DOES THE HIGHER HOUSE RENT MOTIVATE THE LABOR TO WORK LONGER

Report 2 for Microeconometrics

Li Yuanchao

International Undergraduate Experimental Class

SOE, Xiamen University

May 8, 2022

ABSTRACT

Making decision between working and leisure is a very traditional and important topic in economics. In reality, there exist many factors that influence it. This paper wants to explore whether the higher house rents motivate the labor to decide to work longer per week using USA 2020 population census data. To solve the endogeneity problem, this paper has selected the total family income as an instrumental variable and do the 2sls estimation. In addition, to achieve more understanding about the relationship, this paper also adopts the quantile regression method. The results suggest that higher rent could motivate people to work longer significantly, meanwhile there is no big difference between different quantiles. This paper could help people to know more about the working-leisure decision.

Keywords: working-leisure decision; house rent; instrumental variable; quantile regression

1 Introduction

People always face a choice between working and leisure to allocate their time. Working could generate money to increase utility by consumption while leisure could generate happiness which also leads to the increase on utility. There are many factors that would influence this decision-making process. Economists always use the following model to interpret this question. The optimal allocation is achieved when then marginal rate of substitution equals to the wage rate, that is the tangent point of indifference curve and budget constraint.

We could use this model to analyze many factors that influence the decision-making process. The original point of the budget constrain indicates the non-labor income that is the income level if the individual choose to allocate all the time to leisure. The decrease of the non-labor income would shift the budget constraint down. As long as the leisure is the normal good, the individual would allocate more time on working. I want to explore the impact of house rent on working decision in this paper. From the theoretical perspective, the increase of the house rent would lead to the decrease of the non-labor income. If the leisure is normal good, the working hours would increase. I would test this relationship empirically in the following part.

The remainder of this paper would be organized as follows: part 2 reviews the methodology of quantile regression, part 3 would introduce the data and model , part 4 displays the empirical results, finally part 5 would give the conclusions.

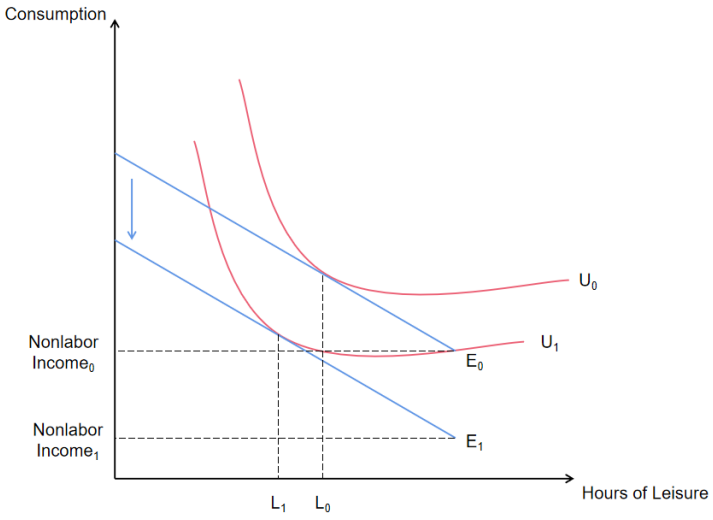

Figure 1: The Decision Model

2 Methodology

In this part, I would make a brief introduction to the quantile regression method. This part is mainly based on Chen Qiang's Advanced Econometrics and Its Stata Application[2] and some other relevant information from the Internet.

2.1 Background

The traditional ols regression is a method based on the average value, however, we always want to know more about the conditional distribution y|x. Hence, to estimate the specific quantile value of the conditional distribution would be helpful. In addition, the loss function of ols contains squares, which make it easy to be affected significantly by some outliers.

To solve the problem, Koenker and Bassett (1978)[1] have proposed the quantile regression method, to estimate the specific quantiles using the the weighted average of the absolute value of residuals($\sum_{i=1}^{n} e_i^2$) as the loss function.

2.2 Estimation

Suppose y is a continuous variable, the cumulative distribution function is $F_{y|x}(\cdot)$, define q-quantile is y_q satisfying the following equation

$$q = P(Y \leq y_q) = F_y(y_q) \tag{1}$$

As for the random variable, we always use sample quantile to substitute the population quantile. When it comes to the regression, we always solve the following optimal solution problem to estimate the sample quantile.

$$min_{\mu} \sum_{i;y_i \geq \mu}^{n} q|y_i - \mu| + \sum_{i;y_i < \mu}^{n} (1-q)|y_i - \mu| \Rightarrow \mu = \widehat{y}_q \tag{2}$$

2.3 Using Stata to do the regression

In this part, I would introduce how to use stata to run the quantile regression.

The most basic command is qreg which could help us do a specific quantile regression while

bsqreg command can do the regression using bootstrap to calculated the covariance matrix. In addition, the sqreg command could help us do several quantile regressions simultaneously.

After achieving the regression results, we may still want to show them graphically, the command grqreg or qregplot could help us to do it. The more details could be checked in Chen Qiang's book or the help files in stata.

3 Data, Model and Variable Selection

3.1 Data

This report has used the 10% sample of 2020 US Decennial Census data containing information collected on persons and households. This data is released by IPUMS USA, the original data set contains 2,641,054 points.I have done the necessary data cleaning and have dropped the point of the military, the points from Hawaii and Alaska and also all the points that have 0 or n/a for the major variables of my interest. Finally, the using data set contains 317,944 points. The descriptive statistics are as follows.

VARIABLES	Obs	Mean	SD	Min	Max
uhrswork	317,944	37.77	12.2	1	99
rentgrs	317,944	1,471	847.5	4	8,228
ftotinc	317,944	70,106	73,389	-10,200	1.54E+06
state	317,944	26.66	16.34	1	56
sex	317,944	1.498	0.5	1	2
age	317,944	38.01	13.88	16	95
trantime	274,282	22.23	22.19	0	165
english	317,944	0.723	0.448	0	1
wkswork1	317,944	44.65	14.52	1	52
incwage	317,944	40,394	49,127	0	743,000
occupation	317,944	4,257	2,571	10	9,820
industry	317,944	624.3	258	10	932
marst	317,944	3.937	2.227	1	6
metro	317,944	2.719	1.271	0	4

Table 1: Descriptive Statistics

3.2 Basic Regression Model

$$uhrs_i = \alpha + \beta * lnre_i + \gamma D_i + \varepsilon_i \tag{3}$$

This is the most fundamental model specification where $uhrs_i$ is the personal usual hours worked per week, $lnre_i$ is the log value of personal monthly gross rent, and D_i are a set of control variables including age, gender, speaking English or not, travel time to work, and the dummy variables for industry, occupation, metropolitan status and marriage status. These variables are all the variables that are likely to be confounding variables from my point of view in the data set. But there may still exist omitted variables and simultaneous causality since the mechanisms of both wokring hours and house rents are complicated, which may make the regression results not plausible.

3.3 The Selection of Instrumental Variable

To solve the potential endogeneity problem and achieve a causal estimation, I decided to choose an instrumental variable and accomplish the 2sls estimation. I have selected the total family income as an instrumental variable. The total family income would have a big impact on housing renting decisions. That is, higher income level may motivate the family to choose a better house with higher rent. By contrast, I trust that there is no direct relationship between the total family income and personal working hours decision. I would discuss more about the instrumental variable's validity in the following part.

3.4 Quantile Regression

Linear regression is the simple average regression function, it could not reveal the impact of x on the conditional distribution of y|x. To obtain more information of the y|x, I decide to use quantile regression to estimate several important conditional quantiles to achieve further understanding of the specific relationship. However, the quantile regression could not give out results if the sample is too large. To solve this problem, I have picked a 1% small sample maintaining the proportion of states to accomplish the quantile regression.

4 Empirical Results

4.1 Basic Results

The above table shows the basic empirical results. The first row is the result of baseline regression, the second-row result is the result of regression adding all the selected control

	(1)	(2)	(3)	(4)
	uhrswork	uhrswork	lnre	uhrswork
lnre	2.008***	1.593***		10.18***
	(47.24)	(30.72)		(43.79)
ftotinc			0.00000165***	
			(89.87)	
sex		-1.963***	-0.00585**	-1.884***
		(-39.94)	(-3.09)	(-36.30)
age		0.00784***	-0.00296***	0.0286***
		(3.89)	(-39.64)	(13.13)
English		0.421***	0.0177***	0.183***
		(8.13)	(8.96)	(3.34)
trantime		0.0155***	(0.00)	0.0161***
		(15.48)	(-0.58)	(15.34)
_cons	23.41***	39.99***	6.752***	-20.42***
	(76.89)	(53.81)	(270.78)	(-11.47)
Industry	No	Yes	Yes	Yes
Occupation	No	Yes	Yes	Yes
State	No	Yes	Yes	Yes
Metro	No	Yes	Yes	Yes
Marst	No	Yes	Yes	Yes
N	317944	274282	274282	274282
adj. R2	0.01	0.17	0.41	0.07

t statistics in parentheses
* p <0.05, ** p <0.01, *** p <0.001

Table 2: Baseline Result

variables. The third row is the first stage's result of 2sls IV estimation and the fourth row is the second stage result. All the control variables also have been added into the IV estimation.

From the table, we could find that all the coefficients of lnre are positive and significant at 0.001 significance level, which implies a robust relationship of gross rent on working hours per week. After adding the instrumental variable, the coefficient is 10.18, implying that 1% increase in the gross house rent could lead to 0.1018 more hours to work per week.

4.2 The Validity of Instrumental Variable

In the above part, we could see that the coefficient for the total family income is statistically significant in the first stage estimation. To prove the validity of the instrumental variable, I have do some further tests and the results are showed as follow. The testing statistics tell us that the instrumental variable is not a weak one.

Variable	R-sq	Adjusted R-sq	Partial R-sq	Robust F(1,2195)	Prob>F
lnre	0.5235	0.4011	0.0751	52.3975	0

Table 3: First-stage regression summary statistics

Minimum eigenvalue statistic = 178.251				
Critical Values	# of endogenous regressors: 1			
H0: Instruments are weak	# of excluded instruments: 1			
2SLS relative bias	5%	10%	20%	30%
	(not available)			
	10%	15%	20%	25%
2SLS Size of nominal 5% Wald test	16.38	8.96	6.66	5.53
LIML Size of nominal 5% Wald test	16.38	8.96	6.66	5.53

Table 4: Weak IV Test

4.3 Robustness Test

To test the robustness of the given relationship, I have replace the dependent variable to another variable that is the working weeks last year. Because I think these two variables are both the evaluations of individual's working decision. In addition, I have deleted all the points with 52 hours, that is the individuals who have worked every weeks last year. In other words, I have evaluated how the rent could motivate people to work more hours but do not consider those people who really like working.

The first row is the baseline result, the coefficient is small and only significant in 5% level. However, after adding the instrumental variable, the coefficient has became much larger and significant in 0.001 level according to the second row, which implies the robustness of the relationship of house rent on working decisions.

	(1)	(2)
	wkswork1	wkswork1
lnre	0.392*	10.51***
	(2.48)	(14.74)
sex	-0.629***	-0.530**
	(-3.83)	(-3.12)
age	0.0224***	0.0539***
	(3.44)	(7.65)
english	1.069***	0.733***
	(6.20)	(4.09)
trantime	-0.0135***	-0.0127***
	(-4.16)	(-3.79)
_cons	24.49***	-45.49***
	(10.36)	(-8.42)
N	54149	54149
adj. R2	0.04	.
Industry	Yes	Yes
Occupation	Yes	Yes
State	Yes	Yes
Merco	Yes	Yes
Marst	Yes	Yes

t statistics in parentheses
* p <0.05, ** p <0.01, *** p <0.001

Table 5: Robustness Test 1

Furthermore, I also do the regression of rent on personal wages. I am not sure whether this is a suitable robustness test or not. I think that higher rents may motivate people to

work much harder to earn higher wages. But I have to admit that the mechanisms between these variables are quite complicated. Anyway, we could see that the relationship of rent of personal wages is significant whether we add the instrumental variable or not.

	(1)	(2)
	lnwage	lnwage
lnre	0.251***	2.788***
	(56.45)	(86.86)
sex	-0.170***	-0.145***
	(-42.23)	(-22.63)
age	0.0103***	0.0164***
	(62.21)	(60.64)
english	0.0802***	0.01
	(19.10)	(1.64)
trantime	0.000471***	0.000593***
	(6.09)	(4.67)
_cons	9.030***	-8.844***
	(151.04)	(-36.60)
N	261659	261659
adj. R2	0.35	.
Industry	Yes	Yes
Occupation	Yes	Yes
State	Yes	Yes
Merco	Yes	Yes
Marst	Yes	Yes

t statistics in parentheses
* p <0.05, ** p <0.01, *** p <0.001

Table 6: Robustness Test2

4.4 Quantile Regression

To explore the different characteristics in different quantiles, I have done the quantile regression for the usual working hours and ln rent. However, to be honest, due to the lack of time and knowledge, there are many problems that I have not solved. For example, I have used Stata and R to run the quantile regression for the whole sample but just cannot get the estimation result. So I have to select 1% sample to do the regression. In addition, the results for the 1% sample quantile regression are very strange and I try to explore the reasons behind these strange results but I fail. Hence, I would just give the conclusions based on my results but I think these results may be not that plausible. I would do more exploration about it and discuss it with some specialists after this report submitting.

From the table, I have found that there is no big differences between 25% sample and 75% sample, however, the significance levels are both only 5%, maybe it is because of the much smaller sample. The graph shows very strange trends, so I would not give any comments to it. Just as what I have said, I would do more exploration and discuss it with others.

	(1)25% Quantile	(2)75% Quantile
	uhrswork	uhrswork
lnre	1.308*	1.612**
	(2.27)	(2.80)
sex	-1.606**	-1.966**
	(-2.88)	(-3.22)
age	0.01	0.0455*
	(0.76)	(2.03)
	(0.47)	(0.31)
	(-0.96)	(-0.57)
trantime	(0.00)	(0.00)
	(-0.27)	(-0.09)
_cons	35.73	46.25*
	(1.22)	(2.28)
Industry	Yes	Yes
Occupation	Yes	Yes
State	Yes	Yes
Marst	Yes	Yes
Merco	Yes	Yes
N	2760	2760

t statistics in parentheses
* p <0.05, ** p <0.01, *** p <0.001

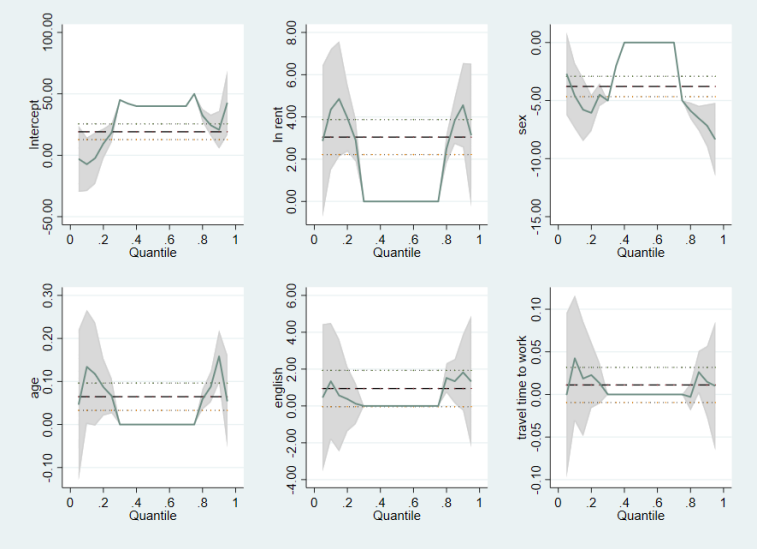

Figure 2: The Decision Model

5 Conclusion

From the article, we could see that house rent have a significant positive impact on working hours no matter what methods and variables we use. However, the differences in the different quantiles are not significantly high. To be honest, there exist many problems for the quantile regression part. So that would be the direction that I would spend much more effort on in the future.

References

[1] Roger Koenker and Gilbert Bassett Jr. Regression quantiles. *Econometrica: journal of the Econometric Society*, pages 33–50, 1978.

[2] Chen Qiang. Advanced econometrics and its stata application. pages 509–517, 2014.