analyze the effect of the statutory labor hours cut policy. This is possible because of the evident rule in Korea that the imposition of the hours cut was a function of the size of the corporation.

In this paper we aim to estimate the impact of the statutory labor hours cut on monthly labor earnings representing workers' welfare, while Yoo and Lee (2014) and Kim and Lee (2012) focused on the actual hours cut. Unlike both of them, we use individual level micro data instead of establishments level micro data. Specifically, Korean Labor & Income Panel Study (KLIPS) data are used for the estimation.[6] We examine the period of 2004-2008 to consider the analysis using the RDD. In section 2, we formulate the RDD model for this problem, and show the estimation results. In section 3, some concluding remarks are provided.

[5]Hunt (1999) estimated the following model

$$h_{ijst} = \alpha \overline{h_{jst}} + D_j + D_s + D_t + D_j D_s + D_j D_s + D_s D_t + u_{ijst},$$

where h_{ijst} is the hours worked of a worker i of an establishment of size s and industry j at time t, and $\overline{h_{jst}}$ is the statutory labor hours a week. If labor hours cut applies, $\overline{h_{jst}} = 40$ and $\overline{h_{jst}} = 44$ otherwise.

[6]For KLIPS, visit "http://www.kli.re.kr/klips/ko/main/main.jsp".

2. ESTIMATION OF RDD MODEL

2.1. LABOR HOURS CUT IN KOREA

The statutory labor hours cut is the treatment determined by the corporation's number of employees. Therefore, the treatment dummy variable is defined as $d_i = I(z_i \geq z_0)$, where the observable z is the number of employees at the worker's corporation, z_0 is the known threshold, and $I(\cdot)$ is the indicator function. Hence, RDD can be useful to estimate the impact of the hours cut on the outcome variable of our interest. Specifically, we consider monthly labor earnings as the outcome variable since the impact of the hours cut on the labor earnings reflects the change in workers' welfare caused by the policy intervention. Table 1 shows the summary of the enforcement of the statutory labor hours cut in Korea. Note that the threshold changed as the policy expanded year by year. We aim to estimate the impact of the policy intervention on those outcome variables during 2004-2008.

Table 1: The enforcement of the statutory labor hours cut

implementation time	number of firm's employees	threshold
2004. 7	$z \geq 1000$	$z_0 = 1000$
2005. 7	$300 \leq z < 1000$	$z_0 = 300$
2006. 7	$100 \leq z < 300$	$z_0 = 100$
2007. 7	$50 \leq z < 100$	$z_0 = 50$
2008. 7	$20 \leq z < 50$	$z_0 = 20$
2009.1∼ 2010.12	$z < 20$	N.A.

† Some industries including finance and insurance and public social security were exempted from the enforcement. Observations from those industries were excluded from the sample for our analysis.

For the analysis, the following comments need to be mentioned. The statutory labor hours cut began in 2004 and was completed at the end of 2010. The policy has applied to all corporations since 2011. Therefore, RDD can be applied to the period 2004-2008 only. For the year 2011, there can be no control group below the threshold since the threshold is zero.[7] As to the type of workers for the analysis, we consider full-time wage workers only since the statutory labor hours

[7] In that case, it is also hard to identify the policy intervention time since the extension of hours cut to all corporations took two years.

cut is applied to full-time workers only. Finance and insurance and public social security industries were exempted from the enforcement of the labor hours cut. Hence, observations from those industries were excluded from the sample for our analysis. For the policy intervention of 2004-2008, the seventh through the eleventh KLIPS surveys are used. We also adjusted the nominal wage income to the real wage income by using the 2010 consumer price index (CPI).[8]

Each year's data is used to estimate the impact of the hours cut in each single year, while the pooled data is used to estimate the overall effect during the entire period of 2004-2008. The specific explanation as to the implementation of both RDD is provided in the following subsection.

2.2. RDD USING CROSS-VALIDATION BANDWIDTH SELECTION

We can consider RDD as follows.

$$Y_i = c_i + \rho \ d_i \tag{1}$$

where $c_i \equiv Y_{0i}$ and $\rho \equiv Y_{1i} - Y_{0i}$. Y_{0i} is the potential outcome when the individual i does not receive the treatment, while Y_{1i} is the potential outcome when the individual i receives the treatment.

Assumption 1 *The decision of the statutory labor hours cut depends solely on the number of employees of the corporation to which the worker belongs.*

Assumption 1 matters. In practice, there is no certain way to recognize whether the worker's corporation adopted the statutory labor hours cut or not, based on KLIPS data.[9] Nevertheless, we can assume that the hours cut was enforced according to the announced schedule in Table 1 by adopting Assumption 1. By Assumption 1, we can define $E[d_i|Z_i] = g(Z_i)$. For RDD, we need the following additional assumptions about the propensity score $E[d|Z=z] \equiv g(z)$.

Assumption 2

[8] In addition, some specific data process can be described as follows. If regular working hours is missing, drop that observation. However, if overtime is missing, treat it as zero. Then, hours worked can be computed as the sum of valid regular working hours and the overtime. We drop the observation if one of values in labor earnings, number of employees and working hours is missing.

[9] There is no other way to recognize actual policy adoption at the level of workers and corporations.

(i) $E[d_i|Z_i = z] \equiv g(z)$ *is not continuous at the threshold* z_0.

(ii) $d^+ \equiv \lim_{z\to z_0^+} E[d_i|Z_i = z] = 1$, *and* $d^- \equiv \lim_{z\to z_0^-} E[d_i|Z = z] = 0$.

(iii) $E[c_i|z_i = z]$ *is continuous at* $z = z_0$.

Under Assumptions 1 and 2, Theorem 1 in Hahn, Todd and Klaauw (2001) implies the treatment effect is

$$\rho = Y^+ - Y^-, \tag{2}$$

where $Y^+ \equiv \lim_{z\to z_0^+} E[Y_i|Z_i = z]$, and $Y^- \equiv \lim_{z\to z_0^-} E[Y_i|Z_i = z]$. The average treatment effect ρ in the neighborhood of the threshold z_0 is identified by the difference of two limits $Y^+ - Y^-$, which is a function of z. Assumption 2 (ii) follows from Assumption 1. Assumptions 1-2 are features of sharp RD since the treatment depends on the running variable z only, and other observed and unobserved characteristics are symmetric among observations whose running variables are around the threshold. If Assumption 1 is violated, Assumption 2 (ii) should be relaxed.[10]

We suggest selecting a bandwidth based on cross-validation proposed by Imbens and Lemieux (2008) combined with local linear nonparametric regression model proposed by Hahn, Todd and Klaauw (2001).[11] . Let $q_{Z,\delta,l}$ be the δ quantile of the empirical distribution of Z for the subsample with $Z_i < z_0$, and let $q_{Z,\delta,r}$ be the δ quantile of the empirical distribution of Z for the subsample with $Z_i \geq z_0$. Then, the optimal bandwidth selection using the cross-validation criterion is

$$\widehat{h}_{CV}^{\delta,\text{opt}} = \arg\min_h CV_Y^{\delta}(h) \tag{3}$$

where

$$CV_Y^{\delta}(h) = \frac{1}{N_\delta}\sum_{i:q_{Z,\delta,l}\leq Z_i\leq q_{Z,1-\delta,r}} (Y_i - \hat{\mu}(Z_i))^2,\ \ N_\delta = \sum_{i=1}^{N} I(q_{Z,\delta,l} \leq Z_i \leq q_{Z,1-\delta,r}),$$

and

$$\hat{\mu}(z) = \begin{cases} \hat{\alpha}^-(z) & \text{if } z < z_0, \\ \hat{\alpha}^+(z) & \text{if } z \geq z_0, \end{cases} \tag{4}$$

[10]RDD under such relaxed assumption is called Fuzzy RDD. In that case, $\rho = \frac{Y^+ - Y^-}{d^+ - d^{-1}}$ with $d^+ - d^{-1} < 1$. Hahn, Todd and Klaauw (2001) consider such general situation.

[11] It is well-known that the local linear nonparametric regression model does not suffer from the boundary problem less seriously than the usual nonparametric kernel regression model does. See Fan and Gujbels (1996) for a local linear nonparametric regression model.

where

$$(\hat{\alpha}^-(z),\hat{\beta}^-(z)) = \arg\min_{\alpha,\beta} \sum\nolimits_{j:z-h<Z_j<z_0} (Y_j-\alpha-\beta(Z_j-z_0))^2,$$
$$(\hat{\alpha}^+(z),\hat{\beta}^+(z)) = \arg\min_{\alpha,\beta} \sum\nolimits_{j:z_0\le Z_j<z+h} (Y_j-\alpha-\beta(Z_j-z_0))^2.$$

Once the bandwidth $\widehat{h}_{CV}^{\delta,\mathrm{opt}}$ in (3) is selected by the cross-validation, the estimator of ρ in (2) can be obtained as

$$\hat{\rho} = \hat{Y}^+ - \hat{Y}^- = \hat{a}^+(z_0) - \hat{a}^-(z_0), \tag{5}$$

where $\hat{a}^+(z_0)$ and $\hat{a}^-(z_0)$ are defined in the following equations (6) and (7)

$$(\hat{a}^-,\hat{b}^-) = \arg\min_{a,b} \sum\nolimits_{j:z_0-h_{CV}^{\delta,\mathrm{opt}}<Z_j<z_0} (Y_j-a-b(Z_j-z_0))^2, \tag{6}$$

and

$$(\hat{a}^+,\hat{b}^+) = \arg\min_{a,b} \sum\nolimits_{j:z_0\le Z_j<z_0+h_{CV}^{\delta,\mathrm{opt}}} (Y_j-a-b(Z_j-z_0))^2. \tag{7}$$

The estimation is implemented for the pooled data of 2004-2008 as well as for each year. For the estimation of each year, independently implement (3) -(7) using Z, z_0 and labor earnings in each year. For the estimation of the pooled data, normalization is used. That is, $Z_j^* = Z_{jt} - z_{0t}$ is used. Then, the normalized threshold z_0^* becomes zero.[12]

In the estimation, a uniform kernel $K(u) = I(|u| \le 1)/2$ is employed. Hence, the asymptotic variance can be obtained by Hahn, Todd and Klaauw (1999). When using the uniform kernel and $h \propto N^{-\delta}$ for $1/5 < \delta < 2/5$, the asymptotic variance of $\sqrt{Nh}(\hat{\rho}-\rho)$ is

$$V_\rho = \frac{4}{f_Z(z_0)}\left(\sigma_{Y_r}^2 + \sigma_{Y_l}^2\right) \tag{8}$$

where

$$\sigma_{Y_l}^2 = \lim\nolimits_{z\uparrow z_0} Var(Y|Z=z), \text{ and } \sigma_{Y_r}^2 = \lim\nolimits_{z\downarrow z_0} Var(Y|Z=z).$$

Here, we suggest using a plug-in estimator of the asymptotic variance.[13] $\delta = 0.5$ is used in the estimation. The estimator of the variance in (8) is presented in detail in the Appendix.

[12] That is, the pooled data is centered data around zero. In addition, for the convenience of the analysis of pooled data, we dropped serious outliers. Specifically, observations are dropped if $|Z_i^*| > 200$.

[13] In fact, this is simple modification of Imbens and Lemieux (2008).

2.3. ESTIMATION RESULTS

The estimation of RDD is implemented in two ways. First, separate estimations are implemented for each single year. Second, the estimation using pooled observations is conducted for the entire period of 2004-2008. For the former, each bandwidth is independently selected for each year, while the common bandwidth is selected by using normalized data for the pooled data. The cross-validation method via grid search is used for both estimations.

The estimation results are presented in Table 2. It is found that the statutory labor hours cut did not significantly affect the welfare of workers receiving the policy intervention on average throughout the period of 2004-2009. However, the impact is estimated to be significantly positive in 2007 and 2008. For those two years, the magnitude of the estimated impacts respectively corresponds to about 21% and 28% of the average monthly labor earnings of the corresponding year. Therefore, the labor hours cut policy is shown to substantially improve the labor earnings of workers receiving the treatment at least in 2007 and 2008. This finding implies that the increase of hourly labor earnings dominated the decrease of statutory labor hours.[14]

Table 2: Impacts of the statutory labor hours cut

year	2004	2005	2006	2007	2008	$2004 \sim 2008$
threshold	1000	300	100	50	20	0
$\hat{\rho}$	0.29	2.01	5.65	32.73^{**}	47.33^{**}	11.10
standard error	12.48	13.89	15.09	15.96	19.48	9.26
$\widehat{h}_{CV}^{\delta,\mathrm{opt}}$	1000.5	198.5	69.5	44.5	14.5	89.5

† * and ** indicate that $H_0 : \rho = 0$ is rejected against $H_1 : \rho \neq 0$ at the 10% and 5% significance level respectively.

†† For 2004-2008, normalized running variable is used.

[14] Presumably, corporations chose higher productivity at the cost of hiring less employees, which can be hinted by observing that the mean of employee number increased from 2005-2008 as can be seen in the summary statistics in the Appendix. As one anonymous referee pointed out, the overtime earning could work to induce the finding. But, we can not verify it since we can not identify labor earnings from regular hours and those from overtime using KLIPS data.

3. CONCLUDING REMARKS

This paper is the first study to analyse the impact of the labor hours cut on the monthly labor earnings by RDD as far as the author knows. In this study, we estimate the impact of the statutory labor hours cut on the real wage income via RDD. It is found that the statutory labor hours cut significantly improved the welfare of the worker receiving the policy intervention via the increase of real wage income at least in 2007 and 2008 even though such finding can not be obtained for the entire period of 2004-2008.

However, admittedly there are still some limitations in this work. First, it is worrisome that the symmetry in terms of observables and unobservables around the threshold may not be guaranteed since the numbers of observations in the right neighborhood from the threshold are much less than those in the left neighborhood.[15] Moreover, observations are not dense around the threshold when the threshold is large. That may be one reason why the performance of RDD is not so good in years 2004, 2005 and 2006.[16] Second, Assumption 1 may look too strong. If we want to relax Assumption 1, then we have to consider Fuzzy RDD. However, we cannot identify whether a worker's firm received the policy intervention based on KLIPS data. If some micro-data makes the identification possible, we can consider the Fuzzy RDD including parametric RDD in Angrist and Pischke (2009), Lee (2008), and Lee and Card (2008) as well as nonparametric RDD in Hahn, Todd, and Klaauw (2001).

[15] See Table 3 in the Appendix.

[16] See Figures 1-2 for the mean of monthly labor earnings around the threshold.

REFERENCES

Angrist, J. D., and J. Pischke (2009), "Mostly Harmless Econometrics ", *Princeton University Press*

Cameron, B. K., and P. K. Trivedi (2004), "Microeconometrics - methods and applications ", *Cambridge University Press*

Crepon, B., and F. Kramarz (2002), Employed 40 Hours or Not Employed 39: Lessons from the 1982 Mandatory Reduction of the Workweek. *Journal of Political Economy* 110: 1355-1389

Fan, J., and I. Gijbels (1996), "Local Polynomial Modelling and Its Applications ", *Chapmand & Hall*

Hahn, J., P. Todd, and W. van der Klaauw (2001), "Identification and estimation of treatment effects with a regression-discontinuity design", *Econometrica* 69: 201-209

Hahn, J., P. Todd, and W. van der Klaauw (1999), "Evaluating the effect of antidiscrimination law using a regression-discontinuity design", *National Bureau of Economic Research*, Working Paper 7131

Hunt, J. (1999), "Has Work-Sharing Worked in Germany?", *Quarterly Journal of Economies* 114: 117-148

Imbens, G., and T. Lemieux (2008), "Regression discontinuity designs: a guide to practice", *Journal of Econometrics*, 142: 615-635

Lee, D. S. (2008), " Randomized experiments from non-random selection in U.S. House elections", *Journal of Econometrics* 142: 675-697

Lee, D. S., and D. Card (2008), "Regression discontinuity inference with specification error", *Journal of Econometrics* 142: 655-674

김형락, 이정민 (2012), "주 40시간 근무제의 도입이 근로시간, 임금 및 고용에 미치는 영향",*노동경제논집* 35: 83-100

유경준, 이진 (2014), "근로시간 단축의 고용효과 분석:기업규모별 추정을 중심으로",*노동경제논집* 37: 1-28

APPENDIX

PLUG-IN ESTIMATOR FOR THE ASYMPTOTIC VARIANCE IN (8)

Note

$$\hat{V}_\rho = \frac{4}{\hat{f}_Z(z_0)}\left(\hat{\sigma}^2_{Y_r} + \hat{\sigma}^2_{Y_l}\right)$$

Compute $\hat{\varepsilon}_i$ as follows.

$$\hat{\varepsilon}_i = Y_i - \hat{\mu}(Z_i) = Y_i - I(Z_i < z_0)\hat{\alpha}_l - I(Z_i \geq z_0)\hat{\alpha}_r,$$

Then, sample counter-parts of $\sigma^2_{Y_r}$ and $\sigma^2_{Y_l}$ are as follows.

$$\hat{\sigma}^2_{Y_l} = \frac{1}{N_{h_l}}\sum\nolimits_{i:z_0-h\leq Z_i<z_0}^{N} \hat{\varepsilon}^2_i, \text{ and } \hat{\sigma}^2_{Y_r} = \frac{1}{N_{h_r}}\sum\nolimits_{i:z_0\leq Z_i\leq z_0+h}^{N} \hat{\varepsilon}^2_i$$

where $N_{h_l} = \sum_{i=1}^{N} I(z_0 - h \leq Z_i < z_0)$ and $N_{h_r} = \sum_{i=1}^{N} I(z_0 \leq Z_i < z_0 + h)$. Moreover,

$$\hat{f}_Z(z_0) = \frac{N_{h_l} + N_{h_r}}{2Nh} \tag{9}$$

since

$$\hat{f}(z_0) \times h = \frac{N_{h_l}}{N} = \frac{N_{h_r}}{N}.$$

ESTIMATION OF (a^+, a^-) AND OBSERVATIONS AROUND THE THRESHOLD

Table 3: $(\hat{a}^+, \hat{a}^-)$ and observations around the threshold

year	2004	2005	2006	2007	2008	2004 $\sim$ 2008
threshold	1000	300	100	50	20	0
$\hat{a}^+$	0.16	1.21	3.32	28.78	34.55	7.02
$\hat{a}^-$	-0.13	-0.80	-2.33	-3.95	-12.78	-4.09
(N_l, N_r)	(824,32)	(96,34)	(234,92)	(723, 74)	(378,89)	(2359,461)

† N_l indicates the number of observations in the left neighborhood from the threshold.

† N_r indicates the number of observations in the right neighborhood from the threshold.

††† $\delta = 0.5$ is used for each year as well as the pooled data. Some outliers are dropped from the pooled data. Here, outliers are observations showing $|Z^*_i| > 200$.

KLIPS DATA USED FOR THE ESTIMATION DURING 2004-2008

Table 4: Summary Statistics

variable (unit)	monthly labor earnings (10^4 Won/month)	number of employees (persons)	weekly hours worked (average hours)
	2004 (# of obs.=1716)		
mean	148.7	389.1	53.0
median	127	10	50
[min, max]	[10,650]	[1,100000]	[4,100]
	2005 (# of obs.=1699)		
mean	130.4	567.9	52.0
median	110.3	10	50
[min, max]	[17.2,775.3]	[1,57997]	[3,114]
	2006 (# of obs.=1871)		
mean	149.9	502.2	51.5
median	131.2	12.0	48.0
[min, max]	[11.4,1585.3]	[1,50000]	[3,144]
	2007 (# of obs.=2019)		
mean	155.6	436.6	51.3
median	135.4	12.0	48.0
[min, max]	[11.7,1613.7]	[1,85000]	[4,180]
	2008 (# of obs.=1358)		
mean	168.5	242.9	51.0
median	41.8	7.0	48.0
[min, max]	[18.9,2079.5]	[1,33000]	[6,116]

† Monthly labor earnings indicates the monthly average real wage income.

†† Weekly hours worked is the sum of regular hours worked and overtime.

††† Observations having missing values in at least one variable are dropped.

MEAN OF MONTHLY LABOR EARNINGS AROUND THE THRESHOLD

Figure 1: Mean of monthly labor earnings around the threshold of each year

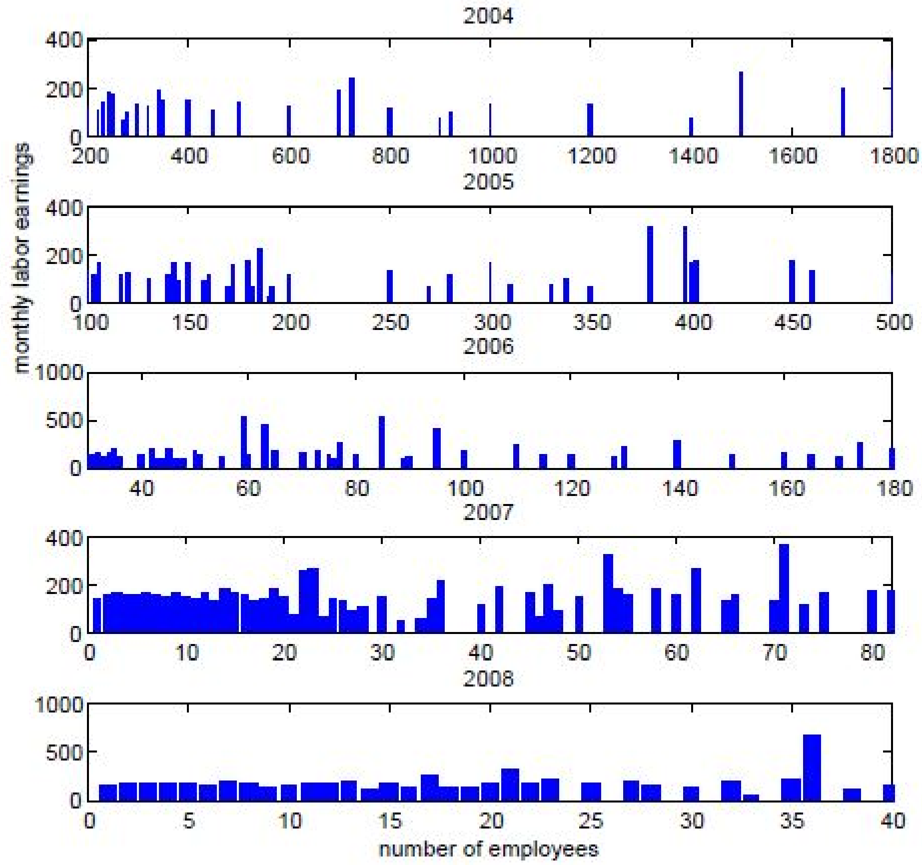

Figure 2: Mean of monthly labor earnings using pooled data