Incentives, Goals and Labor Supply: An Experimental Investigation

Xinchao Zhang [1], Lu Liu [1], Liangcong Fan [2*]
1 College of Economics, Zhejiang University
2 Guanghua Law School, Zhejiang University
*Corresponding Author, flc1984@163.com

Abstract: Incentives are the essence of economics, and the question of how workers' choices of effort and work hours respond to financial incentives is among the oldest questions in labor economics. Many efforts have already been devoted to understanding the determinants of labor supply. This paper aims to investigate the determinants of labor supply under risk by employing a series of real effort lab experiments. Subjects could choose both the work time and effort in one treatment and could only choose the effort in another treatment. The results show that, consistent with predictions of neoclassical model, most subjects in both treatments provided more efforts as the piece rate raised up, indicating a significant substitution effect. However, inconsistent with predictions of neoclassical model, most subjects provided at least no less efforts even when they faced an uncertainty on getting their accumulated piece rate earnings, holding the piece rate constant. Prospect theory cannot explain the result either. However, multiple-reference points theory which suggests that to be success is a more important motivation than loss aversion seems fitted the data better.

Key words: effort supply, Multi-reference Points Model, real effort experiment

JEL: C91, J22

Introduction

Incentives are the essence of economics, and the question of how workers' choices of effort and work hours respond to financial incentives is among the oldest questions in labor economics. Many efforts have already been devoted to understanding the determinants of labor supply. Traditionally, the classical static model of labor-leisure choice offers a positive compensated wage effect on hours of work as its main testable implication. Workers work harder when there is a transitory wage increase, and substitute leisure for labor when the wage is low. This prediction has direct implications for compensation schemes, for example, employers can tie compensation to output, and thus encourage high effort at times when the firm's output is in high demand.

Recently, a kind of model is newly built on the idea that expectations can act as a reference point (e.g., Bell, 1985; Loomes & Sugden, 1986; Gul, 1991; Shalev, 2000; Koszegi & Rabin, 2006, 2007). The reference dependent preferences model starts from the observation from psychology, that people tend to evaluate outcomes as gains or losses relative to a reference point, or put another way, as successes or failures relative to a goal (for a review, see Heath, Larrick, and Wu (1999); Tversky and Kahneman (2000)). The tendency to have goals is pervasive, and could naturally extend to the workplace, where workers may have personal goals in mind, in terms of income or output, when they decide how hard to work. Incorporating reference dependence preference into the standard model is simple, and leads to strikingly different predictions from the standard model in terms of how workers allocate effort over time: workers may actually work less hard on days when the wage is high, or they may work more hard when the wage is low.

This article aims to examine these two popular models of labor supply with controlled lab experiment. Despite their theoretical and intuitive appeal, models of labor supply are inherently difficult to test, as effort is typically hard to observe in the field. Also, in many occupations, workers' labor supply choices are constrained by institutional rules regulating labor time and effort provision. For this particular study, there exists no naturally occurring data set. Such data could possibly be generated, but it is not clear how unbiased data on work effort could be acquired, and the cost could be quite high. Thus, a tightly controlled real-effort experiment was conducted to sidestep this problem.[1]

To the best of our knowledge[2], the earliest economic studies of the effects of material incentives on labor supply in the laboratory were the animal experiments of the early 1980s (Battalio et al., 1981; Battalio and Kagel, 1985). These studies induced hungry animals' effort to test the classic, static economic model of labor

[1] In general, there are always benefits and drawbacks to a specific methodological approach. The most prominent critics faced by controlled experiment is its external validity, while the most impressive strength of controlled experiments is the ability to take a specific theoretical model, where theory says exactly what the equilibrium should be, making comparison of theoretical predictions with experimental results easily. The controlled setting of the experiments greatly reduces the variables that may be confounding the results derived from data generated in more uncontrolled settings. Regarding this study, we cannot easily disentangle the substitute effect with income effect with survey or field data. In contrast, the income effect can be easily controlled in lab. In addition, it seems almost impossible to infer reliably from existing data whether the workers anticipated the wage change. Furthermore, serious endogeneity problems arise, as both supply and demand conditions determine wages. In fact, there is strong evidence suggesting that workers are not free to set their working hours (John C. Ham, 1982; Shulamit Kahn and Kevin Lang, 1991; William T. Dickens and Shelly Lundberg, 1993), rendering the identification of the source of small intertemporal substitution effects difficult. Thus, the typically available data require many auxiliary assumptions when testing the models of labor supply.

[2] There are quite a lot experiments studies concerning how incentives and expectation would impact labor supply. But here we only focus on the the simplest forms of work incentives (a wage per hour worked or an individual piece rate) as we believe that the general principles and models that exist in the literature should be expected to apply with the same force to these simplest laboratory economies as to those economies found in the field (Plott, 1991). Regarding more extensive reviews, please find Charneess and Khun(2011) and List (2011).

supply in which an agent chooses consumption (C) and leisure (L) to maximize a quasi-concave utility function $U(C,L)$, subject to the constraint $C = wL + G$ where w is the wage rate and G is unearned income. The experimenters varied both w and G exogenously and study the animals' reactions, and found that an income-compensated wage decrease would reduce the animals' labor supply and consumption, and the decline of non-labor income (G) would raise labor supply.

Swenson(1988) conduct the first human subjects real-effort laboratory experiment to examine how labor supply responses to wage changes and tax rate. [3] Wages per character typed were fixed, but 'taxed' at rates ranging from 12 to 87 percent. Total tax proceeds from the previous session were randomly distributed to the subjects in the following period, mimicking a balanced government budget but breaking most of the connection between current individual effort and future lump-sum income. Both curves were backward-bending, with tax revenues peaking at the 73 percent tax rate. A decade later, Sillamaa (1999a, 1999b) and Dickinson (1999) conducted similar real-effort experiments.[4] Like Swenson's, Sillamaa's experiments were motivated by questions about the impact of taxation. Sillamaa found that: (a) work effort responds more (positively) to real wage increases in the presence of an (equivalent) linear than a progressive income tax, and (b) introducing a zero top marginal tax rate also increased effort.

Like Sillamaa, Dickinson (1999) paid his subjects a piece rate, but in one treatment allowed his subjects to choose between two types of leisure: on- versus off-the-job. This modification is noteworthy because it provides one of the few empirical links between the types of work decisions that are usually studied in lab (and field) experiments and the traditional application of labor supply theory (to hours worked). Specifically, in the baseline ("intensity") treatment, subjects were required to stay for the entire two-hour experimental period. In the "combined" treatment, subjects could leave at any time during the experimental period. Consistent with theory and with previous research, subjects increased their output in the baseline treatment, substituting on-the-job leisure for effort when incentives were strengthened. In the combined treatment, many subjects responded to higher wages by working more quickly, but reducing their total work time by leaving the experiment early. This substitution of off-the-job for on-the-job leisure is offered as a possible explanation for why econometric estimates of labor supply elasticities are often close to zero.

Moreover, later studies found that different from the prediction of classic theory,

---

[3] In the experiments, subjects were required to repeatedly type "!" and then "enter" on a computer keyboard. Laboratory principal-agent experiments can be divided into those where subjects are paid to perform an actual task ('real-effort') and those where effort decisions are represented by the choice of a decision number that imposes increasing marginal financial costs on the agent ('chosen effort'). Bruggen and Strobel (2007) find little difference between the two methods in a simple gift-exchange labor market game.

[4] In Sillamaa's experiments, workers decoded numerical codes into letters; in Dickinson's workers repeatedly typed paragraphs, with a penalty for mistakes.

the relationship between the piece rate and effort is highly non-monotonic but U-shaped (Gneezy and Rustichini, 2000), and the effect was conditional on subjects' risk aversion level (Cadsby, Song and Tapon, 2009; Ariely, Gneezy, Loewenstein and Mazar, 2009).

Along another line of literature, based on the mass evidences about the important role of expectation-based reference points, some researchers intend to incorporate the literature on violations of expected utility theory in lottery choices into model of labor supply, among which the most prominent one is the prospect theory and the idea of loss aversion (Kahneman and Tversky, 1979; Daniel Kahneman, Jack Knetsch, & Richard Thaler 1990). A series of studies have found evidence consistent with loss aversion around a daily reference income (e.g., Colin Camerer, Linda Babcock, George Loewenstein, and Thaler, 1997; Yuan K. Chou, 2002; Ernst Fehr and Lorenz Goette, 2007; Henry S. Farber, 2008; Vincent Crawford and Juanjuan Meng, 20011), with the exception of Farber (2005). Camerer et al. (1997), collecting data on the daily labor supply decisions of New York City cabdrivers, who unlike most workers in modern economies are free to choose their own hours, found a strongly negative elasticity of hours with respect to realized earnings. They proposed an explanation that drivers have daily income targets and work until the target is reached, and so work less on days when realized earnings per hour is high. Different from Camerer et al. (1997), Farber (2005) found that drivers' stopping probabilities are significantly related to hours but not income. Along this line of research, and followed the model proposed by Koszegi & Rabin (2006) which introduced both targets for hours as well as income, a recent study by Crawford & Meng (2011) used the data collected by Farber (2005, 2008), proxied the rational expectation about a driver's wage by the average wage earned per week day and found evidence for income and hours targeting around this expectation, and thus reconciled Farber's evidences and Camerer et al.'s.

In complementary, and in order to overcome two problems that may appear in dealing with the natural occurring data: endogeneity rising from supply-side shocks and the possible selection effect, which may lead to a downward biased estimate of the wage elasticity, Fehr and Goette (2007) conducted a field experiment. In their experiment, there are two groups of bicycle messengers signing up for shifts—Group A and Group B, who are free to choose hours worked and effort per hour and paid on commission. The treatment is a month-long increase in the commission rate, of 25%. In September, Group A received the treatment and Group B was the control. In November, Group A was the control, and Group B received the treatment. They found a large positive wage elasticity of overall labor supply and an even larger elasticity of hours, which implies that the elasticity of effort per hour is negative. They argued that effort is a more accurate measure of labor supply and concluded that only loss-averse individuals exhibit a negative effort response to the wage increase.

In another real effort lab experimental study, Abeler et al.(2011) exogenously vary rational expectations of subjects regarding earnings and check whether this manipulation influences their effort provision. In their experiment, subjects count the number of zero in tables that consisted of 150 randomly ordered zeros and ones and get a piece rate but receive their accumulated piece rate earnings only with 50 percent probability, whereas with 50 percent probability they receive a fixed, known payment instead. Subjects could decide whether to continue or to stop working anytime. Which payment subjects receive is determined only after they have made their choice about when to stop working. What they manipulate is the amount of the fixed payment, that is, subjects' reference points in expectations. They find that effort provision is significantly different between treatments in the way predicted by models of expectation-based reference-dependent preferences: if expectations are high, subjects work longer and earn more money than if expectations are low.

The main purpose of this paper is to connect these two lines of literature. Our experiment adds to this literature by measuring the impact of reference points as expectations in the domain of real effort choices when a compensated wage effect is also taken into consideration. It means to examine what is the determinant of labor supply under risk, the reference points or wage effect. Moreover, if reference points are important, then what determines the reference point, the status quo or the goals? We employ the real task that was used by Abeler et al. (2011) and conduct two treatments just as what Dickson (1999) did. In one setting, the work time of the subjects is fixed; and in the other setting, subjects could choose both work time and effort freely. Subjects are randomly assigned to either treatment and participate in a three-day experiment. Particularly, they are exposed to a high basic payment for one hour work and a low piece rate in the first day, and then a special basic payment which is calculated by control the income effect and a trifled piece rate in the second day, and finally a piece rate as the second day but receive their accumulated piece rate earnings only with 50 percent probability, whereas with 50 percent probability they receive a fixed, known payment which is set as their first day's earning instead. Our experiment is different from Dichson (1999) in that risk and reference points are taken into account, from Fehr and Goette (2007) in that the changes of incentive schemes are totally unknown to subjects until the experiments start[5], and from Abeler et al. (2011) and other field data studies in that both dimensions of labor supply: hours worked and effort made are taken into consideration. These features make our experiment a useful complement to the existing literature.

The paper is organized as follows. Details of the experimental design are

[5] In their experiments, the variation in the wage rates are told to subjects before the experiments start, which may cast doubt on the variable's exogeneity.

explained in the following section. Section 3 discusses behavioral predictions. Results of the two main treatments are presented in Section 4. Section 5 concludes.

Experimental Design and Procedure

Following Dickson (1999), we conduct two treatments. In one treatment, the work time of the subjects is fixed (short for FT hereafter); and in the other treatment, subjects could choose both work time and effort freely (short for CT hereafter). Each treatment involves a three-day experiment. We start from discussing the features that are common to both treatments.

Subjects are recruited for a one-hour experiment each day for three days. The experimental days are not consecutive primarily due to the recruiting difficulties that arise with students who are scheduled for classes on two or three days of the week, but all experimental days are completed for any given subject within an eight-day period. In the experiment, subjects work on a tedious task as Abeler et al. (2011) employed that counting the number of zeros in tables that consisted of 150 randomly ordered zeros and ones. Figure 1 shows the task screen in the experiment. We choose this task because it does not require any prior knowledge, performance is easily measurable, and there is little learning possibility; at the same time, it is so boring and pointless that entailed a positive cost of effort for subjects. It is also clearly artificial, and output is of no intrinsic value to the experimenter. This minimizes any tendency for subjects to use effort in the experiment as a way to reciprocate for payments offered by the experimenter.

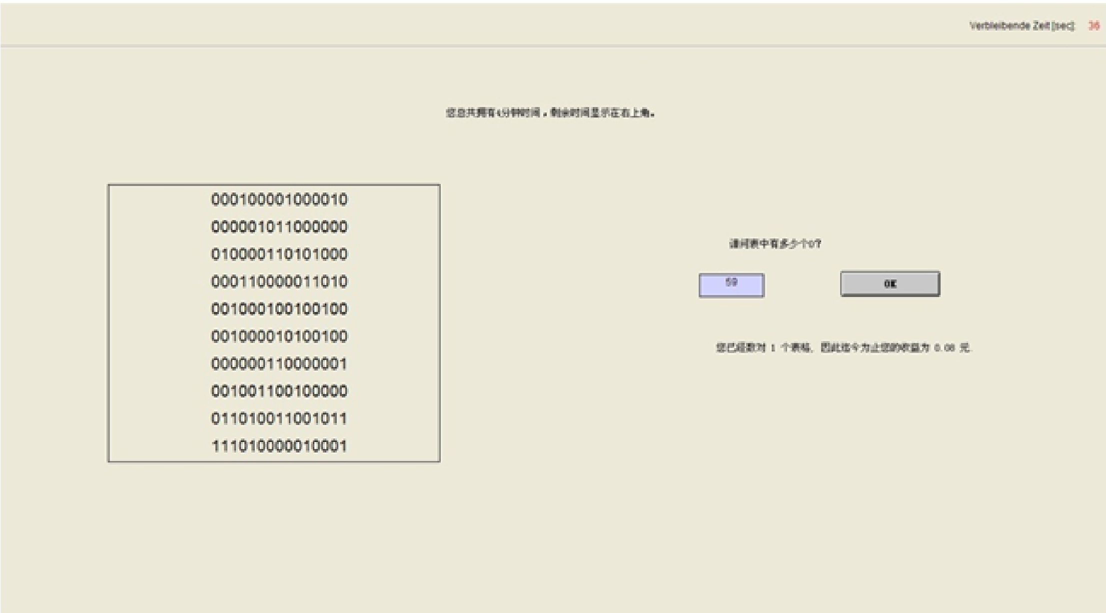

Figure 1 the real effort screen in the experiment

The first day experiment consists of two stages. Prior to the first stage, subjects read the instructions. They are also told that the experiment has a second stage but that details would be provided later. The first stage is a pilot stage, subjects have four minutes to count as many tables as possible. They receive a piece rate of 0.08 RMB per correct answer for sure. This part serves to familiarize subjects with the task. Additionally, we will use subjects' performances in this stage as a productivity indicator. After the pilot stage, subjects read the instructions for the main stage. The task was again to count zeros, but the payoff would involve two parts: a basic payment F_1, which equals to 18 RMB, and a piece rate w_1 of 0.08 RMB per correct answer.

The second day experiment has the same task with the first day, but the piece rate w_2 raises up to 0.24 RMB per correct answer and the basic payment F_2 is determined by the first day's work. Specifically, we adjust the new level of F_2 that making the original (total income) choice point from day one attainable if the subject counts the same number of tables as on day 1, in order to control the income effect.[6] It means that every subject's F_2 may be different from each other.

The final day experiment also consists of two stages. The task of the main stage is the same as before, but subjects may not get their accumulated piece rate earnings for sure. Before they start their work, they have to choose one of two closed envelopes. They know that one of the envelopes contains a card saying "Piece rate earnings plus Basic payment" and that the other envelope contains a card saying "Fixed payment". But they do not know which card is in which envelope. The envelopes remain with the subjects while they are working and are only opened after the subject has stopped working. The subject's payment is then determined by the card in the chosen envelope. If the "Piece rate earnings plus Basic payment" card is chosen , they would be paid with a basic payment F_3 which equals to F_2 and a piece rate w_3 which equals to w_2. Otherwise, no matter how much correct answer they have made, they would be paid with a fixed payment F, which equals to their income of the main stage on day 1. This setup is also for the sake of controlling the income effect. After the main stage, subjects are attended a lottery choice which consists of six choices, each time between a fixed payment of 0 and a small-stakes lottery. The lottery involved a 50/50 chance of winning 6 RMB or receiving Y. Across lotteries, Y was varied from -2 to -7 RMB in steps of 1 RMB. Subjects knew that one of the six choices would be randomly selected and, if they had chosen the lottery, this lottery would be played out for money. These data are collected for measuring subjects' loss aversion level.[7]

[6] In some cases (5 in 60), subjects would get negative F_2 as they made more than 112 correct answer in the first day. In these cases, we set F_2 to 0.

[7] Note that the small stakes mean that rejections of lotteries with positive expected value cannot be explained by standard risk aversion (Rabin 2000). Rather, the number of lotteries that a subject rejects gives an indicator for the individual's degree of loss aversion.

The three-day experimental design is summarized by table 1.

Table 1 Experimental Design

Day 1	Day 2	Day 3	
w_1=0.08 F_1=18	w_2=0.24 F_2<18*	50% w_3=0.24 $F_3 = F_2$	50% f =Day 1 income

* The exact amount of nonwage income on second day depends on the correct number of tables counted on day 1. The compensations are such that if the subject would have counted the same number of tables on day 2 and day 3 as the main stage on day 1, the subject would have earned the same amount of income as the main stage on day 1.

In the FT treatment, subjects are required to stay and work for the full one-hour in each experimental day. Thus, the only choice being made is how much tables they want to count (the effort). On contrast, in CT treatment, they could decide how much and for how long they want to work during each experimental day. At most, they could work for 60 minutes. And at least, they are required to work for 10 minutes.[8] When they wanted to stop, they could press a button on the screen and the experiment was over: subjects would then got paid and leave immediately.

The experiment was conducted in March, 2013 in the Experimental Social Science Laboratory in Zhejiang University, Hangzhou, China. There were 35 subjects in FT treatment and 25 subjects in CT treatment. All the subjects were randomly recruited via online advertisements and were full-time undergraduate students in diverse majors, 48.33% of which were girls. Each subject only participated in one treatment.

Experiments were computerized by Ztree (Fischbacher, 2007). The experiment lasted in each experiment day no more than 70 minutes. The temperature in our experimental days was around 14-19 ℃. The average earnings for each subject in each experiment day were 31.71 RMB, including a 5 RMB show-up fee, specifically, 28.60 RMB on day 1, 33.69 RMB on day2, 32.85 RMB on day 3 and totally 95.14 RMB for all 3 days. At the time of the experiment, 31.71RMB was equal to about $5.1US. For comparison purposes, the wage rate for Zhejiang University undergraduates who had part-time jobs with the university administration was 16 RMB per hour.

[8] The minimum requirement is set to simulate a minimum amount of labor to be supplied in order to survive and continue on to the next day.

Prediction

We examine two categories of models in this section: a canonical model with separable utility, models with expectation-based reference dependence. With regard to the latter, two kinds of expectations are considered: the exogenously introduced reference point and endogenously introduced goal-based reference dependence. Our setup can be described as follows: the subject's choice variable is the number of correctly solved tables e. The subject receives the accumulated piece rate earnings w_1e_1 (w_2e_2) and basic payment B_1 (B_2) in Day 1 (Day 2), where $w_1 > 0$ ($w_2=3w_1 > 0$) is the piece rate per correct table. In Day 3, the subject receives either a fixed payment F or the accumulated piece rate earnings w_2e and basic payment B_2 with probability 0.5 each, where F equals to w_1e_1. $c(e)$ is the subject's cost of effort with $c' > 0$ and $c'' > 0$.

First of all, we integrate the setting in our experiment into a canonical model of intertemporal utility maximization with time-separable utility. Consider a standard model of effort provision with a utility function separable in monetarypayoff and cost of effort. As outcomes in our setup are not very large, we assume consumption utility to be linear and equal to the consumption bundle. Thus we have the subjects' utility function in Day 1: $U_1(e_1, B_1, w_1) = B_1 + w_1e_1 - c(e_1)$, yielding the following first-order condition:

$$\frac{\partial U_1}{\partial e_1} = w_1 - c'(e_1) \quad \Rightarrow \quad c'(e_1^*) = w_1$$

Similarly, first-order condition in Day 2 is:

$$\frac{\partial U_2}{\partial e_2} = w_2 - c'(e_2) \quad \Rightarrow \quad c'(e_2^*) = w_2$$

The utility function in Day 3 is $U_3(e_3, F, B_2, w_2) = 0.5(B_2 + w_2e_2) + 0.5F - c(e_3)$, yielding the following first-order condition:

$$\frac{\partial U_3}{\partial e_3} = \frac{w_2}{2} - c'(e_3) \quad \Rightarrow \quad c'(e_3^*) = \frac{w_2}{2}$$

And we can obtain:

$$e_1^* < e_3^* < e_2^*$$

Models incorporating reference-dependent preferences with the exogenously introduced expectation as the reference point also predict a similar prediction in Day 3. We get the prediction using the model of Kőszegi & Rabin (2007) and the model of Abeler (2011).

In Kőszegi & Rabin (2007), an individual derives "consumption utility" from the consumption bundle c and "gain-loss utility" from comparing c to a reference bundle r. Bundle r is the full distribution of rational expectations, i.e., every outcome that could have happened weighted with its ex-ante probability.

Then, the overall utility is the sum of consumption and gain-loss utility, and is assumed to be separable across the K dimensions of c. Abeler assumes that subjects assess outcomes along 2 dimensions: money and effort costs. The gain-loss utility is defined by the function $\mu(c_k - r_k)$. For small arguments s, Kőszegi & Rabin assume that $\mu(s)$ is piece-wise linear: $\mu(s) = \eta s$ for $s > 0$ and $\mu(s) = \eta\lambda_1 s$ for s < 0 with $\eta > 0$ and $\lambda_1 > 1$; because λ_1 is strictly greater than 1, losses loom larger than equal-sized gains. Abeler assumes that the gain or loss sensation a subject finally experiences depends on their rational expectations about possible earnings amounts held the moment before the envelope is opened. The final piece rate earnings plus basic payment (and the fixed payment amount for the treatment) thus determine the reference point.

If the subject intends to stop at an accumulated earnings and basic payment level below the fixed payment ($B_2 + w_2 e_3' < F$), then the resulting expected utility will be given by

$$U = \frac{B_2 + w_2 e_3' + F}{2} - c(e_3') + \frac{1}{2}\eta\left\{\frac{1}{2}[(B_2 + w_2 e_3') - (B_2 + w_2 e_3')] + \frac{1}{2}\lambda_1[(B_2 + w_2 e_3') - F]\right\} + \frac{1}{2}\eta\left\{\frac{1}{2}[F - (B_2 + w_2 e_3')] + \frac{1}{2}[F - F]\right\}$$

The first two terms are expected consumption utility and cost of effort. The remaining terms are the expected gain-loss utility: the first bracketed term is the gain-loss utility when the outcome is $B_2 + w_2 e_3'$, multiplied by the probability of occurring 0.5 and by η, the strength of gain-loss utility. Inside this term, receiving $B_2 + w_2 e_3'$ feels neutral; but receiving $B_2 + w_2 e_3'$ while expecting the larger F feels like a loss. Since the subject expected to receive F with probability 1, the terms are weighted accordingly. The second bracketed term shows gain-loss utility where the outcome is the fixed payment, applying the same logic. The first-order conditions are then:

$$\frac{\partial U_3}{\partial e_3'} = \frac{w_2}{2} - c'(e_3') + \frac{1}{4}\eta(\lambda_1 - 1)w_2 \quad \Rightarrow \quad c'(e_3'^*) = \frac{w_2}{2} + \frac{w_2}{4}\eta(\lambda - 1)$$

$$\begin{cases} B_2 + w_2 e_3' < F & \Rightarrow \quad e_3' < e_1^* \\ c'(e_3'^*) = \dfrac{w_2}{2} + \dfrac{w_2}{4}\eta(\lambda_1 - 1) & \Rightarrow \quad e_3'^* > e_1^* \end{cases}$$

The two inequalities are inconsistent, so there is no solution under this condition.

If the accumulated earnings and basic payment are higher than the fixed payment ($B_2 + w_2 e_3' \geq F$), the gain-loss utility is different. Receiving the accumulated earnings and basic payment now feels like a gain compared to the lower fixed payment (third term), while receiving the fixed payment now means a loss (terms equal to zero are suppressed here):

$$U = \frac{B_2 + w_2 e_3' + F}{2} - c(e_3') + \frac{1}{2}\eta\left\{\frac{1}{2}[(B_2 + w_2 e_3') - F]\right\} + \frac{1}{2}\eta\left\{\frac{1}{2}\lambda_1[F - (B_2 + w_2 e_3')]\right\}$$

The first-order conditions are then:

$$\frac{\partial U_3}{\partial e_3'} = \frac{w_2}{2} - c'(e_3') - \frac{1}{4}\eta(\lambda_1 - 1)w_2 \quad \Rightarrow \quad c'(e_3'^*) = \frac{w_2}{2} - \frac{w_2}{4}\eta(\lambda_1 - 1)$$

$$\begin{cases} B_2 + w_2 e_3' \geq F & \Rightarrow \quad e_3' \geq e_1^* \\ c'(e_3'^*) = \frac{w_2}{2} - \frac{w_2}{4}\eta(\lambda_1 - 1) & \Rightarrow \quad e_3'^* < e_3^* < e_2^* \end{cases}$$

Thus, we have the prediction:

$$e_1^* \leq e_3'^* < e_3^* < e_2^*$$

Here, e_3 is lower in this model than the prediction of canonical model, much closer to e_1.

In contrast to the exogenously introduced expectation-based model, multiple-reference points theories argue that agents make decisions using more than one reference points (Wang and Johnson, 2012). Evidences from lab and fields provide sufficient support for these theories, showing that they have some advantages in interpreting and predicting choice behavior (Neale and Bazerman, 1991; Wu, 2008). Among others, the theories incorporating an additional reference points, the goals, has attracted many attentions, and has been regarded as a pervasive aspect of human decision making. Heath, Larrick, and Wu (1999) discuss one experiment that is especially germane for labor supply: Workers seem to be twice as willing to provide an additional amount of effort to meet a given goal, than they are willing to provide the same amount of effort to surpass that goal. In other words, if a person falls short of his or her target, he or she is assumed to experience an additional psychological cost. As Goette(2004) summarized, an individual derives “success-failure utility” from comparing performance p to a reference bundle goal g. Although we have not introduced any distinct goal in our experiments, we receive the message from after-experiment survey that many subjects have formed their goals in latter experiment days. They dislike an outcome falling short of their performance at last time. Here, we will introduced a formal model to derive our hypotheses

The overall utility now is the sum of consumption, gain-loss utility and success-failure utility. The success-failure utility is defined by the function $v(p - g)$. For small arguments t, we assume that $v(t)$ is piece-wise linear: $v(t) = \eta t$ for $t > 0$ and $v(t) = \eta\lambda_2 s$ for $s < 0$ with $\eta > 0$ and $\lambda_2 > \lambda_1 > 1$; because λ_2 is strictly greater than 1, failure looms larger than equal-sized success. The last day’s performance thus determines the reference point.

If the subject intends to stop at an accumulated earnings and basic payment level below the fixed payment ($B_2 + w_2 e_3'' < F$), the resulting expected utility will be given by

$$U = \frac{B_2 + w_2 e_3'' + F}{2} - c(e_3'') + \frac{1}{2}\eta\left\{\frac{1}{2}[(B_2 + w_2 e_3'') - (B_2 + w_2 e_3'')] + \frac{1}{2}\lambda_1[(B_2 + w_2 e_3'') - F]\right\} + \frac{1}{2}\eta\left\{\frac{1}{2}[F - (B_2 + w_2 e_3'')] + \frac{1}{2}[F - F]\right\} + \eta\lambda_2(e_3'' - e_2^*)$$

The first-order conditions are then:

$$\frac{\partial U_3}{\partial e_3'} = \frac{w_2}{2} - c'(e_3'') + \frac{1}{4}\eta(\lambda_1 - 1)w_2 + \eta\lambda_2 \Rightarrow c'(e_3''^*) = \frac{w_2}{2} + \frac{w_2}{4}\eta(\lambda_1 - 1) + \eta\lambda_2$$

$$\begin{cases} B_2 + w_2 e_3'' < F & \Rightarrow \quad e_3'' < e_1^* \\ c'(e_3''^*) = \frac{w_2}{2} + \frac{w_2}{4}\eta(\lambda - 1) + \eta\lambda_2 & \Rightarrow \quad e_3''^* > e_1^* \end{cases}$$

The two inequalities are inconsistent, so there is no solution under this condition.

If the accumulated earnings and basic payment are higher than the fixed payment ($B_2 + w_2 e_3'' \geq F$), and the subject intends to stop at an accumulated earnings and basic payment level below the last day ($e_3'' < e_2^*$), the gain-loss utility is different. Receiving the accumulated earnings and basic payment now feels like a gain compared to the lower fixed payment (third term), while receiving the fixed payment now means a loss (terms equal to zero are suppressed here):

$$U = \frac{B_2 + w_2 e_3' + F}{2} - c(e_3') + \frac{1}{2}\eta\left\{\frac{1}{2}[(B_2 + w_2 e_3') - F]\right\} + \frac{1}{2}\eta\left\{\frac{1}{2}\lambda_1[F - (B_2 + w_2 e_3')]\right\} + \eta\lambda_2(e_3'' - e_2^*)$$

Then the first-order conditions are:

$$\frac{\partial U_3}{\partial e_3'} = \frac{w_2}{2} - c'(e_3'') - \frac{1}{4}\eta(\lambda_1 - 1)w_2 + \eta\lambda_2 \Rightarrow c'(e_3''^*) = \frac{w_2}{2} - \frac{w_2}{4}\eta(\lambda_1 - 1) + \eta\lambda_2$$

Now, the relationship between e_1^*, e_2^* *and* $e_3''^*$ is depended on the value of $(\lambda_2 - \frac{w_2}{4}\lambda_1 + \frac{w_2}{4})$. In our experiment, the value of $(\lambda_2 - \frac{w_2}{4}\lambda_1 + \frac{w_2}{4})$ is positive, and λ_2 is relatively larger than w_1 and w_2, thus the order $e_1^* < e_3^* < e_2^* < e_3''^*$ is highly possible. In other words, goal-based reference-dependent preferences may predict a better performance in Day 3 than in Day 2.

Result

The FT Treatment

This section reports the results from our experiment. We will first focus on the FT treatment. Recall that in this treatment, subjects must work for the full one-hour for each experimental day, and thus they can only choose the effort. The best proxy for effort on a given day in our experiment is the output (effort), that is, the number of table they counted correctly. Table 2 reports the nonwage income, output and the sign of the change of output in different mechanisms.

Table 2 FT treatment: *F*, effort and sign of output change

Id	Day 1		Day 2		Day 3		Sign of output (effort) change		
	F_1	e_1	F_2	e_2	F_3	e_3	e_2-e_1	e_3-e_1	e_3-e_2
1	18	63	7.92	60	7.92	67	-*	+	+**
2	18	70	6.8	87	6.8	104	+	+	+**
3	18	64	7.76	103	7.76	105	+	+	+**
4	18	54	9.36	81	9.36	94	+	+	+**
5	18	57	8.88	53	8.88	66	-*	+	+**
6	18	27	13.68	49	13.68	30	+	+	-
7	18	134	0	160	0	170	+	+	+**
8	18	97	2.48	108	2.48	130	+	+	+**
9	18	71	6.64	82	6.64	81	+	+	-
10	18	52	9.68	92	9.68	87	+	+	-
11	18	56	9.04	96	9.04	103	+	+	+**
12	18	85	4.4	106	4.4	108	+	+	+**
13	18	107	0.88	132	0.88	132	+	+	0**
14	18	30	13.2	34	13.2	43	+	+	+**
15	18	57	8.88	67	8.88	70	+	+	+**
16	18	54	9.36	67	9.36	87	+	+	+**
17	18	45	10.8	63	10.8	47	+	+	-
18	18	74	6.16	90	6.16	110	+	+	+**
19	18	59	8.56	80	8.56	87	+	+	+**
20	18	99	2.16	145	2.16	146	+	+	+**
21	18	129	0	167	0	186	+	+	+**
22	18	96	2.64	123	2.64	126	+	+	+**
23	18	113	0	136	0	145	+	+	+**
24	18	87	4.08	96	4.08	99	+	+	+**
25	18	46	10.64	57	10.64	76	+	+	+**
26	18	61	8.24	70	8.24	72	+	+	+**
27	18	77	5.68	95	5.68	98	+	+	+**
28	18	48	10.32	61	10.32	52	+	+	-
29	18	95	2.8	108	2.8	120	+	+	+**
30	18	99	2.16	107	2.16	113	+	+	+**
31	18	43	11.12	58	11.12	63	+	+	+**
32	18	118	0	123	0	129	+	+	+**
33	18	86	4.24	123	4.24	108	+	+	-
34	18	94	2.96	79	2.96	107	-*	+	+**
35	18	91	3.44	99	3.44	91	+	0	-
avg	18	75.37	6.14	93.05	6.14	98.63	+	+	+

Note: all subjects face a wage rate increase from 0.08 in day 1 to 0.24 in day 2 and day 3. The star represents a violation of the prediction of standard model and double star represents a violation of both the standard model and the exogenously introduced expectation-based reference model.

The first important question is whether there is a treatment effect on total effort per subject during the three different experimental days. Table 2 shows the individual statistics and the means. It shows that, during the FT treatment, subjects on average counted correctly 75.37 tables on day 1. However, the amount rises up to 93.05 on day 2, and further to 98.63 on day 3. It suggests that most subjects made higher efforts both in day 2 and day 3 comparing to day 1, indicating a large treatment effect.[9] These results are consistent with the first two categories of models. On the one hand, we see that the substitution effect has the predicted positive sign in most cases, indicating that subjects worked harder when a compensated wage increases and when they cannot choose time worked. On the other hand, loss aversion preference would make subjects strive to maintain above the fixed amount if subjects view revenue below this amount as a loss.

However, when we look at the efforts subject made on day 3, we find that most subjects violate both the predictions of canonical model and exogenous introduced expectation-based reference. Most of them (28 in 35 cases) worked harder even if the expected compensated wage decrease, even if they know they would get nothing for those extra efforts if they have already chose the fixed payment envelope. [10]

We perform a statistical test of the effect of the wage scheme changes on efforts supply. All regressions are of the form:

$$e_{it} = \alpha_i + \beta T_{it} + \gamma_i X_i + \varepsilon_{it} \tag{1}$$

where e_{it} measures the effort made by subject i in day t; α_i is a fixed effect for subject i; T_i is a dummy variable that is equal to 1 if subject is on a given wage scheme; X_i is a sery of control variable, including gender, monthly expenditure, and productivity that ; and ε_{it} is the error term. The results are shown in column 1, Table 4. The results are consistent with non-parameter test. The regression indicates that the treatment effect is highly significant and that the subjects provide significantly more effort on a high piece rate setting compared to a low piece rate setting. Moreover, the subjects provide more effort even if the expected piece rate decreases.[11] Besides, we find from the regression that on average females provide significantly less effort than males, those who consume more provide significantly less effort than those who consume less monthly.[12]

[9] Wilcoxon signrank tests show that, e2 vs. e1, z=-4.833, p=0.000; e3 vs. e1, z=-5.152, p=0.000.
[10] Wilcoxon signrank test shows that, e3 vs. e2, z=-2.942, p=0.003.
[11] Wald test shows that, the coefficient of *Day 2* is significantly smaller than the coefficient of *Day 3*, as X2=6.31, p=0.012.
[12] We add the interaction term of treatment dummy variable and these two control variables, and find no significant difference between different treatments.

The CT Treatment

Next, let's turn to CT treatment. Recall that, in this treatment, subjects are allowed to choose both work hour and work intensity as they are allowed to leave whenever they decide. Therefore, not only the hours they worked, but also the efforts they made should be considered. Table 3 summarize the results, including subjects' efforts, hours worked, productivity per correct answer, sign of output changes, and sign of productivity changes.

Similarly, the first important question is whether there is a treatment effect on total effort per subject during the three different experimental days. Table 3 shows that, during the CT treatment, subjects on average counted correctly 54.44 tables on day 1. However, the amount rises up to 82.08 and 79.64 on day 2 and day 3, respectively. As in FT treatment, except 3 subjects on day 3, almost every subject made higher efforts both in day 2 and day 3 comparing to day 1, which is consistent with the two categories of models. [13]

However, when we look at the efforts subject made on day 3, we find that most subjects (16 in 25 cases) violate both the predictions of canonical model and exogenous introduced expectation-based reference. Most of them worked harder even if the expected compensated wage decrease, even if they know they would get nothing for those extra efforts if they have already chose the fixed payment envelope, although the efforts they made has no statistically significant changes. [14]

A statistical test is presented in regression (2) in table 4, which further confirms the results of non-parameter test. The subjects provide significantly more effort on a high piece rate setting compared to a low piece rate setting, and at least the same level of effort even if the expected piece rate decreases.[15] However, we find from regression (2) that although on average those who consume more provide significantly less effort than those who consume less monthly, different from regression (1), males provide significantly less effort than females when they are allowed to leave before time was ran out. It means males are more responsive to the leave or not choice. [16]Figure 1 provides a vivid image.[17]

---

[13] Wilcoxon signrank tests show that, e2 vs. e1, z=-4.373, p=0.000; e3 vs. e1, z=-3.862, p=0.000.

[14] Wilcoxon signrank test shows that, e3 vs. e2, z=-0.337, p=0.7358.

[15] Wald test shows that, the coefficient of *Day 2* is significantly smaller than that of *Day 3*, as X2=2.07, p=0.157.

[16] This is independent from the treatment effect, as regression model incorporating the interaction term of gender and day shows no significance regarding the interaction term.

[17] We run another regression model in which the interaction term of female and the treatment dummies are included, and found no significant gender effect differences between different treatments.

Table 3　CT treatment: effort, time, productivity and sign of change

id	Day 1			Day 2			Day 3			Sign of output (effort) changes			Sign of work time changes			Sign of productivity change		
	e_1	T_1(s)	P_1(s)	e_2	T_2(s)	P_2(s)	e_3	T_3(s)	P_3(s)	e_2-e_1	e_3-e_1	e_3-e_2	T_2-T_1	T_3-T_1	T_3-T_2	P_2-P_1	P_3-P_1	P_3-P_2
1	52	3600	69.2	53	3600	67.9	67	2868	42.8	+	+	+**	0	-	-	-	-	-
2	113	3600	31.9	150	3600	24	162	3600	22.2	+	+	+**	0	0	0	-	-	-
3	25	1673	66.9	34	2114	62.2	33	1596	48.4	+	+	-	+	-	-	-	-	-
4	28	1131	40.4	60	2389	39.8	60	2054	34.2	+	+	0**	+	+	-	-	-	-
5	60	2703	45.1	82	3092	37.7	85	2705	31.8	+	+	+**	+	+	-	-	-	-
6	29	1642	56.6	100	3600	36	30	1013	33.8	+	-*	-	+	-	-	-	-	-
7	37	3223	87.1	48	3582	74.6	29	1939	66.9	+	-*	-	+	-	-	-	-	-
8	58	3600	62.1	69	3600	52.2	72	3247	45.1	+	+	+**	0	-	-	-	-	-
9	65	3600	55.4	75	3600	48	89	3600	40.4	+	+	+**	0	0	0	-	-	-
10	94	3600	38.3	124	3600	29	100	2577	25.8	+	+	-	0	-	-	-	-	-
11	74	3600	48.6	131	3600	27.5	131	3202	24.4	+	+	0**	0	-	-	-	-	-
12	54	3600	66.7	76	3600	47.4	82	3600	43.9	+	+	+**	0	0	0	-	-	-
13	100	3522	35.2	121	3306	27.3	75	1926	25.7	+	-*	-	-	-	-	-	-	-
14	25	1796	71.9	46	2309	50.2	46	2178	47.3	+	+	0**	+	+	-	-	-	-
15	24	2270	94.6	41	3260	79.5	38	2768	72.8	+	+	-	+	+	-	-	-	-
16	52	3600	69.2	68	3600	52.9	69	3366	48.8	+	+	+**	0	-	-	-	-	-
17	10	765	76.5	30	1789	59.6	15	1039	69.3	+	+	-	+	+	-	-	-	+
18	110	3424	31.1	161	3600	22.4	179	3600	20.1	+	+	+**	+	0	0	-	-	-
19	81	3600	44.4	137	3600	26.3	165	3600	21.8	+	+	+**	0	0	0	-	-	-
20	16	873	54.6	60	2701	45	60	2511	41.9	+	+	0**	+	+	-	-	-	-
21	58	3500	60.3	77	3600	46.8	85	3600	42.4	+	+	+**	+	+	0	-	-	-
22	28	2378	84.9	54	3600	66.7	69	3600	52.2	+	+	+**	+	+	0	-	-	-
23	53	3600	67.9	80	3600	45	73	3600	49.3	+	+	-	0	0	0	-	-	+
24	17	1352	79.5	43	1841	42.8	41	2170	52.9	+	+	-	+	+	+	-	-	+

25	98	3600	36.7	132	3600	27.3	136	3600	26.5	+	+	+**	0	0	0	-	-	-

Note: all subjects face a wage rate increase from 0.08 in day 1 to 0.24 in day 2 and day 3. The star represents a violation of the prediction of standard model and double star represents a violation of both the standard model and the exogenously introduced expectation-based reference model.

Table 4 the treatment effects of different wage schemes on effort under different settings

Dv:	(1) effort	(2) effort	(3) productivity
Day 2	17.686***	27.640***	-13.487***
	(2.262)	(3.783)	(1.527)
Day 3	23.257***	25.200***	-17.781***
	(2.342)	(4.447)	(1.418)
Gender(female=1)	-23.467**	71.289***	-6.874
	(8.994)	(9.268)	(4.388)
Monthly Expenditure	-8.267**	-27.111***	-0.867
	(4.069)	(6.049)	(2.599)
Productivity	5.533***	2.267	-7.611***
	(1.590)	(1.998)	(0.863)
Mean Constant	75.37***	54.44***	48.587***
Individual Fixed effect	yes	yes	Yes
Observations	105	75	75
R-squared	0.95	0.91	0.95

Notes: Robust standard errors are in parentheses. The proxy for productivity is the effort subjects provided during the first stage on day 1. Significance at the 1%, 5%, and 10% level is denoted by ***, **, and *, respectively.

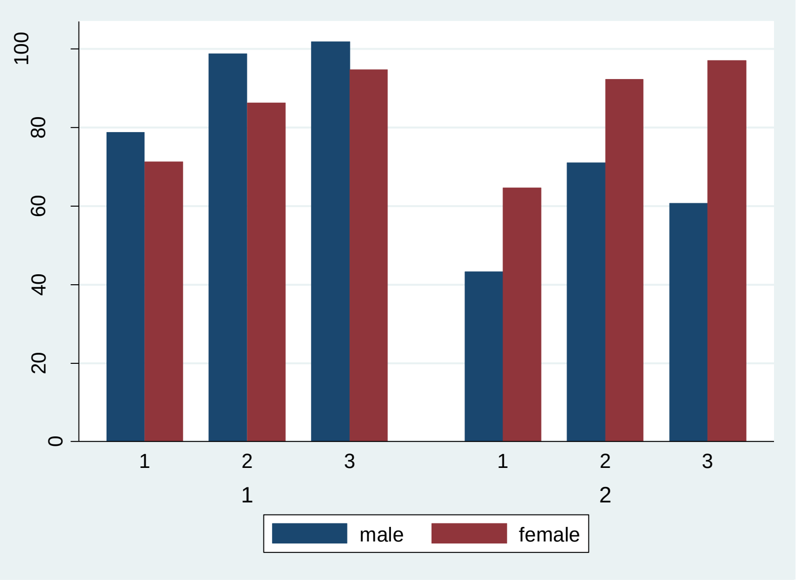

Figure 1　Gender difference in effort supply under different wage schemes

After we documented the strong impact of the wage increase on total labor supply, the natural question is whether both the time worked and the effort per correct answer increased. Firstly, let's consider the impact of the effect of the wage increase on the time worked. The sixth column of table 3 provides an indication of how wage increase affected time worked. It shows that in most cases, subjects worked at least not shorter on both day 2 (24/25) and day 3 (16/25) than on day 1. On average, subjects would spend 2,794.11 seconds on labor on day 1, and the average time worked increased to 3215.30 seconds on day 2, indicating a significantly large substitution effect.[18]

[18] Wilcoxon signrank test shows that, T1 vs. T2, z=-3.261, p=0.0011. Indeed, even then the subjects were free to

However, the average time worked declined to 2782.44 seconds on day 3, which is not significantly different from that on day 1.[19]

We might also be interested in looking at the effect of the wage changes on subjects' productivity during the three experimental days in CT treatment. As the seventh column of table 3 shows, the productivity of each subject increased on both day 2 and day 3 comparing to day 1. On average, it would cost subject 59 seconds to count a correct table on day 1; however, the average time declined to 45.52 seconds on day 2 and further to 41.23 seconds on day 3. [20] These results suggest both the increase of time worked and the productivity are contributed to the increase on total labor supply of subjects on both day 2 comparing to day 1. However, the increase of effort supply on day 3 comparing to that on day 1 is mainly due to the increase of productivity.

Considering the change of time worked and productivity from day 2 to day 3 under risk, however, table 3 shows that, while subjects were free to choose work time, most subjects (16 in 25 cases) chose to leave before the available time was exhausted on day 3, resulting in a fact that they worked significantly shorter than day 2. [21] However, even though the time they worked decreased, subjects achieved higher level of output on day 3 due to the significant increase of productivity.[22] This is consistent with the observation from Dickson (1999) that subjects would work harder, substitute on-job-leisure with off-job-leisure when they could choose work time.

A regression model that uses productivity as dependent variable and independent variable same as function (1) demonstrates the result. The column 4 of Table 4 shows that, subjects worked more efficient on day 2 comparing to day 1, and on day 3 comparing to day 2[23].

Discussion and conclusion

In this paper, we investigate the determinants of labor supply under risk by employing a series of real effort lab experiments. Subjects are recruited for a one-hour real effort experiment each day for three days, facing various incentive schemes in different days: a high basic payment with a low piece rate, a low basic payment with a high piece rate, and a high piece rate but with uncertainty. Subjects could choose both the work time and effort in one treatment and could only choose the effort in another treatment. The results show that, consistent with neoclassical model, most subjects in both treatments provided more efforts in the second day comparing to the first day as the piece rate rose up, indicating a significant substitution effect. However, inconsistent with neoclassical model and Prospect Theory, most subjects in both treatments provided at least no less efforts in the third day comparing to the second

choose their work time, 15 of them rejected to leave before time ran out.

[19] Wilcoxon signrank test shows that, T1 vs. T3, z=0.054, p=0.9567.

[20] Wilcoxon signrank tests show that, p2 vs. p1, z=4.372, p=0.000; p3 vs. p1, z=4.372, p=0.000.

[21] Wilcoxon signrank test shows that, T2 vs. T3, z=3.509, p=0.0005.

[22] Wilcoxon signrank test shows that, p2 vs. p3, z=2.892, p=0.0038.

[23] Wald test show, X2=8.98, p=0.0043.

day, although they faced an uncertainty on getting their accumulated piece rate earnings. Interviews with subjects after the experiments suggest that an important reference points emerged, that is, the goal to surpass previous performances. Such behavioral pattern is consistent with multiple-reference points theory which suggests that to be success is a more important motivation than loss aversion. In a word, although classic model predicts rather well when there is no risk, it fails to make good prediction in risky circumstances. Moreover, although single reference point based model such as prospect theory cannot consistently account for the data when there is a risk, multi-reference point based model could fit the data well.

To sum up, our data support the main prediction of neoclassical model when the increase of piece rate is certain. However, the prediction of neoclassical model fails when the payment is uncertain. Moreover, although single reference point based model such as prospect theory cannot consistently account for the data when there is a risk, multi-reference points based model could fit the data well. We believe that these results both contribute to a deeper understanding of the behavioral foundations of labor supply, and of the open question for reference-dependent preferences: what determines the reference point? This has significant policy implications. On the one hand, if, for example, the intertemporal substitution of labor supply is high, one may interpret the large variations in employment during business cycles as voluntary choices by the workers rather than involuntary layoffs. Intertemporal substitution also plays a crucial role in the propagation of shocks across periods (David Romer 1996; Robert G. King and Sergio Rebelo 1999). On the other hand, a finding that labor supply is reference-dependent means that the usual estimates of wage and income elasticities are likely to be misleading, which may further mislead the evaluation of much government policy regarding tax and transfer programs.

Reference

Abeler, J.; Falk, A.; Götte, L. and Huffman, D. "Reference Points and Effort Provision." *The American Economic Review*, 2011, *101*(2), pp. 470-92.

Ariely, Dan; Gneezy, Uri; Loewenstein, George and Mazar, Nina. "Large Stakes and Big Mistakes." *The Review of Economic Studies*, 2009, *76*(2), pp. 451-69.

Battalio, RC; Green, L and Kagel, JH. "Income-Leisure Tradeoffs of Animal Workers." *The American Economic Review*, 1981, *71*(4), pp. 621-32.

Battalio, RC; Kagel, JH and MacDonald, DN. "Animals' Choices over Uncertain Outcomes: Some Initial Experimental Results." *The American Economic Review*, 1985, *75*(4), pp. 597-613.

Bell, David E. "Disappointment in Decision Making under Uncertainty." *Operations research*, 1985, *33*(1), pp. 1-27.

Brüggen, Alexander and Strobel, Martin. "Real Effort Versus Chosen Effort in Experiments." *Economics Letters*, 2007, *96*(2), pp. 232-36.

Cadsby, C Bram; Song, Fei and Tapon, Francis. "The Impact of Risk Aversion and Stress on the Incentive Effect of Performance Pay." *unpublished paper, University of Guelph*, 2009.

Camerer, Colin; Babcock, Linda; Loewenstein, George and Thaler, Richard. "Labor Supply of New York City Cabdrivers: One Day at a Time." *The Quarterly Journal of Economics*, 1997, *112*(2), pp. 407-41.

Charness, Gary and Kuhn, Peter. "Lab Labor: What Can Labor Economists Learn from the Lab?" in O. Ashenfelt and D. Card (ed.), *Handbook of Labor Economics*, 2011, *4*, pp. 229-330.

Chou, Yuan K. "Testing Alternative Models of Labour Supply: Evidence from Taxi Drivers in Singapore." *The Singapore Economic Review*, 2002, *47*(01), pp. 17-47.

Crawford, Vincent P and Meng, Juanjuan. "New York City Cab Drivers' Labor Supply Revisited: Reference-Dependent Preferences with Rationalexpectations Targets for Hours and Income." *The American Economic Review*, 2011, *101*(5), pp. 1912-32.

Dickinson, D.L. "An Experimental Examination of Labor Supply and Work Intensities." *Journal of Labor Economics*, 1999, *17*(4), pp. 638-70.

Farber, H. S., 'Is tomorrow another day? The labor supply of New York City cabdrivers', *Journal of Political Economy*, 2005, 113(1), 46–82.

Farber, Henry S. "Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers." *The American Economic Review*, 2008, *98*(3), pp. 1069-82.

Fehr, Ernst and Goette, Lorenz. "Do Workers Work More If Wages Are High? Evidence from a Randomized Field Experiment." *The American Economic Review*, 2007, pp. 298-317.

Gneezy, U and Rustichini, A. "A Fine Is a Price." *The Journal of Legal Studies*, 2000, *29*(1), pp. 1-17.

Goette, L.; Huffman, D. and Fehr, E. "Loss Aversion and Labor Supply." *Journal of the European Economic Association*, 2004, *2*(2 鈥?), pp. 216-28.

Gul, F. "A Theory of Disappointment Aversion." *Econometrica: Journal of the Econometric Society*, 1991, *59*(3), pp. 667-86.

Ham, John C. "Estimation of a Labour Supply Model with Censoring Due to Unemployment and Underemployment." *The Review of Economic Studies*, 1982, *49*(3), pp. 335-54.

Heath, Chip; Larrick, Richard P and Wu, George. "Goals as Reference Points." *Cognitive psychology*, 1999, *38*(1), pp. 79-109.

Kahneman, D and Tversky, A. "Prospect Theory: An Analysis of Decision under Risk." *Econometrica: Journal of the Econometric Society*, 1979, *47*(2), pp. 263-91.

Kahneman, D.; Knetsch, J. L. and Thaler, R. H. "Experimental Tests of the Endowment Effect and the Coase Theorem." *Journal of Political Economy*, 1990, *98*(6), pp. 1325.

Kőszegi, Botond and Rabin, Matthew. "A Model of Reference-Dependent Preferences." *The Quarterly Journal of Economics*, 2006, *121*(4), pp. 1133-65.

____. "Reference-Dependent Risk Attitudes." *The American Economic Review*, 2007, pp. 1047-73.

List, John A and Rasul, Imran. "Field Experiments in Labor Economics," in O. Ashenfelt and D. Card (ed.), *Handbook of Labor Economics.* 2011, 103-228.

Loomes, G. and Sugden, R. "Disappointment and Dynamic Consistency in Choice under Uncertainty." *The Review of Economic Studies*, 1986, pp. 271-82.

Loomes, Graham and Sugden, Robert. "Testing for Regret and Disappointment in Choice under Uncertainty." *The Economic Journal*, 1987, *97*, pp. 118-29.

Neale, Margaret Ann and Bazerman, Max H. *Cognition and Rationality in Negotiation*. Free Press New York, 1991.

Plott, CR. "Will Economics Become an Experimental Science?" *Southern Economic Journal*, 1991, *57*(4), pp. 901-19.

Shalev, Jonathan. "Loss Aversion Equilibrium." *International Journal of Game Theory*, 2000, *29*(2), pp. 269-87.

Sillamaa, Mary-Anne. "How Work Effort Responds to Wage Taxation: A Non-Linear Versus a Linear Tax Experiment." *Journal of Economic Behavior & Organization*, 1999b, *39*(2), pp. 219-33.

____. "How Work Effort Responds to Wage Taxation: An Experimental Test of a Zero Top Marginal Tax Rate." *Journal of Public Economics*, 1999a, *73*(1), pp. 125-34.

Swenson, C.W. "Taxpayer Behavior in Response to Taxation: An Experimental Analysis." *Journal of Accounting and Public Policy*, 1988, *7*(1), pp. 1-28.

Wang, XT and Johnson, Joseph G, "A Tri-Reference Point Theory of Decision Making under Risk", *Journal of Experimental Psychology-General*, 2012, 141(4), pp 743.

Tversky, Amos and Kahneman, Daniel. *Choices, Values, and Frames*. Cambridge University Press, 2000.