DISCUSSION PAPER SERIES

IZA DP No. 1002

Do Workers Work More When Wages Are High?

Ernst Fehr
Lorenz Götte

January 2004

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

Do Workers Work More When Wages Are High?

Ernst Fehr
University of Zurich, CEPR and IZA Bonn

Lorenz Götte
University of Zurich and IZA Bonn

Discussion Paper No. 1002
January 2004

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
Email: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of the institute. Research disseminated by IZA may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit company supported by Deutsche Post World Net. The center is associated with the University of Bonn and offers a stimulating research environment through its research networks, research support, and visitors and doctoral programs. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available on the IZA website (www.iza.org) or directly from the author.

IZA Discussion Paper No. 1002
January 2004

ABSTRACT

Do Workers Work More When Wages Are High?

The canonical model of life-cycle labor supply predicts a positive response of labor supplied to transitory wage changes. We tested this prediction by conducting a randomized field experiment with bicycle messengers. In contrast to previous studies we can observe in which way working hours as well as effort respond to a wage increase and we have full control regarding the workers' anticipation of the wage increase. The evidence indicates that workers increase monthly working time and decrease their daily effort but since the working time effect dominates the effort effect overall labor supply increases. The decrease in daily effort contradicts the canonical model of intertemporal labor supply with time separable preferences, since the wage in our experiment directly rewarded effort. We show that a simple model of loss averse, reference dependent, preferences can account for both the increase in working time and the decrease in daily effort. Moreover, we elicit independent individual measures of loss aversion and show that workers who are more prone to loss aversion are more likely to reduce effort in response to higher wages. Our model and our results also reconcile the seemingly contradictory evidence reported in previous studies (Camerer et al. 1997, Oettinger 1999) of high frequency labor supply.

JEL Classification: J22, C93, B49

Keywords: labor supply, intertemporal substitution, loss aversion

Corresponding author:

Lorenz Götte
Institut für Empirische Wirtschaftsforschung
University of Zurich
Blümlisalpstrasse 10
8006 Zürich
Switzerland
Email: lorenz@iew.unizh.ch

The intertemporal substitution of labor supply has far-reaching implications for the interpretation of important phenomena. If, for instance, the intertemporal substitution of labor supply is high, one may interpret the large variations in employment during business cycles as arising from voluntary choices by the workers rather than being caused by job rationing. It also plays a crucial role in the propagation of shocks across periods (Romer, 1996; King and Rebelo, 1999). Previous studies have found little evidence for intertemporal substitution of labor. Often, the estimated elasticities are small and statistically insignificant, and sometimes even negative (see, e.g., Mankiw, Rotemberg and Summers 1985; Pencavel 1986; Altonji 1986; Blundell, 1994; Card 1994 and Blundell and MaCurdy 1999).[1]

However, the low estimates of intertemporal substitution are difficult to interpret because of serious limitations in the available data. The life cycle model of labor supply predicts intertemporal substitution with regard to *transitory* wage changes or wage changes that are *anticipated* by the workers. Yet, the typical wage changes are not transitory; hence they are associated with significant income effects. In addition, it seems almost impossible to reliably infer from existing data whether the workers anticipated the wage change. Another problem arises, if labor markets are characterized by a significant amount of job rationing so that the workers are not unconstrained in their labor supply choices.[2] Moreover, since wages are determined both by supply and demand conditions serious endogeneity problems arise.[3] Thus, if one uses the typically available data one has to make a host of auxiliary assumptions when testing the life cycle model of labor supply.

In this paper, we use an ideal data set to study workers' responses to transitory wage changes. We conducted a randomized field experiment at a bicycle messenger service in Zürich, Switzerland. The bicycle messengers receive no base wage that is independent of effort. They are paid solely a share of the daily revenues they generate. For all the workers we have precise information on the number of shifts they work[4], and the revenues they generate per shift. The key feature of our experiment is the implementation of an *exogenous* and *transitory* increase in the revenue share by 25 percent. Therefore, we can be sure that the wage change was not induced by unobserved supply or demand variations. Each participant in

[1] After reviewing a sizeable part of the literature Card (1994) concludes, for instance, that the "very small magnitude of the estimated intertemporal substitution elasticities" can only account for a tiny fraction of the large person-specific year-to-year changes in labor supply.

[2] In countries like Germany, for instance, firms often reduce the weekly working time during recessions. Typically, this reduction is binding for all individual workers, i.e., individuals have only the choice to accept the reduced working time or to quit.

[3] Oettinger (1999) shows that if one neglects the endogeneity of wage changes, estimates of labor supply elasticities are severely downward-biased.

[4] A shift always comprises five hours. Typically (in 99 percent of the cases), workers work at most one shift per day.

the experiment knew ex-ante the precise duration and the size of the wage (revenue share) increase. Since the wage was increased only for one month, its impact on the workers' lifetime wealth is negligible. In the firm under study, the messengers can freely choose how many shifts (hours) they will work, and how much effort they exert (to generate revenues). Our experiment raises the returns to increasing both the number of shifts and effort per shift. Therefore, we have the unique opportunity to study how hours *and* effort respond to the wage increase.

Our experimental results indicate that workers supply significantly more shifts. The elasticity of shifts with respect to wages is approximately 0.8. However, we also observe a significantly negative effort response. The elasticity of the revenue generated during a shift with respect to wages is roughly –0.3. This means that workers decrease their effort per shift in response to the wage increase. Since the percentage increase in shifts is larger than the decrease in revenue per shift, the overall revenue generated by a worker increases over the experiment, indicating an overall increase in labor supply. The effect on effort per shift remains robustly negative even if we control for individual fixed effects, for daily fixed effects, for workers' fatigue, for competition among the workers and for workers' experience. Thus, the canonical model of intertemporal labor supply based on time separable preferences cannot account for the decrease in effort per shift. This raises the question whether there are alternative models that can explain the simultaneous increase in the number of shifts and the decrease in effort per shift. We show that a simple model of loss averse, reference dependent, preferences can account for both of these facts.

The intuition behind our model is that workers with loss averse preferences have a daily reference income level.[5] Daily incomes below the reference level are experienced as a "loss" and in the loss domain the marginal utility of income is large. In contrast, at and above the reference level the marginal utility of income discontinuously decreases to a lower level. Workers who temporarily earn higher wages are more likely to exceed the reference income level and, hence, their marginal utility of income is low, inducing them to provide less effort. At the same time, however, workers with higher wages have a higher overall utility from working a shift so that they can more easily cover the fixed costs of getting to work. Hence, they are more likely to work.

---

[5] Heath, Larrick and Wu (1999) provide evidence that goals often serve the function of a reference point. Camerer, Babcock, Loewenstein and Thaler (1997), who report a negative labor supply elasticity, also interpret their results in terms of loss aversion.

Loss aversion and reference dependent choices have been documented in many domains (see, e.g. Kahneman and Tversky 2000). Loss aversion has, in particular, been shown to affect behavior under uncertainty. In order to examine the link between loss aversion and the negative effect of wages on effort per shift in more detail we elicited the degree of loss aversion of individual workers in a follow-up experiment. This experiment gives us an independent measure of loss aversion that can be linked to the workers' effort responses. It turns out that a higher degree of individual level loss aversion is associated with a larger reduction in effort per shift. In fact, our results suggest that *only* those workers who exhibit loss aversion in the follow-up experiment generate less revenue per shift. Workers who do not exhibit loss aversion generate the same revenue per shift. Interestingly, our measure of loss aversion is not related to the number of shifts worked, as predicted by our model. This provides fairly strong support for the interpretation of our results in terms of loss aversion.

Our model also reconciles the seemingly contradictory evidence in the two previous studies of intertemporal labor substitution based on high frequency data. Camerer, Loewenstein, Babcock and Thaler (1997) examined how New York City taxi drivers, having decided to work on a given day, vary their daily working time (which is a good proxy for daily effort) in response to wage variations. They report that workers work less hours (provide less effort) on high-wage days indicating a negative effort elasticity. Oettinger (1999) investigated how stadium vendors adjust their probability of working in response to transitory wage variations. He reports a significant increase in the probability of working when wages are high implying a participation elasticity of around 0.6. At first sight it seems that the results of Camerer et al. (1997) and Oettinger (1999) contradict each other. However, when viewed from the perspective of our simple model of loss aversion the contradiction vanishes. A model with loss averse preferences predicts that a wage increase induces taxi drivers to provide less hours per day but it also predicts that stadium vendors are more likely to work on high wage days.

Recent empirical work has shown that – in simple task environments – piece rates have powerful incentive effects compared to flat time-based payment schedules (Lazear 2000, Paarsch and Shearer 2000). It is worthwhile to stress that this effect is also fully consistent with loss aversion. In a flat time-based scheme workers get paid the same amount irrespective of how hard they work during labor time while in a piece rate system they can increase their utility by working harder. The concept of loss aversion does not imply that incentives do not work. Loss aversion "only" means that – once the income target is met – there is a discontinuous drop in the marginal utility of income, which diminishes the incentive to provide effort above the target income. In fact, if the parameters of the piece rate incentive scheme imply that individuals are often below their income target, loss averse preferences predict particularly strong incentive effects of piece rates because the marginal utility of income is very high below the income target. Yet, even if the individuals are above their income target most of the time, piece rates provide stronger incentives than time-based schemes.

The remainder of this paper is structured as follows. Section I describes the institutional environment and the details of the field experiment. Section II presents a canonical life-cycle model of labor supply. We use the model to generate the predictions for our experiment. Section III reports the results from the field experiment. In Section IV we discuss two interpretations of our data and present a simple model of loss averse preferences. This section also describes the follow-up experiment and discusses the link between individual loss aversion and workers effort responses. Section V concludes the paper.

I. Experimental Set-up

In this section, we describe our experimental set-up. Our study is based on the complete delivery records of two relatively large messenger services in Zurich, Veloblitz, and Flash Delivery Services (henceforth “Flash”). Each firm employs between 50 and 60 messengers per month. We first describe the organization of work at a bicycle messenger service. There are three important features. First, as we will explain in more detail below, messengers can freely choose how many shifts they work and how much effort they exert. Second, in both firms messengers receive no fixed wage. Instead, each individual receives a fixed share of the revenue that he or she generates. Third, the demand for messenger services is highly volatile across days. This is important, because it implies that messengers are familiar with substantial variations in daily earnings. Hence, if learning is important to understand the logic of inter-temporal substitution, then our subjects are a well-trained set of subjects and should respond accordingly during the experiment.

At Veloblitz, we implemented a large and fully anticipated temporary variation in revenue shares. Messengers were randomly assigned to one of two treatment groups, A or B. For group A, we implemented a 25 percent increase in the revenue share during four weeks in September 2000, for group B in November 2000. During both treatment periods, this leaves the other messengers at Veloblitz and all messengers at Flash as control group.

A. Work at a Messenger Service

Unless pointed out below explicitly, the arrangements are the same for the two messenger services, Veloblitz and Flash. In order to be accepted by one of the messenger services, an experienced messenger evaluates the applicant with respect to fitness, knowledge of locations and names of streets, courtesy, and skills regarding handling the CB radio.

Hours and Effort

Once accepted as an employee, messengers can freely choose how many five-hour shifts they will work. On each day between Mondays and Fridays, there are about 30 shifts available at Veloblitz, and about 22 at Flash. At the messenger service's office, the shifts are displayed on a shift plan for every calendar week. There are two types of shifts, called "fixed" and "variable". A "variable" shift simply means that a shift is vacant on a particular day. Any messenger can sign up to work that shift, e.g., on Wednesday from 8 am to 1 pm. If a messenger commits to a "fixed" shift, this means that he will work that shift every week. For example, if a messenger chooses Wednesday, 8 am – 1 pm as a fixed shift, he will have to fill that shift on every Wednesday. Fixed shifts can only be cancelled with at least four weeks notice. Roughly two thirds of the shifts are fixed. All other shifts are variable and available for any messenger to sign up for. Two additional items are worth mentioning. First, at both of the messenger services, there is no minimum number of shifts that the messengers have to work. Second, both messenger services have found it difficult to fill the available shifts. Almost always there is at least one unfilled shift and, on average, almost 3 shifts per day remain unfilled. During the period September 1999 – August 2000, approximately 60 shifts remained unfilled every month. This implies that messengers are unlikely to be rationed in the choice of shifts. Figure 1a provides descriptive statistics on the number of shifts the experimental subjects at Veloblitz worked per month. To provide a standard of comparison we also display the number of shifts the bicycle messengers at Flash worked per month (see Figure 1b). The figure shows that prior to the experiment – between September 1999 and August 2000 – the Veloblitz subjects worked on the average 10 shifts per month. However, some messengers worked considerably more shifts, as indicated by the number shifts chosen by the 75th percentile of the workforce, and some worked considerably fewer shifts than average (see the number of shifts at the 25th percentile).

Insert Figures 1a,b here

Messengers' earnings are given solely as a percentage w of their daily revenues. Hence, if a messenger carries out deliveries that generate revenues r, his earnings on that day will be wr. Importantly, messengers have substantial discretion over how much effort to put into work. During a shift they stay in contact with the dispatcher at the messenger service's office only through CB radio. In order to allocate a delivery, say, from location A to location B, the dispatcher will contact the messenger that he thinks is closest to A to pick up the delivery. All messengers can follow the radio. If they believe that they are closer to A than the messenger that was originally contacted, they can get back to the dispatcher and say so and will then be allocated to that delivery. Conversely, if the messenger does not want to carry out the delivery from A to B, he may just not respond to the call. Thus, messengers have several means of increasing the number of deliveries they complete. They can drive at higher speed, follow the radio more actively, lobby for additional deliveries aggressively, or find the shortest possible ways to carry out a delivery.

Thus, work at a bicycle messenger service comes very close to a model where individuals are unconstrained in choosing how many shifts (hours) to work, and how hard to work (how many deliveries they complete during a shift).

The Demand for Messenger Services

As part of the experimental setup, we obtained the complete records of all the deliveries at Veloblitz and Flash between January 1999 and November 2000. These records contain every single delivery that a messenger carried out on a particular date. They allow us to track precisely when a messenger worked a shift and they contain all deliveries and their prices.

Figure 2 displays the evolution of the total number of normalized deliveries per day, carried out by Veloblitz and Flash. The time period spans working days from January 1999 to November 2000, with the exception of a few days in October 2000, for which the Flash records are missing. Since Flash also employs car messengers, we distinguish between total deliveries and deliveries carried out by bicycle messengers at Flash. All three series are normalized by the value of their first observation, because the messenger services requested that the number of deliveries is not made available to their competitors. Figure 2 shows that both firms grew by approximately 50 percent over the two years considered. It is striking how strongly total deliveries by the two messenger services are correlated. The correlation of the total deliveries per day between the two firms is 0.75. The figure also shows that there is a steady decrease in the share of deliveries carried out by bicycle messengers at Flash.

However, the daily deliveries of the bicycle messengers at Flash also display strong swings at exactly the same dates when Veloblitz experiences swings and the correlation is 0.71. Figure 2 makes transparent that Flash and Veloblitz operate in the same market. Hence, in the econometric estimates below, we can use the Flash bicycle messengers as a useful control group to reliably control for daily changes in demand.

Insert Figure 2 here

Volatility in Earnings

Figure 2 also suggests that the messengers' daily earnings are highly volatile. The average hourly earnings for messengers are roughly CHF 25 at Veloblitz and slightly less at Flash. Yet, at both messenger services, the number of deliveries and, hence, aggregate firm earnings fluctuate strongly, as Figure 2 shows. Figures 3a and 3b show that this variation in aggregate earnings is also associated with strong fluctuations of individual messengers' revenues per shift. On the average the experimental subjects at Veloblitz generated roughly CHF 300 revenue per shift between September 1999 and August 2000. The within-subject standard deviation of revenues per shift is, however, rather high. It is rarely below CHF 50 and in several months it comes close to CHF 100. This means that the hourly earnings regularly vary between CHF 18.5 and CHF 32.5. Moreover, the figure also indicates that the between-subject standard deviation in earnings is typically much lower than the within-subject standard deviation. The picture for the control firm exhibits similar features (see Figure 3b). This suggests that demand variations are a strong source of the daily fluctuations in the messengers' earnings (revenues).

Insert Figures 3a,b here

The important point here is that messengers are familiar with substantial variations in earnings opportunities over time. Hence, the wage change implemented by our experiment and described in more detail below, varies wages in a range that is very familiar to the messengers.

B. The Experimental Design

In order to evaluate the impact of an anticipated wage increase on behavior, we conducted the following field experiment at Veloblitz (see also Figure 4 which summarizes the design of the experiment): The messengers who participated in the experiment were randomly assigned to one of two groups, A or B.[6] During Treatment A the members of group A received a large wage increase while in Treatment B the members of group B received the same wage increase. During Treatment A the members of group B did not receive a wage increase, i.e. they served as a control group whose behavior can be compared with the behavior in treatment group A. During Treatment B the members of group A did not receive a wage increase, i.e., they served as a control for the treatment group B. In addition there is also a "field control group" consisting of those messengers at Veloblitz who did not participate in any of the treatments and the bicycle messengers at Flash.

The wage increase took the form of an increase in the revenue share. Recall that the messengers' compensation is a percentage w of their individual daily revenues. Currently, w = 0.39 for males, and 0.44 for females at Veloblitz. Male and female messengers in group A received a (roughly) 25 percent higher revenue share of w = 0.49 and w = 0.54, respectively, during the four weeks between September 11th and October 6th 2000. For members of group B, we increased the share by the same percentage during the four weeks between October 30th and November 24th 2000. The additional earnings were financed by the Swiss National Science Foundation[7] and paid out on December 8th for both groups.

The wage increase and the participation rules were communicated to the messengers on August 29th in a presentation at the Veloblitz office. Moreover, posters at the Veloblitz office and handouts that were placed all over the office ensured that all messengers were informed about the experiment even if they did not attend the presentation. The only constraint for participation in the experiment was that, in order to receive the wage increase, messengers had to fill in a questionnaire at the beginning and at the end of each treatment. After distribution, the messengers had to complete and return the questionnaire roughly within 10 days. If a messenger worked within that period of time, but failed to return the questionnaire, he or she was excluded from the experiment and received no payoff.[8] Hence, in order to participate, messengers had to fill in at most four questionnaires.

Insert Figure 4 here

The messengers did not know that the purpose of the experiment was the study of labor supply behavior. They did not know that we received the full (anonymous) records of each

[6] The randomization was based on the administrative codes that the messenger service uses to identify a messenger in its accounting system. The first messenger that worked at Veloblitz was assigned the number 1, the second 2, and so forth. Messengers with odd numbers were assigned to group A, messengers with even numbers to group B.

[7] Under project number 1214-051000.97.

[8] One of the authors was available for questions regarding the questionnaires every Monday and Friday throughout the experimental period.

messenger about the number of shifts and the number of deliveries completed, either. If pressed, we told the participants that we wanted to study the relation between wages and job satisfaction. The purpose of our study was credible because the questionnaires contained several questions related to job satisfaction.[9]

The experiment as such represents an important innovation to the existing literature for several reasons. First, it implements a fully anticipated, temporary and exogenous variation in the (output based) wage rates of the messengers, which is key to studying the intertemporal substitution of labor. The experimental wage increase was massive. It amounts to a roughly 25 percent higher wage during four weeks, and provides a clear incentive to work more and work harder. Moreover, the participating messengers are experienced, and daily fluctuations in their earnings are common. Hence, we experimentally implement a wage change in an otherwise familiar environment. Second, the data we obtained from Veloblitz allows us to study two dimensions of labor supply: Hours as measured by the number of shifts and effort as measured by the revenues generated per shift or the number of deliveries per shift. No other study that we are aware of can look at these two dimensions simultaneously. Third, we can combine the data set with the full records from a second messenger service operating in the same market. This will prove useful for investigating any effect that the experiment might have had on the control group at Veloblitz, and helps to control for daily fluctuations in demand.

C. Treatment Effects

Three effects will play a key role in our analysis below. We call them the direct treatment effect, the indirect treatment effect, and the announcement effect. Figure 4 helps to illustrate how each of these effects is identified.

1. The **direct treatment** effect is defined as the impact of the wage increase on the treatment group's behavior (shifts worked, revenue generated per shift) relative to the experimental and the field control group during both treatments. To measure this effect econometrically we use a dichotomous variable that equals 1 for all messengers of the treatment group during the

[9] These features of the experiment ensure that our results cannot be affected by the Hawthorne effect. The Hawthorne effect means that subjects behave differently just because they know that the experimenters observe their behavior. Yet, our subjects did not know that we could observe their behavior during the wage increase. Moreover, since both the treatment group and the control group are part of the overall experiment, and since our key results rely on the comparison between these groups we control for a potential Hawthorne effect.

treatment period whereas for the messengers of the experimental and the field control group the variable is set to zero.

2. The **indirect treatment** effect is defined as the impact of the experiment on the behavior of *all* the messengers at Veloblitz relative to all the bicycle messengers at Flash during the treatment periods. To measure this effect econometrically we use a dichotomous variable that equals 1 for all messengers at Veloblitz during the treatment period and which equals zero for the Flash bicycle messengers.

3. The **announcement effect** is defined as the impact of the announcement of the experiment on the participating messengers relative to all the other messengers (non-participating messengers at Veloblitz and all the bicycle messengers at Flash). This effect measures how the behavior of participating messengers differs from all the other messengers as of the announcement of the experiment on August 29th. To estimate this effect we use a dichotomous variable that equals 1 for all the participating messengers at Veloblitz as of August 29th.

The announcement effect may consist of several sub-effects: First, it may capture a possible income effect of the experiment. Note that for a rational messenger who maximizes a time separable intertemporal utility function the income effect becomes operative immediately after we announce the experiment, i.e., immediately after the new information about the future income stream is released. Second, the announcement effect may capture other differences between participants and non-participants. Our experiment offers the advantage of implementing a large anticipated wage change in a real-life setting. However, we cannot force individuals to participate. Out of the 58 messengers at Veloblitz, 45 participated in the experiment. One of the 45 subjects ceased to participate during the experiment. Six messengers were on probationary shifts and, therefore, we did not include them as participants in the experiment.[10] Of the remaining 7 non-participating messengers only one individual explicitly refused to participate. The other 6 non-participants were already quite detached from the company, which is indicated by the low number of shifts they had worked since July 2000.[11] Because of their low frequency of showing up at the firm they may have missed the deadline for the completion of the first survey (which was strictly enforced by us), or did not find it worthwhile to participate. Note, however, that a difference between

[10] Finally, these messengers were not hired by the firm.
[11] Between July and November 2000 they worked roughly one shift per week.

participants and non-participants poses no problem for the main purpose of our study. We are only interested in the comparison of the number of shifts and revenues (deliveries) per shift across treatment conditions for the *participating* messengers. These participants were randomly assigned to treatment and control groups and non-participation affects both groups alike.

II. The Predictions of the Canonical Model of Labor Supply

In this section we integrate the institutional setting at our messenger service into a canonical model of intertemporal utility maximization. This enables us to derive the predictions of the canonical model for our experiment. Not surprisingly, it turns out that messengers who receive a larger share of the revenue (i.e. those in the treatment group) are predicted to work more shifts and work harder during a shift compared to the messengers in the control group. The canonical model, therefore, predicts that messengers in the treatment group generate more revenue than messengers in the control group.

A. The Basic Model

We consider an individual with time-separable utility that maximizes lifetime utility

$$U_o = \sum_{t=0}^{T} \beta^t u(c_t, e_t) \tag{1}$$

where $\beta < 1$ denotes the discount factor, $u(.,.)$ represents the one-period utility function, c_t denotes consumption and e_t is effort in period t. The utility function obeys $u_c > 0$, $u_e < 0$, $u_{cc} < 0$ and $u_{ee} < 0$. The lifetime budget constraint for the individual is given by

$$\sum_{t=0}^{T} p_t c_t (1+\rho)^{-t} = \sum_{t=0}^{T} (w_t e_t + y_t)(1+\rho)^{-t} \tag{2}$$

where p_t denotes the price of the consumption good, w_t the period t wage and y_t non-labor income. For convenience we assume that the interest rate ρ is constant and that there is no uncertainty regarding the time path of prices and wages. The sign of the comparative static predictions is not affected by these simplifying assumptions. Differentiating the Lagrange function for the above maximization problem with respect to c_t and e_t yields the following first order conditions:

$$u_c(c_t, e_t) = \lambda \hat{p}_t \tag{3}$$

$$-u_e(c_t, e_t) = \lambda \hat{w}_t \tag{4}$$

In (3) and (4) λ denotes the Lagrange multiplier for the life-time budget constraint, i.e., λ represents the marginal utility of life-time wealth. $\hat{p}_t$ is defined as $\hat{p}_t = p_t(\beta(1+\rho))^{-t}$ and can be interpreted as the discounted price. $\hat{w}_t$ is defined analogously: $\hat{w}_t = w_t(\beta(1+\rho))^{-t}$. Note that λ is constant along the optimal path of c_t and e_t. This has the important consequence that an *anticipated* temporary wage variation does not affect the marginal utility of life-time wealth. Thus, anticipated temporary variations in wages (or prices) have no income effects. Yet, if there is a non-anticipated temporary increase in the wage λ changes immediately after the new information about the wage increase becomes available and remains constant at this changed level afterwards. For our experiment this means that the income effect stemming from the temporary wage increase has to occur immediately after the announcement of the experiment on 29th August 2000. After that day the marginal utility of life-time wealth again remains constant so that during Treatment A and Treatment B there are no further changes in λ. The difference in behavior between the treatment group and the control group during the two treatments can thus not be due to changes in λ.

In the appendix we show that along the optimal path the within-period decisions of a rational individual maximizing a time-separable concave utility function like (1), subject to constraint (2), can be equivalently represented in terms of the maximization of a static one-period utility function that is linear in income. This static utility function can be written as

$$v(e_t) = \lambda \hat{w}_t e_t - g(e_t, \lambda \hat{p}_t), \tag{5}$$

where $g(e_t, \lambda \hat{p}_t)$ is strictly convex in e_t and measures the discounted disutility of effort while $\lambda \hat{w}_t e_t$ can be interpreted as the discounted utility of income arising from effort in period t. Note also that (5) does not only describe the optimal effort choice in period t but is also based on the optimal consumption decision in period t. For any change in effort the consumption decision also changes in an optimal manner (see appendix).[12]

B. The Response to Temporary Wage Changes

Workers who choose effort according to (5) respond to an anticipated temporary increase in $\hat{w}_t$ with a higher effort e_t. A rise in $\hat{w}_t$ increases the marginal utility returns of effort, $\lambda \hat{w}_t$, which induces the worker to work harder. In our experiment the situation is, however, a bit

[12] Our characterization is inspired by the results in Browning, Deaton and Irish (1985) who show that the within period decisions can be characterized in terms of the maximization of a static profit function. However, the present exposition is more convenient for our purposes.

more complicated because the messengers can choose the number of shifts and the effort during the shift. Theoretically the existence of shifts can be captured by the existence of a minimal effort level $\tilde{e}$ that has to be met by the worker or by the existence of fixed costs of working a shift.

For example, if the worker has to work at least $\tilde{e}$ to receive a positive wage payment, he will never perform below $\tilde{e}$ unless he does not want to work at all in this period. Under these circumstances a worker first determines the optimal temporary effort level e_t^* he would choose in case that he works a shift (i.e., in case he chooses $e_t \geq \tilde{e}$). e_t^* is given by the maximization of (5) subject to the constraint $e_t \geq \tilde{e}$. Next he compares the utility from working a shift at level e_t^*, which is given by $\lambda \hat{w}_t e_t^* - g(e_t^*, \lambda \hat{p}_t)$, with the utility of not working at all in period t. This is given by $u(c_t^*, 0)$ where c_t^* denotes the optimal consumption level at $e = 0$ in period t. If

$$\lambda \hat{w}_t e_t^* - g(e_t^*, \lambda \hat{p}_t) \geq u(c_t^*, 0), \tag{6}$$

the worker prefers to work a shift in period t. On the basis of (5) and (6), it is now easy to see that a wage increase not only induces the worker to work harder, given that he has already decided to work a shift. The wage increase also makes it more likely that the worker chooses to work a shift because it increases the left hand side of (6).[13]

C. Revenues and Effort

The particular setting of our field experiment requires the discussion of two departures from the above model. In the experiment effort is not directly rewarded but workers receive a share of the revenues they generate with their effort. Moreover, we cannot rule out that workers are to some extent competing for the same deliveries. This means that if worker i provides more effort it may be more difficult for worker j to generate revenues. A plausible specification of the revenue function of messenger i is, therefore, given by $r(e_{it}, e_{-it})$ where e_{-it} is the vector of effort levels of the other messengers in the firm. We assume that $r(.,.)$ is strictly increasing and concave in e_{it} and that the marginal revenue of e_{it} is non-increasing in e_{-it}. Under these circumstances a messenger's static objective function is given by

[13] The same result can be obtained if shifts are theoretically captured by fixed costs of working. In this case the worker, once he has decided to work a shift, chooses effort by maximizing $v(e_t)$ subject to $e_t \geq 0$. If the maximized value of $v(e_t)$ minus the fixed costs of working exceeds the utility of not working at all in this period, $u(c_t^*, 0)$, the worker prefers to work a shift.

$$v(e_{it}, e_{-it}) = \lambda \hat{w}_t r(e_{it}, e_{-it}) - g(e_{it}, \lambda \hat{p}_t),$$

where $\hat{w}_t r(e_{it}, e_{-it})$ denotes the discounted income of worker *i* in period *t*. The messenger maximizes $v(e_{it}, e_{-it})$ for given choices of the other messengers. The first order condition for this problem is

$$\lambda \hat{w}_t r'(e_{it}, e_{-it}) = g'(e_{it}, \lambda \hat{p}_t)$$

where *r*' and *g*' denote the partial derivatives with respect to e_{it}. Now compare two identical messengers during the experiment.[14] Assume that *i* is in the treatment group and receives a commission rate of 1.25*w* per unit of revenue whereas *j* is in the control group and receives *w*. This means that the effort choice of *i* is governed by

$$1.25\lambda \hat{w}_t r'(e_{it}, e_{-it}) = g'(e_{it}, \lambda \hat{p}_t) \qquad (7)$$

while *j* chooses effort according to

$$\lambda \hat{w}_t r'(e_{jt}, e_{-jt}) = g'(e_{jt}, \lambda \hat{p}_t). \qquad (8)$$

The obvious difference between the two first order conditions is that the marginal returns to effort are 25 percent higher for *i*, inducing *i* to exert more effort than *j* in equilibrium. Thus, the messengers in the treatment group should exert more effort than the messengers in the control group.[15] Moreover, as in the previous subsection, raising *w* also increases the attractiveness of working additional shifts. Therefore, the members of the treatment group should also work more shifts than the members of the control group.

D. Intertemporal Substitution of Labor

Our field experiment enables us to answer the question how workers respond during treatments A and B to an anticipated exogenous increase in their wage. Before and after our treatments the wages of the workers in the treatment group are kept at the same level as the

[14] Recall that messengers are randomly allocated to the treatment and the control group. Thus, for purposes of comparison we can treat the members of these two groups as identical.

[15] The formal argument is slightly more involved and is borrowed from Athey and Schmutzler (2001) who derive general conditions under which one can compute comparative static results for all equilibria in games with strategic substitutes. Consider the transition of *i*'s and *j*'s wage from (*w*, 1.25*w*) to (1.25*w*, *w*). The increase in *i*'s wage induces *i* to provide more effort and the decrease in *j*'s wage induces *j* to decrease effort. Moreover, the increase in *i*'s effort may reduce the marginal returns to effort for *j* which provides a further incentive for *j* to supply (weakly) less effort. Similarly, the decrease in *j*'s effort may increase the marginal returns to effort for *i*, which provides a further incentive for *i* to supply (weakly) more effort. Thus, the transition from (*w*, 1.25*w*) to (1.25*w*, *w*) induces *i* to increase effort whereas *j* reduces effort. Moreover, since *i* and *j* are identical, subject *i* provides the same effort as subject *j* when they are paid *w*. Likewise, *j* provides the same effort as *i* when they are paid 1.25*w*. Thus, it follows that at wages (1.25*w*, *w*) *i* must exert more effort than *j*.

wages of the workers in the control group during a treatment. Therefore, the comparison of the labor supply between the treatment group and the control group during the treatments provides us also with information about the intertemporal elasticity of labor supply with respect to a temporary wage increase. Let messenger *i* be a representative member of the treatment group while messenger *j* is a representative member of the control group. Since we randomized the assignment of workers to the treatment and the control group they can be viewed as (statistically) identical individuals. Since the revenue generated by a worker is strictly increasing in the worker's effort, the elasticity of the revenue with respect to the wage, which we define as $\varepsilon_{rw} = (\ln r_{it} - \ln r_{jt})/(\ln 1.25w - \ln w) = (\ln r_{it} - \ln r_{jt})/(\ln 1.25)$, gives us a proxy for the intertemporal elasticity of effort.[16]

The question then is under which conditions can we interpret ε_{rw} as (a proxy for) the elasticity of effort of an *individual* worker. This question arises because we raised the wage of many workers during the treatment period and an increase in the effort of one worker may decrease the marginal returns of effort for the other workers. If this is the case ε_{rw} might underestimate the true elasticity of effort of an *individual* worker because it also incorporates the attenuation of incentives arising from other workers' increased effort.[17] It is, however, important to stress that this effect can never induce workers in the treatment group to supply *less* effort than the workers in the control group. Except for the higher wage the representative worker in the treatment group faces the same aggregate environmental conditions (including the effort levels of the other workers) as the representative worker in the control group (see conditions (7) and (8)). Therefore the representative worker in the treatment group is predicted to work harder. This means that, due to potential strategic spillovers across workers, ε_{rw} may underestimate the *size* of the true elasticity of effort but it can never become negative if the assumptions of the canonical model hold.

Moreover, it is possible to show that for a wide class of revenue functions the strategic spillovers across workers do not affect our measure of the elasticity of effort. To show this let

[16] Instead of revenues one could also use the number of deliveries per shift as a proxy for effort. Since we have information on individual revenues per shift as well as on individual deliveries per shift we used both proxies for estimating the elasticity of labor supply.

[17] It seems to us that the two previously published studies of Camerer et al. (1997) and Oettinger (1999) about high frequency („daily") labor supply behavior face the same question. If, on a good day with a high demand for taxi services, many more taxi drivers are on the street, then this could decrease the marginal returns of effort for individual drivers. Likewise, if one can expect a large audience for a particular baseball match, and if, therefore, many more stadium vendors are willing to work during that match, the marginal return for each individual stadium vendor may be smaller.

us assume that the revenue function can be written as $r_{it} = f(e_{it})/F(\boldsymbol{e_t})$ where $\boldsymbol{e_t}$ denotes the whole vector of effort supplies in the firm. As in the previous subsection we assume that r_{it} is strictly increasing and concave in e_{it}. In this case ε_{rw} is given by (ln $f(e_{it})$ – ln $f(e_{jt})$/ln 1.25, i.e., the other workers' behavior does not affect the elasticity of effort. A special case of the above revenue function is, for instance, the function $r_{it} = e_{it}/\Sigma e_{jt}$. In this special case, workers are in extreme competition with each other. Despite this the elasticity of effort is unaffected by the impact of the other worker's effort on *i*'s return.

III. Results

This section reports the results from our field experiment. For most of our estimates we include all the observations between January 1999 and November 2000. We include all the observations where messengers complete more than one delivery per shift, but less than 26. "Shifts" with only one delivery are due to booking errors. Likewise, shifts with more than 26 deliveries also involved erroneous booking in all cases that could be verified. Each restriction excludes roughly two percent of the observations. However, our main results do not change if we include these data in our estimates.

A. The Choice of Shifts

This subsection presents the results for the impact of the experiment on the number of shifts worked. In addition we also examine other determinants of the choice of shifts. We will proceed in the following way. First, we provide a simple comparison of the number of shifts in the treatment group with the number of shifts in the experimental control group. The advantage of this test is that it only compares the choices of participating messengers working under different wage levels. This simple comparison gives a first indication of the direct treatment effect.

The number of participating messengers is 22 in both groups A and B. During treatment A, members of the treatment group worked on the average 13.3 shifts, while the control group worked on the average 8.7 shifts. During treatment B, the treatment group worked 11.4 shifts, and the control group worked 8.7 shifts on average. This means that the intertemporal elasticity of substitution with regard to shifts is equal to 1.6 and is significantly different from zero (t = 2.38, p < 0.05).[18] However this raw elasticity may overstate the direct treatment effect. The reason can be inferred from Figures 5a and 5b, which displays the working hazard, i.e., the probability of working a shift, conditional on the number of days that have elapsed since the latest shift. The figure indicates that the decision to work a shift is strongly duration dependent in both firms. If a messenger worked yesterday he is much more likely to work today, too, compared to a messenger who did not work yesterday. This means that a simple comparison does not give us the pure direct treatment effect because duration dependence artificially amplifies the effect of a wage increase: Individuals who – due to the wage increase – worked yesterday are more likely to work today, even if they did not experience a higher wage today. It is not obvious that this effect should be included in the calculation of the relevant elasticity of substitution.

Insert Figures 5a,b here

To rule out this confound we base our test on the survivor function, i.e., on the share of messengers who have not worked for at least T days. If the direct treatment effect is positive, the survivor function of the treatment group should lie below the survivor function of the control group: For any time interval that elapsed since the last shift, a higher share of the messengers in the control group should choose to not work a shift (hence, more messengers of the control group "survive" in the leisure state). Figure 6 shows that this is indeed the case. The figure plots -ln(-ln (.)) of the survivor function against ln (days since last shifts).[19] The vertical difference between the two curves in Figure 6 can be interpreted as the increase in the probability of working a shift conditional on the time that has elapsed since the latest shift worked. The figure indicates that for most duration levels since the latest shift the probability of working is roughly 20 percent higher in the treatment group. This difference in the survivor functions is significant (log-rank test for equality of the survivor function, $\chi^2(1) = 4.84$, p < 0.05).

Insert Figure 6 here

To examine the determinants of shifts in more detail we use a proportional hazard model, which is also known as a Cox regression (Cox, 1972). It models the probability of working a shift at date t conditional on the characteristics of messenger i and the duration

[18] We define the intertemporal elasticity of substitution with regard to shifts as $(\ln s_T - \ln s_C)/(\ln 1.25w - \ln w)$ where s_T denotes the average number of shifts in the treatment group and s_C the average number of shifts in the control group.
[19] Typically, survivor functions are graphically represented in this way (see e.g., Kiefer 1989).

dependence that specifies how the conditional probability of working varies with the number of days since the latest shift. Formally, we estimate

$$\text{Prob}\ (i \text{ works today}|\text{hasn't worked } T \text{ days}) = \exp\ (\alpha x_{it} + \gamma Treat_{it})\,\psi_i(T). \tag{9}$$

$\psi_i(T)$ is the unknown time dependence, i.e. a function that indicates the baseline probability of working a shift, if the messenger has not worked for T days. As can be seen in Figures 5a and 5b, this time dependence is highly complex. It is an advantage of the Cox regression that there is no need to specify $\psi_i(T)$ (see Cox, 1972). $Treat_{it}$ summarizes the treatment variables that we discussed in section I.C. Finally, x_{it} contains a number of control variables that we discuss below.

We estimate two versions of (9), one in which we stratify by firm, i.e., we allow $\psi_i(T)$ to be different across firms, and one in which we stratify by messenger, i.e., we allow $\psi_i(T)$ to differ across individuals (see Ridder and Tunali, 1999). Our estimation results are presented in Table 1. The first regression in Table 1 allows differences in $\psi_i(T)$ across firms, i.e., it stratifies by firm. The second regression stratifies by messenger. In both regressions we include dummy variables for the different months. To ease the interpretation, Table 1 displays the proportionate change in the working hazard. An estimation coefficient greater than one indicates an increase in the conditional probability of working a shift while a coefficient smaller than one indicates a decrease.

Table 1 shows that the direct treatment effect is positive and significant in both specifications. Moreover, it has roughly the same size in both specifications indicating that even if we allow for different individual patterns of duration dependence the direct treatment effect does not change much. The coefficient of 1.18 in column (2) means that the conditional probability of working is 18 percent higher in the treatment group compared to the control group. This implies that the intertemporal elasticity of substitution with regard to shifts is equal to 0.802.

Insert Table 1 here

The indirect treatment effect measures whether the messengers at Veloblitz behave differently during the treatments compared with the messengers at Flash. Table 1 indicates, however, that with regard to the choice of shifts there are no significant differences between the two firms. The point estimate of the indirect treatment effect is slightly below one but it is far from being significant. This suggests that the implementation of a wage increase at Veloblitz did not cause messengers to be rationed in their choice of shifts.

The announcement effect is positive and highly significant (recall our discussion in section I.C). The other control variables in Table 1 are also interesting. In the regression we distinguish between tenure, i.e., the time elapsed since the messenger joined the company, and experience which is defined as the number of shifts that the messenger has worked during his employment at the messenger service. Longer tenure decreases the probability to work a shift. Conversely, more experience with work increases the probability, holding tenure constant. Both tenure and experience are highly significant. The results also suggest that female messengers work less frequently. Finally, being in the firm for the first month has no significant impact on the probability of working a shift while workers who work for the firm for only one more month (and quit at the end of the month) work significantly fewer shifts than workers who stay with the firm in future months.

Taken together, the main result in this section is that messengers work considerably more shifts when they receive higher wages. The canonical model of labor supply discussed in section II predicts this result. Thus, with regard to the choice of shifts the model does fine. Next, we examine whether this is also the case with regard to effort behavior.

B. The Choice of Effort

During each shift, the messengers have to choose how much effort to exert. Our data provide two measures of effort – the amount of revenue generated per shift and the number of deliveries completed per shift. Yet, longer deliveries command a higher price and require more effort. Hence, the revenue per shift is our preferred measure of effort. In the appendix we also present estimates of the treatment effects that are based on the number of deliveries. It turns out that our estimates are almost identical for either choice of effort measure.

Figure 7 provides a first indication of the direct treatment effect regarding the choice of effort. The figure shows the distribution of revenues for the treatment group and the control group over the two treatment periods. Figure 7 indicates a surprising result. Despite the fact that the treatment group receives higher wages the distribution of revenues for the treatment group lies to the left of the distribution of the control group. The relative frequency of low revenues is higher, and the relative frequency of high revenues is lower in the treatment group compared to the control group. The null hypothesis of equal distributions can be rejected even if we apply a conservative non-parametric Kolmogorov-Smirnov test ($p < 0.05$).

A similar picture arises if we disaggregate the data according to the two treatment periods. In treatment A, the treatment group generated on the average CHF 297 revenue per shift while in the control group the average revenue per shift was CHF 304. In treatment B, the treatment group acquired CHF 293, while the control group gathered average revenues of CHF 314 per shift. Thus, in both treatments the treatment group, even though at a higher wage, acquired lower revenues indicating that the treatment group put forward less effort. This difference in revenues across treatment and control group is significant ($t = 2.328$, $p < 0.05$). Moreover, we get the same conclusions if we use the number of completed deliveries instead of revenues: The members of the treatment group have on average a lower number of deliveries ($t = 2.554$, $p < 0.05$).

Insert Figure 7 here

Although they are informative these simple tests cannot address three important issues. First, we know from the previous section that messengers in the treatment group work more shifts than messengers in the control group. Therefore, if working yesterday hurts today's performance the negative direct treatment effect could be caused by such spillovers. Second, it could be that the two groups worked on different days. Maybe the members of the treatment group just filled any vacant shift, even on days where many other messengers were working or where earnings were expected to be low. Third, there could be a positive correlation between workers' fixed utility costs of working in a shift and their marginal disutility of effort. If this were the case the number of shifts taken by workers with a high marginal disutility of effort would be higher in the treatment group compared to the control group because the higher wage induces relatively more workers with high fixed utility costs to work shifts. To solve these problems we estimated the following regression model:

$$\ln (r_{it}) = \alpha x_{it} + \gamma Treat_{it} + d_t + \varepsilon_{it}. \tag{10}$$

Again, the variables of key interest are the direct and the indirect treatment effect, as well as the announcement effect, summarized in the vector $Treat_{it}$. To address the first problem mentioned above – the issue of fatigue or, more generally, of spillovers across shifts – we include a set of control variables. They are dummy variables indicating whether a messenger has worked on the day before, is going to work on the next day, and an interaction between the two. If exhaustion plays an important role, these variables should have a negative coefficient. The reason is that the messengers should work less hard if they know that fatigue will impact their future productivity. Similarly, there should be evidence of a negative effect of past work on current productivity.

To address the second problem – that messengers in the treatment group work on worse-than-average days – we control for working conditions in two ways. First, we include a daily fixed effect d_t in our regression model. Recall from Figure 2 that the total number of deliveries is highly correlated between Veloblitz and Flash. Hence, daily fixed effects control for variations in demand in a powerful way. Second, we control for the number of competing messengers on a particular day. We include all competing bicycle messengers in both firms plus the number of competing car messengers at Flash. These controls remove all effects stemming from the fact that messengers in the treatment group chose to work also on relatively bad days because their wage was higher.

We address the third problem in several ways because a correlation between fixed and marginal costs of effort may be due to several factors. It may, for instance, be related to the messengers' skill level. Perhaps the wage increase has given unskilled messengers a disproportionate incentive to work more shifts. If this were the case the number of shifts taken by unskilled messengers would be larger in the treatment group compared to the control group and, hence, the negative treatment effect could be due to composition bias. To control for this we add ln(experience) and $\ln(\text{experience})^2$ to control for on-the-job training, as well as a dummy variable indicating whether the messenger has been working at the messenger service for less than a month. We also add a gender dummy that captures possible differences between male and female messengers and a dummy variable indicating whether the messenger is a member of Veloblitz or Flash to capture systematic differences between the two firms. Finally, by adding messenger fixed effects we control for *permanent* individual differences in a powerful way.

However, the correlation between fixed and marginal effort costs may vary over time so that individual fixed effects do not control for this. We can solve this problem by exploiting a particular feature of our data. Recall from section I that there are so-called fixed shifts and variable shifts. A shift is fixed if the worker has committed himself to work during this shift every week. Thus, fixed shifts represent a long-term commitment of the worker. Cancellation of a fixed shift requires one month's notice. In contrast, if a worker decides to work during a variable shift he is not committed in this way. He is just obligated to work that particular shift. The long-term nature of fixed shifts is illustrated by the fact that the fixed shifts did not change during the observation period. Thus, with regard to the fixed shifts there could be no selection of high-marginal-disutility types. Therefore, it makes sense to estimate a direct treatment effect for both the fixed shifts and the variable shifts. If the reduction of effort in the treatment group is indeed caused by the selection of high-marginal-disutility types we should observe a negative direct treatment effect only for the variable shifts but not for the fixed shifts. If, however, the negative direct treatment effect also occurs during the fixed shifts we can exclude selection as the source of this effect.

The results of our regressions are displayed in Table 2. Column 1 presents the estimates without the controls for individual fixed effects. In column 2 we control in addition for individual fixed effects and in column 3 we estimate the direct treatment effect separately for fixed and for variable shifts. In all specifications, we compute robust standard errors that allow for arbitrary correlations between the residuals of the messengers on a particular day. In all specifications we control for daily fixed effects. The important result in Table 3 is that the direct treatment effect is negative and highly significant in all specifications. The estimates in column 1 and 2 imply that the intertemporal elasticity of substitution ε_{rw} is -0.332 or -0.255, respectively. The coefficients for the direct treatment effect in column 1 and column 2 are not significantly different from each other (p = 0.54). Thus, the inclusion of individual fixed effects has no significant impact on the direct treatment effect indicating that the negative effort response to wages cannot be attributed to fixed individual differences. This means, in particular, that less productive messengers do not respond differently to the wage increase compared to more productive messengers. Moreover, when, instead of revenues per shift, we use the number of completed deliveries as a proxy for effort we get basically the same results. In Table A1 of the appendix we present the results of regression 1 and 2 with ln (deliveries) as the dependent variable. These estimates imply almost the same elasticity of substitution as the estimates in column 1 and 2 of Table 2.

The separate estimates for fixed and variable shifts in column 3 yield similar results. Both the coefficient for fixed and for variable shifts is negative and significant. Moreover, the two coefficients are not significantly different from each other (p = 0.61) and of the same order of magnitude as the coefficients in column 1 and 2. They imply that ε_{rw} is between -0.3 and -0.4. This suggests that the negative impact of wages on effort is not due to a selection effect.

Insert Table 2 here

To illustrate the quantitative importance of the messengers' negative effort response to the wage increase it is useful to compare the income of a messenger who provides the same effort irrespective of the wage level and a messenger who exhibits, say, an ε_{rw} of -0.33. The income per shift is given by $z_{it} = r_{it}(e_{it}(w))w$. Therefore, the elasticity of the income per shift with respect to the wage, which we denote by ε_{zw}, is given by $\varepsilon_{zw} = 1 + \varepsilon_{rw}$. This implies that a worker who exhibits ε_{rw} = -0.33 forgoes one third of the income increase that is generated by the wage increase relative to a worker who keeps effort constant. Thus, the workers forgo a substantial fraction of the potential income gains by reducing their effort.

Turning to the control variables, we find that the daily fixed effects are highly significant (F-test, $p < 0.01$). Adding daily fixed effects to the regression reduces the variance of the error term by roughly 60 percent. The number of competitors working a shift also has a large influence on revenues per shift. The point estimate for the number of competing bicycle (car) messengers is around -0.035 (-0.047) in all three specifications and always highly significant. These estimates imply that adding one bicycle (car) messenger to a shift depresses all the other messengers' deliveries by roughly 3.6 (4.7) percent. The number of messengers working per shift varies strongly across days, and its standard deviation is about two. Hence, variations of +/- two messengers are not rare. Our estimates suggest that this produces variations of 14 to 19 percent in the messengers' revenues.

Fatigue does not seem to play a role in determining a messenger's revenues. The point estimates on “will work on next day”, “worked yesterday” and the interaction variable is positive and almost always highly significant. Hence, contrary to the fatigue hypothesis, messengers generate even more revenue when another day of work is ahead, or when they have worked yesterday, too. This is in line with what the messengers told us. Many of them report that having worked yesterday may even make effort less onerous today because it keeps their muscles active.[20]

Finally, we find a large and significant effect of on-the-job training on revenues per shift. The estimated profile is strictly increasing and concave in ln(experience) for the observed sample variation of ln(experience). Even when we control for individual fixed effects, the quantitative impact of on-the-job training on revenues is quite large. The estimated coefficient of 0.095 in column (2) implies that raising experience from 0 shifts to 10 shifts increases revenues by 24.3 percent. Raising experience further to 60 shifts, increases revenues by an additional 18.4 percent.

Taken together, the results of this section show that the negative response of revenues (or completed deliveries) to the wage increase is a robust fact that cannot be accounted for by the various alternative explanations discussed above. Although the messengers in the treatment group have a clear incentive to work harder during a shift they work less hard. This

[20] See also Kulik (1999) who shows that subjects from Veloblitz exhibit a level of maximum oxygen intake far above the average level. High maximum oxygen intake is key to good endurance performance and is very hard to alter by training or lifestyles.

contradicts the canonical model of labor supply presented in section II. This contradiction is even more puzzling in view of the fact that the messengers' choice of shifts is consistent with the model. In the next section we discuss various explanations for this puzzle.

IV. Discussion

Although the messengers in the treatment group work less hard during a shift their overall labor supply increases because the increase in the number of shifts overcompensates the decrease in the effort per shift. The elasticity of the number of shifts with regard to the wage is approximately 0.8 whereas the elasticity of revenues per shift with respect to the wage is roughly –0.3. The challenge for any model of labor supply is to explain why the number of shifts goes up while the effort per shift goes down.

Interestingly, our results are consistent with the two other studies of high frequency labor supply behavior. Camerer et al. (1997) find that taxi drivers in New York – once they have decided to work on a given day – tend to provide less effort (i.e. work less hours) on days with higher wages. This is similar to the reduction of the effort per shift observed in our experiment. Oettinger (1999) finds that stadium vendors are more likely to work on those days on which wages are predictably high. Our messengers behave in the same way because they are more likely to work a shift when the wage is high. This suggests that the behavioral pattern observed in our experiment might be the result of more general behavioral forces.

A. Two Interpretations

One possibility to explain these facts is that – in contrast to the assumption made in section II – the messengers' utility function is not fully separable across periods. If, for instance, time separability only holds for blocks of periods with each block being comprised of J days, the utility function takes the form $u = u(c_1, \ldots, c_J, e_1, \ldots, e_J)$. Even if $u(.,.,..)$ is strictly concave, the only robust prediction then is that $\sum_{t=1}^{J} e_t$ must increase in response to a wage increase, *but not every* e_t is predicted to increase because non-zero cross partial derivatives between labor supply at date t and s may cause spillover effects across periods (see Browning, Deaton, and Irish, 1985; equation 1.19). Thus, working more shifts during the block of J periods but providing less effort per shift is consistent with this prediction. One disadvantage of this model is that it would also be consistent with the fact that workers work less shifts but provide more effort per shift.

Another potential explanation of our facts is that individuals might have preferences that include a daily income target $\tilde{y}$ that serves as a reference point. This explanation is suggested by the large number of studies indicating reference dependent behavior (for a selection of papers on this see Kahneman and Tversky 2000). In terms of the canonical model, the existence of reference dependent behavior can be captured by the following one-period utility function.

$$v(e_t) = \begin{matrix} \lambda(\hat{w}_t e_t - \tilde{y}) - g(e_t, \lambda \hat{p}_t) & \text{if } \hat{w}_t e_t \geq \tilde{y} \\ \gamma\lambda(\hat{w}_t e_t - \tilde{y}) - g(e_t, \lambda \hat{p}_t) & \text{if } \hat{w}_t e_t < \tilde{y} \end{matrix} \qquad (11)$$

The key difference to the model in section II is the presence of a daily income target $\tilde{y}$ that is associated with loss aversion. Loss aversion is indicated by the parameter $\gamma > 1$ which captures the extent to which the income target affects behavior. Previous evidence (see Kahneman and Tversky 2000) suggests that for many individuals $\gamma \approx 2$. Thus, falling short of the income target $\tilde{y}$ imposes a large utility cost on these individuals.[21] In addition, it creates powerful incentives to exert more effort because the marginal utility of effort is high. However, once individuals attain the target $\tilde{y}$, the marginal utility of income drops discretely (from $\gamma\lambda$ to λ), causing a substantial reduction in the incentive to supply effort.

The preferences described in (11) yield a straightforward explanation of our evidence. In addition, they are also consistent with the evidence in Camerer et al. (1997) and Oettinger (1999). A rise in wages increases the utility of working on a given day. Thus, at higher wages it is more likely that the utility of working $v(e_t)$ exceeds the fixed costs of working. Hence, workers are more likely to work on high-wage days. At the same time, however, the increase in wages makes it more likely that the income target is met or exceeded already at relatively low levels of effort. Therefore, compared to the control group the workers in the treatment group are more likely to face a situation in which the marginal utility of income is λ instead of $\gamma\lambda$, i.e., they face lower incentives to work during the shift.[22] To provide an example, suppose that the daily income target of a messenger is CHF 100 per shift. Members of the control group, who receive a revenue share of roughly 40 percent, have to generate revenues of CHF

[21] For this reason γ can be thought of as a measure of the degree of loss aversion.

[22] If γ is sufficiently high relative to the wage increase one may obtain the extreme result that before and after the increase the worker provides effort to obtain exactly $\tilde{y}$. In this case the worker's effort obviously decreases in response to the wage increase because at higher wages $\tilde{y}$ is obtained at lower effort levels. In general, the larger γ, the sharper the kink in the objective function and the more likely the worker's optimal effort choice e^* will be at the kink, i.e., the more likely $\gamma\lambda \hat{w}_t > g'(e^*) > \lambda \hat{w}_t$ holds. Note, however, that even if the worker is not a "perfect" income targeter, i.e., even if before or after the wage increase he does not earn exactly $\tilde{y}$, negative effort responses may occur.

250 per shift to meet that target whereas members of the treatment group, who receive a revenue share of roughly 50 percent, have to generate only a revenue of CHF 200 to meet the target. Thus, members of the treatment group will face a low marginal utility of income (λ instead of $\gamma\lambda$) at much lower levels of effort than members of the control group. As a consequence, members of the treatment group will provide less effort than members of the control group.

There are two other points that deserve to be emphasized. First, loss averse preferences only predict lower effort in response to higher wages in environments where daily earnings are strongly fluctuating around the income target. If the workers in the control and the treatment group were always above (or always below) their target they would always face the same marginal utility of income. In the context of our experiment strong income fluctuations (see Figures 2 and 3) are indeed very frequent. Second, loss averse preferences also predict that piece rate incentives elicit higher effort levels compared to flat time-based payment schemes. In fact, if the parameters of the piece rate incentive imply that individuals are often below their income target, loss averse preferences predict particularly strong incentive effects of piece rates compared to time-based schemes. Moreover, even if the individuals are above their income target most of the time, piece rates provide stronger incentives than time-based schemes. This means that the evidence in Lazear (2000) and Paarsch and Shearer (2000), who show that piece rates elicit more effort than time-based incentives, is fully compatible with loss averse preferences.

B. Individual Differences in Loss Aversion

In our view loss averse preferences as modeled in (11) provide a plausible explanation for the evidence. The plausibility of this explanation would be further enhanced if it were possible to find independent individual measures of loss aversion for the experimental participants and to link these measures to their behavior during the experiment. In other words – if it were indeed the case that messengers who display more loss aversion – as measured by an independent test – are also more likely to respond to higher wages with a lower effort per shift, the credibility of this explanation would be greatly enhanced. Moreover, the existence of loss aversion among the messengers would also constitute direct evidence against the competing explanation based on the assumption that utility is not time separable, because loss aversion is incompatible with standard utility functions irrespective of whether they are time-separable or not.

Loss aversion and reference dependent behavior have implications in a variety of domains. Loss averse choices have been documented, in particular, in the realm of decision-making under uncertainty (Kahneman and Tversky 1979). Therefore, we measured the messengers' loss aversion by observing choices under uncertainty in an experiment that took place eight months after the experimental wage increase. As part of this study, we presented the messengers with the opportunity to participate in the following two lotteries:

Lottery A: Win CHF 8 with probability 1/2, lose CHF 5 with probability 1/2. If subjects reject lottery A they receive CHF 0.

Lottery B: This lottery consists of six independent repetitions of Lottery A. If subjects reject lottery B they receive CHF 0.

Subjects could participate in both lotteries, or only in one lottery, or they could reject both lotteries. Before paying the subjects three weeks later, we provided them with another opportunity of participating in two lotteries. This time the alternative to the lottery choice was a sure gain.

Lottery C: Win CHF 5 with probability 1/2, win nothing with probability 1/2. If subjects reject lottery C they receive CHF 2.

Lottery D: This lottery consists of six independent repetitions of lottery C. If subjects reject lottery D they receive CHF 12.

As before, subjects could participate in both lotteries, they could participate in only one lottery or they could reject both lotteries.

The above lotteries enable us to construct individual measures of loss aversion. In particular, subjects who reject one of our lotteries can be classified as loss averse. This is transparent for lotteries A and B because both in A and B it is obvious that participating in the lottery can lead to losses. However, in case of lottery C and D subjects may also "lose", i.e., fall behind the sure gain in case that the lottery is chosen. In principle, one might think that the rejection of A or B (or of C or D) is also compatible with risk aversion arising from diminishing marginal utility of lifetime income. This interpretation is, however, ruled out by Rabin's calibration theorem (Rabin 2000). Rabin showed that a theory of risk averse behavior based on the assumption of diminishing marginal utility of *life-time* income implies that people essentially *must be* risk neutral for low stake gambles like our lotteries. Intuitively, this follows from the fact that risk averse behavior for low stake gambles implies ridiculously high levels of risk aversion for slightly higher, but still moderate, stake levels. Yet, such unreasonably high levels of risk aversion can be safely ruled out. To illustrate this, if one assumes that the rejection of lottery A is driven by diminishing marginal utility of life time income, then the subject will also reject a lottery where one can lose $100 with probability ½ and win *any* positive prize with probability ½. Thus, there is no finite prize that induces this subject to accept a 50 percent chance of loosing $100. Similar results can be shown to hold for the other lotteries.[23]

Table 3 presents the results of our lotteries. It turns out 62 percent of the subjects reject lottery A or B and 33 percent reject either lottery C or D. Lottery A is rejected in 54 percent of the cases while lottery C is rejected in 28 percent of the cases. Lottery B is rejected in 42 percent of the cases whereas lottery D is rejected in 14 percent of the cases. These results are qualitatively similar to the results obtained in a many other studies (e.g., Read, Loewenstein, and Rabin, 1999; Cubbit, Starmer and Sudgen, 1998; Hogarth and Einhorn, 1992; Keren and Wagenaar, 1987). The only difference is that, overall, messengers are less likely to reject the lotteries than subjects in comparable studies. However, the differences in rejection rates between the different lotteries are quite similar.

Insert Table 3 here

The reason for conducting several lotteries was that we wanted to construct several measures of loss aversion. Since choice behavior is always noisy to some extent it is good to have several measures. This enables us to check the robustness of our conclusions with regard to the different measures. If it can be shown that – irrespective of the details of how we measure loss aversion – the messengers who exhibit loss aversion in the lottery also are more likely to lower their effort during a shift in response to a wage increase we have strong support for the view that it is loss aversion that drives the messengers' effort choices.

[23] Our lotteries shed light on the nature of loss aversion in a variety of ways but, for our purposes, the best indicator is the mere rejection of one of the lotteries. Other indications of loss aversion can be gained, for instance, by a comparison between lottery A and B or between lottery C and D. Note that the likelihood of making a loss (relative to the alternative option of a zero income) is much larger in lottery A than in lottery B. Likewise, the likelihood of "losing", i.e., of falling behind the sure alternative, is much greater for lottery C than for lottery D. Hence, if individuals reject A more frequently than B or if they reject C more frequently than D we have evidence for loss aversion. Note also that an expected utility maximizer who rejects lottery A will never accept lottery B. The same holds for C and D, respectively.

C. The Relation between Loss Aversion and the Choice of Effort

For the purpose of relating individual effort choices to individual measures of loss aversion we constructed three measures. The first measure of loss aversion, LA1, is a dummy variable that is equal to 1 if the messenger rejected either lottery A or lottery C, and zero otherwise. The second measure, LA2, is a discrete variable that takes on zero if both lottery A and C are accepted, that takes on 1 if one of the two lotteries is rejected and the number 2 if both A and C are rejected. Thus, LA2, measures individual differences in loss aversion in more detail. Our third measure, LA3, is analogously constructed as LA2 except that lotteries B and D are the objects of interest. Before we present our results we would like to emphasize that our measures of loss aversion allow us to capture individual differences in the degree of loss aversion but not absolute levels of loss aversion. All we can say is that an individual who accepted lottery A and C, say, is likely to be less loss averse than an individual who rejected A or C. Yet, it may still be the case that an individual who accepts A and C displays some loss aversion if confronted with other choices.

To provide a first indication of the relation between loss aversion and effort per shift we present Figures 8a and 8b. In Figure 8a, we split up the treatment group according to LA1 and display the distribution of revenues per shift for the two subgroups. The figure shows that the distribution generated by the messengers who rejected lottery A or C is shifted leftwards relative to the distribution generated by the messengers who accepted A and C. This difference in distributions is highly significant (p = 0.004) according to a non-parametric Kolmogorov-Smirnov test. Thus, the loss averse messengers of the treatment group generate less revenue compared to those messengers of the treatment group who do not display loss aversion according to LA1. In Figure 8b we compare those members of the treatment group, who do not display loss aversion according to LA1 (accept both lottery A and C), with the members of the control group. The figure shows that the distribution of revenues across the two groups is very similar. In fact, a Kolmogorov-Smirnov test reveals no significant differences (p = 0.32). This suggests that only the loss averse members of the treatment group generated lower revenues in response to higher wages while the other members of the treatment group kept revenues at the same level as the members of the control group.

Insert Figures 8a,b here

To what extent is this result robust to the introduction of other control variables? We examined this by running the same regression as in column 1 of Table 2 except that we also included the interaction between the direct treatment effect with our loss aversion measures.

The results of the regressions are presented in Table 4. In column 1 we present the estimates of the direct treatment effect (as an elasticity) and the interaction term when we use measure LA1. Column 2 reports the results based on measure LA 2 and column 3 displays the results for measure LA3. The striking result in Table 4 is that, for all measures of loss aversion, the direct treatment effect is sizeable and significantly negative for the loss averse members of the treatment group while for those who do not display loss aversion the elasticity is not significantly different from zero. This suggests that only those messengers who exhibited loss aversion in the lottery experiment responded to higher wages with the provision of less revenue per shift.

Recall that LA2 and LA3 allow us to estimate the impact of loss aversion on effort behavior in a more detailed way because these measures distinguish between the case where a messenger rejects only one lottery and rejects both lotteries. The estimates of the interaction term in column 2 and 3 imply that messengers who reject only one lottery exhibit an elasticity of –0.273 or –0.318, respectively, whereas messengers who reject both lotteries exhibit an elasticity that is twice as high in absolute values.

Insert Table 4 here

In the middle panel of Table 4 we present restricted estimates where we have set the direct treatment effect (without the interaction term) to zero. This allows us to estimate the interaction term with more precision. Finally, in the bottom panel of Table 4, we include messenger fixed effects to check whether the interaction term is not picking up systematic differences between the different groups of messengers that are unrelated to the experiment. However, our elasticity estimates remain sizeable and highly significant even if we control for individual fixed effects.

So far, we have shown that differences in loss aversion predict differences in effort responses during a shift. However, loss averse preferences as modeled in (11) also predict that loss averse messengers will work more shifts when they face a higher revenue share. The same is predicted for messengers who exhibit no or little loss aversion. Thus, differences in loss aversion are not associated with systematically different responses in the choice of shifts. It would thus be reassuring for an explanation based on (11) if the loss averse members of the treatment group did not respond differently than the other members of the treatment group.

The impact of loss aversion on the choice of shifts can be tested using the same measures as before. In Table 5 we present the results of Cox regressions for the choice of shifts. The regressions in Table 5 are based on the same controls as the regression in column 2 of Table 1 but, in addition, we also added an interaction between the direct treatment effect and our measures of risk aversion. The table indicates that all three measures add no explanatory power to the model and leave the direct treatment effect essentially unchanged. While the direct treatment effect is always significantly positive the interaction term is close to 1 and never significant.[24] Thus loss aversion has no systematic impact on the choice of shifts.

Insert Table 5 here

V. Summary

This paper reports the results of a randomized field experiment examining how workers, who can freely choose their working time and their effort during working time, respond to a fully anticipated temporary wage increase. We find a strong positive impact of the wage increase on working time but also a sizeable and significantly negative impact on effort during working time. Since the positive impact on working hours (number of shifts) dominates the negative impact on effort per shift the overall supply of effort increases. The effect on effort per shift remains robustly negative even if we control for individual fixed effects, for daily fixed effects, for workers' fatigue, for competition among the workers and for workers' experience. Thus, the canonical model of intertemporal labor supply based on time separable preferences cannot account for the decrease in effort per shift.

We show that a simple model of loss averse, reference dependent, preferences can account for both the increase in working hours and the decrease in effort per hour. Workers with such preferences have a daily reference income level. Daily incomes below the reference level are experienced as a "loss" and in the loss domain the marginal utility of income is large. In contrast, at and above the reference level the marginal utility of income discontinuously decreases to low levels. Workers who temporarily earn higher wages are more likely to exceed the reference income level and, hence, they face a lower incentive to provide effort because their marginal utility of income is low. At the same time, however, workers with higher wages have a higher overall utility of working on a given day so that they can more easily cover the fixed costs of showing up at work. Hence, they are more likely to work. Our model also reconciles the seemingly contradictory evidence in the two previous studies of intertemporal labor substitution based on high frequency data because these studies looked at different dimensions of labor supply. Whereas Camerer et al. (1997) examined how

[24] Recall that a coefficient of 1 in our Cox regressions indicates the absence of an effect.

taxi drivers, who have decided to work on a given day, vary their daily working time (which is a good proxy for daily effort) in response to wage variations, Oettinger (1999) investigated how the probability of working on a given day is affected by wage variations. A model with loss averse preferences predicts that a wage increase induces taxi drivers to provide fewer hours per day but it also predicts that workers are more likely to work on high wage days. This is exactly what Camerer et al. (1997) and Oettinger (1999) observed.

The interpretation of our experimental results in terms of loss aversion is further supported by the fact that workers who are more prone to loss aversion respond to the wage increase with a stronger reduction in hourly effort. Moreover, differences in loss aversion have no impact on the workers' choice of working time. In line with the model, individuals who exhibit relatively more loss aversion show the same working time response as individuals who exhibit little or no loss aversion. Taken together, our results, therefore, suggest that loss averse, reference dependent, preferences should be considered as a serious candidate in the future analysis of labor supply decisions.

References

Altonji, Joseph G. (1986), Intertemporal Substitution in Labor Supply: Evidence from Micro Data, Journal of Political Economy, 94(3), Part 2, S176-S215.

Athey, Susan and Armin Schmutzler (2001), Investment and Market Dominance, RAND Journal of Economics 32(1), 1 – 26.

Blundell, Richard (1994), Evaluating Structural Microeconometric Models of Labor Supply. In: Christopher Sims (Ed.), Advances in Econometrics: Sixth World Congress of the Econometric Society, Vol. II, 3 – 48.

Blundell, Richard; MaCurdy, Thomas E. (1999), Labor Supply: A Review of Alternative Approaches, in Ashenfelter, Orley and David Card (eds.): Handbook of Labor Economics vol. 5.

Browning, Martin, Angus Deaton and Margaret Irish (1985), A Profitable Approach to Labor Supply and Commodity Demands over the Life-Cycle, Econometrica 53(3), 503-543.

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

Card, David (1994), Intertemporal Labor Supply: An Assessment. In: Christopher Sims (Ed.), Advances in Econometrics: Sixth World Congress of the Econometric Society, Vol. II, 49 – 78.

Cox, D. R. (1972). Regression Models and Life-Tables (with discussion). Journal of the Royal Statistical Society Series B 34, 187-220.

Cubbit, Robin, Christ Starmer and Robert Sugden (1998), On the Validity of the Random Lottery Incentive System, Experimental Economics 1(2), 115 – 131.

Heath, Chip, Larrick, Richard and George Wu (1999). Goals as Reference Points, Cognitive Psychology 38, 79-109.

Hogarth, Robin and Hillel Einhorn (1990), Venture Theory: A Model of Decision Weights, Management Science 36, 780 – 803.

Kahneman, Daniel and Amos Tversky (1979), Prospect Theory, Econometrica 47, 263-291.

Kahneman, Daniel and Amos Tversky (2000) Eds., Choices, Values, and Frames. Cambridge University Press.

Keren, Gideon and Willem Wagenaar (1987), Violation of Utility Theory in unique and repeated Gambles, Journal of Experimental Psychology: Learning, Memory, & Cognition. Vol 13(3), 387-391.

Kiefer, Nicholas M. (1989), Economic Duration Data and Hazard Functions, *Journal of Economic Literature* 26(2), 646 – 679.

King, Robert G., Rebelo, Sergio (1999), Resuscitating Real Business Cycles. in: Handbook of Macroeconomics.

Kulik, Karsten (1999), Veränderung des Laktatmetabolismus der Atmungsmuskulatur nach Atmungsausdauertraining, M.A. Thesis, University of Zurich.

Lazear, Edward P. (2000), Performance Pay and Productivity, American Economic Review; 90(5), 1346-61.

Mankiw, Gregory, Julio Rotemberg and Lawrence Summers (1985), Intertemporal Substitution in Macroeconomics, Quarterly Journal of Economics, 100 February, 225 – 251.

Oettinger, Gerald S. (1999), An Empirical Analysis of the Daily Labor Supply of Stadium Vendors, Journal of Political Economy; 107(2), 360-92.

Paarsch, Harry and Bruce Shearer (1999), The Response of Worker Effort to Piece Rates: Evidence from the British Columbia Tree-Planting Industry, Journal of Human Resources 34(4), 643 – 667.

Pencavel, John (1986), Labor Supply of Men: A Survey. in: Ashenfelter, Orley and Richard Layard (Eds.), Handbook of Labor Economics, Amsterdam: North Holland.

Rabin, Matthew (2000), Risk Aversion and Expected Utility Theory: A Calibration Theorem, Econometrica; 68(5), 1281-92.

Read, Daniel; Loewenstein, George; Rabin, Matthew (1999), Choice Bracketing, Journal of Risk and Uncertainty; 19(1 3), 171-97.

Ridder, Geerd and Tunali, Insan (1999). Stratified Partial Likelihood Estimation. Journal of Econometrics 92(2), 193-232.

Romer, David (1996), Advanced Macroeconomics, Upper Saddle River: McMillan.

Appendix

1 The First Order Conditions

The associated Lagrangian to the canonical life-cycle model is

$$L = \sum_{t=0}^{T} \beta^t u(c_t, e_t) - \lambda \sum_{t=0}^{T} (p_t c_t - (w_t e_t + y_t))(1+\rho)^{-t}.$$

The first order conditions to this problem are

$$\beta^t u_c(c_t, e_t) = (1+\rho)^{-t} \lambda p_t$$

and

$$-\beta^t u_e(c_t, e_t) = (1+\rho)^{-t} \lambda w_t.$$

Define the discounted price $\hat{p}_t = (1+\rho)^{-t} \beta^{-t} p_t$ and the discounted wage $\hat{w}_t$ analogously. The first order conditions then have the somewhat simpler form

$$u_c(c_t, e_t) = \lambda \hat{p}_t \tag{A.1}$$
$$-u_e(c_t, e_t) = \lambda \hat{w}_t. \tag{A.2}$$

Equation (A.1) implies that, at every date t, the individual equates the marginal utility of consumption to the marginal utility of lifetime income λ times the discounted price of the consumption good. Similarly, when choosing how hard to work, the individual chooses effort such that the marginal disutility of effort is equal to the marginal utility of lifetime income times the discounted wage per unit of effort $\hat{w}_t$.

2 An equivalent representation of within-period preferences

It is possible to represent within-period preferences in terms of a static objective function. The following derivation is similar to the one in Browing, Deaton, and Irish (1985), but we emphasize a different aspect.

1. Consider equation (A.1) again. Since $u()$ is strictly concave, $u_c()$ is strictly decreasing in c and can be solved for c_t:

   $$c_t = u_c^{-1}(e_t, \lambda\hat{p}_t).$$

   Substitute this into (A.2) to obtain

   $$-u_e\left(u_c^{-1}(e_t, \lambda\hat{p}_t), e_t\right) = \lambda\hat{w}_t. \tag{A.3}$$

   Hence, equation (A.3) expresses e_t as a function of $\lambda\hat{p}_t$ and $\lambda\hat{w}_t$ alone.

2. Now consider the static one-period objective function

   $$v(e_t) = \lambda\hat{w}_t e_t - g(e, \lambda\hat{p}_t) \tag{A.4}$$

   where λ is the lifetime marginal utility of income along the optimal path, and $g(e, \lambda\hat{p}_t)$ is convex and can be interpreted as the disutility of effort. To see this, define

   $$g(e_t, \lambda\hat{p}_t) = -\int_0^{e_t} u_e\left(u_c^{-1}(x, \lambda\hat{p}_t), x\right) dx.$$

   It is obvious that the FOC to (A.4) is identical to (A.2), since.

   $$g_e(e_t, \lambda\hat{p}_t) = -u_e\left(u_c^{-1}(e_t, \lambda\hat{p}_t), e_t\right).$$

3. It remains to be shown that $g(e_t, \lambda\hat{p}_t)$ is convex, i.e., that its second derivative w.r.t. e_t is positive. To see this, we proceed in two steps

   (a) Consider how the individual adjusts c_t in response to a small perturbation of e_t along the optimal path (i.e., λ remains constant):

   $$\frac{dc_t}{de_t} = -\frac{u_{ce}}{u_{cc}}$$

   by (A.1).

   (b) Take the second derivative of $g(e_t, \lambda\hat{p}_t)$ to obtain

   $$g_{ee}(e_t, \lambda\hat{p}_t) = -u_{ee} - u_{ce}\frac{dc_t}{de_t} = -u_{ee} + \frac{u_{ce}^2}{u_{cc}} = \frac{-1}{u_{cc}}\left(u_{cc}u_{ee} - u_{ce}^2\right). \tag{A.5}$$

   To determine the sign, observe that the conditions for concavity of $u()$ are

   $$\begin{aligned} u_{cc} &< 0 \\ u_{ee} &< 0 \\ u_{cc}u_{ee} - u_{ce}^2 &> 0 \end{aligned}$$

   which imply that (A.5) is greater than zero. This establishes the convexity of $g(e_t, \lambda\hat{p}_t)$.

Thus, in the canonical life-cycle model, a rational, forward looking individual behaves as if he maximized the one-period objective function (A.4).

TABLE A1: ROBUSTNESS CHECK

EXPLAINING THE DIRECT TREATMENT EFFECT ON EFFORT

DEPENDENT VARIABLE: LOG(# DELIVERIES PER SHIFT)

OLS REGRESSIONS

	(1)	(2)
Treatment Effects		
Direct Treatment Effect	-0.075*** (-3. 071)	-0.059** (-2.544)
Implied Intertemporal Elasticity of Substitution	-0.336	-0.264
Indirect Treatment Effect	0.04* (1.676)	0.029 (1.332)
Announcement Effect	0.06*** (2.801)	0.031 (1.453)
Control Variables		
Daily Fixed Effects	Yes***	Yes***
Messenger Fixed Effects	No	Yes**
# Competing Bicycle Messengers	-0.038*** (-11.968)	-0.038*** (-13.023)
# Competing Car Messengers	-0.04*** (-9.491)	-0.049*** (-9.966)
Will work on next day (DV)	0.07*** (8.778)	0.043*** (5.526)
Worked yesterday (DV)	0.033*** (4.124)	0.009 (1.159)
Will work on next day, and has worked yesterday (DV)	0.03*** (2.963)	0.024** (2.357)

(continued on next page)

Log(Experience)	0.17*** (17.37)	0.132*** (7.822)
Log(Experience)^2	-0.013*** (-11.197)	-0.011*** (-4.343)
First Month (DV)	-0.022 (-1.267)	-0.031* (-1.903)
Last Month (DV)	-0.021 (-1.41)	-0.028** (-2.035)
Female (DV)	-0.076*** (-5.566)	--
Member of Veloblitz (DV)	0.213 (6.492)	--
Controls for Composition of Deliveries	Yes***	Yes**
Within Days R^2	0.222	0.414
Fraction of unobserved variance due to daily fixed effects	0.666	0.729
Number of Observations	21,737	21,737

Notes:
a. *, **, *** denotes significance at the 10, 5, and 1 percent level, respectively
b. z-statistics in parentheses.
c. constant term omitted.
d. DV indicates dummy variable.

TABLE A2: ROBUSTNESS CHECK:

EXPLAINING THE DIRECT TREATMENT EFFECT ON EFFORT

DEPENDENT VARIABLE: LOG(# DELIVERIES PER SHIFT)

OLS REGRESSIONS

	Model (1) Interaction with: Loss Aversion Measure 1 N =21,560	**Model (2)** Interaction with: Loss Aversion Measure 2 N =21,560	**Model (3)** Interaction with: Loss Aversion Measure 3 N =21,560
Intertemporal Elasticity of Substitution			
Direct Treatment Effect (DTE)	-0.273 (-1.6)	-0.238 (-1.468)	-.282* (-1.677)
DTE × Loss Aversion Measure	-0.21 (-1.057)	-0.175 (-1.369)	-0.134 (-0.995)
Test for joint significance	$p < 0.01$	$p < 0.01$	$p < 0.01$
Interaction alone			
DTE × Loss Aversion Measure	-0.435*** (-2.998)	-0.295*** (-3.033)	-0.291*** (-2.934)

Notes:
a. *, **, *** denotes significance at the 10, 5, and 1 percent level, respectively.
b. z-statistics in parenthesis.
c. same controls as in Table 2, column (1).
d. Tests for joint significance are F-tests testing whether both coefficients equal 0.
e. See text for an explanation of the Loss Aversion measures.

TABLE 1: BASELINE RESULTS FOR THE CHOICE OF SHIFTS

COX REGRESSIONS: PROBABILITY OF WORKING, CONDITIONAL ON DAYS SINCE LAST SHIFT (CHANGES IN HAZARD RATES DISPLAYED)

	(1)	(2)
Treatment Effects		
Direct Treatment Effect	1.205*** (2.875)	1.18** (2.462)
Implied Elasticity of Substitution	0.896	0.802
Indirect Treatment Effect	0.946 (-0.991)	0.921 (-1.385)
Announcement Effect	1.438*** (7.692)	1.381*** (5.889)
Control Variables		
Log(Experience)	1.127*** (14.840)	1.261*** (10.083)
Log(Tenure)	0.854*** (-13.1)	0.831*** (-8.248)
First Month (DV)	0.967 (-.775)	1.01 (0.193)
Last Month (DV)	0.885*** (-3.654)	0.884*** (-2.868)
Female (DV)	0.854*** (-5.26)	
Controls for Months (DVs)	Yes***	Yes***
Stratified according to	Firm	Messenger
Log Likelihood	-182,677.36	-93,408
Number of Failures	21,455	21,455

Notes: a. *, **, *** denotes significance at the 10, 5, and 1 percent level, respectively
b. z-statistics in parenthesis. Z- statistics test whether the change in the hazard rate is significant. A coefficient of 1 implies that the hazard rate has not changed.
c. DV indicates dummy variable

TABLE 2: BASELINE RESULTS FOR THE CHOICE OF EFFORT

DEPENDENT VARIABLE: LOG(REVENUES PER SHIFT)

OLS REGRESSIONS

	(1)	(2)	(3)
Treatment Effects			
Direct Treatment Effect (DTE)	-0.073*** (-3.521)	-0.058*** (-2.689)	--
Implied Elasticity of Substitution	-0.332	-0.255	
DTE × Fixed Shift	--	--	-0.068*** (-3.013)
DTE × Variable Shift	--	--	-0.094** (-1.978)
Indirect Treatment Effect	0.021 (0.994)	0.031 (1.268)	0.02 (0.939)
Announcement Effect	0.046** (2.412)	0.028 (1.201)	0.046** (2.409)
Control Variables			
Daily Fixed Effects	Yes***	Yes***	Yes***
Messenger Fixed Effects	No	Yes***	No
# Competing Bicycle Messengers	-0.036*** (-8.765)	-0.034*** (-8.074)	-0.036*** (-8.806)
# Competing Car Messengers	-0.049*** (-9.181)	-0.045*** (-7.228)	-0.049*** (-9.773)
Will work on next day (DV)	0.075*** (8.778)	0.045*** (5.099)	0.075*** (8.096)
Worked yesterday (DV)	0.028*** (2.89)	0.001 (0.064)	0.028*** (2.97)
Will work on next day, and has worked yesterday (DV)	0.041*** (3.274)	0.051*** (4.223)	0.041*** (3.204)

(continued on next page)

Log(Experience)	0.187*** (14.4)	0.095*** (5.415)	0.187*** (13.956)
Log(Experience)^2	-0.014*** (-9.165)	-0.001 (-0.19)	-0.014*** (-9.324)
First Month (DV)	-0.022 (-1.518)	-0.029* (-1.888)	-0.022 (-1.41)
Last Month (DV)	-0.021 (-1.485)	-0.02 (-1.444)	-0.021 (-1.272)
Female (DV)	-0.085*** (-6.536)	--	-0.085*** (-5.325)
Member of Veloblitz (DV)	-0.035 (-1.098)	--	-0.035 (-1.239)
Within Days R^2	0.121	0.396	0.121
Fraction of unobserved variance due to daily fixed effects	0.585	0.67	0.585
Number of Observations	21,737	21,737	21,737

Notes: a. *, **, *** denotes significance at the 10, 5, and 1 percent level, respectively
b. z-statistics in parentheses, adjusted for clustering on days.
c. constant term omitted.
d. DV indicates dummy variable.

TABLE 3: OUTCOMES OF THE LOTTERY EXPERIMENTS

Percentage of messengers rejecting …		Percentage of messengers rejecting …	
Lottery *A*	54 %	Lottery *C*	28 %
Lottery *B*	42 %	Lottery *D*	14 %
Lotteries *A* or *B*	62 %	Lotteries *C* or *D*	33 %

Notes: a. N = 72 messengers at Veloblitz and Flash Delivery Services.

b. See text for a description of the lotteries.

TABLE 4: EXPLAINING THE DIRECT TREATMENT EFFECT ON EFFORT

DEPENDENT VARIABLE IN ALL MODELS: LOG(REVENUES PER SHIFT)

OLS REGRESSIONS

	Model (1) Interaction with: Loss Aversion Measure 1 N =21,560	**Model (2)** Interaction with: Loss Aversion Measure 2 N =21,560	**Model (3)** Interaction with: Loss Aversion Measure 3 N =21,560
Intertemporal Elasticity of Substitution			
Direct Treatment Effect (DTE)	-0.112 (-0.606)	-0.099 (-0.543)	-0.067 (0.417)
DTE × Loss Aversion Measure	-0.385** (-1.939)	-0.273** (-2.101)	-0.318** (-2.287)
Test for joint significance	$p < 0.01$	$p < 0.01$	$p < 0.01$
Interaction alone			
DTE × Loss Aversion Measure	-0.475*** (-3.66)	-0.323*** (-3.788)	-0.359*** (-3.627)
Interaction alone, including Messenger Fixed Effects			
DTE × Loss Aversion Measure	-0.345*** (-2.588)	-0.229*** (-2.564)	-0.273*** (-2.758)

Notes:
a. *, **, *** denotes significance at the 10, 5, and 1 percent level, respectively.
b. z-statistics in parenthesis, adjusted for clustering on days.
c. same controls as in Table 2, column (1).
d. Tests for joint significance are F-tests testing whether both coefficients equal 0.
e. See text for an explanation of the Loss Aversion measures.

TABLE 5: EXPLAINING THE DIRECT TREATMENT EFFECT ON SHIFTS

DEPENDENT VARIABLE: PROBABILITY OF WORKING, CONDITIONAL ON DAYS SINCE LAST SHIFT

(CHANGES IN HAZARD RATES DISPLAYED)

COX REGRESSIONS

	Model (1) Interaction with: Loss Aversion Measure 1 N =21,278 failures	**Model (2)** Interaction with: Loss Aversion Measure 2 N =21, 278 failures	**Model (3)** Interaction with: Loss Aversion Measure 3 N =21,287 failures
Direct Treatment Effect (DTE)	1.27** (2.261)	1.271** (2.426)	1.263** (2.279)
DTE × Loss Aversion Measure	0.934 (-0.551)	0.948 (-0.68)	0.96 (-.515)

Notes:

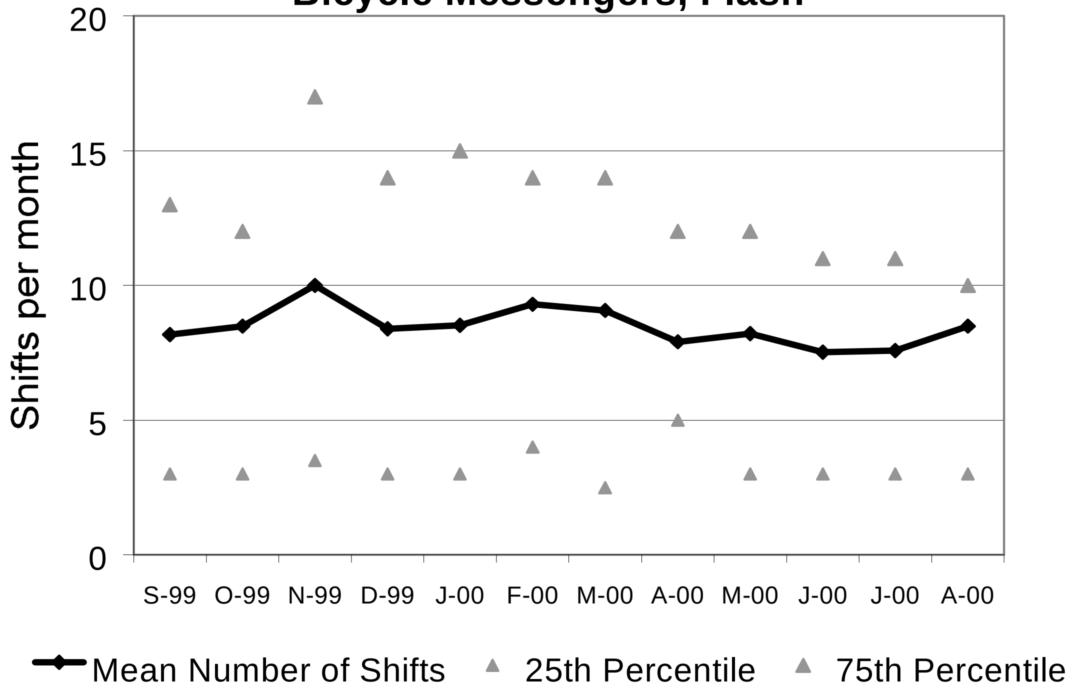
a. *, **, *** denotes significance at the 10, 5, and 1 percent level, respectively.

b. z-statistics in parenthesis. Z- statistics test whether the change in the hazard rate is significant. A coefficient of 1 implies that the hazard rate has not changed. A coefficient above one implies an increase in the hazard rate, a coefficient below one implies a decrease in the hazard rate.

c. same controls as in Table 1, column (2).

Figure 1a: Number of Shifts
Experimental Subjects, Veloblitz

Figure 1b: Number of Shifts
Bicycle Messengers, Flash



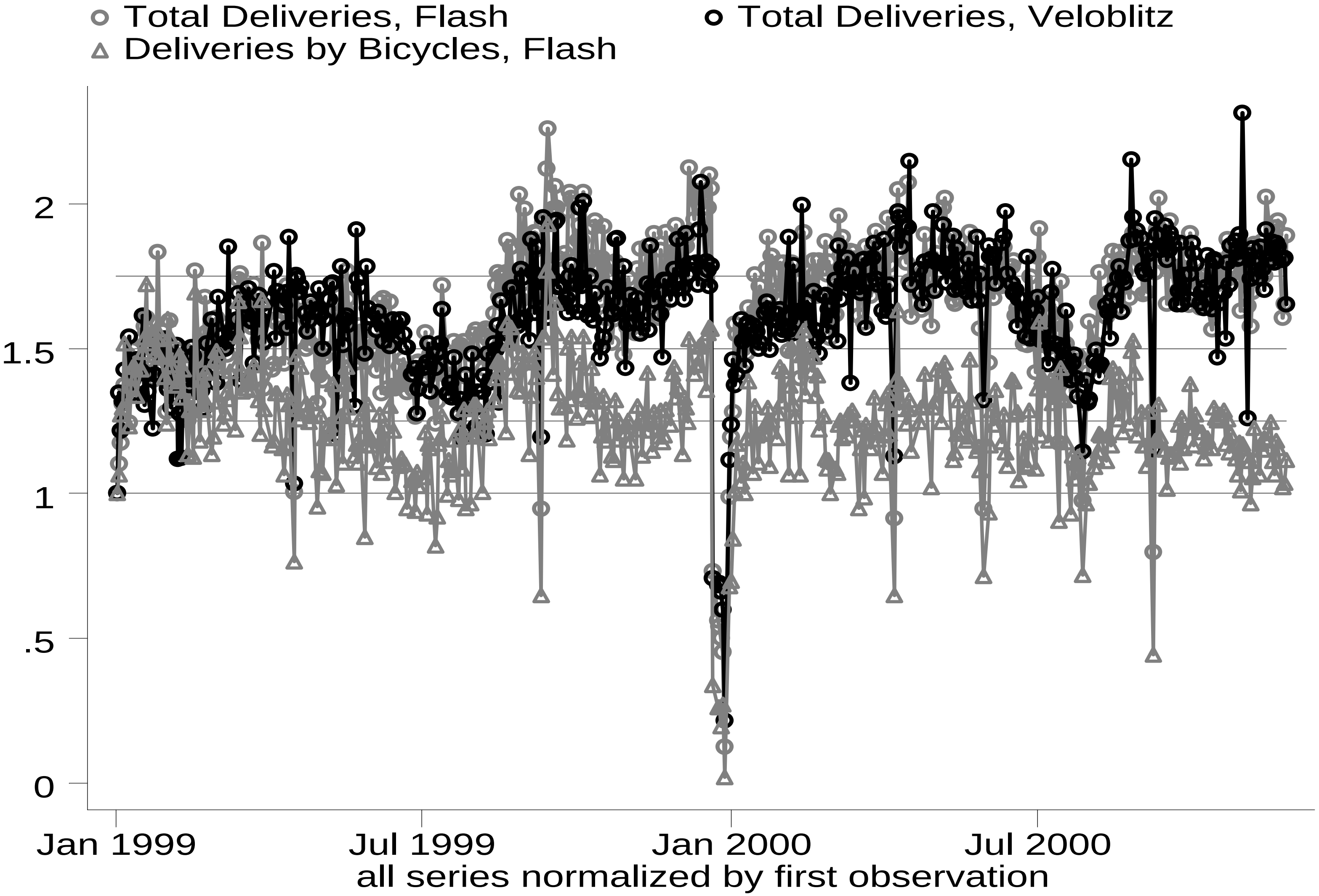


Figure 2: The Demand For Messenger Services

Figure 3a: Mean Revenues and Volatility of Revenues Experimental Subjects only, Veloblitz

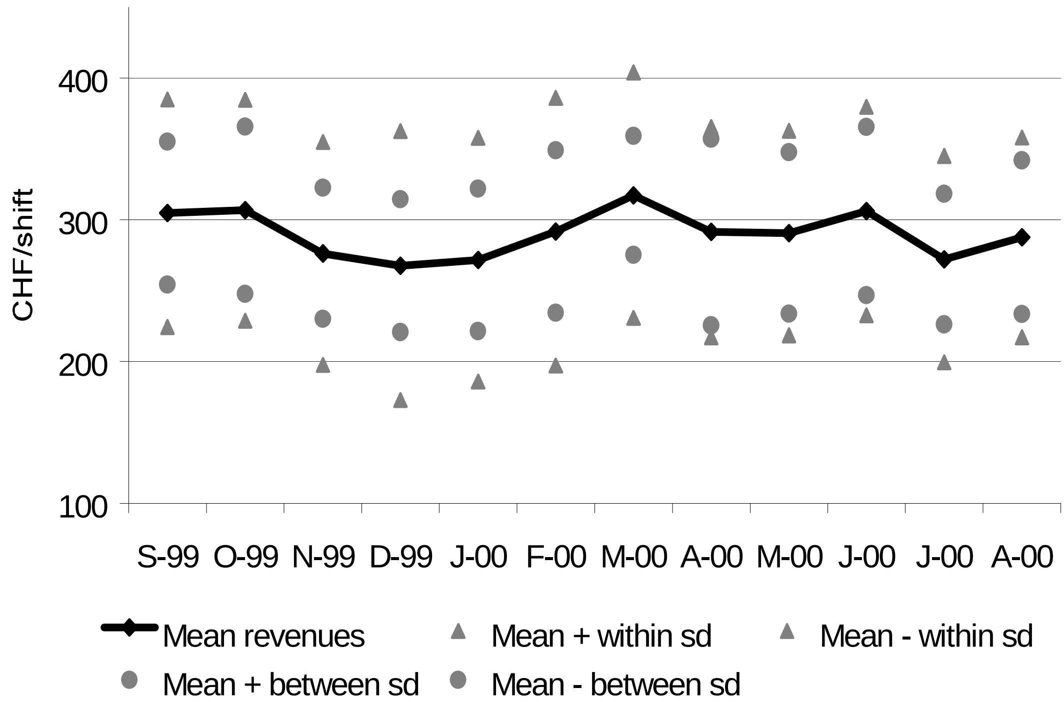
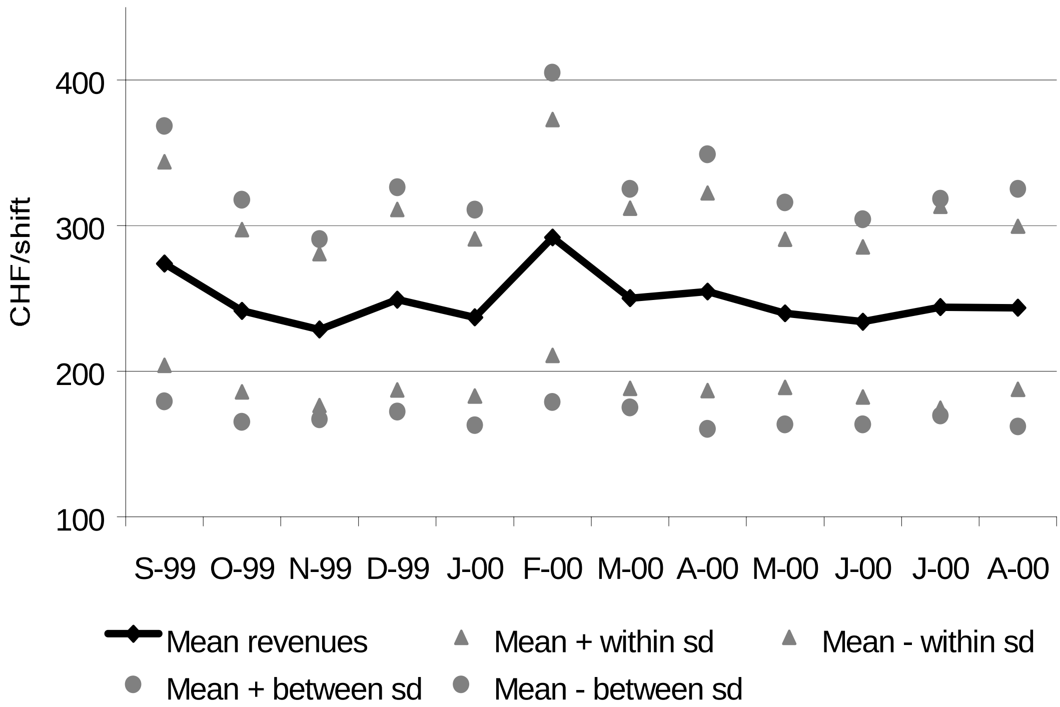


Figure 3b: Mean Revenues and Volatility of Revenues Bicycle Messengers, Flash



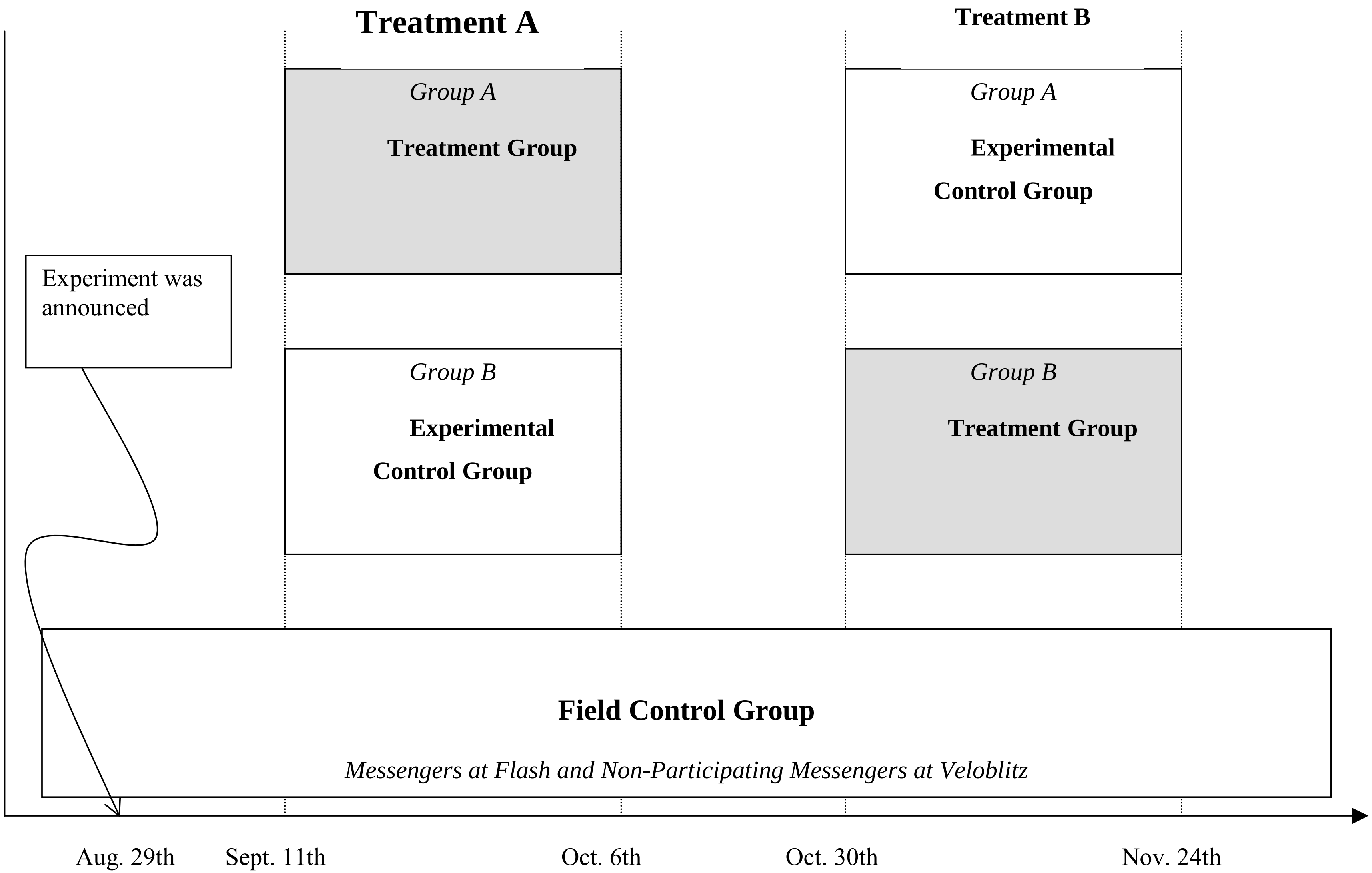


FIGURE 4: THE TIMING OF EVENTS

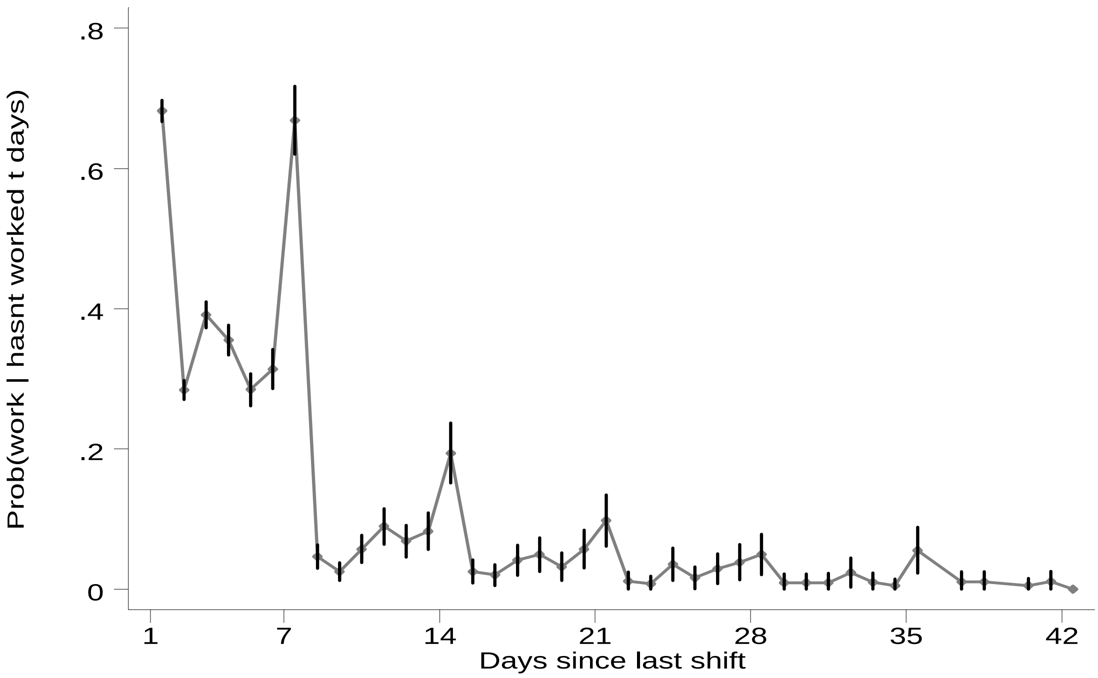


Figure 5a: The Working Hazard at Veloblitz

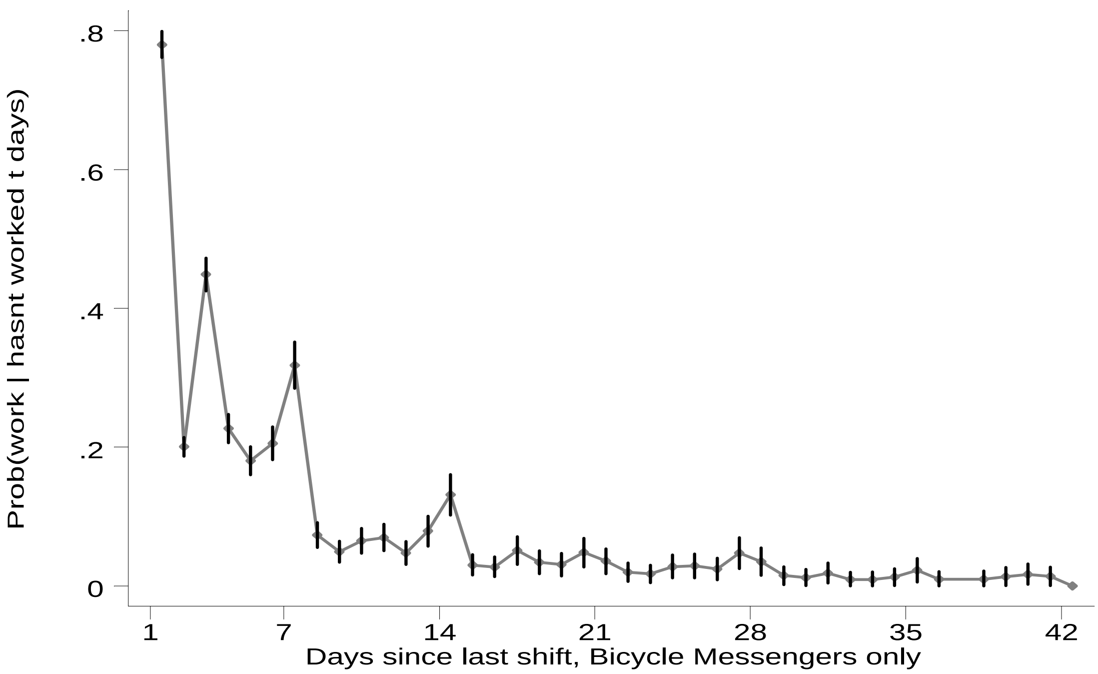


Figure 5b: The Working Hazard at Flash

Figure 6: The Working Hazard during the Experiment

Figure 7: The Distribution of Revenues per Shift during the Field Experiment

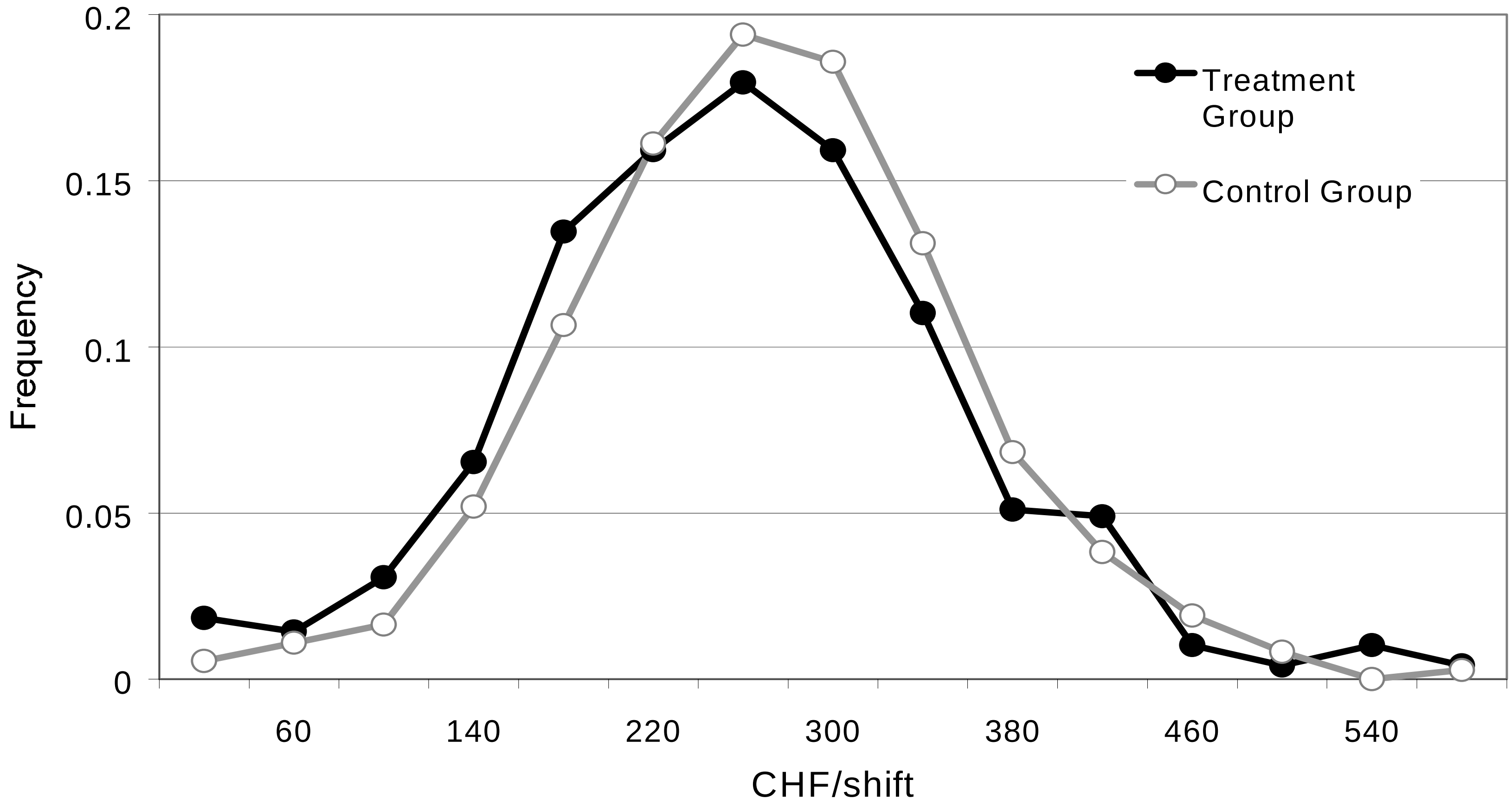


Figure 8a: The Distribution of Revenues per Shift in the Treatment Group

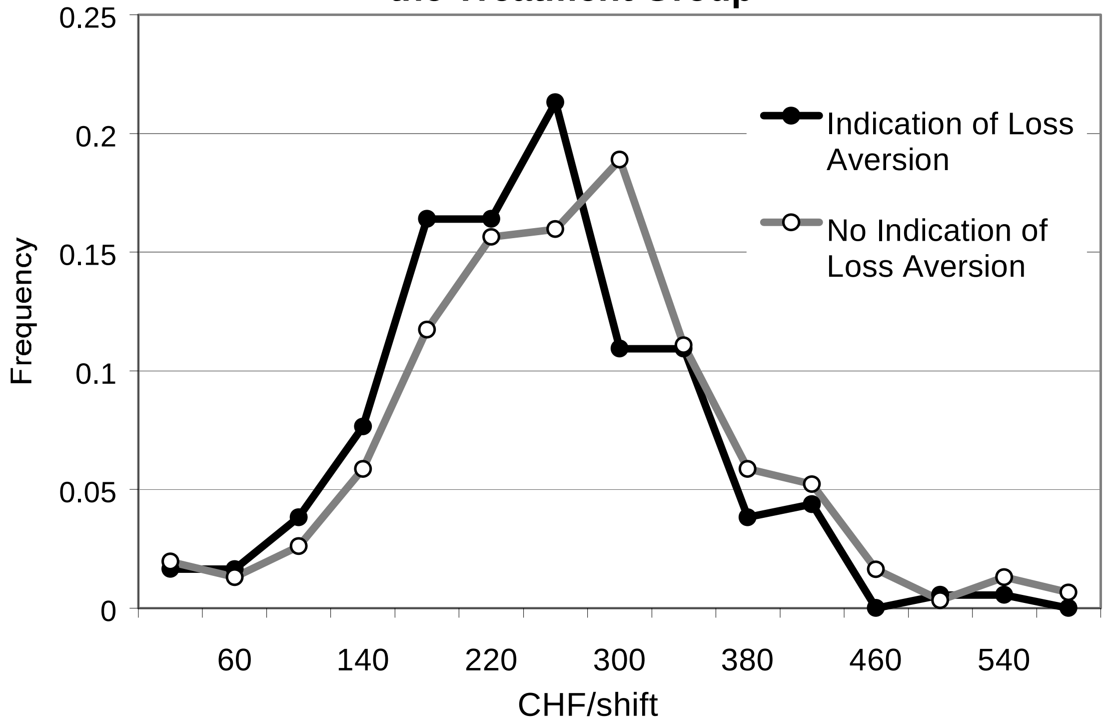


Figure 8b: The Distribution of Revenues per Shift across Control and Treatment Group