DOES THE CURRENT UKRAINIAN INCOME TAX AFFECT SOCIAL WELFARE?

by

Nataliya Lypyavka

A thesis submitted in partial fulfillment
of the requirements for the degree of

Master of Arts in Economics

Economic Education and Research
Consortium, Kyiv-Mohyla Academy

2002

Approved by ____
Chairperson of Supervisory Committee

____

____

____

Program Authorized
to Offer Degree ____

Date____

Kyiv-Mohyla Academy

Abstract

DOES THE CURRENT
UKRAINIAN INCOME TAX
AFFECT SOCIAL WELFARE?

by Nataliya Lypyavka

Chairperson of the Supervisory Committee: Professor Serhiy Korablin,
Institute for Economic Forecasting
National Academy of Sciences of Ukraine

Government policies, such as taxation, subsidization, and the provision of public goods significantly affect the structure of consumption. The comparative evaluation of outcomes of different public policies allows choosing the proper field and degree of government intervention to maximize individual and social welfare. In particular, in this paper I estimate whether the current income tax system helps to improve individual and social welfare in Ukraine. To perform the analysis, I assume the existence of a representative household and constant and homogeneous prices across goods. In the analysis, I use the concept of money metric utility (MMU), which may be derived from knowledge of preferences. I also assume the preferences, given constant prices, are reflected in the total spending for each particular good. As an approximation of utility, I use the level cash expenditures. This choice is supported by two reasons: i) it is ordinality of utility function, ii) MMU is assumed to be increasing in income, that is, more money brings more utility. Fortunately, MMU appears to be concave in income, that is, the appropriately chosen policy aimed to fight inequality in society will reach the target by effectively transferring income from rich to poor.

TABLE OF CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENTS

The author wishes to express the sincere gratitude to Dr. Jonathan Willner, the professor of Oklahoma City University, for his valuable comments and thorough guidance throughout the whole process of thesis writing.

The author is endlessly grateful to Yuriy Gorodnichenko from the University of Michigan for his marvelous comments, permanent constructive critique, and numerous advises.

The author also thanks Professor Stephan Lutz and Professor Iryna Lukyanenko for their useful consultations, Professor Ghaffar Mughal for the help in literature search, and Professor Charles Steele for the help and inspiration in topic selection.

Special thanks to Dmytro Ostanin for the provided technical support in the process of thesis completion, and to Olexandra and Maya Betliy for their help in the process of obtaining the data used in the research.

GLOSSARY

Atkinson Index. The measure of inequality invariant to proportional shifts in income.

Budget Constraint. A requirement that the amount of money spent on the goods is no more than the budget available.

Budget. A total amount of money a consumer has to spend.

Consumer Price Index (CPI). A Laspeyres cost-of-living index calculated as the ratio of the cost of buying a typical bundle of consumer goods and services today in comparison to the cost of buying that bundle in a base period.

Consumer Surplus. The differences between the amount consumers are willing to pay for a good and the amount they actually pay.

Consumption Bundle. A complete list of the goods and services.

Dalton's Measure. The measure of inequality defined as the ratio of the actual level of social welfare to that which would be achieved if income were equally distributed.

Durable Good. A consumption or capital good bought to provide services for a long time.

Engel Curve. The quantity of a good consumed as a function of income.

Equally Distributed Equivalent Level of Income. The level of income per head which if equally distributed would give the same level of social welfare as the present distribution.

Giffen Good. A good whose demand curve slopes upward as a result of a large income effect.

Gini Coefficient. The measure of inequality defined as the ratio of the area between the Lorenz curve and line of equality (45-degree line) to the total area under line of equality.

Ideal Cost-of-Living Index. The cost of attaining a given level of utility at current prices divided by the cost of attaining the same utility at base year prices.

Income Effect. The increase in consumption brought about by an increase in income, when prices of goods are held constant.

Income Elasticity of Demand. The percentage change in the quantity demanded of a good resulting from a 1 percent increase in income.

Income-Expansion Path. A curve that includes the utility-maximizing combinations of goods associated with every income level.

Indifference Curve. A graphical representation of all combinations of market baskets that provide the same level of satisfaction.

Indifference Map. A set of indifference curves that describes the consumer's preferences among various combinations of market baskets.

Inferior Good. A good for which consumption falls as an individual's income rises.

Laspeyres Index. A cost-of-living index that calculates the amount of money at current year prices that an individual requires to purchase a bundle of goods and services that was chosen in the base year, divided by the cost of purchasing the same bundle in the base year.

Lorenz Curve. The graphical representation of cumulative income distribution in the society.

Money Metric Utility (MMU). A concept that measures levels of living by the amount required to sustain them. MMU requires that consumption be adjusted by Paasche price index.

Nonexclusive good. A good that people cannot be excluded from consuming, and for the use of which it is difficult to charge them.

Nonrival good. A good for which the marginal cost of provision to an additional consumer is zero.

Normal good. A good for which consumption increases when income rises.

Paasche price index. A cost-of-living index that calculates the amount of money at current year prices that an individual requires to purchase a bundle of goods and services chosen in the current year, divided by the cost of purchasing the same bundle in the base year.

Public good. A nonrival and nonexclusive good.

Representative household. A concept implies the existence of "average" household.

Revealed preference. An approach to consumer theory in which preferences are determined by observing the choice consumers make.

Utility. The level of satisfaction that a person gets from consuming a good or undertaking an activity.

Welfare Ratio. A concept that measures welfare as multiples of a poverty line.

Chapter 1

INTRODUCTION

The main objective of an economy is to satisfy consumers' needs given the scarcity of resources. This implies that the needs must be satisfied in a way to maximize individual's utility subject to his or her budget constraint. Therefore, the problem narrows to maximization of utility function.

However, utility itself cannot be measured and compared across individuals since agents have different subjective perception of values of commodities. Although the concept of utility is abstract in character, it is still a very useful concept because, in terms of revealed preference theory, it can relate observable choice of an individual with his unobservable preferences. Under the assumption that needs are satisfied by consuming certain physical goods and services and all other aspects like political power and social ranking are irrelevant,[1] the obtained utility function would include only parameters of consumption of physical goods. This approach would allow us to measure welfare more or less precisely. Therefore, the effective measurement of welfare level can provide an efficient evaluation of different public policies. Therefore, there arises the necessity to investigate people's incentives to consume goods or services.

---

[1] These and other things (like pride, relative income, etc.) can matter but this complicates analysis, and so far, there is no operational concept of how such a utility function can be constructed and tested.

According to empirical studies,[2] among many (potentially infinitely many) factors influencing consumption, three factors can be taken as major determinants of consumption levels. These are price level, demographic characteristics, and income level. Assuming all agents face the same price and given single-period data with no information about variation in prices, nothing can be said about the impact of price changes on the consumption of goods or services.

Among demographic characteristics influencing consumption, the most significant are household size, the number of dependants (children, elders) and number of those who earn income. All those characteristics vary from household to household, but they evolve slowly over time and they are not affected exogenously, unless institutions change over time or the government conducts some demographically significant policy.

The situation changes when considering income level effect on consumption. As many theories of consumption assume,[3] income level is a potential factor affecting consumption. However, in contrast to demographic characteristics, it may be regulated relatively easily by the government and therefore is a valuable tool of influencing consumption behavior.

Consumption of different types of goods is determined by different sets of characteristics. For instance, household consumption of food is primarily determined by household size, consumption of durables depends on income level. To capture this feature, goods are usually aggregated into four categories: food, non-food, durables, and housing – such categorization is

---

[2] Deaton, and Grosh, (2000), Slesnick (1998)

[3] Engel's Law, Tornquist curves for necessities, normal, and luxury goods

widely used in Living Standards Measurement Study (LSMS) surveys.[4] I will treat services as an additional separate category.

The purpose of this paper is to investigate structure of consumption, derive the utility measures and then estimate how current income tax affects individual and social welfare. The implication of research is to evaluate the efficiency of income tax.

The paper is organized as follows. Part I considers theories of individual and social welfare measurement developed so far. Chapter 2 briefly discusses relevant papers on individual and social utility functions. Chapter 3 is devoted to other theoretical issues relevant to this study. Part II, and Chapter 4 in particular provides a short discussion of empirical evidence and problems revealed during Living Standards Measurement Study conducted by World Bank in many developing countries. Finally, part III investigates the case of Ukraine. Chapter 5 presents environmental issues such as participants and legislation, and data used in research. Additionally, Chapter 5 specifies the problems, gives the theoretical and practical solutions and provides the obtained results. Concluding Chapter 6 discusses results and suggests fields for further research.

[4] Deaton, and Zaidi (1999)

Chapter 2

LITERATURE REVIEW

Developments on Individual Utility Function

Starting from Adam Smith's "Wealth of Nations", many papers were devoted to researching what makes people wealthier and happier. Though, happiness is subjective and cannot be measured itself. Hence, there was a need to introduce some concept that would allow comparing how good the agent is under different conditions. The utility concept appeared to be appropriate for solving different economic problems.

My analysis starts from estimating household's utility function.

In 1948 Paul Samuelson introduced revealed preference theory. He presented a system of preferences and so-called Weak Axiom of Preferences, which was assumed to explain the consumer behavior. The essence of the Samuleson's proposition is to relate unobserved utility to observed choices.

Two years later Houthakker "proposed the method of ascending and descending sequences for the proof of the fundamental theorem on the numerical representation of preferences"[5] concluding with the Strong Axiom of Preferences, which imposed additional restrictions on the revealed-preference model. During 1950-60s this approach was developed further by Uzawa (1956) and extended in papers by Afriat. Moreover, Afriat (1962)

[5] Afriat (1962), p.306

extended the notion on preferences, utility, and demand by investigating the properties of expenditure functions since "the classical hypothesis of consumer behavior is that expenditures are regulated by a preference order".[6] This work became the groundwork for further investigation of the properties of expenditure and utility functions conducted by Diewert (1973). His works suggested simpler and more elegant proofs to theorems derived by Afriat (1962).

Discussions on Construction of Social Utility Function

In the 1964 Harberger presented his approach to measurement of welfare changes. So-called "Harberger's triangles" rely on assumption that while calculating net social benefits (costs) the surpluses (losses) of all members should be added disregarding any external effects. This work encountered some objections such as "consumer-surplus analysis is valid only when the marginal utility of real income is constant; … does not take into account of changes in income distribution; … it is partial-equilibrium in nature; … is valid only for small changes; … rendered obsolete be revealed preference analysis".[7] Although Harberger (1971) provided some defenses to these critiques, he acknowledged that there are factors that force supply and demand curves to shift and still could not be explained in terms of consumer-surplus analysis. His example referred to exogenous change in resources, technological progress, and a change in trading conditions. So he

---

[6] Afriat (1962), p.305

[7] Harberger (1971), p. 786

admitted that in proposed method's analysis "substitution effects only are involved".[8.]

Despite these limitations, it was shown by Diewert (1976) that this approach has desirable properties required by revealed-preference theory such as antisymmetry and weak preference. Moreover, Diewert mathematically proved that if, as Harberger suggested, normalisation of the price vector is applied, consumer-surplus analysis can serve to represent a broad class of utility functions. In addition to that, Diewert established the consistency of Harberger's indicator with others (like Fisher's) and showed that it is the negative Hicksian measure of consumer surplus (calculated by a shift in compensated demand curve).

Despite the simplicity and, therefore, convenience in use, of Harberger's triangles, the model needed some improvement. Jerry Hausman (1981) suggested replacing the standard approach of derivation of unknown parameters from market demand function after specification of utility functions with starting with the investigation of observed on market demand function, and then deriving unobserved market demand and expenditure function. Therefore, he tried to combine the benefits of Marshallian (observable) and Hicksian demands (correct measure). Hausman found that the smaller part of the good in total budget, the more accurate Marshallian estimates would be. However, while estimating welfare changes, Marshallian demand curve gives less reliable results while the Hicksian approach provides accurate estimates.

[8] Ibid., p. 789

Diewert (1978) developed a model for estimating the effects of infinitesimal changes in the exogenous tax variables. In this model he calculated a weighted sum rather than a simple summation of utility functions to measure losses from taxation implying that the system was Pareto optimal at the beginning. Developing his model, Diewert (1981) admitted that his "local approach" is useful since it requires "only local information about preferences and technologies".[9] However, according to Diewert himself, the approach is not completely satisfactory because "measures of welfare loss or gains may be quite inaccurate for large changes in taxes".[10]

Government intervention into markets deserves special attention. Usually, the purpose of the government is to correct market failure. The measure of effectiveness of such interventions depends on the goal of government program. Therefore, different measures of social welfare may be applied to compare the welfare under different circumstances.[11]

Usually, the objective of any tax system is equality and efficiency of income distribution. For this purpose, comparison may be based on the Gini coefficient, which measures aggregate inequality and can vary anywhere from 0 (perfect equality) to 1 (perfect inequality). The Gini coefficient is closely related to the Lorenz curve, which presents the cumulative income distribution in the society. However, the Gini coefficient does not account for regressive transfers that may occur as a result of government intervention. In those cases, the Gini coefficient remains the same before and after implementing governmental program whereas the Lorenz curves shift

---

[9] Diewert (1981), p. 1242

[10] Ibid.

[11] More on this can be found in "Public finance" by Stiglitz (2000)

reflecting the distributional changes – in particular, the initial and resulting Lorenz curves intersect.[12]

Another approach to measuring inequality in the society was proposed by Atkinson (1970). In his work, Atkinson discussed ranking of income distribution and measures of income inequality. In particular, he criticized the credibility of measure proposed by Dalton. Dalton proposed to use "the ratio of the actual level of social welfare to that which would be achieved if income were equally distributed".[13] Atkinson criticized Dalton's measure and Gini coefficient because of their non-invariance to the linear transformations. Atkinson explained his finding by the fact that both mentioned measures need the same mean for the comparison of distributions under investigation. Instead, Atkinson introduced the concept of the equally distributed equivalent level of income and argued "to measure inequality independently of the mean level of incomes".[14] He supported his suggestion by implicating constant (relative) inequality-aversion for the concave social welfare function. Given these considerations, the Atkinson index is, presumably, more flexible than the Gini coefficient.

Summarizing, over the last 25-30 years a vast number of papers were devoted to technical issues concerning which properties utility functions should possess. Slesnick (1998) provided comprehensive analysis on most important achievements in this field, comparing different approaches.

---

[12] Stiglitz (2000)

[13] Atkinson (1970), p.249

[14] Ibid., p.251

Numerous theoretical developments on consumer welfare were extensively explored by various practical researches. A well-known implementation of developed theories on welfare measurement was started in 1980 by the World Bank.[15] So-called Living Standard Measurement Study (hereinafter LSMS) aimed to investigate poverty in developing countries was applied to a number of countries, among others were the former Soviet Union's republics, e.g. Russia, Kyrgyzstan, Armenia and others. One of the objectives of LSMS is to estimate the distortionary impact of government policies. The research was usually conducted following some standard scheme and using special questionnaires. Deaton, and Grosh (2000) discussed the technique of creating appropriate questionnaire. They make some arguments about why consumption measures living standard better than income does, and give some examples on this. The main advantage of using consumption is its smoother and less-variable behavior. Overall, the paper is devoted to what consumption data is required to make some conclusions about efficiency of government policies, how that data can be obtained and analyzed. Usually the standard questionnaire is taken and then updated to the needs of some particular country. Deaton and Zaidi (1999) discussed data source and provided some explanations of theoretical background. In particular, they discussed the advantages and disadvantages of applying Money Metric Utility versus Welfare Ratios[16] while estimating whether households became better or worse after some particular government policy is applied. Some additional issues on time and leisure, durables and housing, public goods, and home production are discussed as those that change and, therefore, affect overall

---

15 Deaton, and Grosh (2000)

16 The concepts of Money Metric Utility and Welfare Ratios will be discussed in the theoretical part (ch. 3)

welfare in reality, although in theories they are frequently considered as "other things being equal".

LSMS studies have been conducted in many CIS countries. Unfortunately, Ukraine is not one of them. The purpose of proposed thesis is to estimate distortionary impact of taxes on individual and social welfare given available data on household survey for 2000 year. Although analysis is likely to be spatial, some important relationships may be revealed.

Chapter 3

THEORY CHAPTER[17]

The theory of consumer choice implies derivation of demand functions under utility maximization given budget constraints.

To simplify the analysis, it is a common practice to make some assumptions about consumer preference ordering. The ordering is also necessary for setting bundles that are to be compared. "It can be shown that if the preference ordering is complete, reflexive, transitive, and continuous, then it can be represented by a continuous utility function".[18] Consequent assumption is that the consumer will choose the best basket affordable given his budget constraint.

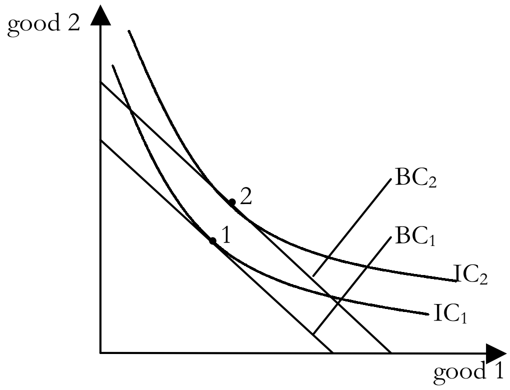
Graphically, given the well-behaved utility function, each level of utility is reflected by convex indifference curve. For goods, the higher utility curve, the better the individual is. The indifference curves corresponding to the different levels of utility do not intersect. The interior solution to utility maximization problem is the tangent point of the indifference curve and the budget constraint line. 1 provides the illustration for the two-good space.

---

[17] The microeconomic part of this chapter is based mainly on "Microeconomic Theory" by Hal Varian

[18] Varian (1992), p. 95

Figure 1. CONDITION FOR OPTIMAL CONSUMPTION BUNDLE



Given budget constraint BC_1, highest achievable utility is represented by indifference curve IC_1. To achieve IC_1, individual should choose bundle 1. Given BC2>BC1, IC2, which represents higher utility that IC1 does, is achievable. In this case, bundle 2 should be chosen.

Let x be bundle from consumption set X; $u(x)$ – utility obtained from bundle x; m – the amount of money available to the consumer; $p=(p_1, \ldots, p_n)$ – the vector of prices of goods 1,…,n; then the utility maximization problem is be written as:

$$u(x) \rightarrow max \quad (3.1)$$
$$such\ that\ px \leq m \quad (3.2)$$
$$x \in X \quad (3.3)$$

The solution to this optimization problem is the observable demand function and known as Marshallian demand function.

Closely related to this concept is the concept of *indirect utility function* (IUF) $v(p,m)$ – function that gives maximum attainable utility given p and m. The properties of IUF are the following: *i)* it is non-increasing in p; *ii)* non-decreasing in m; *iii)* homogeneous of degree 0 in (p,m); *iv)* quasi-convex in p; *v)* continuous at all $p>>0$, $m>0$.

Since $v(p,m)$ is non-decreasing in m, it can be inverted it and solved for m as a function of utility. The obtained function is called *expenditure function* (EF) and is denoted as $e(p,u)$. This function estimates the minimum cost of achieving a fixed level of utility and possesses the following properties: *i)* it is homogeneous of degree *1* in p; *ii)* concave in p; *iii)* continuous in p, $p>>0$.

Assuming that all $p>0$ and that $e(p,u)$ is differentiable,[19] then if $h(p,u)$ is compensated demand function also known as Hicksian function. $h(p,u)$ provides the expenditure-minimizing bundle necessary to achieve utility level u at prices p:

$$h_i(p,u)=[\partial e(p,u)/\partial p_i] \qquad (3.4)$$

(Shephard's Lemma). $h(p,u)$ is not directly observable since it depends on utility function which is not observable.

Assuming further that prices are fixed, *direct money metric utility* is be derived as

$$m(p,x)=e(p,u(x)) \qquad (3.5)$$

which would provide the level of income required to get the same utility as at $u(x)$ given constant p. Substituting IUF into EF will result in *money metric indirect utility function*

$$\mu(p,q,m)=e(p,v(q,m)) \qquad (3.6).$$

This $\mu(p,q,m)$ measures how much one would need at prices p to be as well off as one would facing prices q and having income m.

---

19 first partial derivatives exist

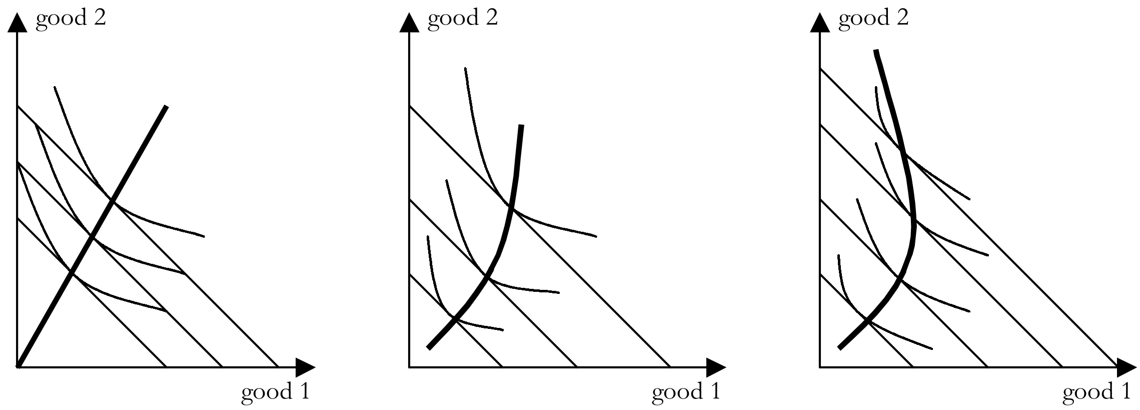
Under fixed prices and varying income, it is possible to derive an *income-expansion path*. This income expansion path allows investigation of the relationship between income and demand for the good – *the Engel curves.* There are three stylized generalizations of how the Engel curve may behave:

i) the ray from the origin is the straight line: with the increase of income the demand increases proportionally (unitary income elasticity of utility);

ii) the income expansion path bends toward one good: with the income increase consumer buys more of one good (luxury good) than of another (necessary good); and

iii) the income expansion path bends backward: with the increase of income, consumer buys less of inferior good.

Figure 2. INCOME-EXPANSION PATHS[20]



i) unitary income elasticity ii) good 2 is luxury good iii) good 1 is inferior good

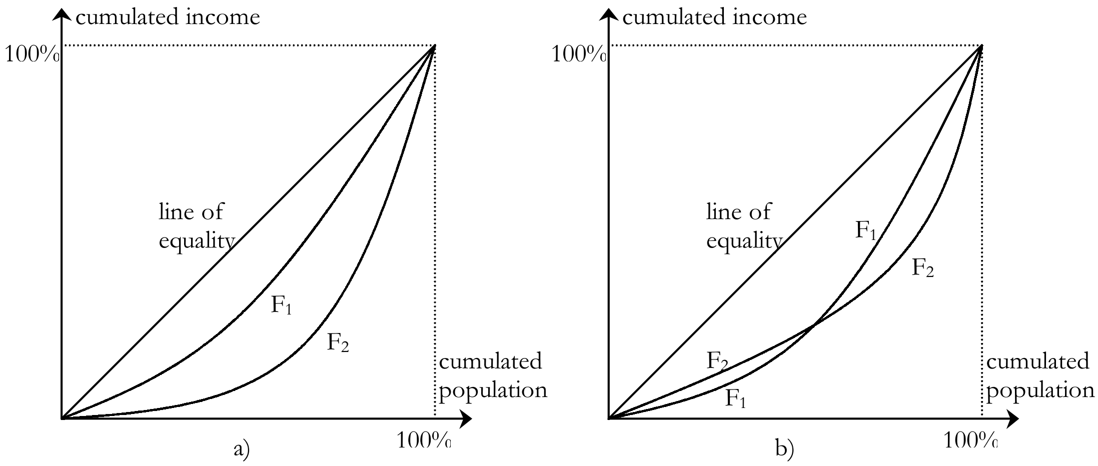
[20] Varian (1992), p.117

The affordable level of utility may vary with taxes imposed on particular goods (excise tax) or lump sum income tax. In this paper, I will discuss the effect of lump sum income tax. Neither form of excise tax is considered in the research since that tax is included in the price and prices themselves are assumed to be constant and homogeneous among consumers. Labor market decisions as well as distortionary impact of income tax on labor supply is also beyond the scope of this paper.

In the paper, the social welfare will be estimated with the help of two most common approaches: the Gini coefficient supported by the Lorenz curves and Atkinson's approach.

The Gini coefficient is based on a Lorenz curves (Figure 3), which reflect the cumulative income distribution within society.

Figure 3. LORENZ CURVES



The vertical axis measures the percentage of income going to income recipients, who are arrayed in percentiles on the horizontal axis. Income recipients are ordered from poorer to richer.

The Gini coefficient is the ratio of area between the Lorenz curve and line of equality to total area under the line of equality. Therefore, the Gini coefficient lies in the range between zero (perfect equality) and 1 (perfect inequality). Graphically, the closer Lorenz curves to the line of equality (45-degree line) – the higher income equality in the society.

The Gini coefficient may be calculated as:

$$Gini = \frac{1}{\overline{Y}}\sqrt{\sum_{i=1}^{n}\frac{n_i}{n}\left(Y_i - \overline{Y}\right)^2} \qquad (3.7)$$

where $\overline{Y}$ is the actual mean income, Y_i – income of i^{th} agent (here, household), n_i – the number of such agents, and n – the total number of agents.

However, Gini coefficient accurately measures the change in income distribution if transfers satisfy *the principle of transfers*, defined by Dalton: "if we make a transfer of income d from a person with income y_i to a person with a lower income y_2 (where $y_2 \leq y_1 - d$), then the new distribution should be preferred".[21] How this principle works is shown graphically in Figure 3a): here the distribution F_1 (with smaller value of Gini coefficient) is preferred to the distribution F_2 (where Gini coefficient is larger).

Although, in case of regressive transfers (depicted in Figure 3b).), Lorenz curves shift reflecting the change in distribution of income, but the how changes the value of Gini coefficient is ambiguous. Therefore, to estimate the effect of governmental intervention, some other measures are to be employed.

---

[21] Atkinson (1970), p.247

One of the prospect candidates to measure and to compare income inequality within different income distributions is the Atkinson's Index (AI) that can be calculated as:

$$AI = 1 - \left(\frac{Y^e}{\overline{Y}}\right) = 1 - \left[\frac{1}{n}\sum_{i=1}^{n}\left(\frac{Y_i}{\overline{Y}}\right)^{1-e}\right]^{\frac{1}{1-e}} \qquad (3.8)$$

where e is an attitude towards inequality (the higher e – the higher inequality-aversion) and Y^e is the equally distributed equivalent of income such that:

$$W(Y^e) * n = \sum_{i=1}^{n} W(Y_i) \qquad (3.9)$$

According to Atkinson, social welfare function W is assumed to be constant (relative) inequality-aversion:

$$W = \begin{cases} \sum_{i=1}^{n} \frac{Y_i^{1-e}}{1-e}, \text{ if } e \neq 1 \\ \sum_{i=1}^{n} \ln(Y_i), \text{ if } e = 1 \end{cases} \qquad (3.10)$$

If AI falls, the distribution become more equal, since the equally distributed equivalent of income becomes closer to actual mean of income.

Although Atkinson's Index is more flexible,[22] if social welfare function exhibits increasing (relative) inequality-aversion, both AI and Gini coefficient reflect the same effect direction of changes.

[22] for the reasons discussed in Chapter 2 in subsection "*Discussions on Construction of Social Utility Function*"

Chapter 4

EXPERIENCE OF LSMS STUDIES

The World Bank started its Living Standard Measurement Study (LSMS) program in the early 1980s aiming to evaluate nations' welfare (and poverty). The remedy for that evaluation is the estimation of their consumption level and not income level, although both measures can be used as approximation to utility. The reasons why the concepts are similar and why they differ are the following.[23]

Under the "rights approach" income measures potential claims. At the same time, consumption measures acquired assets. Therefore, from this point, both may be defended. As measures of utility, for the purposes of study, the measure of real welfare provided by consumption level is more appropriate than the measure of potential welfare. On the other hand, under the life-cycle permanent income hypothesis, a person's consumption at any point in time is proportional to his or her lifetime resources. However, even when this is true, time horizons matter, since individuals are not immortal and, therefore, government policies should take into account the current state of the poor part of population. Since the consumption is smoother than income,[24] it appears to be more suitable for welfare measurement.

---

[23] Deaton, and Grosh (2000)

[24] Intuitively: people do not receive payments every day, but they do consume every day; formally, consumers have concave utility functions (i.e. risk averse) and therefore in the course of maximizing expected utility

From a practical stand point, income is more attractive for researchers to evaluate welfare since it is easier to collect than consumption. However, this is more applicable to developed countries with a large formal sector. In contrast, in developing countries, the informal sector plays a major role, complicating the estimation of wages and incomes of "self-employed". Moreover, survey respondents may be reluctant to give honest answers about their real incomes. This may be due to both intended partial hiding of income (tax evasion) and/or the respondent's inability to assess real income of the household.

Although LSMS have numerous standardized procedures of gathering data, there are still some problems with collecting basic data. The first one is evaluation of consumption that does not come through the market. The estimated value should be added to "official" transactions. This procedure will give total consumption. It is worth noting that the level of aggregation matters in gathering consumption data. As the past LSMSs[25] show, the broader the categories used (that is the shorter the questionnaire) the more underestimated the consumption and vice versa.[26]

The second problem is the recall period problem, wherein the respondent reports purchases but he or she is not sure whether this particular purchase was made during the reference period. There are a few ways to solve this problem. One method is to visit the household at the beginning of the reference period to make a benchmark for the respondents and then after the

---

they tend to smooth consumption. Complete smoothness is not achieved because of many reasons: aggregate uncertainty, institutional constrains, asymmetric information, etc

[25] Deaton, and Grosh (2000)

[26] Short questionnaire include about 10 consumption items, long questionnaire more than 200 items; the appropriate number of consumption items appears to be 70-100

reference period is over. This appears to be an unsuccessful method, since usually the respondents remember the visit, but not the date. Therefore, the concept of diaries given to the household proved to be more effective.[27]

Finally, there is a problem of obtaining prices that households face. There are two alternatives to collect information on prices: at the household level, asking about cash spending and quantities purchased, or at the community level. Both approaches have their advantages and disadvantages. The advantages of the first method are that it allows observing actual demand for each particular good and avoiding assumption about where the purchase was made. The main disadvantage is that product unit value varies across households. The way to solve this is to use averaged across households price. The advantage of the other approach is that it provides the prices the consumers actually face. However, there are some difficulties in gathering data in this way. It is impossible to observe the parameters such as price and place of the market transaction. Unclear definition of local market complicates the analysis.

Experience of countries, where LSMS have been conducted, shows the consistency between the results of LSMS and official statistics provided by the state statistical committees.[28] These are two major sources for empirical data on households' income and consumption. Since LSMS has not been conducted in Ukraine yet, the data from Statistical Committee of Ukraine is used in this research.

---

[27] Deaton, and Grosh (2000)

[28] Ibid.

Chapter 5

UKRAINIAN CASE STUDY

Environment: Participants, Markets, and Institutions

The objective of this study implies that the main actor is a *household* (hereinafter the assumption of existence of representative household will be exploited). The household's consumption depends on many factors. Among others are the number of people in the household, the number of children and of elderly (not working) members, and the number of workers.[29]

Consider the effect of the number of people in the household. Obviously, the larger this number, the more the household is expected to consume, other things being equal. However, this does not mean that each member in the household consumes more. Therefore, it would be reasonable to deflate the household consumption by the number of people in the household. Although this deflating is convenient, it does not reflect the effect of economies of scale, which can be substantial in, for example, per capita housing spending or consumption of durable goods. Another effect, missed by simply deflating household consumption by the number of people in the household, is the effect of dependent members (children and elderly) on consumption. Their needs may substantially differ from those of workers, but are still involved in the demand function.[30] On the other hand, large

---

[29] Deaton, and Zaidi (1999)

[30] Ibid.

families are usually those with a larger number of children, and they would benefit more from economies of scale. To simplify the analysis, I assume that these effects are negligible.

This point relates closely to income sources. Wages are implied to be the major source of cash income – according data provided by Ukrainian-European Policy and Legal Advice Center, wages constituted about 50% of households' monetary income in year 2000.[31] Thus, the number of workers is also considered as important factor in determining the level of consumption.

Another important participant is the government. For the purpose of this research, I assume that the government acts as a re-distributor of income among the population and as provider of public goods. Both actions are performed by means of different taxation and subsidization policies, that is, the tax revenues are transferred to consumers in the form of subsidies but not cash payments themselves.   The aim of such behavior is to increase social welfare by smoothing welfare over cohorts of the population.

In this context, it is necessary to consider income tax legislation. According to Decree of Cabinet of Ministers of Ukraine "About income tax imposed on citizens", income tax is a progressive tax with brackets determined as multiples of minimum wage. Income tax consists of two parts – flat and proportional. Both parts are determined separately for each level of income. Income tax should be paid by "Ukrainian citizens, foreign citizens, and people without any citizenship, disregarding whether they have permanent placement in Ukraine or not". To have permanent placement implies to live on Ukraine's territory "not less than 183 days in a calendar year". For people

[31] *Ukrainian Economic Trends*, December 2000

with permanent placement in Ukraine, the tax is imposed on total income disregarding where it was earned; for those without permanent placement, income tax is collected only from incomes earned in Ukraine. The Decree also regulates the situation of double taxation of incomes earned by Ukrainians abroad. Separate articles are devoted to tax-exempted transfers and to factors, that decrease tax base. Among them are tax-exempted pensions (to retired, veterans, disabled workers, Chornobyl victims, etc.), child benefits and unemployment benefits. Therefore, the income tax base is primarily at the wages earned on major place of employment excluding benefits (such as for child benefits paid for each child under 16) and other one-time transfers. Income tax is calculated progressively using the concept of "non-taxable minimum" ("neopodatkovuvanyy minimum").[32] The non-taxable minimum (17UAH in 2000[33]) is not the same as minimum wage (that changed during year 2000 three times: at the beginning of the year it was 74UAH, during April – June 90UAH, and starting from July 2000 it was 118 UAH).[34]

Data Description

The data used in this research is a raw data for households' consumption for the year 2000; the data is available at the State Committee of Statistics of Ukraine.

The total number of observations is 9318.[35] However, for the whole set, the average propensity to consume appeared to be greater than 1 (see Appendix

---

[32] Table for exact calculations are presented in Appendix 1.

[33] The President's Decree "On fiscal reform in Ukraine"

[34] "The Law about the value of minimum wage for the year 2000"

[35] Some summary statistics are presented in the Appendix 3

4). This can be explained by either overall dissaving or underreporting income or overestimating expenditures. To correct for those measurement errors, two criteria were applied to sample selection. First, in order to exclude cash overspending as a result of probable dissaving and/or credits, the observations, for which expenditures are greater then income and reported dissaving and credits are zero, were dropped. Second, for technical reasons observations, with reported positive wage earned but zero number of workers in the household, were also excluded. The total sample after truncation consists of 4541 observations.

Each observation has weight – the number of households it represents. Summing those weights provides the number of households in the survey – 8,709,082.

The data also contains the geographical parameters (region, rural/urban area), the household's characteristics (size, number of dependants and workers). For the truncated dataset, the average size of household equals three people.

Expenditures and incomes in the dataset are reasonably categorized. The data on expenditures contains 52 food items and 70 non-food items. By Deaton, Grosh and Plewwe (2000), on this level of disaggregation consumption data provides accurate results.

For ease, *goods* are aggregated into several subcategories. Aggregation techniques developed during LSMS studies conducted by the World Bank, use subcategories of food, non-food items, durables, housing, and services.[36]

Consumption of food is represented by 52 categories and 3 ways of obtaining it: food bought for cash, home-produced food, and presented. "Food consumed outside the household" is another one-dimensional variable. Food consumption presented dually: in monetary and in quantitative terms. This would allow calculation of a price index for food.

Unfortunately, there is no such an option for other categories, which are given in the dataset. Spending for non-food items are given without any division. The categorization was done taking into account the purpose, the character, and the timing (durability) of the purchase.

Finally, the dataset contains the data on wages, benefits, value of home-produced goods, and total income.

Since the purpose of the paper is to investigate the impact of income taxes, I consider cash income and cash expenditures as directly affected by changes in the income tax base and rate.

Individual Utility and Maximization Problem

As may be derived from the expenditure function, the higher the level of income implies higher value of the money metric utility. Taking into account assumed ordinality of utility and uniformity of prices across all households, cash expenditures (***ce***) will be used as the proxy for utility level (***u***):

[36] Deaton, and Zaidi (1999)

$u_i = ce_i$ (6.1)

The utility is assumed to be of Cobb-Douglas form with five variables: food (f), non-food (nf), durables (d), housing (h), and services (s):

$u(f, n, d, h, s) = f^{a1} n^{a2} d^{a3} h^{a4} s^{a5}$ (6.2)

However, the only data given is the total spending by each household on each category, which is to be deflated by relevant price index in order to obtain the real demands. Under assumption of homogeneity of prices (all households faces the same price level), spending to be deflated by some constant terms ($\boldsymbol{x} = \boldsymbol{X} / \boldsymbol{P_x}$, $\boldsymbol{x} \in \{\boldsymbol{f}, \boldsymbol{nf}, \boldsymbol{d}, \boldsymbol{h}, \boldsymbol{s}\}$, $\boldsymbol{X} \in \{\boldsymbol{F}, \boldsymbol{NF}, \boldsymbol{D}, \boldsymbol{H}, \boldsymbol{S}\}$):

$U(F, N, D, H, S) = AF^{a1} N^{a2} D^{a3} H^{a4} S^{a5}$ (6.3)

where $\boldsymbol{A} = (\Pi_i \boldsymbol{P_i^{\alpha i}})^{-1}$ is a constant reflecting the effect of prices on utility and:

F – spending on food

N – spending on non-food items

D – spending on durables

H – spending on housing

S – spending on services

The prices of the composite goods ***F***, ***NF***, ***D***, ***H***, and ***S*** are equal to one.

Therefore, the maximization problem for the individual may be written as:

$u(F, N, D, H, S) \rightarrow max$ (6.4)

$s.t.\ F + N + D + H + S \leq ci$ (6.5)

where ***ci*** is cash income.

The optimal spending is obtained by Lagrange multiplier method. Once optimal spending on every category of goods are found, these spending are deflates by appropriate price indices. This would provide the optimal demand functions necessary to maximize utility.

Theoretically, derived expressions for total spending would be equal to the shares of income spent on particular good, normalizing the shares in utility function so that they would add up to one:

$$X=[a_x/\sum_{over\ all\ x}a_x]*ci \tag{6.6}$$

The optimal pattern of expenditure shares is estimated and presented in the Appendix 5. By construction, the sum of coefficients should sum up to one. Despite the expectations, the sum of shares is less than one. This may be explained by concavity of utility in all variables. The other possible explanation is that some factors that affect utility were not included.

Substituting the found expression for expenditure shares (6.6) into the utility function will give the indirect utility function as a function of cash income:

$$U(ci)=A(\Pi_{over\ all\ i}a_i^{ai})ci^{\Sigma ai} \tag{6.7}$$

where, again, ***ci*** is cash income.

The expenditure function (and therefore MMU) then is:

$$ci= (u/[A(\Pi a_i^{ai})])^{1/\Sigma ai} \tag{6.8}$$

Which in logarithmic form would be as follows:

$$lci = const + (\sum a_i)^{-1} * lu \qquad (6.9)$$

where $\boldsymbol{lu} = \ln(\boldsymbol{u})$ and $\boldsymbol{lci} = \ln(\boldsymbol{ci})$.

Denoting $\beta = (\sum a_i)^{-1}$, the regression equation would look like

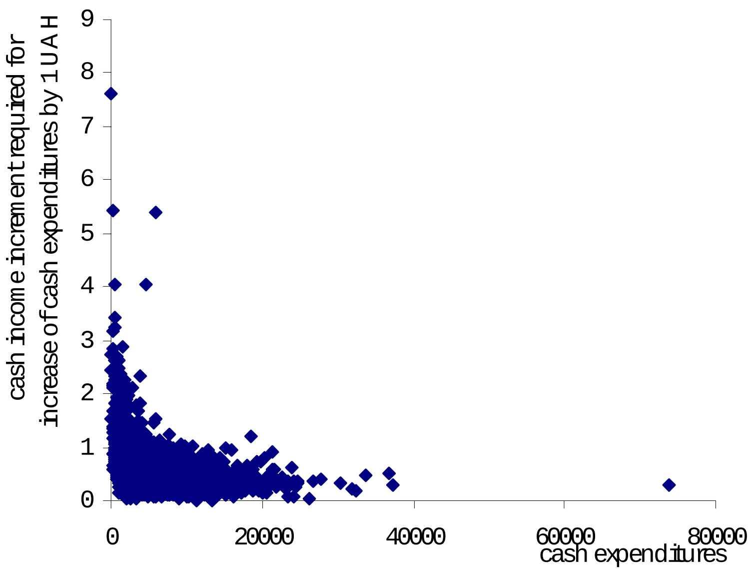
$$lci_j = const + \beta * lu_j \qquad (6.10)$$

The coefficient β will show how should change cash income in order to satisfy increase in cash expenditures. Since β is the inverted sum of shares, by theory it is supposed to be equal to one. Therefore, it becomes possible to estimate whether the sum of shares is statistically significantly differs from 1. Additionally if β appears to be less then one, then the cash becomes more valuable for the poor, and social welfare may be improved by progressive transfers from rich to poor. Otherwise, if β appears to be greater than 1, the rich would be assumed to value cash more, and therefore social welfare would be maximized with the regressive transfers.

Since according to the theory the scalar measure for utility level used in this equation is indirect utility, there is a possible endogeneity problem between the cash requirements ***ci*** and desired level of utility ***u***. Therefore, it would be reasonable to check for endogeneity. As shown in table A6.1 of Appendix 6, the Hausman test suggests the desirability of using an instrumental variables (IV) approach. The overidentifying restrictions test (OIR) shows the validity of the chosen instruments.

As can be seen in the Appendix 6, the estimated coefficient at ***lu*** is positive, that is, in order to get the higher cash expenditures, the average Ukrainian should spend more than he currently spends. Furthermore, the estimated coefficient at ***lu*** is less than 1, and statistically significantly differs from 1 (both OLS and IV estimates are significantly greater than 1: 10% and 23% respectively). That is, the cash increment requirement for increasing cash expenditures by 1UAH decreases as it gets higher cash expenditures levels (see Figure 4). This implies at least two consequences. First, since utility increases as cash income increases at a diminishing rate, MMU is concave in cash income and, therefore, people are risk-averse. Second, money is more valuable for the poor than for the rich. The appropriate government transfers policy should decrease the income inequality.

Figure 4. REQUIRED CASH INCOME INCREMENT



Taxes

Since the only source of cash income that is exposed to taxation is wage, to complete the analysis, pre-tax wages (***amwgt***) and implied tax payments (***tx***) should be restored and shown explicitly. Under currently established tax system, the income tax is progressive: higher incomes are exposed to higher tax rates. Therefore, in order to determine the amount of income tax paid by each household, their average after-tax wages (***amwg***) should be arranged and pooled into several intervals, implied by income tax legislation.

Certain tax rate, consisted of flat and proportional part, corresponds for each wage interval. The upper bound for the interval can be determined in the following way. From the highest pre-tax income for certain interval (known from the legislation) flat part of income tax and then proportional part are subtracted. Obtained value is the highest after-tax income for the given interval. At the same time, it is the lower income bound for the higher wage interval.

$$amwg_i = amwgt_i - [flat]_i - [proportional]_i \qquad (6.11)$$

where "flat" and "proportional" represent flat and proportional parts of income tax respectively.

$$[proportional]_i = (amwgt_i - [lower\ bound]_i) * [rate]_i \qquad (6.12)$$

where "lower bound" implies lower bound of corresponding pre-tax income interval, and "rate" means tax rate corresponding to that income interval.

Therefore, plugging (6.12) into (6.11) and rearranging for ***amwgt***, average monthly pre-tax wage is obtained:

$$amwgt_i = (amwg_i + [flat]_i - [lower\ bound]_i * [rate]_i) / (1-[rate]_i) \quad (6.13)$$

The implied tax outlays are calculated as a difference between pre-tax and after-tax wages:

$$tx_i = amwgt_i - amwg_i \quad (6.14)$$

The detailed information on income intervals and corresponding to them tax rates can be found in Appendix 1. The direct calculations based on the tax legislation are presented in the do-file and LOG-file (Appendices 7 and 8 respectively).

Calculated taxes are added to the cash income, generating total (taxable) income (***cit***):

$$cit_i=ci_i+tx_i \quad (6.15)$$

It would be logical to assume cash expenditure increasing in cash income. Assuming average propensity to consume (***apc***) to be constant, imputed tax outlays (***tx***) would only be partially spent (***tx****apc*). Since the actual cash expenditures are not observable, I use their fitted values (***cet***) obtained as the sum of actual cash spendings (***ce***) and of imputed tax outlays multiplied by average propensity to consume.

$$cet_i=ce_i+tx_i*apc_i \quad (6.16)$$

Projected cash expenditures give a new proxy for utility (***ut***):

$$ut_i=cet_i \quad (6.17)$$

Taking logarithms of non-taxed cash income (***lcit***=ln(***cit***)) and of utility (***lut***=ln(***ut***)) and replicating the equation (6.10) for this new variables, the following regression should be estimated:

$$lcit_j = const + \beta * lut_j \qquad (6.18)$$

where both *const* and coefficient β have the same meaning, but presumable not the exactly the same value as before.

The results of estimation are presented in the table 6.2 of Appendix 6. In general, they are quite close to the results obtained from the equation (6.10) for the taxed levels of cash income (***ci***) and utility (***u***).

Finally, to estimate the direct effect of income tax on utility, I calculate the difference of logarithms of taxed cash income and non-taxed cash income, obtaining the difference in utility levels corresponding to each tax amount paid. The results of IV are presented in the table 6.3 of Appendix 6. The sign of the coefficient before taxes supports the prediction that the taxes put the individual on the lower indifference curve.

Social Welfare

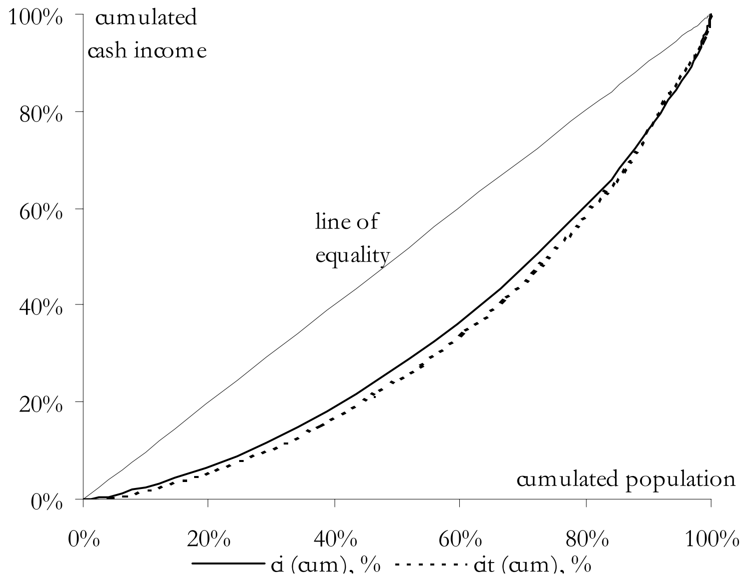
To complete the analysis of income tax on welfare of Ukrainian consumers, social welfare changes should be estimated in addition to evaluation of the income tax impact on individual utility. There exist many estimation procedures for welfare evaluation as well as there are many conventional measures for welfare changes estimations.[37] In this paper, I assume that the main concern of the income tax is to provide more equal distribution of

[37] Discussions on some of them are provided in Chapter 2.

income. The most commonly used measures for a comparison of two frequency distributions of income are Gini coefficient and Atkinson's Index.[38]

To start the analysis of the income tax impact on social welfare, consider firstly the approach that employs Gini coefficient and Lorenz curves. Figure 5 presents two cumulative income distributions: for the current income tax regime and for the zero-tax regime.

Figure 5. LORENZ CURVES FOR UKRAINE



It can be noticed from this graph that the Lorenz curve corresponding to current income tax regime mostly lies above the Lorenz curve which corresponds to zero-tax regime. This means that current income tax provides more equal distribution of income in Ukraine than it could be under zero-tax regime. Gini coefficients for both distributions are estimated according to the

---

38 Stiglitz (2000)

equation (3.7). The results are presented in the Table A7.1 of the Appendix 7. Generalizing the results, I can say that if welfare changes are estimated by the means of Gini coefficient and Lorenz curves, income tax does improve social welfare in terms of inequality aversion. However, the mean income under tax regime is less than if there were no income tax. The other conclusion that may be driven from this estimation is that the absolute impact of income tax on social welfare is very small – it is only 2.23% improvement of Gini coefficient.

Another point worth noting is the intersection of Lorenz curves in the right upper part of the graph. From the Figure 5 seen, that the situation certainly improves with taxes comparing to the zero-tax situation. Though, to be sure that Gini coefficient accurately reflects the reality in the case of Ukraine, Atkinson's approach should be employed in evaluation of these two distributions.

Following Atkinson (1970), I assume social welfare function to be constant (relative) inequality-aversion and be defined as in equation (3.10):

$$W = \begin{cases} \sum_{i=1}^{n} \frac{Y_i^{1-e}}{1-e}, \text{ if } e \neq 1 \\ \sum_{i=1}^{n} ln(Y_i), \text{ if } e = 1 \end{cases} \qquad (6.19)$$

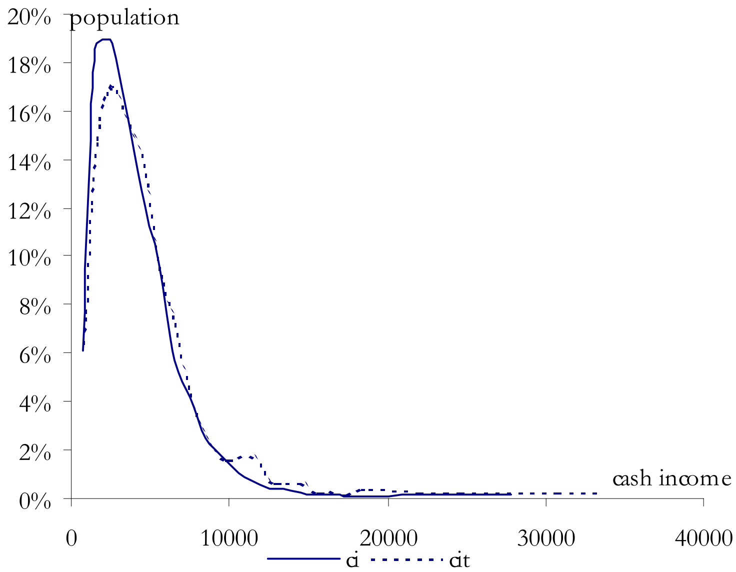
As it was already stated, a higher parameter ***e*** implies higher inequality aversion. Atkinson (1980) used ***e*** to be between one and two.[39] Atkinson's index itself was calculated as given in equation (3.8). Table A7.2 of Appendix 7 presents the estimations of equally distributed equivalent levels of income as well as values of Atkinson's Index AI for the parameters $e=1$ and $e=2$. From

[39] Stiglitz (2000)

these estimations, if incomes were identically distributed, it would require about 80% of current national income for *e=1* and about 60% of current national income for *e=2* to achieve current level of social welfare. Yet, comparing the indices for current income distribution and for zero income tax distribution of income, the imputed improvement provided by tax system is very small: it equals to 2% if *e=1*, and 3% if *e=2*. The equally distributed equivalent level of income is less under current tax regime, no matter what values parameter *e* takes.

Figure 6 presents the current income distribution and projected income distribution under zero income tax regime for the case *e=1*.

Figure 6. INCOME DISTRIBUTION FOR UKRAINE



In sum, it is possible to conclude for Ukraine's case that all applied in analysis measures show positive although negligible impact of income tax on social welfare.

Chapter 6

CONCLUSIONS

The income tax is assumed to be the influential instrument for the improvement in social welfare in case of market failure. However, as shown in the research, it is a disputable question whether income tax is an efficient instrument for governmental intervention.

The analysis of income tax impact on welfare started from the household's utility function investigation. Considering the obtained results, it is possible to conclude that Ukrainian population is rather poor, since the lion's share of cash income goes for food. Moreover, household's utility appeared to be increasing and concave in income. The implication of this finding is that money is more valuable for poor.

Given the defined properties of household's utility, incorporation of taxes into the welfare analysis allows concluding that taxes do decrease household's utility. However, this caused by tax decrease in utility diminishes as cash income rises. In other words, the higher the household's cash income, the less sensitive is the household to the tax payments.

The major implication from the so far derived conclusions is that the appropriate (progressive) income tax system could improve overall welfare in the society. Consequently, there arises the necessity to evaluate whether the current income tax system in Ukraine is appropriate for effective welfare improvement from the societal point of view.

The main concern of any tax system is redistribution of income so that to improve social welfare. In this research, I employed egalitarian approach and assumed the tax policy to be oriented on more equal distribution of income. Three tools were used to perform the analysis of current income tax impact on social welfare: Gini coefficient, Lorenz curves, and Atkinson Index. All three measures produced the consistent results of positive income tax impact towards equality of income distribution. However, this improvement in social welfare is of negligible size. This finding reflects low effectiveness of income tax. Therefore, to make income tax more powerful instrument for improvement of social welfare, the tax system should be revisited.

It is worth noting that according to the obtained results, people spend more cash, than they receive. This may signal about either about national cash dissaving or about possible measurement error due to either underreporting cash income or overestimated cash expenditures. Therefore, further aspects of this research can be assessed after obtaining data for earlier periods. This would allow investigating the reasons for reported overspending, incorporating intertemporal choice of households, and estimating dynamics of social welfare.

BIBLIOGRAPHY

Afriat, S.N. 1962. Preference Scales and Expenditure Systems, *Econometrica*, 30:2, pp.305-327

Atkinson, A.B. 1970. On the Measurement of Inequality, *J. Econ. Theory*, 2, pp.244-263

Atkinson, A.B., and Stiglitz, J.E. Lectures in Public Economics. 1980. McGraw-Hill.

Deaton, A., and Grosh, M. Chapter 6: Consumption in Grosh, M., and Glewwe, P. eds., Designing Household Survey Questionnaires for Developing Countries: Lessons from 15 Years of LSMS. 2000. World Bank

Deaton, A., and Muellbauer, J. Economics and consumer behavior. 1999. Cambridge University Press.

Deaton, A., and Zaidi., S. Guidelines for Constructing Consumption Aggregates for Welfare Analysis (processed). 1999. Princeton University and Development Research Group, World Bank: Washington DC

Diewert, W. E. 1973. Afriat and Revealed Preference. *Rev. Econ. Stud.*, 40:3, pp. 419-425

Diewert, W. E. 1976. Harberger's Welfare Indicator and Revealed Preference Theory. *Amer. Econ. Rev.*, 66:1, pp. 143-52

Diewert, W. E. 1981. The Measurement of Deadweight Loss Revisited. *Econometrica*, 49:5, pp. 1225-1244

Harberger, A.C. 1971. Three Basic Postulates for Applied Welfare Economics. *J. Econ. Lit.*, 9:3, pp. 785-97

Hausman, J.A. 1981. Exact Consumer's Surplus and

Deadweight Loss. *Amer. Econ. Rev.*, 71: 4, pp. 662-76

Houthakker, H.S. 1970. Revealed Preference Theory and the Utility Function. *Economica*, 17, pp. 159-174

Mas-Colell, A., Whinston, M.D., and Green, J.R. Microeconomic Theory. 1995. Oxford University Press.

Nicholson, W. Microeconomic theory: basic principles and extensions. 7th ed. 1998. The Dryden Press.

Samuelson, P.A. 1948. Consumption Theory in Terms of Revealed Preference. *Economica*, 15:60, pp.243-53

Slesnick, D.T. 1998. Empirical Approaches to the Measurement of Welfare. *J. Econ. Lit.*, 36, pp. 2108-2165

Stiglitz, J.E. Public Finance. 3rd ed. 2000. W.W. Norton & Company, Inc.

The Law "About the value of minimum wage for the year 2000" from 01.06.2000 #1766-III

The President's Decree "On fiscal reform in Ukraine" from 25.08.96 #762/96 and The President's Decree "About amendments to The President's Decree from 13.09.1994 #519" from 21.11.1995

Ukrainian-European Policy and Legal Advice Centre. Ukrainian Economic Trends. Quarterly Issue December 2000.

Uzawa, H. 1956. "On the Logical Relation Between Preference and Revealed Preference". *Technical Report* #38, Dept. Of Economics, Stanford University.

Varian, H. Microeconomic theory. 3rd ed. 1992. W.W. Norton & Company, Inc.

A P P E N D I C E S

Appendix 1. INCOME TAX CALCULATIONS

Income tax in Ukraine is progressive: higher incomes are exposed to higher tax rates. Table A1.1 presented the official tax rates correspondent to different level of incomes.

Table A1.1

Monthly aggregate taxable income (in non-taxable minimums, NTM)	**Tax rates and amounts**	**Calculated tax rates and amounts**
One NTM (17 UAH)	Not taxed	Zero
From 1 NTM + 1UAH (18UAH) up to 5 NTMs (85UAH)	10% from the amount that exceeds 1 NTM	10% from the amount exceeds 17 UAH
From 5 NTMs + 1UAH (85UAH) up to 10 NTMs (170UAH)	The tax from 5 NTMs + 15% from the amount that exceeds 5 NTMs	6.8UAH + 15% from the amount that exceeds 85UAH
From 10 NTMs + 1UAH (171UAH) up to 60 NTMs (1020UAH)	The tax from 10 NTMs +20% from the amount that exceeds 10 NTMs	19.55UAH + 20% from the amount that exceeds 170UAH
From 60 NTMs + 1UAH (1021UAH) up to 100 NTMs (1700UAH)	The tax from 60 NTMs + 30% from the amount that exceeds 60 NTM	189.55UAH + 30% from the amount that exceeds 1020UAH
More than 100 NTM (1701UAH)	The tax from 100NTMs + 40% from the amount that exceeds 100 NTMs	393.55UAH + 40% from the amount exceeds 1700UAH

Source: The President's Decree "About amendments to The President's Decree from 13.09.1994 #519" from 21.11.1995

Appendix 2. LIST OF VARIABLES*

AMTX – average monthly taxes paid, wage (calculation is given in the DO-file as well as in the LOG-file), UAH
AMWG – average monthly wage (calculation is given in the DO-file as well as in the LOG-file), UAH
AMWGT – pre-tax average monthly wage (calculation is given in the DO-file as well as in the LOG-file), UA
CE – total cash expenditures, UAH
CENTER – dummy for regions of Central Ukraine
CET – fitted values of cash expenditures for CIT, UAH
CH – number of children in the household
CI – total cash income, UAH
CIT – cash income with taxes, UAH
CODE – code of the household
CONSUM – consumption cash expenditures, UAH
CRED – total amount of credits used, UAH
D – cash expenditures for durables (not discounted), UAH
DCR – dummy for credits
DISSAVE – total amount of dissaving, UAH
DLU – difference in logarithms of utility (cash expenditures) levels with and without taxes
DSAV – dummy for dissaving
DTTLEXP – difference between total expenditures of two particular households, UAH
DTTLINC – difference between total incomes of two particular households, UAH
DWG – dummy for wage payments
DWORK – dummy for workers' presence
EAST – dummy for regions of Eastern Ukraine
ELD – number of elders in the household
F – total cash expenditures for food (FDBT and FDOUT), UAH
FDBT – cash expenditures for food, UAH
FDFR – value of food obtained for free (e.g. as a help from relatives), UAH
FDHP – value of home-produced food, UAH

* Full list of dataset variables contains 457 items and is available upon request

FDOUT – cash value of food consumed outside of the household, UAH
H – expenditures for housing, sum of the cash expenditures for housing and of the “latent” values such as subsidies, UAH
KIEV – dummy for KIEV
LCI – logarithm of CI
LCIT – logarithm of CI
LTX – logarithm of taxes (TX)
LU – logarithm of U
LUT – logarithm of UT
NF – cash expenditures for non-food items, UAH
NORTH – dummy for regions of Northern Ukraine
PEN – pensions, UAH
PLACE – type of placement (1-big city, 2-small city, 3-rural area)
REG – region (oblast)
RESAV – total resources available (TTLEXP minus help to the relatives), UAH
S – expenditures for services, sum of the cash expenditures and of the “latent” values such as subsidies and priviledges, UAH
SEV – dummy for Sevastopol
SIZE – the number of people in the household
SOUTH – dummy for regions of Southern Ukraine
TTLEXP – total expenditures (CE and payments on loans), UAH
TTLINC – total income (sum of VHP and CI), UAH
TTLRES – total resources (TTLINC and DISSAVE), UAH
TX – total amount of taxes paid during the year, UAH
TYPE – type of the household (1-with children, 2-without children)
U – utility level, rank of CE
UT – utility level without taxes, CET
VHP – total value of home-produced goods, UAH
WEIGHT – weight of the household in the sample (number of households similar to the particular one in the sample)
WEST – dummy for regions of Western Ukraine
WG – total amount of wages received, net of taxes, UAH
WG1 – wage received from the main place of employment, net of income tax, UAH
WG2 – wage received from other place of employment, net of tax, UAH
WG3 – income from individual activity, UAH
WORK – number workers in the household

Appendix 3. SUMMARY STATISTICS

The data for the tables in this appendix is monthly and deflated by size of households[40]

Table A3.1

	Total Expenditures							
Total Income	<100	100-200	200-300	300-400	400-500	500-700	700+	Grand Total
<100	768	1286	235	45	10	4	3	2351
100-200	154	2730	1524	363	86	46	4	4907
200-300	2	197	627	416	146	87	17	1492
300-400		6	65	135	91	60	24	381
400-500		1	2	18	43	29	19	112
500-700				4	5	23	23	55
700+		1		1		4	14	20
Grand Total	924	4221	2453	982	381	253	104	9318

Table A3.2

	Total Cash Expenditures							
Total Cash Income	<100	100-200	200-300	300-400	400-500	500-700	700+	Grand Total
<100	2408	2312	460	81	22	11	3	5297
100-200	293	1661	866	212	87	38	5	3162
200-300	10	112	217	164	74	31	12	620
300-400		8	25	38	29	30	20	150
400-500	1		3	9	12	16	6	47
500-700			1	1	3	11	11	27
700+		1		2		1	11	15
Grand Total	2712	4094	1572	507	227	138	68	9318

40 Assuming the cost of child equals to cost of adult (according to Deaton and Zaidi (1999) estimation on developed and developing countries)

Table A3.3

	Size Of Household												
Total Cash Expenditures	1	2	3	4	5	6	7	8	9	10	11	12	Grand Total
<100	377	567	557	646	328	155	44	24	5	1	4	4	2712
100-200	835	1244	986	725	208	73	19	3	1				4094
200-300	453	563	346	171	29	8	2						1572
300-400	177	186	95	40	8	1							507
400-500	90	80	40	13	4								227
500-700	69	45	20	4									138
700+	45	17	4	1	1								68
Grand Total	2046	2702	2048	1600	578	237	65	27	6	1	4	4	9318

Table A3.4

	Size Of Household												
Total Expenditures	1	2	3	4	5	6	7	8	9	10	11	12	Grand Total
<100	59	143	202	275	126	68	26	14	4		3	4	924
100-200	662	1137	1002	880	356	138	30	12	2	1	1		4221
200-300	670	810	547	319	71	27	8	1					2453
300-400	316	362	193	91	15	4	1						982
400-500	163	136	54	20	8								381
500-700	113	85	40	14	1								253
700+	63	29	10	1	1								104
Grand Total	2046	2702	2048	1600	578	237	65	27	6	1	4	4	9318

Table A3.5

	Size Of Household												
Total Cash Income	1	2	3	4	5	6	7	8	9	10	11	12	Grand Total
<100	1115	1363	1034	1035	457	198	54	27	5	1	4	4	5297
100-200	676	1014	829	489	108	35	10		1				3162
200-300	157	239	145	62	12	4	1						620
300-400	63	56	23	8									150
400-500	18	14	12	2	1								47
500-700	10	10	5	2									27
700+	7	6		2									15
Grand Total	2046	2702	2048	1600	578	237	65	27	6	1	4	4	9318

Table A3.6

	Size Of Household												
Total Income	1	2	3	4	5	6	7	8	9	10	11	12	Grand Total
<100	290	559	537	573	233	101	30	17	4		3	4	2351
100-200	1115	1376	1093	847	306	126	30	10	2	1	1		4907
200-300	412	544	342	144	35	10	5						1492
300-400	141	157	51	29	3								381
400-500	53	40	15	3	1								112
500-700	24	20	9	2									55
700+	11	6	1	2									20
Grand Total	2046	2702	2048	1600	578	237	65	27	6	1	4	4	9318

Appendix 4. PROPENSITY TO CONSUME

Table A4.1 Average Propensity To Consume

Source	SS	df	MS
Model	5E+10	1	5E+10
Residual	4.45E+10	4539	1E+07
Total	9.45E+10	4540	

Number of obs	= 4541
F(1 ,4539)	= 5103.74
Prob > F	= 0.0000
R-squared	= 0.5293
Adj R-squared	= 0.5292
Root MSE	= 3130.2

ttlexp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ttlinc	0.972924	0.0136187	71.44	0.0000	0.9462246	0.9996231
_cons	1600.227	87.9373138	18.197	2E-71	1427.82721	1772.62741

Table A4.2 Marginal Propensity To Consume[41]

Source	SS	df	MS
Model	7.94E+10	1	8E+10
Residual	7.21E+10	4539	2E+07
Total	1.52E+11	4540	

Number of obs	= 4541
F(1 ,4539)	= 5001.67
Prob > F	= 0.0000
R-squared	= 0.5242
Adj R-squared	= 0.5240
Root MSE	= 3985.3

ttlexp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
ttlinc	0.950755	0.01344346	70.722	0.0000	0.9243988	0.9771103

41 Assuming homogeneous households

Appendix 5. EXPENDITURE PATTERN

Table A5.1.

Expenditures	OLS		IV*		OIR	Hausman
	coef.	st.err.	coef.	st.err.	p-value of Ho: instruments are good	p-value of Ho: difference in coefficients not systematic
F	0.5793	0.0059	0.6878	0.0076	0.1573	0.0000
NF	0.1877	0.0021	0.1996	0.0026	0.7364	0.0000
D	0.0672	0.0023	0.0610	0.0029	0.9351	0.0003
H	0.0954	0.0013	0.1139	0.0016	0.4192	0.0000
S	0.1653	0.0023	0.1686	0.0029	0.5902	0.0548
sum	1.0950		1.2308			

*Instruments used: dummy for workers (dwork) and dummy for wage (dwg)

Appendix 6. MMU

Table A6.1. Taxed MMU

	OLS			IV			OIR	Hausman
Dependent variable: lci*	coef.	st.err.	R^2	coef.	st.err.	R^2	p-value of Ho: instruments are good	p-value of Ho: difference in coefficients not systematic
lu	0.6579	0.0097	0.5057	0.5317	0.0306	0.4871	0.6458	0.0000
_cons	2.5964	0.0802		3.6404	0.2534			

*Instruments used: dissaving (dissave) and credits (cred)

Table A6.2 Not-taxed MMU

	OLS			IV			OIR	Hausman
Dependent variable: lcit*	coef.	st.err.	R^2	coef.	st.err.	R^2	p-value of Ho: instruments are good	p-value of Ho: difference in coefficients not systematic
lut	0.6959	0.0094	0.5494	0.5438	0.0302	0.5232	0.5521	0.0000
_cons	2.2988	0.0783		3.5659	0.2516			

*Instruments used: dissaving (dissave) and credits (cred)

Table A6.3 Differences in utility levels

	OLS			IV			OIR	Hausman
Dependent variable: dlu*	coef.	st.err.	R^2	coef.	st.err.	R^2	p-value of Ho: instruments are good	p-value of Ho: difference in coefficients not systematic
ltx	0.0163	0.0001	0.7400	0.0187	0.0002	.	0.0004	0.0000

**Instruments used: dummy for kiev (kiev) and wage received (wg)

Appendix 7. SOCIAL WELFARE

Table A7.1 Gini Coefficients

Tax Regime	Mean	Gini
Current Tax	4048.65	0.7533
Zero Tax	4428.89	0.7759
Improvement With Taxes		-0.0226

Table A7.2 Atkinson's Indices

Tax Regime	e=1*		e=2**	
	Y^e	AI	Y^e	AI
Current Tax	3207.65	0.207724	2459.75	0.392453
Zero Tax	3415.76	0.228755	2507.61	0.433807
Improvement With Taxes		-0.02		-0.04

Social Welfare Functions:

$$* \ W = \sum_i log(Y_i)$$

$$** \ W = \sum_i -\frac{1}{Y_i}$$

Appendix 8. DO-file

```
log using "\path\name_of_log.log"
use "\path\name_of_database.dta"
* adjust dataset
*  i)  exclude  households  with  positive  wage  AND  zero  workers
reported
g wg= wg1+ wg2+ wg3
label var wg "wage"
g dwg=0
replace dwg=1 if wg>0
label var dwg "dummy for wage"
g dwork=0
replace dwork=1 if work>0
label var dwork "dummy for work"
keep if (dwg-dwork)<1
*  ii)  exclude  overspending  households  with  zero  dissaving  and
credits reported
g doversp=0
replace doversp=1 if ttlres<=ttlexp
label var doversp "dummy for overspending"
g dsav=0
replace dsav=1 if dissave>0
label var dsav "dummy for dissaving"
g dcr=0
replace dcr=1 if cred>0
label var dcr "dummy for credits"
keep if (doversp-dsav-dcr)<1
*
* calculate average propensity to consume (apc):
regress ttlexp ttlinc
test ttlinc=1
* coefficient near the ttlinc is average propensity to consume
g apc=ttlexp/ttlinc
label var apc "average propensity to consume"
*
*  calculate  demand  functions  (assuming  Cobb-Douglas  Utility
Function):
* 1) for food
regress f ci, noc
hausman, save
ivreg f (ci= dwg dwork), noc
predict resids, residual
matrix accum ZZ=dwg dwork, noc
matrix vecaccum EZ=resids dwg dwork, noc
matrix StatChi=EZ*syminv(ZZ)*EZ'
matrix Chi=StatChi/(e(rmse)^2)
matrix list Chi
* p-value Ho: instruments are good
```

```
display chiprob(1, Chi[1,1])
hausman, less
drop resids
matrix drop EZ
matrix drop StatChi
matrix drop Chi
* 2) for non-food
regress nf ci, noc
hausman, save
ivreg nf (ci= dwg dwork), noc
predict resids, residual
matrix vecaccum EZ=resids dwg dwork, noc
matrix StatChi=EZ*syminv(ZZ)*EZ'
matrix Chi=StatChi/(e(rmse)^2)
matrix list Chi
* p-value Ho: instruments are good
display chiprob(1, Chi[1,1])
hausman, less
drop resids
matrix drop EZ
matrix drop StatChi
matrix drop Chi
* 3) for durables
regress d ci, noc
hausman, save
ivreg d (ci= dwg dwork), noc
predict resids, residual
matrix vecaccum EZ=resids dwg dwork, noc
matrix StatChi=EZ*syminv(ZZ)*EZ'
matrix Chi=StatChi/(e(rmse)^2)
matrix list Chi
* p-value Ho: instruments are good
display chiprob(1, Chi[1,1])
hausman, less
drop resids
matrix drop EZ
matrix drop StatChi
matrix drop Chi
* 4) for housing
regress h ci, noc
hausman, save
ivreg h (ci= dwg dwork), noc
predict resids, residual
matrix vecaccum EZ=resids dwg dwork, noc
matrix StatChi=EZ*syminv(ZZ)*EZ'
matrix Chi=StatChi/(e(rmse)^2)
matrix list Chi
* p-value Ho: instruments are good
display chiprob(1, Chi[1,1])
hausman, less
drop resids
matrix drop EZ
matrix drop StatChi
matrix drop Chi
```

```
* 5) for services
regress s ci, noc
hausman, save
ivreg s (ci= dwg dwork), noc
predict resids, residual
matrix vecaccum EZ=resids dwg dwork, noc
matrix StatChi=EZ*syminv(ZZ)*EZ'
matrix Chi=StatChi/(e(rmse)^2)
matrix list Chi
* p-value Ho: instruments are good
display chiprob(1, Chi[1,1])
hausman, less
drop resids
matrix drop EZ
matrix drop StatChi
matrix drop Chi
*
* => IV to be used while calculating demand functions
* the sum of income shares spent on goods is: 1.230774
*
* define proxy to utility:
* since MMU (expenditure function that gives the minimum cost of
being at some particular indifference curve) is increasing in
income (the more money, the higher utility)
g u=ce
label var u "utility"
* utility is concave in cash income
g lu=ln(u)
label var lu "ln_utility"
g lci=ln(ci)
label var lci "ln_cash income"
regress lci lu
hausman, save
test lu=1
ivreg lci (lu= dissave cred)
test lu=1
predict resids, residual
matrix accum ZZ= dissave cred , noc
matrix vecaccum EZ=resids dissave cred , noc
matrix StatChi=EZ*syminv(ZZ)*EZ'
matrix Chi=StatChi/(e(rmse)^2)
matrix list Chi
* p-value Ho: instruments are good
display chiprob(2, Chi[1,1])
hausman, less
drop resids
matrix drop EZ
matrix drop StatChi
matrix drop Chi
matrix drop ZZ
*
* => IV to be used
* => to get utility by 1 level, cash income should rise by:
* ci(i+x)/ci(i)=(1+1/u(1))^(coefficient near the lu=0.6579015)
```

```
* => with higher cash expenditure levels, income increment required
for increase of cash expenditures by 1 UAH decreases
* => MMU is concave
*
* calculate taxes
g amwg=0
replace amwg=wg/(work*12) if work>0
label var amwg "avg monthly wage"

g lowwg=(amwg-17*0.1)/(1-0.1)
replace lowwg=0 if amwg>(85-(85-17)*0.1)
replace lowwg=0 if amwg<=17

g midwg=(amwg+(85-17)*0.1-85*0.15)/(1-0.15)
replace midwg=0 if amwg>(170-((85-17)*0.1+(170-85)*0.15))
replace midwg=0 if amwg<=(85-(85-17)*0.1)

g highwg=(amwg+(85-17)*0.1+(170-85)*0.15-170*0.2)/(1-0.2)
replace  highwg=0  if  amwg>(1020-((85-17)*0.1+(170-85)*0.15+(1020-
170)*0.2))
replace highwg=0 if amwg<=(170-((85-17)*0.1+(170-85)*0.15))

g  hiwg=(amwg+(85-17)*0.1+(170-85)*0.15+(1020-170)*0.2-1020*0.3)/(1-
0.3)
replace   hiwg=0   if   amwg>(1700-((85-17)*0.1+(170-85)*0.15+(1020-
170)*0.2+(1700-1020)*0.3))
replace   hiwg=0   if   amwg<=(1020-((85-17)*0.1+(170-85)*0.15+(1020-
170)*0.2))

g          hwg=(amwg+(85-17)*0.1+(170-85)*0.15+(1020-170)*0.2+(1700-
1020)*0.3-1700*0.4)/(1-0.4)
replace   hwg=0   if   amwg<=(1700-((85-17)*0.1+(170-85)*0.15+(1020-
170)*0.2+(1700-1020)*0.3))

g amwgt= lowwg+ midwg+ highwg+ hiwg+ hwg
label var amwgt "pre-tax avg wage"
g amtx=amwgt-amwg
label var amtx "avg monthly tax levied"
g tx=amtx*12*work
label var tx "tax levied"
drop lowwg midwg highwg hiwg hwg
*
* calculate the nontaxed income
g cit=ci+tx
label var cit "nontaxed cash income"
g lcit=ln(cit)
label var lcit "ln_nontaxed cash income"
g cet=ce+tx*apc
label var cet "nontaxed cash expenditures"
g ut=cet
label var ut "nontaxed utility"
g lut=ln(ut)
label var lut "ln_nontaxed utility"
*
```

```
* recalculate MMU for nontaxed income:
regress lcit lut
hausman, save
test lut=1
ivreg lcit (lut= dissave cred)
test lut=1
predict resids, residual
matrix accum ZZ=dissave cred, noc
matrix vecaccum EZ=resids dissave cred, noc
matrix StatChi=EZ*syminv(ZZ)*EZ'
matrix Chi=StatChi/(e(rmse)^2)
matrix list Chi
* p-value Ho: instruments are good
display chiprob(1, Chi[1,1])
hausman, less
drop resids
matrix drop EZ
matrix drop StatChi
matrix drop Chi
matrix drop ZZ
*
*  calculate  the  effect  of  income  tax  on  individual  utility
(welfare):
g dlu=lut-lu
label var dlu "difference in ln_utility caused by tax"
g ltx=0
replace ltx=ln(tx) if tx>0
label var ltx "ln_tax"
*
regress dlu ltx, noc
hausman, save
ivreg dlu (ltx= kiev wg ), noc
predict resids, residual
matrix accum ZZ= kiev wg , noc
matrix vecaccum EZ=resids kiev wg, noc
matrix StatChi=EZ*syminv(ZZ)*EZ'
matrix Chi=StatChi/(e(rmse)^2)
matrix list Chi
* p-value Ho: instruments are good
display chiprob(1, Chi[1,1])
hausman, less
drop resids
matrix drop EZ
matrix drop StatChi
matrix drop Chi
matrix drop ZZ
```

Appendix 9. LOG-file

```
. * adjust dataset
. * i) exclude households with positive wage AND zero workers reported
. g wg= wg1+ wg2+ wg3
. label var wg "wage"
. g dwg=0
. replace dwg=1 if wg>0
(6469 real changes made)
. label var dwg "dummy for wage"
. g dwork=0
. replace dwork=1 if work>0
(6098 real changes made)
. label var dwork "dummy for work"
. keep if (dwg-dwork)<1
(395 observations deleted)
. * ii) exclude overspending households with zero dissaving and credits reported
. g doversp=0
. replace doversp=1 if ttlres<=ttlexp
(6986 real changes made)
. label var doversp "dummy for overspending"
. g dsav=0
. replace dsav=1 if dissave>0
(1327 real changes made)
. label var dsav "dummy for dissaving"
. g dcr=0
. replace dcr=1 if cred>0
(2604 real changes made)
. label var dcr "dummy for credits"
. keep if (doversp-dsav-dcr)<1
(4382 observations deleted)
. *
. * calculate average propensity to consume (apc):
. regress ttlexp ttlinc
  Source |       SS       df       MS                  Number of obs =    4541
---------+------------------------------               F(  1,  4539) = 5103.74
   Model |  5.0008e+10     1  5.0008e+10               Prob > F      =  0.0000
Residual |  4.4474e+10  4539  9798242.59               R-squared     =  0.5293
---------+------------------------------               Adj R-squared =  0.5292
   Total |  9.4482e+10  4540  20810993.3               Root MSE      =  3130.2
------------------------------------------------------------------------------
  ttlexp |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  ttlinc |   .9729238   .0136187     71.440   0.000       .9462246     .999623
   _cons |   1600.227   87.93731     18.197   0.000       1427.827    1772.627
------------------------------------------------------------------------------

. test ttlinc=1
 ( 1)  ttlinc = 1.0
```

```
       F(  1,  4539) =    3.95
            Prob > F =    0.0469
. * coefficient near the ttlinc is average propensity to consume
. g apc=ttlexp/ttlinc
. label var apc "average propensity to consume"
. *
. * calculate demand functions (assuming Cobb-Douglas Utility Function):
. * 1) for food
. regress f ci, noc
Source |       SS       df       MS                  Number of obs =    4541
---------+------------------------------              F(  1,  4540) = 9497.62
   Model |  3.6227e+10     1  3.6227e+10              Prob > F      =  0.0000
Residual |  1.7317e+10  4540  3814290.92              R-squared     =  0.6766
---------+------------------------------              Adj R-squared =  0.6765
   Total |  5.3544e+10  4541  11791139.5              Root MSE      =  1953.0
------------------------------------------------------------------------------
       f |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      ci |   .5793461   .0059447     97.456   0.000       .5676915    .5910006
------------------------------------------------------------------------------
. hausman, save
. ivreg f (ci= dwg dwork), noc
Instrumental variables (2SLS) regression
Source |       SS       df       MS                  Number of obs =    4541
---------+------------------------------              F(  1,  4540) =       .
   Model |  3.4958e+10     1  3.4958e+10              Prob > F      =       .
Residual |  1.8586e+10  4540  4093779.90              R-squared     =       .
---------+------------------------------              Adj R-squared =       .
   Total |  5.3544e+10  4541  11791139.5              Root MSE      =  2023.3
------------------------------------------------------------------------------
       f |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      ci |   .6877722   .0076229     90.225   0.000       .6728277    .7027167
------------------------------------------------------------------------------
Instrumented:  ci
Instruments:   dwg dwork
------------------------------------------------------------------------------
. predict resids, residual
. matrix accum ZZ=dwg dwork, noc
(obs=4541)
. matrix vecaccum EZ=resids dwg dwork, noc
. matrix StatChi=EZ*syminv(ZZ)*EZ'
. matrix Chi=StatChi/(e(rmse)^2)
. matrix list Chi
symmetric Chi[1,1]
           resids
resids  1.9995487
. * p-value Ho: instruments are good
. display chiprob(1, Chi[1,1])
.15734605
. hausman, less
            ---- Coefficients ----
         |       (b)          (B)           (b-B)   sqrt(diag(V_b-V_B))
         |      Current     Prior        Difference        S.E.
```

```
---------+-------------------------------------------------------------
      ci |   .6877722     .5793461         .1084261     .0047716
---------+-------------------------------------------------------------
            b = less efficient estimates obtained from ivreg.
            B = more efficient estimates obtained previously from regress.
Test:  Ho:  difference in coefficients not systematic
            chi2(  1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                      =   516.34
            Prob>chi2 =     0.0000
. drop resids
. matrix drop EZ
. matrix drop StatChi
. matrix drop Chi
. * 2) for non-food
. regress nf ci, noc
Source |       SS       df       MS                  Number of obs =    4541
---------+------------------------------               F(  1,  4540) = 8344.44
   Model |  3.8038e+09     1  3.8038e+09               Prob > F      =  0.0000
Residual |  2.0695e+09  4540  455847.825               R-squared     =  0.6476
---------+------------------------------               Adj R-squared =  0.6476
   Total |  5.8733e+09  4541  1293402.93               Root MSE      =  675.17
------------------------------------------------------------------------------
      nf |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      ci |   .1877294   .0020551     91.348   0.000       .1837004    .1917584
------------------------------------------------------------------------------
. hausman, save
. ivreg nf (ci= dwg dwork), noc
Instrumental variables (2SLS) regression
Source |       SS       df       MS                  Number of obs =    4541
---------+------------------------------               F(  1,  4540) =       .
   Model |  3.7887e+09     1  3.7887e+09               Prob > F      =       .
Residual |  2.0847e+09  4540  459181.394               R-squared     =       .
---------+------------------------------               Adj R-squared =       .
   Total |  5.8733e+09  4541  1293402.93               Root MSE      =  677.63
------------------------------------------------------------------------------
      nf |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      ci |   .1995709    .002553     78.172   0.000       .1945658     .204576
------------------------------------------------------------------------------
Instrumented:  ci
Instruments:   dwg dwork
------------------------------------------------------------------------------
. predict resids, residual
. matrix vecaccum EZ=resids dwg dwork, noc
. matrix StatChi=EZ*syminv(ZZ)*EZ'
. matrix Chi=StatChi/(e(rmse)^2)
. matrix list Chi
symmetric Chi[1,1]
           resids
resids  .11328212
. * p-value Ho: instruments are good
. display chiprob(1, Chi[1,1])
.73643799
```

```
. hausman, less
            ---- Coefficients ----
         |       (b)          (B)           (b-B)   sqrt(diag(V_b-V_B))
         |      Current     Prior        Difference        S.E.
---------+-------------------------------------------------------------
      ci |    .1995709     .1877294         .0118415     .0015147
---------+-------------------------------------------------------------
            b = less efficient estimates obtained from ivreg.
            B = more efficient estimates obtained previously from regress.
Test:  Ho:  difference in coefficients not systematic

            chi2(  1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                      =    61.12
            Prob>chi2 =     0.0000
. drop resids
. matrix drop EZ
. matrix drop StatChi
. matrix drop Chi
. * 3) for durables
. regress d ci, noc
Source |       SS       df       MS                  Number of obs =    4541
---------+------------------------------               F(  1,  4540) =  828.34
   Model |   487243139     1   487243139               Prob > F      =  0.0000
Residual |  2.6705e+09  4540   588216.04               R-squared     =  0.1543
---------+------------------------------               Adj R-squared =  0.1541
   Total |  3.1577e+09  4541  695385.149               Root MSE      =  766.95
------------------------------------------------------------------------------
       d |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      ci |   .0671888   .0023345     28.781   0.000        .062612    .0717655
------------------------------------------------------------------------------
. hausman, save
. ivreg d (ci= dwg dwork), noc
Instrumental variables (2SLS) regression
Source |       SS       df       MS                  Number of obs =    4541
---------+------------------------------               F(  1,  4540) =       .
   Model |   483059415     1   483059415               Prob > F      =       .
Residual |  2.6747e+09  4540  589137.565               R-squared     =       .
---------+------------------------------               Adj R-squared =       .
   Total |  3.1577e+09  4541  695385.149               Root MSE      =  767.55
------------------------------------------------------------------------------
       d |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      ci |   .0609628   .0028918     21.081   0.000       .0552935    .0666321
------------------------------------------------------------------------------
Instrumented:  ci
Instruments:   dwg dwork
------------------------------------------------------------------------------
. predict resids, residual
. matrix vecaccum EZ=resids dwg dwork, noc
. matrix StatChi=EZ*syminv(ZZ)*EZ'
. matrix Chi=StatChi/(e(rmse)^2)
. matrix list Chi
symmetric Chi[1,1]
```

```
           resids
resids  .00663977
. * p-value Ho: instruments are good
. display chiprob(1, Chi[1,1])
.93505645
. hausman, less
            ---- Coefficients ----
         |       (b)          (B)           (b-B)   sqrt(diag(V_b-V_B))
         |      Current     Prior        Difference        S.E.
---------+-------------------------------------------------------------
      ci |   .0609628     .0671888        -.0062259     .0017066
---------+-------------------------------------------------------------
            b = less efficient estimates obtained from ivreg.
            B = more efficient estimates obtained previously from regress.
Test:  Ho:  difference in coefficients not systematic
            chi2(  1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                      =    13.31
            Prob>chi2 =     0.0003
. drop resids
. matrix drop EZ
. matrix drop StatChi
. matrix drop Chi
. * 4) for housing
. regress h ci, noc
Source |       SS       df       MS                  Number of obs =    4541
---------+------------------------------                F(  1,  4540) = 5715.19
   Model |   982041254     1   982041254                Prob > F      =  0.0000
Residual |   780108280  4540  171830.018                R-squared     =  0.5573
---------+------------------------------                Adj R-squared =  0.5572
   Total |  1.7621e+09  4541   388053.19                Root MSE      =  414.52
------------------------------------------------------------------------------
       h |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      ci |   .0953869   .0012617     75.599   0.000       .0929132    .0978605
------------------------------------------------------------------------------
. hausman, save
. ivreg h (ci= dwg dwork), noc
Instrumental variables (2SLS) regression
Source |       SS       df       MS                  Number of obs =    4541
---------+------------------------------                F(  1,  4540) =       .
   Model |   945015722     1   945015722                Prob > F      =       .
Residual |   817133812  4540  179985.421                R-squared     =       .
---------+------------------------------                Adj R-squared =       .
   Total |  1.7621e+09  4541   388053.19                Root MSE      =  424.25
------------------------------------------------------------------------------
       h |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      ci |   .1139083   .0015984     71.266   0.000       .1107747    .1170419
------------------------------------------------------------------------------
Instrumented:  ci
Instruments:   dwg dwork
------------------------------------------------------------------------------
. predict resids, residual
. matrix vecaccum EZ=resids dwg dwork, noc
```

```
. matrix StatChi=EZ*syminv(ZZ)*EZ'
. matrix Chi=StatChi/(e(rmse)^2)
. matrix list Chi
symmetric Chi[1,1]
           resids
resids  .65260271
. * p-value Ho: instruments are good
. display chiprob(1, Chi[1,1])
.41918368
. hausman, less
            ---- Coefficients ----
         |       (b)          (B)            (b-B)   sqrt(diag(V_b-V_B))
         |      Current     Prior        Difference        S.E.
---------+-------------------------------------------------------------
      ci |   .1139083     .0953869         .0185214     .0009812
---------+-------------------------------------------------------------
            b = less efficient estimates obtained from ivreg.
            B = more efficient estimates obtained previously from regress.
Test:  Ho:  difference in coefficients not systematic
            chi2(  1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                      =   356.32
            Prob>chi2 =     0.0000
. drop resids
. matrix drop EZ
. matrix drop StatChi
. matrix drop Chi
. * 5) for services
. regress s ci, noc
Source |       SS       df       MS                  Number of obs =    4541
---------+------------------------------               F(  1,  4540) = 5104.93
   Model |  2.9498e+09     1  2.9498e+09               Prob > F      =  0.0000
Residual |  2.6233e+09  4540  577823.943               R-squared     =  0.5293
---------+------------------------------               Adj R-squared =  0.5292
   Total |  5.5731e+09  4541  1227278.54               Root MSE      =  760.15
------------------------------------------------------------------------------
       s |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      ci |   .1653166   .0023138     71.449   0.000       .1607805    .1698527
------------------------------------------------------------------------------
. hausman, save
. ivreg s (ci= dwg dwork), noc
Instrumental variables (2SLS) regression
  Source |       SS       df       MS                  Number of obs =    4541
---------+------------------------------               F(  1,  4540) =       .
   Model |  2.9486e+09     1  2.9486e+09               Prob > F      =       .
Residual |  2.6245e+09  4540   578074.00               R-squared     =       .
---------+------------------------------               Adj R-squared =       .
   Total |  5.5731e+09  4541  1227278.54               Root MSE      =  760.31
------------------------------------------------------------------------------
       s |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      ci |   .1685598   .0028645     58.845   0.000        .162944    .1741756
------------------------------------------------------------------------------
Instrumented:  ci
```

```
Instruments:   dwg dwork
------------------------------------------------------------------------------
. predict resids, residual
. matrix vecaccum EZ=resids dwg dwork, noc
. matrix StatChi=EZ*syminv(ZZ)*EZ'
. matrix Chi=StatChi/(e(rmse)^2)
. matrix list Chi
symmetric Chi[1,1]
           resids
resids  .28999711
. * p-value Ho: instruments are good
. display chiprob(1, Chi[1,1])
.59022238
. hausman, less
            ---- Coefficients ----
         |       (b)          (B)            (b-B)   sqrt(diag(V_b-V_B))
         |     Current     Prior        Difference        S.E.
---------+-------------------------------------------------------------
      ci |   .1685598     .1653166         .0032432     .0016887
---------+-------------------------------------------------------------
            b = less efficient estimates obtained from ivreg.
            B = more efficient estimates obtained previously from regress.
Test:  Ho:  difference in coefficients not systematic
            chi2(  1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                      =     3.69
            Prob>chi2 =     0.0548
. drop resids
. matrix drop EZ
. matrix drop StatChi
. matrix drop Chi
. *
. * => IV to be used while calculating demand functions
. * the sum of income shares spent on goods is: 1.230774
. *
. * define proxy to utility:
. * since MMU (expenditure function that gives the minimum cost of being at som
> e particular indifference curve) is increasing in income (the more money, the
>  higher utility)
. g u=ce
. label var u "utility"
. * utility is concave in cash income
. g lu=ln(u)
. label var lu "ln_utility"
. g lci=ln(ci)
. label var lci "ln_cash income"
. regress lci lu
Source |       SS       df       MS                  Number of obs =    4541
---------+------------------------------              F(  1,  4539) = 4644.27
   Model |  1122.93814     1  1122.93814              Prob > F      =  0.0000
Residual |   1097.4847  4539  .241789975              R-squared     =  0.5057
---------+------------------------------              Adj R-squared =  0.5056
   Total |  2220.42283  4540  .489079919              Root MSE      =  .49172
------------------------------------------------------------------------------
     lci |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
```

```
---------+--------------------------------------------------------------------
      lu |   .6579015   .0096539     68.149   0.000       .6389752    .6768278
   _cons |   2.596394   .0801758     32.384   0.000        2.43921    2.753577
------------------------------------------------------------------------------
. hausman, save
. test lu=1
 ( 1)  lu = 1.0
       F(  1,  4539) = 1255.74
            Prob > F =    0.0000
. ivreg lci (lu= dissave cred)
Instrumental variables (2SLS) regression
Source |       SS       df       MS                  Number of obs =    4541
---------+------------------------------               F(  1,  4539) =  301.42
   Model |  1081.59578     1  1081.59578               Prob > F      =  0.0000
Residual |  1138.82705  4539  .250898227               R-squared     =  0.4871
---------+------------------------------               Adj R-squared =  0.4870
   Total |  2220.42283  4540  .489079919               Root MSE      =   .5009
------------------------------------------------------------------------------
     lci |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
      lu |   .5316663   .0306233     17.361   0.000       .4716297    .5917029
   _cons |    3.64043   .2533811     14.367   0.000        3.14368    4.137181
------------------------------------------------------------------------------
Instrumented:  lu
Instruments:   dissave cred
------------------------------------------------------------------------------
. test lu=1
 ( 1)  lu = 1.0
       F(  1,  4539) =  233.89
            Prob > F =    0.0000
. predict resids, residual
. matrix accum ZZ= dissave cred , noc
(obs=4541)
. matrix vecaccum EZ=resids dissave cred , noc
. matrix StatChi=EZ*syminv(ZZ)*EZ'
. matrix Chi=StatChi/(e(rmse)^2)
. matrix list Chi
symmetric Chi[1,1]
          resids
resids  .8744607
. * p-value Ho: instruments are good
. display chiprob(2, Chi[1,1])
.64582265

. hausman, less
            ---- Coefficients ----
         |       (b)          (B)           (b-B)   sqrt(diag(V_b-V_B))
         |      Current     Prior        Difference        S.E.
---------+-------------------------------------------------------------
      lu |   .5316663     .6579015        -.1262352     .0290618
---------+-------------------------------------------------------------
            b = less efficient estimates obtained from ivreg.
            B = more efficient estimates obtained previously from regress.
Test:  Ho:  difference in coefficients not systematic
```

```
            chi2(  1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                      =    18.87
            Prob>chi2 =     0.0000
. drop resids
. matrix drop EZ
. matrix drop StatChi
. matrix drop Chi
. matrix drop ZZ
. *
. * => IV to be used
. * => to get utility by 1 level, cash income should rise by:
. * ci(i+x)/ci(i)=(1+1/u(1))^(coefficient near the lu=0.6579015)
.  *  =>  with  higher  cash  expenditure  levels,  income  increment  required  for
increase of cash expenditures by 1 UAH decreases
. * => MMU is concave
. *
. * calculate taxes
. g amwg=0
. replace amwg=wg/(work*12) if work>0
(3504 real changes made)
. label var amwg "avg monthly wage"
.
. g lowwg=(amwg-17*0.1)/(1-0.1)
. replace lowwg=0 if amwg>(85-(85-17)*0.1)
(2475 real changes made)
. replace lowwg=0 if amwg<=17
(1139 real changes made)
.
. g midwg=(amwg+(85-17)*0.1-85*0.15)/(1-0.15)
. replace midwg=0 if amwg>(170-((85-17)*0.1+(170-85)*0.15))
(1336 real changes made)
. replace midwg=0 if amwg<=(85-(85-17)*0.1)
(2066 real changes made)
.
. g highwg=(amwg+(85-17)*0.1+(170-85)*0.15-170*0.2)/(1-0.2)
. replace highwg=0 if amwg>(1020-((85-17)*0.1+(170-85)*0.15+(1020-170)*0.2))
(16 real changes made)
. replace highwg=0 if amwg<=(170-((85-17)*0.1+(170-85)*0.15))
(3205 real changes made)
.
. g hiwg=(amwg+(85-17)*0.1+(170-85)*0.15+(1020-170)*0.2-1020*0.3)/(1-0.3)
. replace hiwg=0 if amwg>(1700-((85-17)*0.1+(170-85)*0.15+(1020-170)*0.2+(1700-
> 1020)*0.3))
(0 real changes made)
. replace hiwg=0 if amwg<=(1020-((85-17)*0.1+(170-85)*0.15+(1020-170)*0.2))
(4525 real changes made)
.
. g hwg=(amwg+(85-17)*0.1+(170-85)*0.15+(1020-170)*0.2+(1700-1020)*0.3-1700*0.4
> )/(1-0.4)
. replace hwg=0 if amwg<=(1700-((85-17)*0.1+(170-85)*0.15+(1020-170)*0.2+(1700-
> 1020)*0.3))
(4541 real changes made)
.
. g amwgt= lowwg+ midwg+ highwg+ hiwg+ hwg
```

```
. label var amwgt "pre-tax avg wage"
. g amtx=amwgt-amwg
. label var amtx "avg monthly tax levied"
. g tx=amtx*12*work
. label var tx "tax levied"
. drop lowwg midwg highwg hiwg hwg
. *
. * calculate the nontaxed income
. g cit=ci+tx
. label var cit "nontaxed cash income"
. g lcit=ln(cit)
. label var lcit "ln_nontaxed cash income"
. g cet=ce+tx*apc
. label var cet "nontaxed cash expenditures"
. g ut=cet
. label var ut "nontaxed utility"
. g lut=ln(ut)
. label var lut "ln_nontaxed utility"
. *
. * recalculate MMU for nontaxed income:
. regress lcit lut
  Source |       SS       df       MS                  Number of obs =    4541
---------+------------------------------               F(  1,  4539) = 5534.36
   Model |  1368.48581     1  1368.48581               Prob > F      =  0.0000
Residual |   1122.3625  4539  .247270874               R-squared     =  0.5494
---------+------------------------------               Adj R-squared =  0.5493
   Total |   2490.8483  4540  .548645001               Root MSE      =  .49726
------------------------------------------------------------------------------
    lcit |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     lut |   .6959084   .0093545     74.393   0.000       .6775691    .7142476
   _cons |   2.298778   .0782766     29.367   0.000       2.145318    2.452238
------------------------------------------------------------------------------
. hausman, save
. test lut=1
 ( 1)  lut = 1.0
       F(  1,  4539) = 1056.75
            Prob > F =    0.0000


. ivreg lcit (lut= dissave cred)
Instrumental variables (2SLS) regression
  Source |       SS       df       MS                  Number of obs =    4541
---------+------------------------------               F(  1,  4539) =  324.43
   Model |  1303.11335     1  1303.11335               Prob > F      =  0.0000
Residual |  1187.73496  4539  .261673267               R-squared     =  0.5232
---------+------------------------------               Adj R-squared =  0.5231
   Total |   2490.8483  4540  .548645001               Root MSE      =  .51154
------------------------------------------------------------------------------
    lcit |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     lut |   .5438083   .0301914     18.012   0.000       .4846185    .6029982
   _cons |   3.565858   .2516262     14.171   0.000       3.072549    4.059168
------------------------------------------------------------------------------
```

```
Instrumented:  lut
Instruments:   dissave cred
------------------------------------------------------------------------------
. test lut=1
 ( 1)  lut = 1.0
       F(  1,  4539) =  228.31
            Prob > F =    0.0000
. predict resids, residual
. matrix accum ZZ=dissave cred, noc
(obs=4541)
. matrix vecaccum EZ=resids dissave cred, noc
. matrix StatChi=EZ*syminv(ZZ)*EZ'
. matrix Chi=StatChi/(e(rmse)^2)
. matrix list Chi
symmetric Chi[1,1]
           resids
resids  .35362103
. * p-value Ho: instruments are good
. display chiprob(1, Chi[1,1])
.55207048
. hausman, less
            ---- Coefficients ----
         |       (b)          (B)            (b-B)   sqrt(diag(V_b-V_B))
         |      Current     Prior        Difference        S.E.
---------+-------------------------------------------------------------
     lut |    .5438083     .6959084          -.1521     .0287057
---------+-------------------------------------------------------------
            b = less efficient estimates obtained from ivreg.
            B = more efficient estimates obtained previously from regress.
Test:  Ho:  difference in coefficients not systematic
            chi2(  1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                      =    28.08
            Prob>chi2 =     0.0000
. drop resids
. matrix drop EZ
. matrix drop StatChi
. matrix drop Chi
. matrix drop ZZ
. *
. * calculate the effect of income tax on individual utility (welfare):
. g dlu=lut-lu
. label var dlu "difference in ln_utility caused by tax"
. g ltx=0
. replace ltx=ln(tx) if tx>0
(3402 real changes made)
. label var ltx "ln_tax"
. *
. regress dlu ltx, noc
  Source |       SS       df       MS               Number of obs =    4541
---------+------------------------------            F(  1,  4540) =12919.75
   Model |  28.7416562     1  28.7416562            Prob > F      =  0.0000
Residual |  10.0998192  4540   .00222463            R-squared     =  0.7400
---------+------------------------------            Adj R-squared =  0.7399
   Total |  38.8414753  4541  .008553507            Root MSE      =  .04717
```

```
------------------------------------------------------------------------------
     dlu |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     ltx |   .0162501    .000143    113.665   0.000       .0159698    .0165303
------------------------------------------------------------------------------
. hausman, save
. ivreg dlu (ltx= kiev wg ), noc
Instrumental variables (2SLS) regression
  Source |       SS       df       MS                  Number of obs =    4541
---------+------------------------------               F(  1,  4540) =       .
   Model |  28.0956917     1  28.0956917               Prob > F      =       .
Residual |  10.7457836  4540  .002366913               R-squared     =       .
---------+------------------------------               Adj R-squared =       .
   Total |  38.8414753  4541  .008553507               Root MSE      =  .04865
------------------------------------------------------------------------------
     dlu |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     ltx |   .0186862   .0001722    108.535   0.000       .0183487    .0190237
------------------------------------------------------------------------------
Instrumented:  ltx
Instruments:   kiev wg
------------------------------------------------------------------------------
. predict resids, residual
. matrix accum ZZ= kiev wg , noc
(obs=4541)
. matrix vecaccum EZ=resids kiev wg, noc
. matrix StatChi=EZ*syminv(ZZ)*EZ'
. matrix Chi=StatChi/(e(rmse)^2)
. matrix list Chi
symmetric Chi[1,1]
           resids
resids  12.563298
. * p-value Ho: instruments are good
. display chiprob(1, Chi[1,1])
.0003934
. hausman, less
            ---- Coefficients ----
         |       (b)          (B)            (b-B)   sqrt(diag(V_b-V_B))
         |      Current     Prior         Difference        S.E.
---------+-------------------------------------------------------------
     ltx |   .0186862     .0162501         .0024361     .0000959
---------+-------------------------------------------------------------
            b = less efficient estimates obtained from ivreg.
            B = more efficient estimates obtained previously from regress.
Test:  Ho:  difference in coefficients not systematic
            chi2(  1) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                      =   644.89
            Prob>chi2 =     0.0000
. drop resids
. matrix drop EZ
. matrix drop StatChi
. matrix drop Chi
. matrix drop ZZ
end of do-file
```