# EC 831

# The responsiveness of labour supply to varying pay rates and tax rates and the micro-foundations of the Laffer curve in a real-effort experiment

Jakub Jilek

University of Essex

Department of Economics

April, 2014

Word count: 9980

## Abstract

This paper examines the responsiveness of labour supply to varying incentives in real-effort experiments. The focus lies on the effects of taxation on individual effort. Previous experimental literature attempted to answer a wide array of research questions by manipulating the pay-rates and tax rates imposed on participants. Some of the experiments had one or several features that may be problematic for their external validity. A new experiment was conducted in EssexLab at University of Essex in an attempt to resolve these issues and obtain new data. Subjects were paid a piece-rate for carrying out a real-effort task and their revenue was taxed. The pay-rates and tax rates varied. All participants had an opportunity evade taxes by underreporting their income. In contrast to previous literature, the results of OLS regression show that participants were responsive to their net wages. Taxes, on the other hand, had no significant effect on their own. No evidence was found for the existence of a Laffer curve. Almost none of the participants attempted to evade taxes.

# Contents

# 1. Introduction

## 1.1. Motivation

The aim of income tax research is to understand the effects of income taxation on individual and aggregate behaviour. Taxation makes labour more expensive and lowers personal income, thereby affecting the economic incentives faced by workers and firms. Understanding these effects enables the government to assess the impact of its policies on the economy. Taxes are also the chief source of government revenue. If they are set too high, government revenue may decrease, as high tax rates disincentivise work and motivate people to find ways to avoid paying them. Governments with a good understanding of the incentives produced by taxation can thus learn to better optimise their revenue. (Hausman, 1985)

The first objective of this paper is to empirically test the prediction of supply-side economics that taxation has a disincentive effect on labour supply (Swenson, 1988). If this prediction is true, governments aiming to increase production should lower income taxes. The second objective is to provide the micro-foundations for the Laffer curve, which is the functional relationship between tax level and tax income. Revenue-optimising governments should set the tax rate that maximises their revenue. Finding the optimal rate is thus of importance. A supplementary goal of this paper is to investigate the relationship between tax rates and tax evasion.

There are two primary methods of conducting empirical research – experimentally or through observational (econometric) studies. Econometric studies analyse data from naturally occurring markets, which may contain a lot of noise. Taxpayer behaviour may be affected by confounding factors, which cannot be measured or modelled. Since variables in the real world cannot be easily manipulated (e.g. tax policy cannot be experimentally changed), it may be difficult to perform critical tests of a theory in a real-world setting. In contrast, experiments allow the researcher to create a well-defined micro-economy tailored to a particular theory (Davis, et al., 1987). Whereas in observational studies, variables are often determined endogenously, experiments allow the researcher to exogenously manipulate selected variables, enabling her to infer causality rather than just correlation. Therefore, experiments make it possible to simulate an institutional setting of interest and directly measure the effects of specific institutional arrangements. For these reasons, experiments are a very useful tool for tax policy research (Charness, 2010).

There are two approaches to operationalizing the labour supply variable in experiments. One is to present subjects with a task and ask them to indicate how much effort they would be willing to provide under different conditions. This is referred to as chosen-effort. The other is to have the subjects carry out the actual task, called real-effort. There is an on-going debate as to which method is superior. Brüggen and Strobel (2007) find evidence for the equivalence of the two methods. Charness (2010) contends that real-effort experiments are a good response to the criticism that tasks may be too abstract or artificial. As will be discussed later, using a real-effort design is a good way to satisfy the necessary conditions for a well-defined experimental micro-economy.

A major concern in experimental Economics is about the external validity of results. Whereas observational studies collect data from the real world, laboratory experiments attempt to simulate a situation from the real world in a laboratory. Due to practical reasons, experiments provide less complex incentives and institutional structures than the real world. This is potentially problematic. Are experiments sufficiently similar to the real world to warrant real-world inferences? This question has no clear solution and is also subject to on-going debate. Charness (2010) recommends that design should be made as complex and similar to the real world as possible. As Bardsley (2010) contends, experiments are not intended to replace observational studies. Rather, they are to complement them in areas where relationships hard to measure or model outside the laboratory.

Vernon Smith (1976) defined 5 essential conditions for a well-defined laboratory micro-economy. The first one is the assumption of non-satiation, which guarantees consistent preferences: Subjects prefer more not less of income. Secondly, subjects must fully understand the link between effort and reward payments (salience). Thirdly, the payment must dominate any other preference the subjects may have (dominance). Fourthly, the experiment must not be too long lest the subjects become bored, which would lead to a loss of control over preferences. The fifth condition is an assumption about subjects' preferences: Subjects abide by the expected utility model when making decisions.

## 1.2. Overview

There is an existing body of experimental literature where subjects were paid to carry out real-effort tasks and their income was taxed. The researchers attempted to answer various questions, including the investigation of disincentives, tax evasion and the Laffer curve. However, the design of most experiments had aspects that may be problematic for their external validity. These aspects are described in detail in the Literature Review. For this reason, a new experiment was conducted in EssexLab at University of Essex, which attempted to address the problems with an improved design.

The rest of this paper unfolds as follows: Firstly, a review is presented of the existing literature on taxation real-effort experiments. Secondly, a new experiment is presented with improvements in design that aim to address the concerns with previous literature. This part also explains how the conditions for a well-defined experiment are satisfied. Thirdly, two models for optimising behaviour and one for the Laffer curve are introduced. Fourthly, the results of the experiment are analysed using OLS regressions. Finally, the paper summarises the findings and draws lessons for further research.

## 2. Literature Review

### 2.1. Literature sources

This part presents an exhaustive summary and analysis of all existing real-effort taxation experiments. Most research papers were found using Google Scholar, JSTOR, Springer and ScienceDirect. A thorough search was accompanied by a request on the ESA Experimental Methods Discussion Forum, where academics in the field of experimental Economics kindly recommended missing articles that could not be found by search engines.[1]

### 2.2. Linear and non-linear taxes

The first real-effort taxation experiment was conducted by Swenson (1988). The same design was later replicated by Sillamaa (1999a) with a different real-effort task.[2] Sillamaa paid participants to carry out a task under 5 tax treatments: 12%, 28%, 50%, 73% and 87% tax. Each subject was exposed to one out of 13 different sequences consisting of a random combination of these treatments. Participants also received a free exogenous income. Controlling for the free income, all taxation treatments had a significant negative effect on labour supply ($p<0.02$). Two other experiments by Sillamaa (1999b and 1999c) imposed various types of non-linear taxes such as regressive, progressive[3] and zero-tailed tax.[4]

[1] I cannot guarantee that there does not exist a paper which I did not find. However, the search took several weeks and I used all possible avenues of finding the papers, thus it is likely that almost all papers are included here.

[2] Swenson's task consisted of hitting keys on a keyboard. Since this task had a high error rate (>30%), Sillamaa replaced it with a letter-decoding task with an error rate (<2%), allowing a more precise analysis. Due to the similarity of these experiments in other respects, I further refer only to Sillamaa.

[3] Regressive tax has a decreasing marginal rate, whereas progressive tax has an increasing marginal rate.

Sillamaa found evidence that all types of non-linear taxes disincentivise effort, but the effect varied.

In the non-linear tax experiments, Sillamaa observed a logarithmically rising trend in average output with a peak in the middle of the experiment. She hypothesised that the trend was due to learning and fatigue effects. In order to control for these effects, a period variable was added and used as one of the regressors. It is hard to establish how well the trend variable fitted the effects. Given that the adjustment was aggregate, not individual, it was likely to be imprecise if participants had different learning-fatigue curves.

Sillamaa also noted that participants differed in ability, which had an impact on labour supply decisions. In order to control for ability, Sillamaa used the rank variable instead of effort as the regressand, which took the value of 1 for the lowest work effort of an individual participant, the value of 2 for the second lowest effort, etc.[5] While controlling for ability, using rank also prevents the researcher from measuring the changes in actual labour supply from the changes in taxation level. It seems that a much better way of controlling for both the learning effects and ability is to measure the time each participant spends on one unit of task. This idea is elaborated on in Experimental Design.

### 2.3. The welfare state

Guido Ortona et al. (2008) criticised the experimental evidence from Sillamaa and Swenson as displaced. In their experiments, subjects were exposed to a Leviathan state, which only imposed taxes and provided no benefits or public goods. According to Ortona, the results are therefore valid only under strong leviathan-state assumptions. If taxation were to be used as insurance, protecting subjects against risk, their expected income would in fact rise, inducing more labour supply.

To test this hypothesis, participants were randomly selected into one of the two contracts: the State of Nature and the Welfare State. Both contracts exposed participants to the risk of losing income. In contrast to the State of Nature, The Welfare State imposed a proportional tax, which was used to fund insurance and benefits. The tax revenue minus transaction costs and insurance payments was then redistributed equally among subjects

[4] Zero-tailed tax is regressive up to a certain point, after which the marginal rate is zero.

[5] To illustrate, consider that a participant worked under three treatments, with the following effort levels: 12, 7, 28. His/her rank values would therefore be 2, 3, 1.

irrespective of their contribution. Ottone & Ponzano (2008) used an identical experimental design to test for the existence of the Laffer curve.[6]

Both experiments manipulated several variables, such as state's efficiency, transaction costs, and risk. The taxation treatments imposed 30%, 50% or 70% tax. The Welfare State contract did not decrease output, compared to the State of Nature. On the contrary, it was output-enhancing, as long as the tax did not exceed 50%. Labour supply significantly decreased when tax rate grew from 50% to 70% but not from 30% to 50%. Consequently, tax revenue significantly increased only when the tax increased from 30% to 50%. The results thus provided evidence for the Laffer curve.

It is possible that both experiments rested on too strong assumptions. It is not obvious that citizens of democratic countries perceive taxation as insurance and a payment for public goods, as the recipients of public goods and benefits are not necessarily the ones who pay for them from their taxes. There is no direct relationship between the amount of taxation a particular taxpayer pays and the amount of public goods they receive. It seems more likely that a marginal increase in taxation is understood as lower income. For these reasons, Swenson's and Sillamaa's experiments may not be so far from reality in their assumptions as was suggested by Ortona et al.

## 2.4. Behavioural effects

Sutter (2002) examined whether people punish tax setters who set 'unfair'[7] taxes by producing less output. Subjects played a two-player game, one of them being the tax setter and the other the taxpayer. In the certainty treatment, subjects knew their role beforehand. In the uncertainty treatment, they were uncertain about their final position. Average effort levels were higher in the uncertainty treatment and tax revenues were highest at the 50% tax; whereas in the certainty treatment they peaked at 65%. Sutter concluded that in the uncertainty treatment, participants chose a 'fair' tax, because they did not know whether they would end up as taxpayers or tax setters. Consequently, the taxpayers produced more output. Lévy-Garboua (2008) obtained identical results in a similar experiment.

[6] Laffer curve is a proposed non-linear relationship between taxation level and taxation revenue. As the taxation level increases, so do government revenues. However, this continues until a certain point, after which a further increase in taxes lowers the revenue. This may happen for several reasons, such as people being disincentivised from working, or moving into the black economy. A formal model is provided in a later section.

[7] An unfair tax is defined as a tax that violates a generally accepted social norm. A fair tax is not necessarily lower than an unfair tax.

Both experiments provide some evidence that labour supply decisions may be influenced by psychological factors. Tax rates perceived as unfair lead to a larger reduction in labour supply than equivalent 'fair' rates. However, it is not clear that citizens would try to punish the tax-setter (i.e. the state) by providing less labour supply than is required by income-leisure optimisation, especially in the long run. Firstly, it would be against their financial self-interest. It is possible that if the game were played many times during the experiment, subjects would converge to the optimal level of effort, eliminating the behavioural effect. Secondly, in the experiment, the tax setter was a known tangible person. Hence the taxpayer may have had a personal motivation of retribution towards that person when they set the tax too high. In the real world, the tax setter is not a tangible person. Therefore, personal motivation may not play such an important role.

## 2.5. Tax Evasion

Another important factor that could have an impact on labour supply is the opportunity to evade taxes. Kirchler et al. (2009) hypothesised that tax compliance may be affected by the amount of effort exerted. Two experiments were conducted to test the hypothesis. Congruent with the Prospect Theory, the results showed that tax evasion was more common in low-effort conditions, since higher-income individuals face the risk of losing more money than lower-income individuals.[8] Risk-averse individuals will therefore be more compliant when they exert more effort. Bühren & Kundt (2013) obtained the same results in a similar experiment.

Doerrenberg & Duncan (2012) tested a hypothesis that labour supply elasticity depends on tax evasion opportunities. The experiment had one control and one treatment condition, both of which imposed three tax levels (15%, 35%, 50%). In the tax evasion treatment, participants could underreport their income with a 10% chance of being caught. Upon detection a fine was imposed equal to 2 times the tax revenue from the participant. Subjects had significantly stronger responses under the tax evasion treatment, implying that tax evasion makes labour supply more elastic.

Doerrenberg & Duncan paid subjects for one randomly selected round, in order to avoid satiation.[9] This is potentially problematic, as it could blunt the incentives. If risk-averse participants know that they will be paid only for one of the rounds, they might aim to spread

[8] See Daniel Kahneman, 'Prospect Theory: An Analysis of Decision under Risk' (1979)

[9] Satiation occurs when the participant has earned enough to satisfy their aspiration level.

their effort more evenly across all treatments in order to minimise risk. As a result, there may be less work effort under low taxes and more effort under high taxes than otherwise.[10]

## 2.6. Methodology

There are three major methodological differences among the experiments that deserve a discussion. Firstly, except for Sillamaa (1999b, 1999c) and Swenson (1988), all the experiments applied the strategy method. Participants were informed about all tax treatments in advance and they had to choose how much effort to provide under each, prior to being randomly selected to one. In the direct-response method, subjects were assigned a treatment first and then they decided about effort. Brandts and Charness (2011) found mixed evidence on the differences between the two methods. Whereas the treatment effect is always observable regardless the method, its strength differs, contingent on what is being measured. It appears that the strategy method makes the tax rates and pay rates more salient but at the same time risks causing the demand effect.

The second difference lies in the design of the income-leisure trade-off. Some experiments, e.g. Swenson (1988) and Sillamaa (1999a, 1999b, 1999c), used fixed time intervals for each treatment, during which participants could either work or do a leisure activity (such as reading a magazine or playing a computer game). Other experiments, e.g. Ottone & Ponzano (2008), allowed subjects to proceed to the next treatment whenever they wished and leave the laboratory as soon as they finished. The second method seems to be superior. It is likely that the researcher is not a very good judge of what participants consider as leisure. If the leisure activity is not what the participants would choose for themselves[11] and if they have to spend a predefined time in the laboratory, they may decide to spend it on earning income rather than the enjoying leisure. Therefore, using fixed time intervals can lead to a loss of control over preferences.

## 2.7. Framing and reference points

The third dissimilarity concerns framing. While Sillamaa (1999a, 1999b, 1999c) familiarised the subjects with their net wage and never mentioned the word ‘tax’, all the other authors informed participants about their gross wage and tax rate. Fochmann (2010)

[10] This will happen assuming that participants come to the experiment with a certain aspiration level in the form of a target income.

[11] The activity can be boring or uninteresting.

conducted an experiment in order to test whether workers base their work-leisure decisions on their net wage or the gross wage. Keeping the net wage constant, Fochmann manipulated the gross wage and taxation rate. There were significant differences ($p<0.05$) between the treatments, suggesting that workers are influenced by the gross wage. Fochmann dubbed this effect a 'gross wage illusion'. This result poses a question about the external validity of Sillamaa's experiments. It is possible that they merely measured subjects' responsiveness to different pay schemes, rather than different taxes. The results may have been different if the subjects had known that they were being taxed.

Fochmann's experiment highlights the importance of reference points. Researchers in the field of decision-making have demonstrated in various experiments that the distribution of reference points has an impact on subjects' evaluation of a stimulus. For instance, Hsee (1996) had participants evaluate two dictionaries:

1) 50,000 words, slightly damaged 2) 30,000 words, new

When evaluated separately, subjects placed a higher value on the second dictionary. When evaluated jointly, the first dictionary was valued more highly. It is likely that joint evaluation made it easier to evaluate what counts as a 'large' or 'small' number of entries in the dictionary. Analogously, when deciding how much effort to provide in Fochmann's experiment, subjects made an evaluation of the contract they were working under.[12] It is possible that joint evaluation of different contracts would make it easier for participants to evaluate what constituted a 'high' or 'low' tax. Consequently, the gross-wage illusion may disappear, as subjects will take into consideration both the gross wage and the tax rate, instead of focusing mainly on the gross wage.

## 2.8. Behaviour under complexity

All the experiments described above share one particular feature of design that may be problematic for their external validity – a lack of complexity. Abeler (2013) conducted an experiment to test whether the degree of complexity has any influence on behaviour. Subjects faced either a simple (SS) or a complex set (CS) of rules. Whereas 40% of participants chose the payoff-maximising output under the SS, only 1.7% did so under the CS. Furthermore, when one extra tax rule was added to both the simple and complex treatments, subjects in the

[12] Contract is understood as the pay rate and tax rate.

complex treatment underreacted to the new incentive, compared to their counterparts in the simple treatment ($p<0.01$). Therefore, complexity may have an impact on optimisation behaviour, producing a 'status-quo' bias. Considering that people's behaviour is affected by complexity, it seems appropriate to ask how much can be inferred from simple-rules experiments that is applicable onto the real world. Although a valid concern, it ignores the possibility of adjustment over time. Through processes such as trial and error, citizens may converge to an optimising level of labour supply over time, despite complexity.

## 3. Experimental Design

### 3.1. Overview

In the light of the previous literature, a new experiment was conducted in EssexLab at University of Essex. Participants were recruited by email from the usual student pool a week in advance. The experiment was paper-based and had only one session, due to financial limitations. The author of this paper funded the experiment from his personal resources.

There were 28 subjects in total. Each subject received a participation fee of £2.5 in addition to the earnings made during the session. The average payment per session was guaranteed to be at least £8, in accord with the laboratory regulations. Paying participants above the current minimum wage[13] was necessary to ensure that the dominance condition be satisfied.

Participants carried out a paid real-effort task, which consisted of calculating the sums of sequences of numbers, such as the following:

8 + 3 + 9 + 2 + 3 + 2 + 4 + 7 + 1 + 2 + 2 + 4 + 8 + 7 + 3 + 4 + 9 + 1 + 5 =

Subjects were paid a gross pay-rate per each 3 correctly calculated sequences and their income was taxed. The pay-rate and the tax rate varied during the experiment. The real-effort task had 3 work-periods (from now on called rounds). In each round, participants received one worksheet with 27 sequences. Subjects were instructed at the beginning that their pay-rate and wage-rate would differ in each round and that there would be 3 rounds in total.

[13] The current minimum wage in the UK is £6.31 (https://www.gov.uk/national-minimum-wage-rates)

### 3.2. Part A

The first 10 minutes were devoted to instructions. After the instructions, part A began. Each Participant received their Trial Sheet and their Contract Sheet. They had 15 minutes to attempt the task and read through the contracts. The Trial Sheet featured 9 sample sequences with correct answers. Subjects had an opportunity to attempt the task and learn how fast and precise they were. Providing a trial period served to minimise learning effects and helped participants make a better optimising decision. In the contract sheet, subjects familiarised themselves with 9 possible contracts. An example is shown below:

| Contract | Pay Rate | Tax Rate |
|---|---|---|
| A | 0.3 | 83% |

Each contract imposed a different pay-rate and tax-rate. Although the focus of the experiment lay on measuring the effects of taxation, varying the pay rate was important for examining whether the effect of taxation remains constant with wage.

Participants were instructed that they would be randomly selected to work under 3 of these contracts in 3 rounds and that they had 35 minutes in total. Since previous experiments lasted about an hour, 35 minutes was considered short enough time to keep control over preferences. Subjects could decide how much time to spend on each round. Whenever they wished to proceed to the next round, they raised their identification sheet above their head and waited for the supervisor to bring them their next sheet. Instead of reading magazines or playing games in the laboratory, which some subjects may not consider as leisure, they could leave the experiment at any point and do whatever they wished in their free time. As discussed earlier, this type of design seems to more realistically represent the income-leisure trade-off, satisfying the fourth condition proposed by Smith (1976).

The Contract Sheet was used in order to address the issues with previous experiments. The first concern regards framing. Sillamaa and Swenson did not inform their subjects that they were being taxed. Rather, they informed their subjects of their net pay-rate. To enhance external validity, subjects in this experiment were informed of their pay rate and their tax rate.

The second concern is closely related to the first. A considerable body of psychological literature has shown that the reference point matters for evaluation.[14] If

[14] See for instance the Range Frequency Theory by A. Parducci (1963)

subjects have no reference point provided in the experiment, they are likely to evaluate their contract (i.e. decide how much effort to provide) by its most salient feature, whether it be the tax rate or pay rate. Further, subjects may use their individual knowledge of tax levels as a subjective reference point. If they come from different countries, the reference points will differ. For these reasons, the researcher may lose control over preferences. Providing each subject with the same contract holds the reference point constant. Contracts A, C, E, F, G, I thus had exactly the same pay-rate and tax-rate for each participant. Also, each participant worked under contract B in the first round, contract D in the second round; and contract H in the third round. All treatments can be found in Table 1.

This experiment used a design lying somewhere between the strategy method and the direct response method. Participants were not asked to indicate how much effort they would provide under each of the contracts. Rather, they were instructed to rank them by preference (from 1 – the best; to 9 – the worst). The ranking was used to motivate participants to think about the different contracts, making the reference point more salient. Similar to the direct-response method, subjects decided on their labour allocation after they learnt about their contract. Using a combined design served to preserve the advantage of the strategy method in making contracts more salient, while mitigating its potential downside of causing the demand effect.

The welfare state was somewhat present in the experiment in the form of the lump-sum participation fee. Any connection between tax payments, insurance against risk and provision of public goods was deemed unrealistic, hence not included. Considerations of fairness were completely excluded from the experiment.[15]

### 3.3. Part B

After the first 15 minutes passed, all subjects received their worksheet for round 1. The worksheet had 27 sequences of numbers as described above. When a participant received their worksheet, the supervisor took down the exact minute and second of time. Measuring the time each participant spent on each round seemed to be a simple yet innovative way of completely controlling for different ability, learning and fatigue effects in a single measure.

[15] See Literature Review for a discussion

### 3.4. Part C

After the 3rd round of part A, participants waited several minutes until all their worksheets had been corrected and then received the results of their work. In addition, they were asked to fill in 3 Tax Reports, one for each round. Since calculators were not allowed during the session, instead of their tax payment, participants filled in the number of tasks they calculated correctly, their pay-rate and tax-rate. The lower the number of tasks reported, the lower would be the tax payment. Subjects were informed at the start that the supervisor would check 3 out of 28 'most suspiciously looking' Tax Reports in each period for dishonest behaviour. 'Suspiciously looking' Tax Report was defined as that which reported a very low number of tasks. Participants were informed that if they cheated on their tax payment, they would pay a fine equivalent to 30% of their income, in addition to their tax payment.

This type of design was considered to be more realistic as tax reports are evaluated by computers and only those with a mismatch between the income reported and records from organisations are investigated, instead of a random selection. Examining tax evasion was not the primary goal of this experiment. However, including the opportunity to evade taxes brought the experiment closer to a complex real-world scenario.

### 3.5. Part D

After filling out the Tax Reports, participants received the Questionnaire. The Questionnaire was designed to control for demographic factors, experience with taxation and other variables. The demographics consisted of gender, age and nationality. The other questions included the hourly wage, income tax, experience with tax forms; how much the participant took into consideration the gross wage and the net wage and whether they came to the experiment in order to earn an income or for other reasons. The last question was of particular importance since one of the axioms of income-leisure optimisation is that work is a bad and leisure is a good. If participants came to the laboratory for other reasons than making money, the experimenter would lose control over preferences and the result would be skewed.

## 4. Theoretical Model

### 4.1. Taxation and labour supply

Individual labour supply can be derived from the optimisation of indifference curves. The indifference curve describes all the combinations of leisure and income that give the worker-consumer an equal amount of utility. The optimal point with the highest utility is found where the constraint given by the wage and price level is tangent to the highest indifference curve. The effect of a change in the tax rate (implying a change in the net wage) can be decomposed into the income and substitution effects. The graph below[16] shows the effect on labour supply from an increase in the wage rate from $W_A$ to $W_B$ and $W_C$. Firstly, an increase in the wage rate makes leisure more expensive, inducing a substitution effect from leisure to work, represented by the movement from point A to J. As a result, the individual devotes more time to work. Secondly, the participant becomes better off. Assuming that leisure is a normal good, the participant will consume more of it. This is represented by a movement from J to B. As long as the substitution effect dominates the income effect, the labour supply curve will slope upwards (Morgan, Katz & Rosen, 2009). [17]

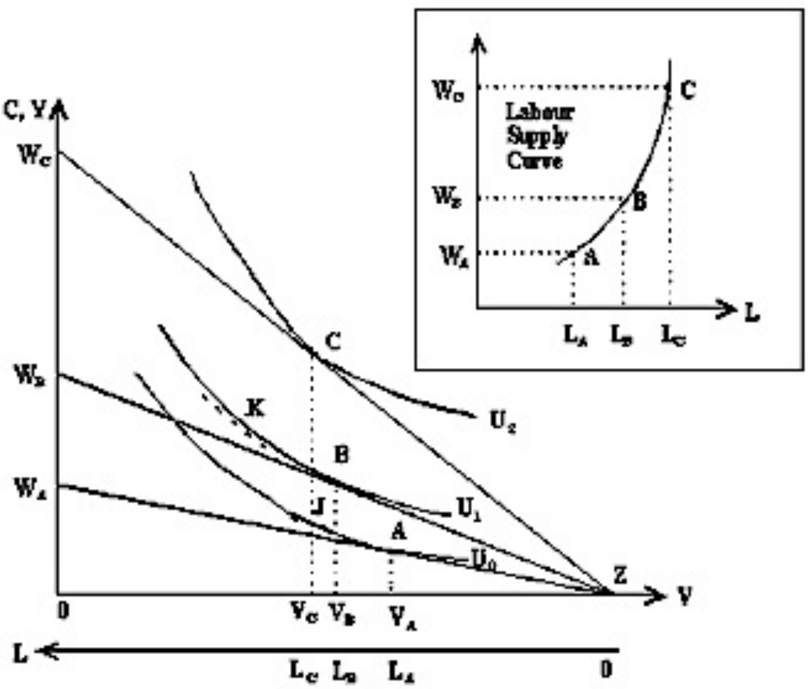



**Chart 1: Derivation of individual Labour Supply**

[16] Borrowed from (http://highered.mcgrawhill.com/sites/0070891540/student_view0/chapter2/chapter_notes.html)

[17] Since participants were paid £8 on average, the income effect was likely negligible compared to the substitution effect

We can thus estimate the following model of aggregate labour supply:

$$\text{effort} = \beta_0 + \beta_1\text{wage} + \beta_2\text{tax} + \beta_3\text{tptask},$$

where effort is the number of calculated sequences, wage is the pay-rate, tax the taxation rate; and tptask stands for the average time the participant spend on one task in a given round. Tptask is a control variable that captures variation in individual ability and learning and fatigue effects. The coefficient $\beta_3$ should negative – the longer it takes a participant to carry out one task, the less effort they will provide. We would expect $\beta_1$ to have a positive sign, as increasing the wage makes leisure more expensive compared to goods and services, incentivising more effort. Similarly, $\beta_2$ should have a negative sign as taxation makes leisure less expensive.

### 4.2. Intertemporal labour supply under risk

There is an alternative way of modelling labour supply in this experiment. Assuming that subjects participated with the primary goal of earning money and they set aside time for this purpose,[18] their optimisation problem did not involve choosing between leisure and income. Rather, it was about maximising the expected income within the already allocated time. A rational risk-neutral agent with sufficient computational capability who is maximising their income will abide by the following normative model:

$$\max y = e_1 * EV_1 + e_2 * EV_2 + e_3 * EV_3,$$

where $e_1$ stands for the number of tasks calculated in round 1 and $EV_1$ is the expected value of the net wage in round 1. Participant makes an intertemporal Labour Supply decision at the start of Round 1, when they are familiarised with their first contract. Therefore, $EV_1$ is simply the net wage in the first round.[19] $EV_2$ is the expected value of the net wage in the second round of the experiment. Since each of the contracts has an equal probability of being drawn in the second round, $EV_2$ is the average of the net wages of the remaining 8 contracts. $EV_3$ is the same as $EV_2$. Participants face 4 constraints in total. The number of sequences available

[18] Indeed, in the questionnaire, 93% of the participants indicated that their primary reason for attending the experiment was to earn money. Also, 26 out of 28 participants used up all 35 minutes.

[19] Note that the participants could not use calculators, thereby having to estimate their net wages. The degree to which their estimates were correct is verified by regression analysis in the next chapter.

on each worksheet (27) determines the first 3 constraints. The fourth constraint is given by the maximum time allowed – 35 minutes (2100 seconds).

$$\max\ y = e_1\, n_1 + e_2 \frac{1}{8}\sum_{j=2}^{9} n_j\ +\ e_3 \frac{1}{8}\sum_{j=2}^{9} n_j\ \ with\ respect\ to\ e_1, e_2, e_3$$

$$subject\ to\ \ e_1 \leq 27$$

$$e_2 \leq 27$$

$$e_3 \leq 27$$

$$e_1 + e_2 + e_3\ \leq \frac{2100}{T}$$

Non-negativity restrictions: $e_1, e_2, e_3 \geq 0$

T is time per task, i.e. how many seconds it takes a participant to calculate one sum. For simplicity, we assume that each participant has a constant speed in each round. This is not a realistic assumption, due to the learning effect and fatigue. However, it is a functionally correct assumption with regards to the subjective maximisation problem. Participants had a chance to attempt the task and learn about their individual speed prior to the start of round 1. They did or did not realise that their speed would change throughout the course of the experiment. Either way, this change could not be known in advance. Mathematically speaking, each participant would have to adjust their T by a constant for each of the three rounds to account for learning and fatigue. Since these three constants are unknown, the best that can be achieved by the agent is to use the T they learn in the Trial Period and apply it to all three rounds.

The equation above represents the optimal decision with respect to the first round. However, it does not provide the optimal solution for the second and third rounds. Participants face an intertemporal decision under risk. Once they proceed to the second round and find out what contracts they are working under, the expected values of the contracts in the second and third rounds change and so will the participant's Labour Supply Decision for the second and third rounds. Therefore, the first equation only gives the optimal solution for $e_1$; as $EV_2$ and $EV_3$ change after the first round. The optimal second-round and third-round decisions are given by:

$$\max\ y = e_2\, n_2 + \ e_3 \frac{1}{8}\sum_{j=3}^{9} n_j \ \ with\ respect\ to\ e_2, e_3$$

$$subject\ to$$

$$e_2 \leq 27$$

$$e_3 \leq 27$$

$$e_2 + e_3 \ \leq \left(\frac{2100}{T} - e_1\right)$$

$$e_2, e_3 \geq 0$$

Non-negativity restrictions: $e_2, e_3 \geq 0$

These optimisation problems require linear programming. The result is given by the simplex algorithm, which iteratively finds the optimal point. The optimal point is represented as maxef in the estimated regression:

$$\text{effort} = \ \beta_0 + \beta_1 \text{maxef} + \beta_2 \text{tptask}$$

This model assumes that the agent is risk-neutral, income-maximising and capable of calculating their net wage. In reality, participants were likely to be agents with bounded rationality, different preferences towards risk and insufficient computational capability. However, the aim of this model is not to describe actual behaviour. Rather, it is to obtain sets of optimal actions that would on average maximise income; and to test to what degree participants abided by them.

### 4.3. Laffer curve

Laffer curve denotes the non-linear relationship between taxation and government revenue. The essential premise holds that when taxation is 100%, people have no incentive to work as the government keeps all of their income. Consequently, people do not work and there is no tax revenue. No taxation also implies no revenue. The optimal tax rate therefore lies somewhere between 0 and 100%. In order to capture for the non-linear relationship, the following model will be used:

$$\text{revenue} = \ \beta_0 + \beta_1 \text{tax} + \beta_2 \text{tax}^2 + \beta_3 \text{tptask}$$

We would expect the coefficient $\beta_1$ to be positive and $\beta_2$ to be negative. Tptask is again a control variable for ability, learning and fatigue.

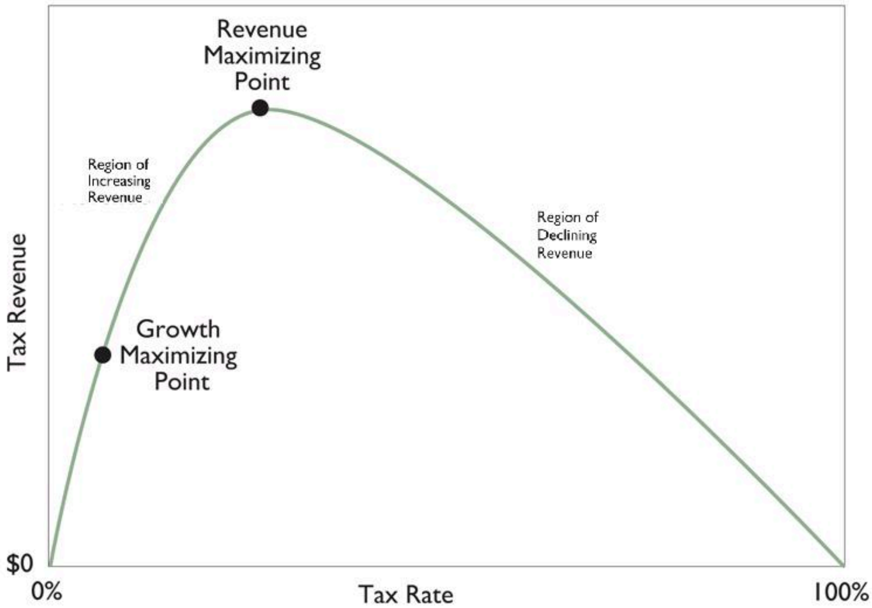


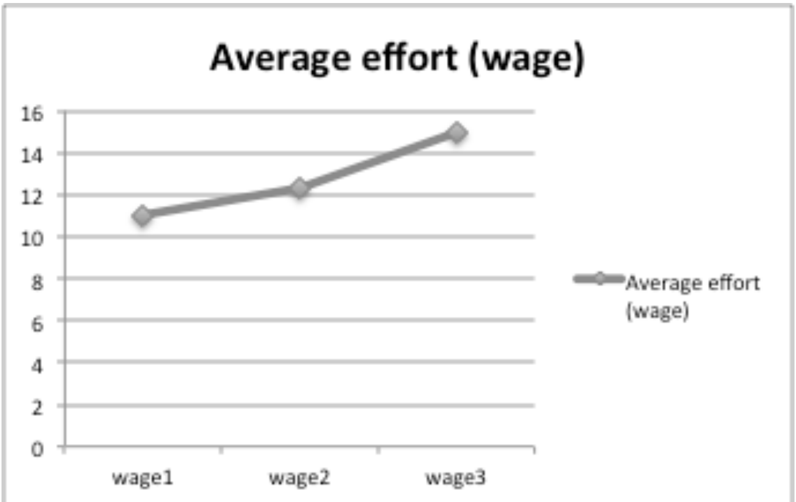

**Chart 2: Laffer Curve**

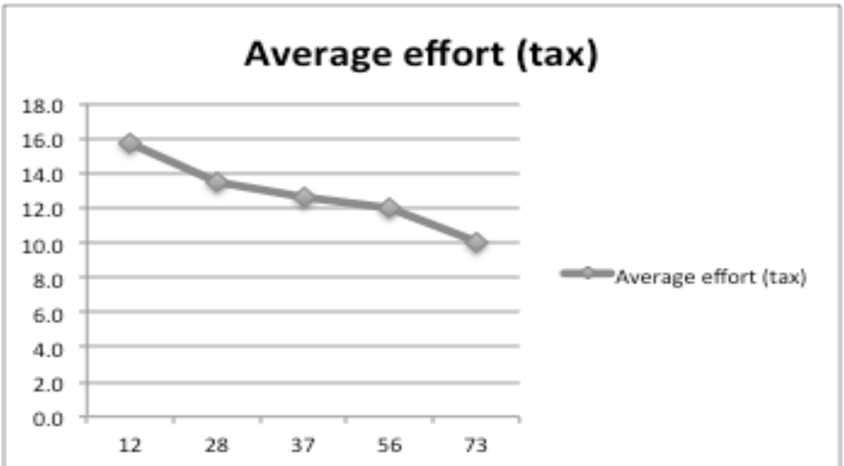
## 5. Analysis and Results

### 5.1. Graphs and Outliers

The data were analysed using the Stata package and Microsoft Excel. In the first instance, a search was conducted for possible outliers. Since the experiment worked with a small sample, outliers could have a large impact on the regression analysis. After a thorough search one possible candidate was found – participant number 9, who worked under the worst contract on their sheet in the first round. She also rated the contract as the worst by preference. However, She devoted 29 out of 35 minutes to it, carrying out the complete 27 tasks. It is highly likely that the subject misunderstood the task and thought that they were supposed to calculate as many sequences as they could in each round. Given the small sample, it seems reasonable to exclude participant 9 from the analysis.[20]

The graphs below show the average effort per wage and per tax.




[20] All regressions shown below were run excluding participant 9 as his inclusion made all variables except tptask and round1 in all regressions insignificant.

Subjects 18, 19, 20 and 25 ran out of time before reaching the third round. There were 28 participants in the experiment. Without participant 9, there were 77 observations.

## 5.2. Ability, learning and fatigue

Labour supply decisions do not depend solely on participants' preferences, but also on their ability. Sillamaa (1999) further observed that learning and fatigue also influence output. With a large enough sample, the learning and fatigue effects together with differences in ability should spread out evenly across treatments. However, given the sample of 28 subjects, it is possible that the effects of some treatments might be biased. For example, a group of participants of above-average ability could work under the same treatment, biasing the average output upwards. The zero conditional assumption requires that the expected value of the residual is independent of any combination of the dependent variables:

$$E(u|X) = E(u) = 0$$

Not controlling for learning, fatigue and ability could violate the zero conditional mean assumption if any of these were correlated with the dependent variables, resulting in biased estimators. In order to isolate the effects of taxation and enable a ceteris paribus inference, these effects must be kept under control. For this purpose, the time that each participant spent on each worksheet was measured. This measure was subsequently divided by the number of tasks attempted by the participant. The resulting variable was called time per task (tptask) and used in OLS regressions as one of the predictors of output.

## 5.3. Risk-aversion

Participants faced a constrained maximisation problem. They had to decide how much time to allocate to each of the three contracts. As the contracts were unknown to the participants, there was a degree of risk present during the experiment. Prior to conducting statistical analysis, it is important to account for possible risk-aversion. If there were sufficient randomisation and no risk-aversion, the average efforts in each round should be roughly the same. However, in the first round, participants carried out more tasks than in the other two rounds, possibly out of fear of receiving worse conditions in their second and third rounds. This may imply risk-aversion.

| | Average effort | Average tptask |
|---|---|---|
| round 1 | 15.3 | 64.5 |
| round 2 | 11.8 | 68 |
| round 3 | 11.9 | 63 |

An alternative explanation might be that participants underestimated the time needed for the remaining two rounds. Since they were regularly informed about time during the session, this seems unlikely. Another possible causal factor might be a skewed distribution of treatments in the first round in favour of high net pay-rates. However, the skew is positive (.95), indicating that most of the net rates in the first round were below the average. The tptask variable shows that the average time per task was roughly the same in all rounds. As a consequence, average effort was most probably higher in the first round due to risk-aversion. In order to control for the effect, a dummy was created for round 1.

## 5.4. Preferences

As participants were asked to rate the contracts by preference, it was of interest to establish to what degree their preferences reflected the net wages. If their preferences do not reflect the net wages,[21] participants either had non-financial preferences or bounded rationality. In regression 4.1, the null hypothesis is that the coefficient on net wage is zero, against a two-sided alternative.

$$\beta_0 = 0$$
$$\beta_1 \neq 0$$

The hypothesis is rejected at p<0.001. Net wage is a significant predictor of subjects' preferences, explaining 66% of the variation. This result provides preliminary evidence for the hypothesis that Fochmann's gross-wage illusion was caused by a missing reference point. When subjects have a number of different tax rates and pay-rates to compare, they make better decisions and the gross-wage illusion disappears, although perhaps not completely, since there is still 34% of variation left unexplained.

[21] The causes may include participants not being able to calculate their net wage, or failing to grasp the concept of taxation

## 5.5. Labour Supply

The first model of labour supply was borrowed from Sillamaa (1999), who regressed work effort on dummies for each taxation treatment and a variable capturing the learning effect. For this experiment, dummies were generated for each taxation rate and each wage rate. Regression A is the following:

$$\text{effort} = \beta_0 + \beta_1 \text{dwage}_1 + \beta_2 \text{dwage}_2 + \beta_3 \text{tax}_1 + \beta_4 \text{tax}_2 + \beta_5 \text{tax}_3 + \beta_6 \text{tax}_4 + \beta_7 \text{tptask} + \beta_8 \text{round1}$$

Only dwage1, tptask and round1 were found significant ($p<0.05$). An F-test was conducted on all the dummies and failed to reject the hypothesis that they are all equal to zero ($p>0.4$).

Regression B used dummies that were generated for each net wage. The coefficients were also found insignificant for each net wage, both separately and jointly. When the specification is based on dummies, there are insignificant differences in effort among the treatments. Rhemtulla et al. (2012) found that for more than 5 categories, treating categorical variables as continuous yields acceptable performance. For these reasons, instead of dummies, the following regressions treat the tax and wage rates as continuous variables.

A Breusch-Pagan test fails to reject homoscedasticity in all of the following regressions; hence it is possible to use standard hypothesis testing. However, since there were three observations per participant, the command ‘cluster (id)’ was used in Stata to adjust the standard errors for clustering. This procedure necessitates the use of robust standard errors.

Effort and log(effort) were regressed on several combinations of explanatory variables obtained in the experiment. Apart from tptask, round1 and net, only two other variables were found significant, namely impwage and imptax. Both are self-reported measures from the Questionnaire, where participants rated how much they took into consideration the tax rate and the wage rate when making decisions about their labour supply. Both measures have a scale of 1 – 10. No other variables generated from the questionnaire were found significant. Regression 2.3. enables control for participants’ intentions and preferences. Participants who came to the experiment for other than financial reasons were less responsive to the tax rate and wage rate. Indeed, impwage and imptax are both significant at ($p<0.05$).

Regressions 2.1., 2.2. and 2.3. did not find significant coefficients on the variables of interest: tax, wage and net rate. Ramsey RESET test was therefore conducted on these regressions to test for functional misspecification. The null hypothesis (the model has no omitted variables) was rejected at ($p<0.05$). Therefore, we need to add more variables or try a different specification. Regressions 3.1., 3.2. and 3.3. use a logarithm of effort as the dependent variable. Regression 3.3. provides the best fit:

$$\log(\text{effort}) = 3.76 + 0.38\,\text{round1} - 0.02\,\text{tptask} + 0.21\log(\text{net}) + 0.06\,\text{impwage} - 0.06\,\text{imptax}$$

All coefficients are significant at ($p<0.05$), except for impwage. However, when conducting a joint significance F-test, where

$$H_0: \quad \beta_4 = 0,\ \beta_5 = 0;$$
$$H_A: \quad \beta_4 \neq 0,\ \beta_5 \neq 0;$$

The two variables are jointly significant at ($p=0.05$). The coefficient on log(net) measures the elasticity of Labour Supply with regards to the net wage. On average, a 5% increase in the net wage leads to an approximate 1% increase in labour supply. Labour supply is therefore inelastic. One unit change in tptask (i.e. how many seconds a participant spent on one task) leads to a 2% change in labour supply. Working in round 1 leads to a 38% increase in output. Regression 3.3. provides further evidence against the gross-wage illusion. Participants in this experiment were responsive to the net wage, rather than separate wage and tax rates.

## 5.6. Intertemporal Labour Supply

The variable maxef represents the optimal labour supply allocation given by the solution to the linear programming problem from the theoretical model. Regression 2.4. estimated to what degree participants adhered to the formal income-optimising model:

$$\text{effort} = 18.24 + 3.2\text{round1} - 0.14\text{tptask} + 0.165\text{maxef} + 1.07\text{impwage} - 0.9\text{imptax}$$

The coefficient on maxef is significant at ($p<0.05$). However, its absolute value is rather small. A 6-unit change in optimal effort leads approximately to a 1-unit change in actual effort. The theoretical model thus explains a small part of behaviour.

## 5.7. Laffer curve

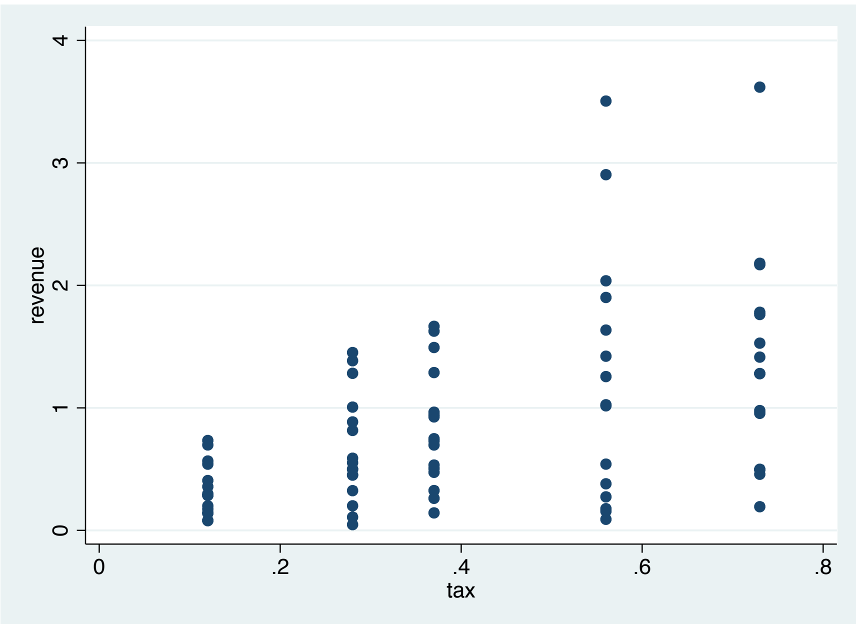
The second main goal of the experiment was to provide micro-foundations for the Laffer curve. The Laffer curve is a functional relationship between taxation and government revenue. As the government increases taxation, it also increases the revenue. However, after a certain point the revenue starts decreasing.

Regression 5.1. utilises a purely linear specification. All coefficients are statistically significant. Regression 5.2. provides a better fit ($R^2 = 0.47$). A 1% increase in net wage leads to a 0.87% increase in government revenue.

Regression 5.3. tests the theoretical model:

$$\text{revenue} = \beta_0 + \beta_1 \text{tax} + \beta_2 \text{tax}^2 + \beta_3 \text{tptask}$$

The coefficient $\beta_2$ is not significant. When revenue is plotted in a graph, there appears to be an outlier at the 73% level of tax (the highest point the upper right corner). Participant 3 carried out 27 tasks in round 1, 24 in round 2 and 21 in round 3. He was also the fastest of all participants. Hence his labour supply responsiveness to incentive strength did not vary much with the taxation level and pay-rate. The participant calculated many sequences even under the highest tax. For this reason, participant 3 was excluded from Regression 5.4.



Despite the exclusion of participant 3, the coefficient on $tax^2$ remains insignificant. However, the P-Value is p=0.090, as opposed to previous p=0.188. It seems reasonable to conclude that the insignificance is primarily due to the sample being very small. A much larger sample of participants would be required in order to make any meaningful inferences.

If we assume that $tax^2$ would have a significant coefficient in a larger sample of participants, we could calculate the maximisation point of government revenue.

$$revenue = 0.557 + 3.72\ tax - 2.35\ tax^2 - 0.01\ tptask$$

$$\frac{\delta revenue}{\delta tax} = 0$$

$$\frac{\delta revenue}{\delta tax} = 3.72 - 2 * 2.35 tax = 0$$

$$tax^* = 0.79$$

Therefore, the government would maximise its revenue at 79% tax. However, this number is not meaningful, not only because the coefficient is insignificant, but also because there is no information on subjects' behavioural response to a 79% tax from the experiment. What can be concluded is that the government in this scenario would have an incentive to set the tax-rate at 73%.

## 5.8. Tax Evasion

During the experiment, participants had an opportunity to evade taxes under the risk of being caught and having to pay a fine. Unexpectedly, only 3 participants cheated on their tax payments. Therefore, there was not enough variation to conduct an analysis. The may be several reasons for this result. The first one is that people are honest and do not usually cheat on their tax payments. Given the data from previous literature, however, this seems highly unlikely. The second possible explanation is that the sample size was too small and the distribution skewed in favour of more honest participants.

The third explanation is that the concept of Tax Reports and the instructions were not clear to the participants. Subjects had to fill in the number of correctly calculated sequences. It is possible that they failed to understand that if they filled in a smaller number they would pay less in taxation. As a result, they almost always filled in the true number according to the instructions. Furthermore, there may have been a flaw in the design itself. Participants were told that the supervisor would examine the most suspiciously looking tax report. It is possible that they had an incentive not to underreport their income, as that would increase the likelihood of getting caught. A better design is necessary for future research.

## 6. Conclusion

This paper analysed a new experiment, which measured the effects of taxation on individual labour supply. The first goal was to test the prediction of supply-side Economics that taxation disincentivises effort. The taxation treatments, whether treated as a categorical or continuous variable, had no statistically significant effect on their own. The same applied to wage rates. In contrast, the net wage was a significant predictor of Labour supply. This suggest that subjects took into consideration their net wages, instead of separate tax rates or wage rates.

The second goal of this paper was to provide micro-foundations for the Laffer curve. Regression analysis failed to provide any evidence. Government taxation in this scenario did not exhibit negative returns, as proposed by the Laffer curve. In contrast, there was a significant positive relationship between tax and revenue. This result is in accord with the first one, as taxes on their own do not appear to disincentivise effort.

Does the result imply that taxes have no effect on Labour supply? The net wages are calculated as the product of tax rates and pay-rates. Since lower net wages led to less effort; and given that wage rates were insignificant on their own, taxes do seem to have had an effect on effort. The effect is, however, impossible to separate from the effect of wages.

As the reference point was held constant during the experiment, we cannot infer that it had an impact on the evaluation of payment conditions. In spite of it, since the result contradicts Fochmann (2010), where participants exhibited the 'gross-wage' illusion, it is possible that providing a reference point can have an impact on behaviour. If the reference point is not provided during the experiment, the researcher may lose control over a key variable, depending on the characteristics of the sample.

The normative income-maximisation model successfully predicted a part of labour supply decisions. Income-maximisation therefore played a role in subjects' decision making, albeit a small one, as the coefficient on maxef is rather small.

Participants did not evade taxes. As discussed above, this was likely due to a flaw in experimental design. The supervisor stressed too much that 'suspicious' tax reports would be reassessed, unintentionally disincentivising underreporting of income.

The sample of 28 participants was far too small to enable any meaningful inferences. The sample size could have been the causal factor behind some of the results that contradict previous literature. It is possible that the results would have been different with a larger sample. For example, the taxation treatments may have been significant and the Laffer curve may have reached a peak within the range of imposed tax rates. This is, however, impossible to tell without replicating the experiment.

Together with previous literature, this experiment indicates that results may depend heavily on the fine details of design and framing. The design and framing of the experiment interact with cognitive programs of participants. As such, any differences may produce different results, even if the tax rates imposed are the same. For example, Sillamaa (1999a) used similar tax rates as this experiment but obtained a different result. Future research should thus place more emphasis on the cognitive processes underlying decision-making in Labour supply situations, in congruence with the modern approaches of behavioural Economics. Fore example, reference points should be considered as an important factor for evaluation. Most importantly, the perception of taxation itself should be investigated.

In conclusion, the concern about external validity of experiments seems to be extremely relevant. The debate about what constitutes the most realistic experimental taxation scenario continues and is by no means resolved with this or any other experiment. Hopefully, this paper brings the debate another small step closer to its conclusion.

## 7. Bibliography


J. Abeler and S. Jäger (2013), "Complex Tax Incentives - An Experimental Investigation," CESifo Working Paper

N. Bardsley (2010), "*Experimental economics: Rethinking the rules.* ", Princeton University Press

J. Brandts and G. Charness (2011), "The strategy versus the direct-response method: a first survey of experimental comparisons," *Journal of Experimental Economics*, vol. 14, no. 3, pp. 375–398

A. Brüggen, M. Strobel (2007), "Real effort versus chosen effort in experiments*", Economics Letters 96,* pp. 232–236

G. Charness and P. J. Kuhn (2010), "Lab Labor: What Can Labor Economists Learn from the Lab?," *National Bureau of Economic Research Working Paper Series*, vol. No. 15913

J. S. Davis, C. Swenson (1987), "The role of experimental economics in tax policy research", Faculty working paper no.1414, University of Southern California

M. Fochmann, et al., (2010) "Gross-wage illusion in a real effort experiment," FEMM working paper n.09, Otto-von-Guericke University Magdeburg, Faculty of Economics and Management

J. A. Hausman (1985), "Taxes and Labour Supply", *Handbook of Public Economics*, North-Holland Publishers, vol. 1, chapter 4

D. Kahneman (1979), "Prospect Theory: An analysis of decision under risk"; *Econometrica*, vol. 47, no.2, 263-292

L. Lévy-Garboua, D. Masclet, and C. Montmarquette (2009), "A behavioral Laffer curve: emergence of a social norm of fairness in a real effort experiment," *Journal of Economic Psychology*, vol. 30, no. 2, pp. 147–161

W. Morgan, M. Katz and H. Rosen (2009): *Microeconomics, 2nd European edition*, McGraw-Hill, ch.4 & 5.1

G. Ortona, S. Ottone, F. Ponzano, and F. Scacciati (2008), "Labour supply in presence of taxation financing public services. An experimental approach," *Journal of Economic Psychology*, vol.29, no. 5, pp. 619–631

S. Ottone and F. Ponzano (2008), "How People perceive the Welfare State. A real effort experiment," Working paper n.591, Institute of Public Policy and Public Choice – POLIS

Parducci, A. (1963), "The range-frequency compromise in judgment, *Psychological Monographs*, vol. 2, no. 565, p. 77

Rhemtulla (2012), "When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions", *Psychological Methods,* vol. 17, pp. 354-372

M. A. Sillamaa (1999a), "Taxpayer behavior in response to taxation: comment and new experimental evidence," *Journal of Accounting and Public Policy*, vol. 18, no. 2, pp. 165–177

M. A. Sillamaa (1999b), "How work effort responds to wage taxation: A non-linear versus a linear tax experiment," *Journal of Economic Behavior & Organization*, vol. 39, no. 2, pp. 219–233

M.A. Sillamaa (1999c), "How work effort responds to wage taxation: An experimental test of a zero top marginal tax rate*," Journal of Public Economics*, vol. 73, no. 1, pp. 125–134

M. Sutter and H. Weck-Hannemann (2003), "Taxation and the Veil of Ignorance–A real effort experiment on the Laffer curve," *Journal of Public Choice*, vol. 115, no. 1–2, pp. 217–240

C. W. Swenson (1988), "Taxpayer behavior in response to taxation: An experimental analysis," *Journal of Accounting and Public Policy*, vol. 7, no. 1, pp. 1–28

D. J. Zizzo (2008), "Experimenter Demand Effects in Economic Experiments", *Springer,* vol. 13, pp. 75-89

## 8. Appendix

Table 1: Treatments

| Treatment n. | Pay-rate | Tax-rate |
|---|---|---|
| 1 | 0.3 | 0.12 |
| 2 | 0.3 | 0.28 |
| 3 | 0.3 | 0.37 |
| 4 | 0.3 | 0.56 |
| 5 | 0.3 | 0.73 |
| 6 | 0.5 | 0.12 |
| 7 | 0.5 | 0.28 |
| 8 | 0.5 | 0.37 |
| 9 | 0.5 | 0.56 |
| 10 | 0.5 | 0.73 |
| 11 | 0.7 | 0.12 |
| 12 | 0.7 | 0.28 |
| 13 | 0.7 | 0.37 |
| 14 | 0.7 | 0.56 |
| 15 | 0.7 | 0.73 |

Table 2: Estimates of the effects of taxation and wage on effort

| Dependent Variable | Effort | | | |
|---|---|---|---|---|
| Explanatory Variables | 2.1 | 2.2 | 2.3 | 2.4 |
| Constant | 19.85***<br>(3.189) | 19.54***<br>(2.329) | 19.43***<br>(1.906) | 18.24***<br>(2.057) |
| Round1 | 3.256*<br>(1.392) | 3.105*<br>(1.335) | 2.922*<br>(1.313) | 3.200*<br>(1.298) |
| Tptask | -0.158***<br>(0.027) | -0.157***<br>(0.029) | -0.164***<br>(0.0236) | -0.141***<br>(0.0227) |
| Tax | -3.631<br>(2.951) | | | |
| Wage | 7.573<br>(3.714) | | | |
| Net | | 8.819<br>(4.530) | 8.227<br>(4.636) | |
| Maxef | | | | 0.165**<br>(0.0577) |
| Impwage | | | 1.119*<br>(0.441) | 1.073**<br>(0.370) |
| Imptax | | | -0.947**<br>(0.340) | -0.895**<br>(0.287) |
| $R^2$ | 0.48 | 0.47 | 0.51 | 0.55 |
| Adjusted $R^2$ | 0.45 | 0.45 | 0.47 | 0.51 |
| F-value | 16.40 | 21.40 | 14.58 | 17.01 |
| Observations | 77 | 77 | 77 | 77 |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

## Table 3: Estimates of the effects of taxation and wage on log(effort)

| Dependent Variable | Log(effort) | | |
|---|---|---|---|
| Explanatory Variables | 3.1 | 3.2 | 3.3 |
| Constant | 3.128*** (0.182) | 3.673*** (0.199) | 3.757*** (0.284) |
| Round1 | 0.379** (0.111) | 0.387** (0.112) | 0.376** (0.114) |
| Tptask | -0.017*** (0.0026) | -0.017*** (0.00263) | -0.018*** (0.00239) |
| Net | 0.800* (0.353) | | |
| Log(net) | | 0.226* (0.0912) | 0.213* (0.094) |
| Impwage | | | 0.058 (0.0355) |
| Imptax | | | -0.063* (0.0266) |
| $R^2$ | 0.55 | 0.56 | 0.57 |
| Adjusted $R^2$ | 0.53 | 0.54 | 0.54 |
| F-value | 30.10 | 30.42 | 18.82 |
| Observations | 77 | 77 | 77 |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

## Table 4: Preferences

| Dependent Variable | Preference |
|---|---|
| Explanatory Variables | 4.1 |
| Constant | 8.570*** (0.379) |
| Net | -13.23*** (1.203) |
| Preference | |
| $R^2$ | 0.66 |
| Adjusted $R^2$ | 0.66 |
| F-value | 147.9 |
| Observations | 77 |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

## Table 5: Estimates of the Laffer Curve

| Dependent Variable | Revenue | Log(revenue) | Revenue | Revenue |
|---|---|---|---|---|
| Explanatory Variables | 5.1 | 5.2 | 5.3 | 5.4 |
| Constant | 0.858*** | 1.549*** | 0.609** | 0.557** |
| | (0.153) | (0.233) | (0.180) | (0.190) |
| Tax | 1.934*** | | 3.561** | 3.719** |
| | (0.393) | | (1.184) | (1.197) |
| Tptask | -0.012*** | -0.018*** | -0.012*** | -0.011*** |
| | (0.002) | (0.003) | (0.002) | (0.002) |
| Log(Tax) | | 0.869*** | | |
| | | (0.130) | | |
| $Tax^2$ | | | -1.885 | -2.347 |
| | | | (1.394) | (1.332) |
| $R^2$ | 0.42 | 0.47 | 0.43 | 0.4 |
| Adjusted $R^2$ | 0.40 | 0.46 | 0.40 | 0.37 |
| F-value | 26.36 | 33.04 | 18.02 | 15.33 |
| Observations | 77 | 77 | 77 | 74 |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

## Table 6: List of Variables

| variable | source | description |
|---|---|---|
| id | defined | identification number of participant |
| effort | measure | number of calculated sequences |
| tptask | measure & | time spent per task |
| wage | treatment | pay-rate |
| tax | treatment | taxation rate |
| net | calculation | net pay-rate (net = wage * tax) |
| round1 | defined | dummy for round 1 (1 if round 1, else 0) |
| impwage | questionnaire | rating of how important wage was for participants' decision (1-not at all, 10 - |
| imptax | questionnaire | rating of how important tax was for participants' decision (1 - not at all, 10 - very |
| preference | Contract Sheet | preference for the given contact (1-9) |
| revenue | measure & | government tax revenue; revenue = effort/3 * tax |
| maxef | calculation | optimal effort as given by the normative income-maximising model |

## Table 7: Summary Statistics

| Variable | Obs | Mean | Std. Dev. | Min | Max |
|---|---|---|---|---|---|
| effort | 74 | 12.62 | 6.84 | 1.00 | 27.00 |
| tptask | 74 | 66.75 | 26.51 | 32.56 | 186.33 |
| revenue | 74 | 0.86 | 0.70 | 0.05 | 3.50 |
| wage | 74 | 0.51 | 0.16 | 0.30 | 0.70 |
| tax | 74 | 0.41 | 0.21 | 0.12 | 0.73 |
| net | 74 | 0.30 | 0.15 | 0.08 | 0.62 |
| round1 | 74 | 0.35 | 0.48 | 0.00 | 1.00 |
| impwage | 74 | 6.68 | 2.18 | 1.00 | 10.00 |
| imptax | 74 | 7.03 | 2.53 | 1.00 | 10.00 |
| preference | 74 | 4.58 | 2.41 | 1.00 | 8.00 |
| Log(effort) | 74 | 2.34 | 0.71 | 0.00 | 3.30 |
| Log(net) | 74 | -1.33 | 0.56 | -2.51 | -0.48 |