Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

Precautionary Savings and Household Consumption in Slovakia

Michal Gramblička Bratislava 2011

COMENIUS UNIVERSITY IN BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

PRECAUTIONARY SAVINGS AND HOUSEHOLD CONSUMPTION IN SLOVAKIA

Master's thesis

MICHAL GRAMBLIČKA

Department of Applied Mathematics and Statistics

9.1.9 Applied Mathematics

Mathematics of Economics and Finance

AIS Code: f022f334-27d8-4cfa-9ce5-a2c159cc23c1

Supervisor:

doc. Dr. Jarko Fidrmuc

Bratislava, 2011

UNIVERZITA KOMENSKÉHO V BRATISLAVE
FAKULTA MATEMATIKY, FYZIKY A INFORMATIKY

BEZPEČNOSTNÉ ÚSPORY A SPOTREBA DOMÁCNOSTÍ NA SLOVENSKU

Diplomová práca

MICHAL GRAMBLIČKA

Katedra aplikovanej matematiky a štatistiky
9.1.9 Aplikovaná matematika
Ekonomická a finančná matematika

AIS Kód: f022f334-27d8-4cfa-9ce5-a2c159cc23c1

Vedúci diplomovej práce:

doc. Dr. Jarko Fidrmuc

Bratislava, 2011

Univerzita Komenského v Bratislave
Fakulta matematiky, fyziky a informatiky

ZADANIE ZÁVEREČNEJ PRÁCE

Meno a priezvisko študenta: Bc. Michal Grambličk a
Študijný program: ekonomická a finančná matematika (Jednoodborové štúdium, magisterský II. st., denná forma)
Študijný odbor: 9.1.9. aplikovaná matematika
Typ záverečnej práce: diplomová
Jazyk záverečnej práce: anglický

Názov : Precautionary Savings and Household Consumption in Slovakia

Cieľ : Microeconomentric analysis of household consumption using consumption expenditure surveys in Slovakia, application of instrumental estimations (two stage least squares)

Vedúci : doc. Ing. Jarko Fidrmuc, Dr.

Dátum zadania: 15.02.2010

Dátum schválenia: 07.04.2011

....
prof. RNDr. Daniel Ševčovič, CSc.
garant študijného programu

....
študent

....
vedúci práce

Dátum potvrdenia finálnej verzie práce, súhlas s jej odovzdaním (vrátane spôsobu sprístupnenia)

....
vedúci práce

Declaration on Word of Honor

I declare on my honor that this work is written on my own knowledge, references and consultation with my supervisor.

Michal Gramblička

Acknowledgement

I would like to express special thanks to my supervisor doc. Dr. Jarko Fidrmuc, for all of the support and guidance he offered throughout the elaboration of this thesis.

Abstract

GRAMBLIČKA, Michal: Precautionary Savings and Household Consumption in Slovakia [Master's thesis].
Comenius University in Bratislava, Faculty of Mathematics, Physics and Informatics, Department of Applied Mathematics and Statistics.
Supervisor: doc. Dr. Jarko Fidrmuc
Bratislava, 2011

In our work we investigate consumption pattern of households in Slovakia. In the first part, we explore background for our empirical work and draw conclusions based on literature. Additionally, we describe dataset by descriptive statistics to illustrate composition of households in Slovakia. In the second part, we focus on the econometric methods, which are used throughout empirical section. We explain use of the two-stage least squares method by possible presence of endogeneity of regressors. In the final part, we move to empirical results of our work where models of consumption for different household samples and income types are analyzed.

Keywords: consumption, savings, propensity to consume, endogeneity, instrumental variables, two-stage least squares method.

Abstrakt

GRAMBLIČKA, Michal: Bezpečnostné úspory a spotreba domácností na Slovensku [Diplomová práca].
Univerzita Komenského v Bratislave, Fakulta matematiky, fyziky a informatiky,
Katedra aplikovanej matematiky a štatistiky.
Vedúci diplomovej práce: doc. Dr. Jarko Fidrmuc
Bratislava, 2011

V našej práci sa zaoberáme spotrebou domácností na Slovensku. V prvej časti skúmame pozadie pre empirickú časť práce, pričom vyvádzame predpoklady s ohľadom na literatúru, ktorá sa danej problematike venuje. Taktiež popisujeme dáta pomocou deskriptívnych štatistík na ilustrovanie zloženia domácností na Slovensku. V druhej časti sa zameriavame na ekonometrické metódy, ktoré sú použité v empirickej časti. Zdôvodnenie použitia dvojstupňovej metódy najmenších štvorcov spájame s problémom endogenity použitých regresorov. V poslednej časti sa dostávame k výsledkom našej empirickej práce, kde analyzujeme výsledky modelov spotreby na rôznych vzorkách domácností a príjmových typov.

Kľúčové slová: spotreba, úspory, sklon k spotrebe, endogenita, inštrumentálne premenné, dvojstupňová metóda najmenších štvorcov.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Household consumption represents very substantial part of gross domestic product of each country. Therefore, patterns in consumer behavior driven by changes in income hold important information for analysts and policy makers to evaluate impact of policy changes. To understand movements in consumer behavior one should include analysis of changes in propensity to consume throughout life-cycle together with influence of other regressors which might shed some light on rationality behind those changes.

The goal of this work is to scrutinize consumption of households in Slovakia throughout life of a consumer and different samples to understand behavior which drives it. We also take a look at impact of the shift from the ordinary least squares method to the two-stage least squares method due to endogeneity of regressors.

The thesis is organized as follows. In the first chapter, we focus on the background for our empirical work. Basic consumption function developed by Keynes is introduced, together with different approaches arisen from this function. Based on reviewed literature we draw possible conclusions for our empirical work. Additionally, dataset is described by descriptive statistics to illustrate composition of households in Slovakia. In the second part, we focus on the econometric methods which are used throughout empirical section. Firstly, classic regression is introduced following by possible errors which might cause its bias. Consequently, we explain use of the two-stage least squares method by possible presence of endogeneity of regressors. Endogeneity is treated by instrumental variables and their use and conditions which are needed to be fulfilled we explain in the next section. Moreover, we focus on test of overidentifying restrictions and test of endogeneity to help us evaluate correctness of use of instrumental variables and the two-stage least squares method. In the final chapter, we move to empirical results of our work. Models of consumption for different household samples and income types are analyzed. Firstly, Keynes' consumption function extended by other regressors is estimated, followed by results for different age and household type samples. Secondly, we distinguish regular and irregular income and treat them as endogenous variables.

Chapter 2

Consumption and Savings

Household consumption represents very substantial part of GDP of each country, particularly the final consumption of households has corresponded to 57% of GDP [1] in Slovakia in 2010. Therefore, patterns in consumer behavior driven by changes in income for each household hold important information for analysts and policy makers to evaluate impact of policy changes.

A quick look at the dependency between income and consumption should lead us to a general idea that change in income goes hand in hand with disruption of consumption pattern. Drawback of this condition is firstly in assuming that people do not care if their consumption is changed every time they receive paycheck or their salary is corrected. More natural behavior is smoothing[2] lifetime consumption. To protect against constant income volatility an agent uses savings and loans to smooth expenditures on consumption in long term horizon.

We would like to express that throughout this work we exchange terms household and

[1] Based on data from the Statistical Office of the Slovak Republic (SUSR) `http://portal.statistics.sk/showdoc.do?docid=24139`, final household consumption in a year 2010 was EUR 37.928bil while GDP in current prices was EUR 65.906bil.

[2] Term consumption smoothing is used in a way, that an agent is supposed to keep the expected marginal utility of expenditures constant. Consumption itself is not smooth, people do not consume every single moment same fraction of their income. For example at a weekly basis their consumption rises during the weekend or yearly during Christmas period.

family. This notion describes group of people inhabiting same home and sharing expenses and earnings.

2.1 Consumption Function

Basic consumption function was developed and published by John Maynard Keynes [16] in a year 1936. This function tells us how dependent consumption is on disposable income.

Consumption function:

$$C = c_0 + c_1 Y^d, \tag{2.1}$$

where

- c_0 is autonomous consumption;
- c_1 is marginal propensity to consume;
- Y^d is disposable income; income reduced by tax payments $W - T$.

Parameter c_0 describes autonomous consumption, independent of income; even in a situation that consumer does not have an income, he or she needs to consume elementary goods as food, water, clothes, etc.

The second parameter c_1 shows what amount of disposable income is spent on consumption and what proportion of income is saved. Therefore, difference $1 - c_1$ is called marginal propensity to save. Hence, we assume typical household to spend substantial part of their disposable income on consumption, based on the information from National Bank of Slovakia (NBS) [3] this number should be around 94%. Although, this corresponds to average propensity to consume, we assume that it should not differ substantially from marginal propensity, which will be estimated. However, person could be satisfied with her/his level of consumption and decide to use this income for an additional precautionary

[3] National Bank of Slovakia: BIATEC- Banking journal No. 7/2005, http://www.nbs.sk/_img/Documents/BIATEC/BIA07_05/12_15.pdf

motive in savings, to build up reserves higher than needed for natural life-cycle requirements. In average 6% of family disposable income is transformed into savings.

2.2 Different approaches

Flaws of Keynesian type of consumption are obvious; our present consumption depends only on current net income. There is no weight on future possibility of better paid job (because of higher qualification, better skills gained). On the other hand, nobody thinks about being unemployed or sick either.

This absolute theory, as it is called, has encouraged economists to think about more realistic models. Among others, two out of the developed theories stood up, the permanent income hypothesis (PIH) and the life-cycle hypothesis by Milton Friedman and Fisher, Harrod, Ando, Modigliani respectively.

They both take into account future uncertainty. In Friedman‘s PIH consumption is determined by income expectations in longer period. Short-term swings in income do not have significant impact on consumer‘s behavior.

However, there is a general distinction between different types of income change. This difference is based on anticipation. When the agent undergoes income shock or unanticipated boost of income their reaction should be different as when they are facing anticipated change. The basic Life-Cycle/Permanent Income Hypothesis (LCH/PIH) tells us that consumption expenditures of the agent are supposed not to react to the anticipated change in income. The change has no effect on the growth rate of expenditures. This hypothesis is important to be understood by policy makers for optimal timing of change in fiscal policy, and the effects of income fluctuations on the growth rate of the economy. The agent is expected to use the impact of income fluctuations to immunize volatility in consumption, smooth the consumption, therefore, their response should be limited if any, to this type of change. As in Shea [20] LCH/PIH is typically tested by equation

$$log(C_t) - log(C_{t-1}) = \alpha + \beta X_{t-1} + \gamma Z_{t-1} + \epsilon_t, \tag{2.2}$$

where C is a real consumption in a period t, $t-1$ respectively. X_{t-1} is a vector of controls such as the expected real after-tax rate of return between $t-1$ and t, Z_{t-1} is a variable in the information set at $t-1$ correlated with income growth between income growth between $t-1$ and t. Under the LCH/PIH γ should equal zero, hence anticipated income movements should not cause consumption to change. Household expecting income to change based on this hypothesis, tends to save or borrow to smooth the consumption path.

However, to identify situations in which income change is anticipated or not is difficult, if not impossible from questionnaires from which data are obtained. Hence, we have to state our point here. We will assume throughout this work that irregular income is an income change, which is spent or saved during the period covered by questionnaire. This change will be treated as anticipated.

The main implication of life-cycle and permanent income hypothesis is that household consumption growth should be orthogonal to variables in the household's current information set; the agents should not react to short-term income changes. Friedman's permanent income hypothesis has the key conclusion in statement that unexpected short-term income changes have very limited effect on consumer spending behavior. On the other hand, long-term income expectations are crucial. Hence, consumption pattern is derived from an individual's real wealth, not from disposable current income as it was in (2.1). The agent takes into account transitory effect of future improvement of his experiences and consequently income. Therefore, he or she is aware of current real assets (experience, skills, wealth, shares) and is able to estimate future gains from those assets in anticipated lifetime income.

2.3 Literature Review

The responses of consumption to changes in income have been tested in a number of papers. Among studies that consider evidence from individual household expenditures are John Shea [20], M. Browning and D. Collado [2], Chang-Tai Hsieh [12] and A. Parker [18]. Consumption path and dependency on income is illustrated in Pierre-Olivier Gourinchas and A. Parker [7]. The interesting summary can be found in Jappelli and Pistaferri [13], however, these analyses mostly work with intertemporal changes and not with cross-sectional data. Therefore, their contribution to our work is limited, but we can take these results figuratively to help us draw boundaries for our models.

Browning and Collado [2] studied panel data evidence of impact of extra payments for workers in Spanish economy. These extra payments are received by majority of full-time workers included in bonus scheme. Every month they receive one-fourteenth of their annual pay and one-seventh twice a year (usually June/July and December). Those payments varies during the year in a predictable way. However, if we look at them only in a month period, consistent with our approach, we might interpret normal pay as regular income and these bonus payments as irregular income, forasmuch as their value varies. Based on amount of bonus received, two types of behavior arise. An agent in normal times tends to smooth income with large and well-understood income changes, and spend small changes. This is summarized as bounded rationality, where utility from small change is lesser than afford spent on recalculation of smoothed consumption on higher value. On the other hand, large anticipated income change pushes the consumption for whole period on a level with higher utility. They found no evidence of synchronization between income flow during the year and expenditures. According to their explanation this is caused mostly because of their selection and amount of money disbursed. The benefits of making the optimal expenditure calculations in their income scheme are large and the costs small. Hence, we would assume, based on these results, that small change in income which reflects the insignificant irregular income payment would be transferred into consumption in the current months, on the other hand, large change will be used for smoothing consumption in the following long-term horizon, therefore, in this case majority of the irregular income should be saved.

Additionally, very similar results might be found in Hsieh [12]. His research took evidence from natural experiment in Alaska, where Alaskan residents receive significant payments once a year from The Alaska Permanent Fund. The Alaska Permanent Fund was established in 1976, when 25% of the state government's oil royalties was deposited into this trust fund. Since 1982, approximately half of the annual dividend income from the fund has been distributed among the eligible residents (those who live in the state for a minimum of one calendar year and has applied for the payment) every year. The size of dividend payments has varied from $331 in 1984 to $3269 in 2008. The formula used to calculate the amount to be disbursed is widely known and newspapers in Alaska publish estimates for each year's payments regularly, therefore, it is anticipated income and obviously substantial. Although, he found no evidence that Alaskan residents overreact to these payments, the very same households are excessively sensitive to their income tax refunds, which are significantly smaller than fund payments. This again can be interpreted as the way before. Large and anticipated income is taken into account when deciding over lifecycle consumption, but small and irregular is not. However, there might be another more genuine reason behind agents' behavior. They evaluate resources given them back from tax-refunds as once already lost, therefore they take them as effort-free, hence spend significant part of it. Contrary, resources from Alaska Permanent Fund are in their perception earned, they had to apply for them and fulfill criteria that might be reason why they do not tend to spend all of them.

On the other hand, Parker [18] was analyzing whether expenditures on nondurable goods increase contemporaneously with predictable changes in Social Security tax withholding. He assumed that small change in tax rates, which was legislated and announced well ahead, should not influence household consumption. However, his study showed that households changed their consumption expenditures as a response to the fluctuations in income induced by the tax system. One-percent increase in after-tax income contemporaneously increased expenditures on nondurable consumption by around a half of a percent.

Therefore, consumers do not perfectly smooth their demand for goods across expected income changes. His work can be interpreted as evidence against consumption smoothing. If expected changes in taxes influence significantly consumption behavior, then fiscal stabilization plays important role influencing the household consumption.

Complementary to Parker, John Shea [20] tested LCH/PIH on set of households[4]. In this paper expected wage growth is significantly correlated with consumption growth, contrary to the LCH/PIH. Consequently, we can take this result as a support for our research to assume that irregular income will be spent on consumption.

Based on those diverse results it is hard to take a stand and acclaim what to expect. Since we try to obtain results from cross-sectional data with consumption function consequently differentiated to regular and irregular income, we naturally expect regular income to be strongly correlated with current consumption and irregular income not to play important role in consumption.

During life-cycle persons are supposed to build up reserves due to rising income based on their improvement of skills and stabilizing of consumption, these savings can be according to John M. Keynes [16] and Browning [3] generalized into two main categories. People are cutting their current consumption to enjoy brighter tomorrows and insure for retirement (the improvement and retirement motive) or to build up reserves, a buffer, against uncertainty of the future (the precaution).

The second leads people to build up a certain amount of reserves, which is named as a buffer-stock. The buffer-stock or precautionary savings are therefore additional savings to those, which are needed because of unpredictable future risk, for example loss of employment or health problems. It is closely related to risk aversion and prudence of each agent. The more risk averse agent is, bigger buffer-stock is he or she going to build.

Based on Gourinchas and Parker's [7] results from consumption during lifecycle typical young consumers behave as buffer-stock agents. Around age 43 typical household

[4]He broadened set of data usually used before, data from Panel Study of Income Dynamics (PSID). He isolated a subset of households whose heads could be matched to particular labor-union contracts to extent set by expected wage growth.

starts with accumulating assets for retirement. Figure 2.1 of household consumption and income over the life cycle is very similar in our dataset to their illustration.

As can be seen, income of young households rises steeply together with consumption as their situation improves along with possible better paid job based on better qualification and higher expenditures connected with more spending on children or better lifestyle. This peeks in early 40's. From that point on expenditures on kids decline since they reach late teenage years and are supposed to be more independent. Eventually, income falls under level of consumption as persons reach retirement age, consequently, elderly take advantage from accumulated capital (imbursement for retirement) to support their expenses.

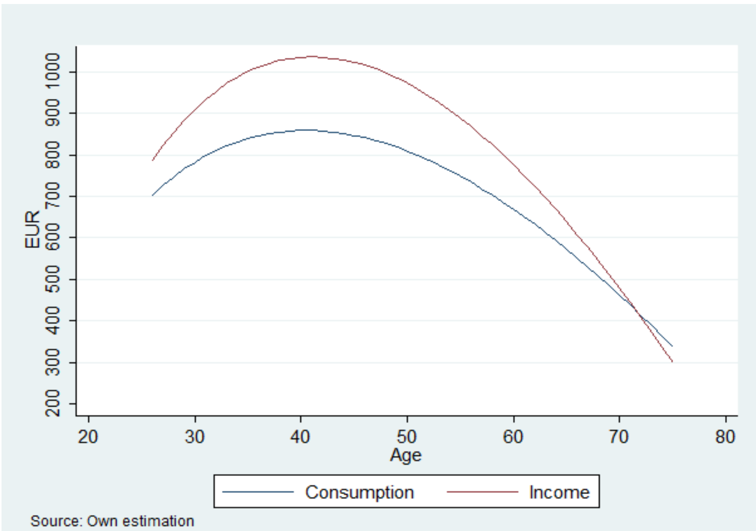
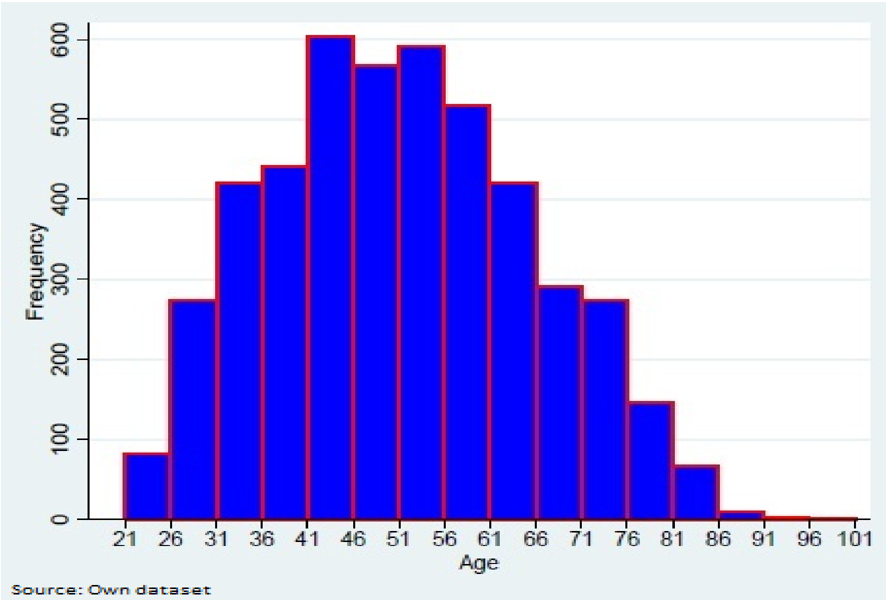


Figure 2.1: Household consumption and income over the life cycle.

We have excluded observations where head of a family is older than 75 years or younger than 25, because insufficiency and possible outlier property of those observations. In the Figure 2.2 is showed histogram and frequency of age of household heads represented in our data.



Age	Obs.
16-20	3
21-25	82
26-30	274
31-35	420
36-40	441
41-45	603
46-50	567
51-55	591
56-60	517
61-65	421
66-70	291
71-75	274
76-80	146
81-	80

Figure 2.2: Age Distribution

2.3.1 Precautionary Savings

The precautionary saving theory based on Lusardi [17] basically asserts that people save to insure against income risk. Saving rate of farmer should be higher than doctor's, because farmer is worried his income may drop dramatically in case of a bad season. Therefore, motivation behind his saving is precaution. Theoretically, income uncertainty boosts saving. Problem of empirical work is how to measure this risk. Hence, saving and wealth are related not only to nominal value of income but also to its variance. Consumers accumulate capital not only to offset future declines in income, but also against income risk associated with loss of employment.

Karazosian [14] showed strong precautionary motive for saving with positive effect of income uncertainty on the wealth to permanent income ratio in his panel study. On the other hand, Guiso [9] found that only 2% of accumulated savings can be explained by precautionary motive. Hence, empirical results are rather mixed, and not yielding satisfactory conclusion.

Since our dataset does not possess information about actual wealth of each household or its development throughout period of time, we are not able to make typical analysis where change of wealth to income ratio is studied. Alternative approach might be done in our dataset by estimating propensity to consume in households with different samples of income risk. This risk might be associated with type of work, better to say if head of a family is self-employed or an employee. Our assumption is that households with self-employed head face higher income risk and therefore need to accumulate more precautionary savings than households with employed heads whose income might be considered as safe.

2.4 Data Description and Descriptive Statistics

In our work we use data from Household Budget Survey (HBS) with harmonized expenditures recorded in Classification Of Individual Consumption by Purpose adapted to the needs of HBS (COICOP-HBS) together with income for individual household estimated for a typical month in a year 2005. The HBS is a national survey and focuses on household's expenditure on goods and services. These data were collected by the Statistical Office of the Slovak Republic (SUSR) from interviews and logs maintained by each household on a behalf of the European parliament. This set consists results from 4710 households. Expenditures and incomes of each unit (household) are precisely recorded together with socioeconomic details of each member, where whole dataset is collected on 5 year basis. Prices are in EUR for a period of one month.

Collected data might be distinguished into 3 main categories:

1. Basic information;

2. Expenditures;

3. Earnings.

Each of these categories contains detailed information, for example in the first you can find in which region household is situated or size of settlement, size of household and socio-economic characteristics; year of birth, gender, family status, education, common economic activity of each member.

Expenditures are recorded at price actually paid, therefore it includes indirect taxes. It consists of expenses on non-durable goods as food, beverages, clothes, durable goods as household appliances and vehicles together with expenses on services and maintenance costs.

In the last category you can find all types of income, gross earnings from employment, irregular income from employment, earnings from self-employment, royalties, bonuses, pensions, etc.

We use throughout this work variables summarized in Table 7.4 with descriptive statistics in Table 7.1, Table 7.2 and Table 7.3 found in Appendix.

The descriptive statistics shown in Table 7.1 display main characteristics for typical household in a sample. In average, every household spends on nondurable goods EUR696.2 and has disposable income EUR742.2 monthly. Highest income and consumption on a family member is in Bratislava region EUR329.4 and EUR312.1, respectively. Contrary in Prešov on each member falls only EUR223.0 of income and EUR205.5 on consumption. We can compare these values with those found in regional database of the Statistical Office of the Slovak Republic [5] summarized in Appendix in Table 7.3. We cannot compare nondurable consumption due to lack of details in data from SUSR. As can be seen there is a big difference in nondurable consumption for each member in Košice and value from SUSR, this difference should be allocated specifically to durable consumption. Substantial expense on vehicle in Košice region contributes to this difference, particularly it adds EUR16.1 to nondurable consumption of every member in this region. Hence, we add expenses on vehicles and big household appliances to compare data with those from SUSR, as can be noted, they are very similar. Nevertheless, in empirical part we have excluded nondurable expenses because as can be seen one extensive payment could bias results.

Propensity to consume in our dataset is estimated to be 0.94, hence in average households consume 93.8% of their disposable income. This result is in line with findings of

[5] Regional statistics of the Statistical Office of the Slovak Republic (SUSR) http://px-web.statistics.sk/PXWebSlovak/index_en.htm

the National Bank of Slovakia (NBS), which estimated average propensity to consume of Slovakian households in the year 2004 to be 94.2% [6].

Average family has slightly less than three members 2.93. However, there is significant impact of region, where household is situated. Largest households are located in Žilina and Prešov region with 3.3 and 3.4 members, respectively. Those bigger households might be associated with two regional specifications. In Prešov region high density of Roma minority is situated, statistically Roma families tend to live in bigger composition with more kids. Secondly, higher natality in Žilina region and bigger households are typical due to strong Christian tradition with tight rules for birth-control. Households consisting of one person account for 16.2% of all households. These findings are similar to results published in OECD Family database [7] where average Slovakian household has 3.04 members and 19.4% are only single personal. Moreover, based on The Statistical Yearbook of the Economic Commission for Europe of UN [8] average size of Slovakian family is 2.9. In our set there is 0.64 member under 18 years per household. Head of a family with college education is in 13% of observations and head or husband or wife in 17%.

Approximately every fifth household lives in one of the eight regional cities (Bratislava, Trnava, Trenčín, Nitra, Žilina, Banská Bystrica, Prešov or Košice). In 69% of households is head of a family man, only 2% of households heads are employed as a manager or in a leading position and 18% of heads work in public sector. Additionally, 36.5% of all employed persons in our set operates in public sector. This number is similar to 39.1% obtained from the SUSR [9].

[6] National Bank of Slovakia: BIATEC- Banking journal No. 7/2005, http://www.nbs.sk/_img/Documents/BIATEC/BIA07_05/12_15.pdf

[7] OECD Family database, http://www.oecd.org/document/4/0,3746,en_2649_34819_37836996_1_1_1_1,00.html

[8] The UN Statistical Yearbook of the Economic Commission for Europe, http://www.unece.org/stats/trends2005/Sources/114_Average%20household%20size.pdf

[9] Statistical Office of the Slovak Republic, http://portal.statistics.sk/showdoc.do?docid=16968

Chapter 3

Econometric Methods

In this chapter we describe and summarize econometric methods used in our work for estimating and analyzing results. We have found significant help for writing this part in Wooldridge [22], Davidson and MacKinnon [5], Hayashi [11], Greene [8], Baum, Shaffer and Stillman [1], Cameron and Trivedi [4] and Kennedy [15].

In the beginning, we would like to gently introduce the econometrics. Define this scientific discipline strictly is difficult, because it is widely used. Nowadays, it gets into fields no one would expect before[1].

Econometrics is based on use of statistical methods for estimating economic relations, testing economic theories and assumptions. Moreover, evaluating and implementing government and entrepreneurial actions. The most common application of econometrics is forecasting macroeconomic models, e.g., inflation, gross domestic product, unemployment rate. The econometric methods are not strictly linked with these types of data; they are also frequently used in areas which are not tightly connected with macroeconomics. Econometrics evolved as an individual discipline from mathematics and statistics. In both these disciplines the method of regression analysis is widely used.

The core of econometric modeling is the classical linear regression model (CLR), where the most frequently used method for estimating is the ordinary least squares method

[1]The credit card company was recently accused from use of personal data from purchases to forecast change in life of the customers (getting divorced) [27].

(OLS). Basic form of regression model is the linear single-equation model, which is interpreted as

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + u_i, \tag{3.1}$$

where

y_i	is dependent variable for i^{th} observation, $i = 1, 2, \ldots, n$;
x_{ik}	k^{th} explanatory variable for i^{th} observation, $i = 1, 2, \ldots, n$;
β_k	k^{th} unknown parameter;
u_i	unobserved error term for i^{th} observation, $i = 1, 2, \ldots, n$.

We get n equations for n observations ($i = 1, 2, \ldots, n$). Econometric methods are used to estimate parameters of regressors $\beta_0, \beta_1, \ldots, \beta_k$. The model (3.1) can be rewritten in matrix form as

$$y = X\beta + u, \tag{3.2}$$

where each member has following form

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1k} \\ 1 & x_{21} & \ldots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \ldots & x_{nk} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}.$$

Estimate of β is denoted as $\hat{\beta}$ and enumerated value of dependent variable is labeled $\hat{y}$.

Consequently, we obtain

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik}. \tag{3.3}$$

Hence, unobserved error term u_i from (3.1) acquires accurate value of error e_i; for the dependent variable it yields

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \cdots + \hat{\beta}_k x_{ik} + e_i,$$

that can be rewritten as

$$y = X\hat{\beta} + e. \tag{3.4}$$

Several assumptions need to be fulfilled to achieve best linear unbiased regression (BLUE) for estimating vector β using OLS:

- expected value of unobserved error term is zero, i.e., the mean of the distribution from which is error term drawn equals zero;
- error terms are normally distributed;
- the variance of disturbance terms is the same (homoskedasticity assumption);
- the disturbance terms are not correlated with one another $Cov(u_i, u_j) = 0$, for $\forall i \neq j$;
- the observations on the independent variables can be considered as fixed in repeated samples;
- independent variables are not linearly dependent (multicollinearity problem) and their number is smaller than the number of observations.

3.1 Ordinary Least Squares

If the previous assumptions hold OLS is BLUE and estimation of parameters is based on minimalization of sum of squared residuals. Equation for residuals is taken from (3.4)

$$u = y - X\beta.$$

This minimalization can be reported as

$$(y - X\beta)^T(y - X\beta) \to \min_{\beta},$$

which after multiplying yields

$$y^T y - y^T X\beta - \beta^T X^T y + \beta^T X^T X\beta \to \min_{\beta}. \tag{3.5}$$

To get the minimum one needs the derivative (3.5) with respect to x

$$-y^T X - X^T y + 2(X^T X)\hat{\beta}_{OLS} = 0$$
$$(X^T X)^{-1}\left(\frac{X^T y + y^T X}{2}\right) = \hat{\beta}_{OLS}.$$

Estimating parameters using OLS method is based on

$$\hat{\beta}_{OLS} = (X^T X)^{-1}(X^T y). \tag{3.6}$$

3.2 Endogeneity

One of the possible problems, which might occur in econometric modeling is inconsistency of the parameter estimation. This inconsistency arises when the first assumption does not hold $E(u_i|X_i) \neq 0$. If one of the regressors is endogenous then in general OLS estimates of all regression parameters are inconsistent. Hence, nonzero mean conditional is associated with situations in which error terms are correlated with the regressors. The major reasons for this endogeneity are omitted variable bias, sample selection bias, measurement error bias and simultaneity. Broad discussion about each of these problems and possible solutions can be found in Cameron and Trivedi [4] and Kennedy [15].

A general approach to control for endogeneity is the instrumental variables method. This method cannot always be applied, as necessary instruments may not be available.

The problem of endogeneity in our dataset might be mainly associated with two causes; simultaneity and the measurement error bias. Simultaneity arises when one or more of the explanatory variables is determined with the dependent variable. In our case endogeneity of income may be caused by reverse causality. Although, it looks more natural to denote consumption to be following income, opposite approach might arise. Consumer might have determined level of consumption for which he or she needs/wants to obtain resources. Therefore, he or she seeks a job which will provide determined consumption. Unfortunately, people usually do not work in field which provides them with personal satisfaction and happiness but in that supporting their financial needs. The bias linked with the reverse causality should be positive. If we consider simple model where y is consumption and x disposable income (for simplicity we drop i for identification the number of observation, other exogenous regressors and intercept), then two equations are denoted as

$$y = \beta_1 x + u_1, \tag{3.7}$$

$$x = \beta_2 y + u_2. \tag{3.8}$$

Hence, we have equation for estimating impact of income on consumption, and vice versa. When we substitute for y in the second equation we obtain

$$x = \beta_2\beta_1 x + \beta_2 u_1 + u_2, \tag{3.9}$$

putting x on left-side and dividing by $(1 - \beta_2\beta_1)$, where we suppose that $\beta_2\beta_1 \neq 1$, since both parameters should be less than 1, we get

$$x = \frac{\beta_2 u_1 + u_2}{1 - \beta_2\beta_1}. \tag{3.10}$$

The question is whether x and u_1 are correlated and if so, what bias does it produce. From 3.10 we can see that they are correlated whenever $\beta_2 \neq 0$ and this correlation's bias should be positive based on

$$Cov(x, u_1) = [\frac{\beta_2}{1 - \beta_2\beta_1}]E(u_1^2) = [\frac{\beta_2}{1 - \beta_2\beta_1}]\sigma_1^2, \tag{3.11}$$

where $\sigma_1^2 = Var(u_1) > 0$, also $\beta_2 > 0$, which together with $\beta_2\beta_1 < 1$ yields positive bias.

Secondly, the measurement error bias may be the reason of inconsistency of OLS in our dataset, because many economic variables are collected with flaws. For example, lots of data is collected from surveys and imperfect questionnaires. Hence, measurement error biases data from the true value which means that independent variables (collected with errors) can be correlated with the error terms.

Considering CLR model, problem with measurement errors may be stated as

$$y_i^* = \beta_0 + \beta_1 x_i^* + u_i^*, \quad u_i^* \sim iid(0, \sigma^2), \tag{3.12}$$

where iid is abbreviation for independent and identically distributed, the variables y_i^* and x_i^* are not directly observed. Instead, we collect

$$\begin{aligned} x_i &\equiv x_i^* + v_{1i}, \\ y_i &\equiv y_i^* + v_{2i}. \end{aligned} \tag{3.13}$$

Measurement errors are denoted as v_{1i} and v_{2i}, they are assumed to be independent of y_i^*, x_i^* and u_i^*. It is easy to see from (3.13) that $x_i^* = x_i - v_{1i}$ and $y_i^* = y_i - v_{2i}$. Substituting into (3.12), we find that

$$\begin{aligned} y_i - v_{2i} &= \beta_0 + \beta_1(x_i - v_{1i}) + u_i^* \\ y_i &= \beta_0 + \beta_1(x_i - v_{1i}) + u_i^* + v_{2i} \\ &= \beta_0 + \beta_1 x_i + u_i^* + v_{2i} - \beta_1 v_{1i} \\ &= \beta_0 + \beta_1 x_i + u_i, \end{aligned}$$

where $u_i = u_i^* + v_{2i} - \beta_1 v_{1i}$. The measurement error of the error term increases the variance of u_i, however greater variance should not be a problem in OLS method, but especially estimating consumption function might be clumsy thanks to problem pointed out by Davidson and MacKinnon [5]:

> Measurement errors are a potential problem whenever we try to estimate a consumption function, especially if we are using cross-section data. Many economic theories suggest that household consumption will depend on "permanent" income or "life-cycle" income, but surveys of household behavior almost never measure this. Instead, they typically provide somewhat inaccurate estimates of current income. If we think of y_i as measured consumption, x_i^* as permanent income, and x_i as estimated current income, then the

above analysis applies directly to the consumption function. The marginal propensity to consume is β_1, which must be positive, causing the correlation between u_i and x_i to be negative. The probability limit of $\hat{\beta}_1$ is less than the true value β_1. In consequence, the OLS estimator $\hat{\beta}_1$ is biased downward, even asymptotically.

Since our data are cross-sectional, therefore it is natural to assume that income does not take into account future imbalances or precautionary motive, presented current income is biased, since β_1 is positive and correlation between u_i and x_i negative. The estimated parameter $\hat{\beta}_1$ is biased downwards. Hence, we have to find a way to diminish correlation between error terms and measured income.

We have pointed out two problems which might have effect on estimation of parameters, since their expected influence on bias is contradictory, we cannot predict final bias of OLS regression. Consequently, we have to treat this bias, which might be done by methods which include Instrumental Variables (IV).

3.3 Instrumental Variables

Instrument is an additional variable for regressor, which is contemporaneously [2] correlated with the disturbance term. Denote matrix of instruments Z. Each instrument needs to fulfill two characteristics.

1. The instrument has to be uncorrelated with the error term

$$E(u) = E(u|Z) = 0. \tag{3.14}$$

2. It needs to be (highly) correlated with regressors, for which it is supposed to be used as the instrument.

[2] By contemporaneously is meant, that j^{th} regressor observation is correlated with j^{th} error term.

Standard regression has an assumption that disturbance terms are not correlated with regressors, hence effect of x on y is only direct (3.15):

$$\begin{array}{ccc} x & \longrightarrow & y \\ & \nearrow & \\ u & & \end{array} \tag{3.15}$$

In some situations, as was mentioned earlier, regressors are correlated even with errors (3.16).

$$\begin{array}{ccc} x & \longrightarrow & y \\ \uparrow & \nearrow & \\ u & & \end{array} \tag{3.16}$$

Suppose, the correlation between x and u is negative and growth of these variables has positive impact on y. Result of this indirect effect of x on u is that OLS estimation $\hat{\beta}$ will be lower than the real value β. Higher value of x leads directly into higher y, but there is also unestimated effect of lower u on dependent variable. To see this effect we need to differentiate with respect to x the model. Then it is obvious to see in (3.17)

$$\begin{aligned} y &= \beta_0 + \beta_1 x + u(x), \\ \frac{\mathrm{d}y}{\mathrm{d}x} &= \beta_1 + \frac{\mathrm{d}u}{\mathrm{d}x}. \end{aligned} \tag{3.17}$$

The effect of change $\mathrm{d}x$ is not in this case only β_1 but $\beta_1 + \frac{\mathrm{d}u}{\mathrm{d}x}$. Difference leads into bias of OLS, which is needed to be treated by a method where this additional influence is taken care of. This method is the mentioned IV estimation. In this method one needs to find an instrument z which is correlated with regressor but not with disturbance terms (3.18). Change in instrument is not associated with change in y, only indirectly via x.

$$\begin{array}{ccccc} z & \longrightarrow & x & \longrightarrow & y \\ & & \uparrow & \nearrow & \\ & & u & & \end{array} \tag{3.18}$$

3.3.1 Identification

There are usually several candidates to be used as instruments. For example for disposable income, which tends to be endogenous regressor we might choose age of a family head, education, gender, family size, living in city or urban area, working status, etc. as instrumental variable.

If the number of instruments is the same as the number of endogenous regressors, then the model is just-identified. In case there are more instruments than endogenous regressors we have to deal with overidentified model. Overidentified models can be solved as they were just-identified by choosing linear combination of instruments, since both instruments are uncorrelated with error term their linear combination will be also uncorrelated. Difficulty might arise in choosing proper weights. This approach is called generalized instrumental variable estimation (GIVE).

If an instrument fails to succeed in holding the first and the second condition, model might end as unidentified. Particularly, if an instrument is correlated with error term then it is invalid instrument. In case it is uncorrelated with endogenous variable then we have irrelevant instrument. Finally, if the instrument is not strongly correlated with endogenous regressor the instrument is called a weak instrument and the model is weakly indentified.

3.3.2 Instrumental Variables Estimation

First assumption of OLS says $E(u_i) = 0$. It implies also $E(u_i|X_i) = E(X_{ij}u_i) = 0$, in case regressors are not correlated with disturbance terms, what eventually arises into (3.6). Violation of this assumption in sense that regressors X are correlated with error terms leads into different estimation.

Denote Z matrix of instruments. Hence, value of each instrument for particular observation. Original regressor (from matrix X) which is not contemporaneously correlated with the error term uses own values as instrumental variable in the matrix Z. Since instruments are not correlated with disturbance term we can write $E(u_i) = E(u_i|Z_i) =$

$E(Z_{ij}u_i) = 0$, hence

$$E(Z_{ij}u_i) = E(Z_{ij}(y_i - X_i\beta) = 0$$
$$\frac{1}{n}\sum_{i=1}^{n}(Z_{ij}(y_i - X_i\beta)) = 0$$
$$Z^T(y - X\beta) = 0.$$

$$\hat{\beta}_{IV} = (Z^T X)^{-1} Z^T y \tag{3.19}$$

This method falls into the Generalized Method of Moments (GMM) framework, where additional moment conditions gained from instrumental variables can be used for estimation.

3.3.3 Two-stage least squares method

The drawback of IV estimator is that it needs precisely as much instruments as the number of regressors. In overidentified models it can be used by not taking spare instruments into account to obtain just-identified model with number of instruments equal to number of endogenous regressors or by taking linear combination of instruments, since instruments and error terms are uncorrelated, their linear combination is also.

On the other hand, 2SLS can be used with more instruments than is the number of regressors without exception. If the number of instruments is equal to number of endogenous regressors than the 2SLS estimator is simplified into IV estimator (3.19) with instruments Z. In an overidentified model these two estimators equal if the instruments are $\hat{X}$, where $\hat{X} = Z(Z^T Z)^{-1} Z^T X$ is the estimated value of X using OLS of Z on X.

This method is called as two-stage least squares, because in the days without powerful statistic software it was obtained by two consecutive OLS steps. In the first one, OLS regression was used to obtain estimates by an $\hat{X}^*$ regressing endogenous X^* by Z. In the second part of 2SLS, OLS was used for X together with $\hat{X}^*$ as regressors on y to obtain $\hat{\beta}_{2SLS}$.

If a regressor x_i is valid instrument, it is included in the matrix Z. Since, $P_Z x_i = Z(Z^T Z)^{-1} Z^T x_i = x_i$, no first step is needed and such regressor serves as its own instrument.

The fitted values obtained from the first-stage plus the values of exogenous regressors which serve as their own instruments, are together recorded in matrix $P_Z X = Z(Z^T Z)^{-1} Z^T X$. Then the second-stage equation is

$$y = P_Z X \beta + u, \tag{3.20}$$

and the 2SLS estimator is expressed

$$\hat{\beta}_{2SLS} = (X^T Z(Z^T Z)^{-1} Z^T X)^{-1} X^T Z(Z^T Z)^{-1} Z^T y. \tag{3.21}$$

3.3.4 Test of Overidentifying Restrictions

Picking the right instruments might not be trivial exercise. As it was said earlier, every instrument has to fulfill two basic rules. It must not be correlated with the error term and it needs to be (preferably strongly) correlated with the endogenous regressor. Behind every taken instrument should be rational explanation, why it should be correlated and therefore, taken into account in the regression. Every aspect has the other side, use of too many instrumental variables is not applicable. Consequently, we need a test which is able to determine, if the regression is not overidentified, in a sense that some of the instruments are either not satisfying the orthogonality condition required for their use (they are not uncorrelated with the error process) or the model is misspecified. Misspecification is meant, that the instruments are incorrectly excluded from the "main" regression.

The Sargan-Hansen test[3] of overidentifying restrictions has the null hypothesis that the instruments are valid instruments, i.e., uncorrelated with the error term and that the excluded instruments are correctly excluded from the estimated equation. Under the null, the test J statistic is distributed as χ^2_r, where $l - k = r$ with l instrumental variables and

[3] Named after Hansen for the GMM estimation in Hansen [10] and Sargan who used test of overidentifying restrictions in the context of GIVE in Sargan [19].

k number of regressors. Hence, the number of degrees of freedom equals to the number of overidentifying restrictions, diminished by k degrees of freedom, which are used up in estimating the vector of coefficients $\hat{\beta}$. A rejection of null hypothesis infers that the instruments do not satisfies the orthogonality condition and, therefore, they are not fully exogenous or are incorrectly excluded from the "main" regression.

In the context of GMM estimator, the test statistic is the minimized value of the GMM criterion function under the null hypothesis for heteroskedastic errors, the covariance matrix is estimated as $\hat{\Omega}$, which is the same as in the beginning of the GMM section n x n diagonal matrix with i^{th} diagonal element e_i^2,

$$J(\beta_{pGMM}) = eZ^T(Z^T\hat{\Omega}Z)^{-1}Ze^T \sim^A \chi^2_{l-k} \tag{3.22}$$

For the 2SLS estimator, the statistic is typically calculated as $N * R^2$ from a regression of the IV residuals on the full set of instruments.

3.3.5 Test of Endogeneity

Test of endogeneity in regression estimated via instrumental variables has null hypothesis which states that an OLS estimator of the same equation would yield consistent estimates; any endogeneity among the regressors would not derogate quality of OLS estimates, hence only efficiency should be lost by turning to 2SLS and estimates should be qualitatively unaffected. A rejection of the null indicates that endogenous regressors' effects on the estimates are meaningful, and instrumental variables techniques are required.

This Durbin-Wu-Hausman (DWH) test is numerically equivalent to the standard Hausman test, which compares two different estimators and their resulting coefficient vectors. A quadratic form in the differences between the two coefficient vectors, scaled by the precision matrix, gives rise to a test statistic for the null hypothesis that the OLS estimator is consistent and fully efficient.

Denote $\hat{\beta}^c$ the estimator that is consistent under both the null and the alternative hypotheses and $\hat{\beta}^e$ the estimator that is fully efficient under the null but inconsistent if the null is rejected. Then the Hausman takes quadratic form

$$H = n(\hat{\beta}^c - \hat{\beta}^e)^T (V(\hat{\beta}^c) - V(\hat{\beta}^e))^{-1} (\hat{\beta}^c - \hat{\beta}^e), \tag{3.23}$$

where $V(\hat{\beta})$ denotes a consistent estimate of the asymptotic variance of β.

A Hausman statistic for a test of endogeneity in an 2SLS regression is done by choosing OLS as the efficient estimator $\hat{\beta}^e$ and 2SLS as the inefficient but consistent estimator $\hat{\beta}^c$. The test statistic is distributed as Chi-squared with m degrees of freedom, where m is the number of regressors specified as endogenous in the original instrumental variables regression. The test is perhaps best interpreted not as a test for the endogeneity or exogeneity of regressors, but rather as a test of the consequence of employing different estimation methods on the same equation.

If the Hausman statistic is formed using the 2SLS estimate of the error variance, then matrix $(V(\hat{\beta}^c) - V(\hat{\beta}^e))$ becomes

$$\sigma^2_{2SLS}((X^T P_Z X)^{-1} - (X^T X)^{-1}). \tag{3.24}$$

3.4 The Generalized Method of Moments

In this section we will take a look at a general method for obtaining estimations. This method is the Generalized Method of Moments (GMM). This technique consists of universal process after which special cases of estimations arise. It is also possible to use GMM for estimation of previously mentioned OLS.

Each model is usually represented by data-generating processes (DGP), which has actually generated the data. Every process is characterized by a parameter vector β. In the beginning of this method we need to specify set of moment conditions which involve data and unknown parameters. Therefore, these functions depend on the generated data and the parameters. In case, they are estimated at the parameters that correspond to the DGP, their expectation has to be zero[4].

[4]In the classic method of moments population mean is estimated when y is iid with sample mean μ

$$E[y - u] = 0.$$

As was mentioned earlier in OLS model, the error terms have mean zero. Although, the errors are not observed, because the true parameters β are unknown, we can define residuals as a function of observed data and unknown parameters $u_i(\beta) \equiv y_i - X_i\beta$. These functions are used for development of GMM estimation. If residuals are estimated with true vector parameter β, they have mean zero, but with $\hat{\beta}$ different, this condition will not hold. This fact is used for estimating the parameter vector $\hat{\beta}$.

3.4.1 GMM in CLR model

Let us use this approach in a CLR model $y = X\beta + u$, and $E(u^T u) = \Omega$ with n observations and covariance matrix Ω with dimension n x n. Some of k explanatory variables from matrix X (size(X)=n x k), might be endogenous. It is assumed that instrumental variables which explain the endogenous regressors exist and are observed. Therefore, we obtain the matrix of instrumental variables noted as Z. This matrix satisfies the condition (3.14), instrumental variables are independent of error term. Consequently, this assumption yields for all $i = 1, \ldots, n$,

$$E(Z_i^T(y_i - X_i\beta) = 0. \tag{3.25}$$

Any predetermined column of matrix X is also column of Z. Additionally, for all $i, j = 1, \ldots, n$, it yields $E(u_i u_j | Z_i, Z_j) = \omega_{ij}$, where ω_{ij} is the element of matrix Ω.

By replacing mean operator $E[.]$ with the average we obtain corresponding sample moment

$$\frac{1}{N}\sum_{i=1}^{N}(y_i - \mu) = 0.$$

Putting μ on the right side yields an estimator $\hat{\mu} = \frac{1}{N}\sum_i y_i$. This is a classic case of method of moments for population mean.

Therefore,

$$\begin{aligned} Var(\frac{1}{\sqrt{n}}Z^Tu) &= \frac{1}{n}E(Z^Tuu^TZ) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}E(u_iu_jZ_i^TZ_j) \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}E(E(u_iu_jZ_i^TZ_j|Z_i,Z_j)) \\ &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}E(\omega_{ij}Z_i^TZ_j) \\ &= \frac{1}{n}E(Z^T\Omega Z). \end{aligned} \tag{3.26}$$

Equation (3.25) is a set of moment conditions as was defined earlier for derivation of methods of moments. Each of these theoretical conditions corresponds to a sample moment, empirically gained from data

$$\frac{1}{n}\sum_{i=1}^{n}Z_{ij}^T(y_i - X_i\beta) = \frac{1}{n}z_j^T(y - X\beta),$$

where j=1,. . . ,l and z_j is the j^{th} column of matrix Z. When dealing with just-identified model, what is the model where number of endogenous regressors is equal to number of instruments, we can set these sample moments equal to zero and afterwards solve the resulting system of k equations, which yields IV estimator (3.19). But in a more common situation when $l > k$ another way is needed to be found.

Let us denote J the l x k matrix with full column rank k. If we consider the estimator obtained by using k columns of ZJ as instruments, we get estimator which solves k equations

$$J^TZ^T(y - X\beta) = 0. \tag{3.27}$$

These moment conditions are also sometimes named orthogonality conditions as it is easy to see vector of residuals is required to be orthogonal to the columns of ZJ.

Consider the estimation error of the vector of coefficients. It is

$$\hat{\beta} - \beta = (J^TZ^TX)^{-1}J^TZ^Tu. \tag{3.28}$$

If $\hat{\beta}$ is consistent, (3.28) goes to 0 when we send $n \to \infty$. To be able to use asymptotic theory we need to add well-chosen powers of n. Consequently, analysis with estimator $\hat{\beta}$ which solves (3.27)

$$n^{1/2}(\hat{\beta} - \beta) = (n^{-1} J^T Z^T X)^{-1} n^{-1/2} J^T Z^T u. \tag{3.29}$$

Using result from (3.26) and substituting $Z = ZJ$ we find asymptotic covariance matrix of $\hat{\beta}$ is

$$(\text{plim}_{n\to\infty} \frac{1}{n} J^T Z^T X)^{-1} (\text{plim}_{n\to\infty} \frac{1}{n} J^T Z^T \Omega Z J)(\text{plim}_{n\to\infty} \frac{1}{n} X^T Z J)^{-1}. \tag{3.30}$$

To reduce this sandwich form we need to choose proper estimation of matrix J to minimize the covariance matrix (3.30). The simplest choice which will be useful and eliminates the form is

$$J = (Z^T \Omega Z)^{-1} Z^T X. \tag{3.31}$$

Hence, using (3.31) in (3.30) yields

$$(\text{plim}_{n\to\infty} \frac{1}{n} X^T Z (Z^T \Omega Z)^{-1} Z^T X)^{-1} (\text{plim}_{n\to\infty} \frac{1}{n} X^T Z (Z^T \Omega Z)^{-1} \ldots$$
$$\ldots Z^T \Omega Z (Z^T \Omega Z)^{-1} Z^T X)(\text{plim}_{n\to\infty} \frac{1}{n} (X^T Z (Z^T \Omega Z)^{-1} Z^T X)^{-1}$$
$$= (\text{plim}_{n\to\infty} \frac{1}{n} X^T Z (Z^T \Omega Z)^{-1} Z^T X)^{-1}.$$

The efficient GMM estimator is

$$\hat{\beta}_{GMM} = (X^T Z (Z^T \Omega Z)^{-1} Z^T X)^{-1} X^T Z (Z^T \Omega Z)^{-1} Z^T y. \tag{3.32}$$

So far we have been working with a strong assumption that covariance matrix Ω is known and the errors are restricted to be homoskedastic. Then $\Omega = \sigma^2 I$ and $Z^T \Omega Z = \sigma^2 Z^T Z$. Useful generalization is that this covariance matrix appears only in l x l matrix product $Z^T \Omega Z$ as in (3.31) and (3.32). But this generalization does not take possible heteroskedasticity into account. Therefore, we want to work with heteroskedasticity-robust standard errors. Homoskedasticity assumption might be limiting in our case. It speaks about equality of variance of error terms based on regressors. Income is different depending on regions, in which is household situated, therefore discrepancy of those earnings is higher and also variance of residuals might grow based on the regions.

Therefore, we would like to obtain the heteroskedasticity consistent estimate for the covariance matrix $\hat{\Omega}$. For this estimation we need a preliminary consistent estimate of the parameter vector β, which are used to obtain consistent estimates of the error terms.

Those estimates of the parameter vector β needs to be consistent, but not asymptotically efficient. This is done by choosing proper form of weightening matrix W instead of $(Z^T \Omega Z)^{-1}$, one choice is simply $W = (Z^T Z)^{-1}$. Consequently, the preliminary estimator will be the generalized IV estimator (3.21). The estimates of $\hat{\beta}$ are then used to evaluate the residuals $e_i \equiv y_i - X\hat{\beta}$. Hence, the typical element of the matrix $\frac{1}{n} Z^T \Omega Z$ can be estimated by

$$\frac{1}{n} \sum_{i=1}^{n} e_i^2 Z_{ij} Z_{ik}. \tag{3.33}$$

The covariance matrix with element (3.33) can be rewritten as $\frac{1}{n} Z^T \hat{\Omega} Z$, where $\hat{\Omega}$ is an n x n diagonal matrix with i^{th} diagonal element e_i^2. Then the GMM estimator obtained with use of the preliminary estimator is

$$\hat{\beta}_{pGMM} = (X^T Z (Z^T \hat{\Omega} Z)^{-1} Z^T X)^{-1} X^T Z (Z^T \Omega Z)^{-1} Z^T y. \tag{3.34}$$

It is easy to see that (3.34) is the (3.32) only with $\hat{\Omega}$ instead of Ω. Since $\frac{1}{n} Z^T \hat{\Omega} Z$ estimates consistently $\frac{1}{n} Z^T \Omega Z$, the $\hat{\beta}_{pGMM}$ is asymptotically equivalent to (3.32), therefore, statistics used for testing (3.32) are also valid for (3.34).

3.4.2 The GMM criterion function

Finding least squares estimates is more difficult in general cases, we need to minimize the sum of squared residuals function with respect to parameter β. Since this function is not generally quadratic as it was in OLS estimator case (3.6) there is not analytic solution. Therefore we need smooth criterion function $Q(\beta)$, which will be consequently minimized.

Criterion function in GMM estimation is

$$Q(\beta, y) \equiv (y - X\beta)^T Z(Z^T \Omega Z)^{-1} Z^T (y - X\beta), \tag{3.35}$$

after multiplying we get

$$\begin{aligned}
&(y - X\beta)^T Z(Z^T \Omega Z)^{-1} Z^T (y - X\beta) \\
&= y^T Z(Z^T \Omega Z)^{-1} Z^T y - y^T Z(Z^T \Omega Z)^{-1} Z^T X\beta \\
&\cdots - \beta^T X^T Z(Z^T \Omega Z)^{-1} Z^T y + \beta^T X^T Z(Z^T \Omega Z)^{-1} Z^T X\beta.
\end{aligned}$$

Minimizing this criterion function with respect to vector of parameters β yields first-order condition

$$\begin{aligned}
-2X^T Z(Z^T \Omega Z)^{-1} Z^T y + 2X^T Z(Z^T \Omega Z)^{-1} Z^T X\beta = 0 \\
X^T Z(Z^T \Omega Z)^{-1} Z^T (y - X\beta) = 0,
\end{aligned}$$

which is the efficient GMM estimator (3.32).

However, there is a problem with the GMM instrumental variable approach if associated with weak instruments. In those cases finite sample bias can occur and transfer this difficulty into GMM. There is no well-established test for weak instruments, therefore results estimated by GMM might not be the best out of possible.

Chapter 4

Empirical results

We would like to present empirical results of our work in this chapter. For all estimates OLS, Instrumental variable GMM or 2SLS is used with help of statistical software STATA/SE version 10.1. Commands in STATA were *regress* for OLS, *ivregress gmm* for GMM and *ivreg2* for 2SLS. We are going to proceed in the following order.

4.1 Strategy

Firstly, we estimate basic consumption function (2.1) developed by Keynes. Then treat disposable income according to Davidson and MacKinnon [5] as previously mentioned as endogenous variable. Consequently, use of 2SLS is desirable to estimate equation to obtain results and deal with endogeneity which occurs because of imperfection of observed data, together with reverse causality as was mentioned earlier. The reverse causality might be linked with income following consumption. Hence, people are searching for a job which will provide them with financial support and stability to fulfill their consumption needs. Addtionaly, consumer might have determined level of consumption for which he or she needs/wants to obtain resources. Let us present this idea on an example of a cab driver. He or she might have set a level of income for each day and after gaining this amount driver is satisfied and not working any more. On the other hand, if experiencing unsuccessful day, he or she will try to find customers to fulfill daily revenue. More frequently this backward pattern should be associated with side jobs and irregular income which is earned from them, however, it is difficult to know whether respondents denote their irregular income in questionnaires to be regular. Therefore, we focus on regressing income with instrumental variables throughout this part to obtain unbiased results. We try to rationalize use of each instrument and use various age, head's work status and single household samples to study consumption pattern.

Secondly, we are differentiating disposable income into two categories; regular income and irregular income and continue with finding applicable variables for regression. Irregular income is defined as irregular income from employment or bonuses and royalties. Regular income is disposable income reduced by irregular income.

Last but not least, we try to confront results of Browning and Collado [2] and Hsieh [12], who found out that an agent in normal times tend to smooth income with large and anticipated income changes, and spend smaller ones. Therefore, we would assume, based on their results, that small irregular income in proportion to regular income would be transferred into consumption in the current month, on the other hand, large change will be used to smooth consumption in the following long-term horizon, hence, most of the income from irregular income should be saved. We have changed our approach in this specification. In this case, income is differentiated into three categories:

- Regular income- regular income of household;
- Significant irregular income- irregular income, which accounts for more than 33% of regular income;
- Insignificant irregular income- irregular income, which is equal or smaller than 33% of regular income.

4.2 Basic Consumption Function

In this part we try to analyze findings from estimation of results of the Keynes' consumption function extended by other regressors to disposable income.

Table 4.1: Basic consumption function

est: Consumption | method	OLS		GMM		2SLS			
Stage:					1 stage[1]		2 stage[2]	
Disposable income	0.745	***	0.865	***			0.867	***
	(0.011)		(0.028)	e			(0.028)	e
Number of kids	0.030	***	0.020	***	0.057	***	0.020	***
	(0.007)		(0.007)		(0.009)		(0.007)	
Education	0.088	***	0.052	***	0.219	***	0.050	***
	(0.015)		(0.017)		(0.017)		(0.017)	
Age	0.014	***	0.008	***	0.041	***	0.008	***
	(0.002)		(0.003)		(0.003)		(0.003)	
Age sq	-0.000	***	-0.000	***	0.000	***	0.000	***
	(0.000)		(0.000)		(0.000)		(0.000)	
Results of the first stage equation for instruments for endogenous regressor - disposable income								
Leading job					0.175	***		
				iv	(0.049)	iv		
Public sector					0.221	***		
				iv	(0.015)	iv		
Gender					0.382	***		
				iv	(0.016)	iv		
Big town					0.062	***		
				iv	(0.017)	iv		
Number of observations	4710		4710		4710		4710	
Wald χ^2/F-statis[3]	1430.09		3549.27		740.45		708.48	
p-value of Hansen J statistics			0.51		0.51		0.51	
p-value of DWH statistics			0.00		0.00		0.00	
F test of excluded instrument:					214.23			
R^2	0.60		0.59		0.48		0.59	

Note: Coefficients report the average marginal probability effects. ***,**, and * denote significance at 1 percent, 5 percent, 10 percent level, respectively. e-Variable taken as endogenous in regression estimated by instrumental variable. iv- Instrumental variable. [1]- First stage OLS for the endogenous regressor- Disposable income. [2]- Second stage for endogenous variable- consumption. [3]-F-statistics for 2SLS, Wald for GMM. Used Robust option in STATA command to report SEs and statistics that are robust to the presence of heteroskedasticity in the error terms.

As can be seen from table 4.1 we have used all three methods for the estimation. Tables are going to be organized similarly with four parts separated by line. In the first part the income regressor is situated, which is treated as endogenous in GMM and 2SLS regressions. In the second box exogenous regressors for main regression are arranged. In the third section instrumental variables for GMM and 2SLS are selected, for 2SLS we also have statistics from the first step of regression. However, we will simplify 2SLS section in the following tables, where results from the first stage for exogenous regressors will not be covered, since their added value is limited. Therefore, results from the first stage for instrumental variables and second stage for exogenous regressors are illustrated in one column as in Table 7.5. In the last part statistics as number of observations, Wald χ^2 for GMM and F-statistics for OLS and 2SLS estimations, p-value of Hansen J statistics of overidentifying restrictions, p-value of Durbin-Wu-Hausman test of endogeneity, together with F test of excluded instrument in the first stage of 2SLS and R^2 are recorded. We use robust option in STATA command to report standard errors and statistics that are robust to the presence of heteroskedasticity in the error terms.

All the regressors are significant and results from GMM and 2SLS are basically equal, however, OLS estimates are biased downwards and hence we could assume from discussion in Section 3.2 that influence of the measurement error is stronger than the reverse causality, therefore endogeneity of disposable income might be caused by imperfection of questionnaires and their estimates of income, which might be biased because of life-cycle access and future uncertainty. Value of Hansen's J statistics indicates that instrumental variables are well-chosen and rejection of the null in DWH tells that endogenous regressors' effects on the estimates are meaningful, and instrumental variables techniques are required. Therefore, we assume 2SLS results to be the most appropriate with disposable income instrumented by four variables Leading job, Gender, Big town and Public sector. Reason why we take these variables is based on histograms shown in Appendix in Figure 7.1 and assumptions driven for those variables explained in the following part.

The first instrument considers whether respondents are working in public sector or not. It is surprising that working in public sector is based on this variable better off, as can be seen dispersion of disposable income in public sector is lower and nicely distributed. Additionally, public sector provides usually secure job with social benefits and steady wage increase and strong position of unions. Since high proportion of population is employed in public sector it provides also potential earnings and opportunities to find job for persons which are not able to be employed in private sector.

The second instrument is gender of family head. As can be seen there is major difference between subsets of men and women as heads of the households. In our dataset households with female heads have only 66% of disposable income of those where heads are men. This difference is supported by gender wage gap. Gender wage gap is according to Simón [21] very significant in Slovakia, particularly 31.3%. In average women earn 31% less than men in the same profession. Therefore, we assume that in households where woman is taken as head of a family, which means she should have higher income than man in this household and still be earning less in average based on the wage gap, disposable income of whole family is lower and gender of a head should, consequently, be strong instrument.

The third instrument is living in one of the eight regional cities (Bratislava, Trnava, Trenčín, Nitra, Žilina, Banská Bystrica, Prešov, Košice). There is not strong contrast in two types of placement of household, but living in big city is usually associated with better opportunities to find employment, forasmuch as unemployment in big cities is lower than in rural areas and smaller towns and disposable income of family is higher, since business centers and bigger entrepreneurial activities are situated in cities with higher population. Moreover, typical city jobs as auditors, consultants together with government sector are dominantly matter of those cities.

The last instrument is leading position, this dummy variable equals one, if head or husband or wife works as legislator, manager or team-leader. All those types of employment falls into better paid jobs, therefore shape of histogram is expected.

From 2SLS results we can deduce that households spend 87% of income growth; it is convenient to assume that households live on tight budget, since major part of growth is spent and only 13% can be transferred into savings. Age effect and effect of number of kids are not economically interesting. Based on effects of age and age squared every 10 years household adds or cuts 2% of consumption and every child adds 2%. Impact of college education of head or wife or husband is estimated to be 5.0%. Although, a more natural way might be seen in including education into the set of instruments, we believe that there is reasonable explanation into putting it into main regression and hence using it as a regressor for consumption. Education should influence level of income, based on Ferrer and Riddell [6] this impact should be approximately 9% for females and 6% for males, supporting the well-known quote of Benjamin Franklin, who said that an investment in knowledge pays the best interest. Although, the strong positive relationship between education and earnings is one of the most well-established relationships in social science, there is reluctance among the scientists to interpret this correlation as evidence that education exerts a causal effect on earnings. They argue that schooling raises earnings because it enhances workers' skills, hence making employees more productive and valuable to employers. However, the positive relationship between earnings and schooling could arise because both these factors are correlated with unobservable factors such as ambition, perseverance, etc. Generally, person who is highly motivated may be more likely successful, even in the absence of education. Therefore, we do not include education into set of instrumental variables, since our instruments are based on Hansen's statistics well-chosen, but into main regression. Rationality behind rise in consumption caused by education might be seen in force driven by environment. If person is employed and has college education, he or she might be under pressure of surroundings. Exposed to factors, which will not occur without college education; e.g., roommate bought new laptop, econometric book, or all colleagues are heading to ski trip, therefore, environment and personal intention will drive the agent to work more to be able to consume products and services he or she will otherwise not.

Additionally, we use the subsamples of various age groups to take a look on different consumption patterns throughout lifecycle together with subsample of households with one member. In the Appendix in the Table 7.5 results for simple consumption function without other exogenous regressors are displayed with disposable income treated as endogenous variable and in the Table 4.2 are shown results for specific models with additional exogenous variables as number of kids and education of head of a family or husband or wife. Since results from 2SLS and GMM are in this case very similar, we have decided to project only 2SLS in simplified version, where results for instruments from the first stage are noted in the same column with results for exogenous regressors from the second stage.

Table 4.2: Basic consumption function for different age groups and single households

est: Consumption	(1)	(2)	(3)	(4)	(5)	(6)
Sample- age/single:	<35	35-44	45-54	55-64	>64	Single
Disposable income	1.024 ***	0.712 ***	0.846 ***	0.806 ***	1.074 ***	0.816 ***
	(0.090) e	(0.066) e	(0.066) e	(0.047) e	(0.058) e	(0.170) e
Number of kids	-0.007	0.034 ***	0.033 **	0.066 ***	-0.023	
	(0.018)	(0.012)	(0.015)	(0.024)	(0.049)	
Education	0.035	0.065 *	0.053 *	0.058	0.034	0.089
	(0.046)	(0.036)	(0.033)	(0.039)	(0.040)	(0.068)
Age						0.012 *
						(0.006)
Age sq						0.000 ***
						(0.000)
Results of the first stage equation for instruments for endogenous regressor - disposable income						
Leading job	0.201 **	0.099	0.147 *	0.217 **		0.261 *
	(0.101) iv	(0.087) iv	(0.083) iv	(0.104) iv		(0.161) iv
Public sector	0.191 ***	0.180 ***	0.156 ***	0.337 ***		0.345 ***
	(0.035) iv	(0.028) iv	(0.027) iv	(0.034) iv		(0.052) iv
Gender	0.357 ***	0.310 ***	0.344 ***	0.443 ***	0.459 ***	0.034
	(0.048) iv	(0.041) iv	(0.035) iv	(0.032) iv	(0.076) iv	(0.038) iv
Big town	0.024	0.141 ***	0.050	0.010	0.079 ***	0.074 **
	(0.041) iv	(0.040) iv	(0.032) iv	(0.040) iv	(0.030) iv	(0.034) iv
Number of observations	694	984	1187	961	884	763
F-statistics	62.75	85.90	111.93	164.59	165.74	63.26
p-value of Hansen J stat.	0.73	0.47	0.75	0.67	0.65	0.57
p-value of DWH statistics	0.00	0.90	0.03	0.04	0.00	0.74
F test of excluded instr.:	23.43	28.43	36.94	88.22	183.09	18.95

Note: Coefficients report the average marginal probability effects. ***,**, and * denote significance at 1 percent, 5 percent, 10 percent level, respectively. e-Variable taken as endogenous in regression estimated by instrumental variable. iv- Instrumental variable; statistics are taken from the first stage of 2SLS. Used Robust option in STATA command to report SEs and statistics that are robust to the presence of heteroskedasticity in the error terms.

Significance of living in big town on income for young households, column (2), does not come as a surprise, particularly this set should be the most influenced by decision to move or stay in regional town. Since the Velvet Revolution and early 90's, differences between rural areas and towns have risen, hence graduates or young couples who have decided to live in big town with better opportunities and probably higher salaries during those times fall in 2005 into the second age category.

Impact of college education of either head of a family or husband or wife on consumption is 6.5% for column (2) and 5.3% in column (3). As can be seen we experience propensity to consume to move in two cycles. Greater expenditures from additional income in the youngest set (1). Based on lifecycle hypothesis young consumers tend to smooth their income by consuming more than earn during early stage of work life because they assume income to rise sharply after improvement of their professional abilities and skills, consequently their current higher consumption will be balanced by savings later. Since persons establish family and pay a rent or a mortgage on recently bought residence, their consumption is promoted by loans against future income, furthermore parents might still be helping children in the beginning of their independent life. In the next stage, column (2), agents take into account future uncertainty by accumulation of precautionary savings for them and their children and do not follow present consumption target so strongly what can be seen from endogeneity test. After mid-40's, column (3), propensity to consumption is not far from the average level and again falls with being closer to the retirement age, column(4), with stronger impact of tendency to accumulate assets. From this assets retirees, column(5), dissolve capital since their consumption is higher than disposable income. Elderly tend to run accumulated assets down due to lower levels of uncertainty and increased probability of death. We have excluded leading jobs and public sector dummy from set of instruments, because only small number of observations where dummy equals 1 is in this age group. Particularly leading job is observed only in one household, therefore we face singleton dummy variable, which might cause problem in estimating covariance matrix. Nevertheless, we can see from endogeneity test, that only in the second age group, column (2), and sample of single households, column (6), regression does not benefit from switching into 2SLS since the result shows that the estimates are qualitatively unaffected. Since we experience the lowest propensity to consume within young age group in column (2), we assume that the reason might be connected with this finding. The agents do save bigger part of their additional income, because their income is during this stage of life usually peaking and they tend not to follow the current consumption so tightly.

Moreover, results for single households, column (6), look that consumers spend 82% of additional income, however, in this set age regressor plays an important role, since major part of this set consists of older people and if you look into Table 7.5 where additional regressors are excluded, propensity to consume for column (5) and (6) are similar.

If we compare results from households where head of a family is self-employed to those where he or she is an employee, we experience higher propensity to save in self-employed sample. This can be seen from results in the Table 4.3 and associated to stronger precautionary motive of self-employed heads due to higher income risk. Additionally, self-employed persons are also covered by social security scheme, hence this motive should be linked mainly to the future uncertainty not to retirement assurance. Significance of education only in employees sample might mean, as was said earlier, that consumers with college education probably share same workplace, hence their consumption might be influenced by each other. Switching into 2SLS in self-employed sample does not make big contribution to accuracy since instruments are not as strong as hoped for in this case.

Table 4.3: Basic consumption function for different work statuses

Sample- work type:	Self-employed						Employee					
est: cons | method	OLS		GMM		2SLS		OLS		GMM		2SLS	
Disposable income	0.683	***	0.751	***	0.772	***	0.739	***	0.878	***	0.869	***
	(0.049)		(0.111)		(0.115)		(0.017)		(0.042)		(0.042)	
Number of kids	0.004		0.001		0.000		0.032	***	0.024	***	0.024	***
	(0.021)		(0.020)		(0.020)		(0.008)		(0.009)		(0.009)	
Education	0.061		0.046		0.039		0.063	***	0.043	**	0.043	**
	(0.049)		(0.057)		(0.058)		(0.021)		(0.021)		(0.021)	
Age	0.048	***	0.042	**	0.045	**	0.006		0.002		0.001	
	(0.018)		(0.018)		(0.018)		(0.006)		(0.006)		(0.006)	
Age sq	-0.001	***	-0.001	**	-0.001	**	0.000		0.000		0.000	
	(0.000)		(0.000)		(0.000)		(0.000)		(0.000)		(0.000)	
Results of the first stage equation for endogenous regressor - disposable income												
Leading job					0.174						0.194	***
				iv	(0.130)	iv				iv	(0.057)	iv
Public sector					0.145	***					0.075	***
				iv	(0.046)	iv				iv	(0.017)	iv
Gender					0.109	*					0.298	***
				iv	(0.075)	iv				iv	(0.021)	iv
Big town					0.015						0.032	*
				iv	(0.060)	iv				iv	(0.019)	iv
Nb. of observations	363		363		363		2462		2462		2462	
Wald χ^2/F-statistics[1]	47.28		94.90		17.55		426.53		709.98		134.21	
p-value of Hansen J statistics			0.65		0.65				0.30		0.30	
p-value of DWH statistics			0.54		0.54				0.00		0.00	
F-test of excluded instrument					10.45						133.15	
R^2	0.39		0.40		0.39		0.46		0.45		0.45	

Note: Coefficients report the average marginal probability effects. ***,**, and * denote significance at 1 percent, 5 percent, 10 percent level, respectively. e-Variable taken as endogenous in regression estimated by instrumental variable. iv- Instrumental variable; statistics are taken from first stage of 2SLS. Used Robust option in STATA command to report SEs and statistics that are robust to the presence of heteroskedasticity in the error terms. [1]-F-statistics for 2SLS, Wald χ^2 statistics for GMM.

4.3 Differentiated income

Our next approach is to divide income into regular and irregular to help us observe how do households use irregular income from employment, royalties and bonuses. If they use significant part of them for short term improvement of consumption or save to utilize their effect for longer period. Irregular income is, therefore, the one obtained from royalties, tantiems or irregular income from employment. Regular is equal to disposable subtracted by irregular. Results of regressions using OLS, GMM and 2SLS method are summarized in Table 4.4, where regular income in treated in GMM and 2SLS method as endogenous.

Table 4.4: Consumption function with different income types and regular income as endogenous

est: cons | method	OLS		GMM		2SLS	
Regular income	0.703	***	0.832	***	0.833	***
	(0.011)		(0.029)	e	(0.029)	e
Irregular income	0.057	***	0.061	***	0.060	***
	(0.006)		(0.008)		(0.009)	
Number of kids	0.033	***	0.022	***	0.022	***
	(0.007)		(0.008)		(0.008)	
Education	0.099	***	0.061	***	0.060	***
	(0.015)		(0.017)		(0.017)	
Age	0.016	***	0.008	**	0.009	***
	(0.003)		(0.003)		(0.003)	
Age sq	-0.000	***	-0.000	***	-0.000	***
	(0.000)		(0.000)		(0.000)	
Results of the first st. eq. for instr. for endogenous regressor - regular income						
Gender					0.397	***
				iv	(0.016)	iv
Big town					0.054	***
				iv	(0.017)	iv
Public sector					0.231	***
				iv	(0.015)	iv
Leading job					0.173	***
				iv	(0.048)	iv
Nb. of observations	4710		4710		4710	
Wald χ^2/F-statistics[1]	1102.21		3448.24		573.69	
p-value of Hansen J statistics			0.47		0.47	
p-value of DWH statistics			0.00		0.00	
F-test of excluded instrument					215.29	
R^2	0.58		0.57		0.57	

Note: Coefficients report the average marginal probability effects. ***,**, and * denote significance at 1 percent, 5 percent, 10 percent level, respectively. e-Variable taken as endogenous in regression estimated by instrumental variable. iv- Instrumental variable. [1]-F-statistics for 2SLS, Wald χ^2 statistics for GMM. Used Robust option in STATA command to report SEs and statistics that are robust to the presence of heteroskedasticity in the error terms.

This differentiation obviously resulted into similar results as we obtained in Table 4.1.

We have taken only regular income as endogenous, nevertheless in the next approach we compare these results with those in Table 4.5 where irregular income is treated also as endogenous. Results from Table 4.4 suggest that growth in household's irregular income is significant and contributes in average by 6% to consumption.

As can be seen from both tables, differences between estimations obtained by GMM and 2SLS are negligible. This might be caused by properties of instruments, which occur to be well-chosen. Although, irregular income is significant when treated as exogenous, we want to emphasize the results from the second table, where is taken as endogenous. Rationality behind treating irregular income as endogenous we see in obtaining resources for designated level of consumption. Families have regular income, however, to acquire additional means, they might participate in irregular types of activities, side jobs. Therefore, agents have certain amount of irregular income to obtain to satisfy their consumption. Thus, it should be taken care of possible endogenous effect. The only major difference between regressions with and without endogenous irregular income is in its significance.

Instrumental variables are picked with the same approach as previously and histograms for each category are shown in Figure 7.2 and Figure 7.3 in Appendix. Gender of head of a household plays significant role as instrument for regular income, but not for irregular. Since this type of income is related to short run employment, we can infer that wage gap is not significant and employer does not distinguish between sexes. Additionally, living in regional town picked as instrumental variable and its significance for irregular income should be associated with type of employment and higher occurrence of side jobs in bigger settlements. Some services are typical only for cities, where people spend less time in their households, hence cleaning services are made use of. Artists are usually living in cities, where opportunities are easily found for their creative activities, hence irregular income should cover their income. Consequently, these side jobs might be used as a source of extra income to obtain additional resources. The last instrument is dummy variable which equals one if head of a family has only casual, short-term job without contract, and it is only added to second regression, where irregular income is treated as endogenous,

Table 4.5: Consumption function with different income types and both regular and irregular income as endogenous

est: cons | method	OLS		GMM		2SLS			
Regular income	0.703	***	0.829	***	0.826	***		
	(0.011)		(0.031)	e	(0.028)	e		
Irregular income	0.057	***	0.029		0.011			
	(0.006)		(0.082)	e	(0.081)	e		
Number of kids	0.033	***	0.025	***	0.023	***		
	(0.007)		(0.008)		(0.008)			
Education	0.099	***	0.068	***	0.067	***		
	(0.015)		(0.019)		(0.019)			
Age	0.016	***	0.009	***	0.009	***		
	(0.003)		(0.003)		(0.003)			
Age sq	-0.000	***	-0.000	***	-0.000	***		
	(0.000)		(0.000)		(0.000)			
Results of the first stage equation for endogenous regressors					- regular income		- irregular income	
Gender					0.387	***	0.005	
				iv	(0.016)	iv	(0.031)	iv
Big town					0.049	***	0.158	***
				iv	(0.017)	iv	(0.041)	iv
Public sector					0.240	***	0.024	
				iv	(0.015)	iv	(0.036)	iv
Leading job					0.136	***	-0.019	
				iv	(0.050)	iv	(0.111)	iv
Casual work					0.164	***	-0.084	*
				iv	(0.024)	iv	(0.043)	iv
Nb. of observations	4710		4710		4710		4710	
Wald χ^2/F-statistics[1]	1102.21		3259.28		568.21		568.21	
p-value of Hansen J statistics			0.17		0.32		0.17	
p-value of DWH statistics			0.00		0.00		0.72	
F-test of excluded instrument					187.60		10.18	
R^2	0.58		0.57		0.57		0.57	

Note: Coefficients report the average marginal probability effects. ***,**, and * denote significance at 1 percent, 5 percent, 10 percent level, respectively. e-Variable taken as endogenous in regression estimated by instrumental variable. iv- Instrumental variable. [1]-F-statistics for 2SLS, Wald χ^2 statistics for GMM. Used Robust option in STATA command to report SEs and statistics that are robust to the presence of heteroskedasticity in the error terms.

since this dummy variable is significant for irregular income. Table 4.5 is organized differently, while two regressors are treated as endogenous. We have included results from first-stage equation from 2SLS for regular and irregular income into third part of the table.

Although, p-value of DWH statistics indicates that contribution of switch between OLS and 2SLS is not important, insignificance of irregular income states otherwise, but better instruments should be fined to obtain more trustworthy results. We may deduce that irregular income does not play important role in consumers consumption, which might be also caused by the fact that bonuses and royalties are not so frequently disbursed.

4.4 Further Analysis

Our motivation was to study assumption of Browning and Collado [2] and Hsieh [12], where big irregular income is not consumed, however agents are excessively sensitive to much lower income from tax refunds. We have divided irregular income into two different types. The first one is income, where fraction of irregular on regular is higher than 33% and the second is smaller. Nevertheless, we were unable to obtain statistical relevant results since irregular income is observed only in less than 200 observations and this differentiation does diminish our ability to estimate good results.

We were unable to find evidence due to lack of households, where royalties and bonuses are paid. This might be Slovakian specific, because royalties and bonuses are not so usually paid out, or it is a problem of questionnaires and ability of people to distinguish between regular and irregular part of their salary. Although, their paycheck hold information about core and variable fraction of salary, people do not make a difference after being paid, hence they observe only nominal amount which finds its way to their account.

Consequently, we recommend further research of this hypothesis, which might be done based on following releases of another datasets for another year and linking them together to obtain more observations for irregular income.

Chapter 5

Conclusion

In this work we have studied Slovakian households and their consumption behavior. Treating endogeneity of income by use of instrumental variables we have obtained results which indicate that consumption of average Slovakian family moves in two cycles. The first one starts with overconsumption of young families connected with smoothing this overconsumption later after possible improvement of professional background and lesser effect of children on consumption, since they enter late teenage years. This is followed by securing for retirement by accumulating capital and finally repeated overconsumption by retirees in a sense they consume more than earn, dissolving capital since future uncertainty is lower. Additionally, taking type of employment into account has yielded results with possible stronger precautionary motive for self-employed heads. Since we have worked with cross-sectional data we were unable to analyze development of wealth to income ratio by which precautionary savings are usually illustrated, however it can be noted that self-employed households tend to save more due to higher income risk of their work status.

Differentiating and treating both partitions of income; irregular and irregular, led into finding that irregular income is not statistically significant for consumption. Therefore, households include only regular part into their decision about consumption.

Attempt to confront Hsieh's assumption by differentiating irregular income on significant and insignificant did not lead to applicable results due to lack of observations on irregular income, which might be associated with inability to fill questionnaires reasonably or Slovakian atypism of not paying bonuses as often as is typical for the western countries. Consequently, we would recommend further study with extended dataset by more recent data, where intertemporal changes of consumption and saving behavior might be recorded. We believe that this research might be done by linking datasets together based on similar household characteristics to obtain panel data to focus on those changes. Together with wider base for possible study of consumption of irregular income to confront results of Hsieh and Browning and Collado.

Chapter 6

Résumé

V našej práci študujeme spotrebu domácností na Slovensku. Keďže spotreba domácností zastupuje podstatnú časť hrubého domáceho produktu krajiny, zmeny správania spotrebiteľov počas ich životného cyklu poskytujú informácie pre analytikov a politikov ako hodnotiť vplyv možných zmien.

Našim cieľom je preskúmať spotrebu domácností na Slovensku, pričom sa zameriavame na rôzne skupiny domácností vzhľadom na vek hlavy rodiny ako aj typu práce. Endogenita príjmov v regresii je zohľadnená použitím inštrumentálnych premenných a následne dvojstupňovej metódy najmenších štvorcov.

Práca je rozdelená nasledovne. V prvej kapitole sa zameriavame na pozadie empirickej práce, predstavujeme základnú funkciu spotreby odvodenú Keynesom, na ktorú nadväzujú ďalšie prístupy. Ďalej z literatúry, ktorá sa venuje spotrebe vyvodzujeme závery, ktoré by mali platiť pre naše modely, taktiež približujeme čitateľovi naše dáta vďaka deskriptívnym štatistikám, pomocou ktorých predstavujeme zloženie domácností na Slovensku. V druhej časti sa sústredíme na ekonometrické metódy použité v empirickej časti. Na úvod predstavujeme klasickú regresiu pomocou metódy najmenších štvorcov spolu s možnými problémami, ktoré môžu viesť k vychýleniu výsledkov tejto metódy. Následne sa snažíme zdôvodniť použitie dvojstupňovej metódy najmenších štvorcov pre endogenitu v regresoroch. S touto endogenitou sa vysporiadavame pomocou inštrumentálnych premenných, kde podmienky na ich použitie sú dané. Taktiež sa sústredíme na test

prílišných reštrikcií a endogenity na ohodnotenie správnosti použitia inštrumentálnych premenných ako aj prechodu z metódy najmenších štvorcov na dvojstupňovú metódu najmenších štvorcov. V poslednej časti sa dostávame k výsledkom empirickej časti. V nej sú rozobrané modely spotreby domácností pre rôzne špecifikácie domácností ako aj príjmov. Výsledky v tabuľke 4.2 naznačujú, že spotreba priemernej rodiny sa pohybuje v dvoch cykloch. Prvý začína pre mladé rodiny, stĺpec (1), keď spotreba prevyšuje príjmy, teda vyhladzovanie spotreby ide na úkor budúcich príjmov, ktoré sú očakávané vyššie z dôvodu rastu pracovných skúseností v ďalšej fáze spolu s nižšími výdavkami na deti pri ich dospievaní a osamostatňovaní. Nasleduje fáza, kde spotrebitelia berúc do úvahy neistotu akumulujú úspory pre seba, prípadne deti. Od kategórie domácností s vekom medzi 40 a 50 začína na sile naberať zabezpečovanie na dôchodok a teda akumulovanie kapitálu, ktorým je opäť vyššia spotreba ako príjem v dôchodkovom veku zapezpečovaná. Dôchodcovia tento kapitál využívajú, keďže neistota z budúcnosti je nižšia. Pri zohľadnení či je hlava domácnosti samostatne zárobkovo činná osoba (SZČO) alebo zamestnanec, tabuľka 4.3, dostávame silnejší motív akumulovania bezpečnostných úspor pre samostatne zárobkové osoby, čo je spojené s neistotou v ich príjmoch. Motív akumulovania úspor by mal byť odôvodnený iba neistotou, keďže SZČO taktiež platia odvody a sporia na dôchodok. Keďže máme k dispozícii iba cross-sectional dáta nemôžeme skúmať vývoj podielu bohatstva k príjmu, ktorého vývojom sa obyčajne bezpečnostné úspory analyzujú, avšak rozdelením domácností podľa typu činnosti dostávame tento motív silnejší u SZČO.

Rozdelením príjmu na pravidelný a nepravidelný, tabuľka 4.4 a 4.5, dostávame výsledok, v prípade, že obe časti berieme ako endogénne, že nepravidelný príjem nie je štatisticky signifikantný, teda nezohráva pri rozhodovaní domácností o spotrebe dôležitú úlohu. Pri pokuse o overenie hypotézy vyplývajúcej z práce Hsieh [12] a Browning a Collado [2] rozdelením nepravidelného príjmu na väčší ako tretina pravidelného a menšieho sme narazili na problém nedostatku pozorovaní. Z týchto prác vyplýva, že očakávaný veľký nepravidelný príjem by mal byť využitý na upravenie spotreby v dlhodobom horizonte, a na druhej strane menší očakávaný nepravidelný príjem by mal byť spotrebovaný v aktuálnom období. Bonusy, prípadne odmeny nie sú až tak časté na Slovensku prípadne respondenti ich zahrnuli v dotazníku do pravidelného príjmu, z tohto dôvodu by sme odporúčali pri ďalších vydaniach dát tieto spojiť a skúmať či sa spotrebitelia naozaj správajú podľa tohto predpokladu.

Bibliography

[1] Christopher F. Baum, Mark E. Schaffer and Steven Stillman *Instrumental Variables and GMM: Estimation and Testing* Boston College, Department of Economics, 2003

[2] Martin Browning and M. Dolores Collado, *The Response of Expenditures to Anticipated Income Changes: Panel Data Estimates*, The American Economic Review, Vol. 91, 2001

[3] Martin Browning and Annamaria Lusardi, *Household Saving: Micro Theories and Micro Facts*, Journal of Economic Literature, Vol. 34, 1996

[4] A. Colin Cameron and Pravin K. Trivedi, *Microeconometrics, Methods and Applications*, Cambridge University Press, 2005

[5] Russell Davidson and James G. MacKinnon *Econometric theory and methods*, Oxford University Press, 2002

[6] Ana Ferrer and W. Craig Riddell, *The Role of Credentials in the Canadian Labour Market*, Canadian Journal of Economics 35, 2002

[7] Pierre-Oliver Gourinchas and Jonathan A. Parker, *Consumption Over the Life Cysle*, Econometrica, Vol. 70, 2002

[8] William H. Greene *Econometric Analysis*, Prentice Hall, 2002

[9] Luigi Guiso, Tullio Jappelli and Daniele Terlizzese, *Earnings uncertainty and precautionary saving*, Journal of Monetary Economics 30, 1992

[10] Lars Peter Hansen, *Large Sample Properties of Generalized Method of Moments Estimators*, Econometrica, Vol. 50, 1982

[11] Fumio Hayashi *Econometrics*, Princeton University Press, 2000

[12] Chang-Tai Hsieh, *Do Consumers React to Anticipated Income Changes? Evidence from the Alaska Permanent Fund, The American Economic Review*, Vol. 93, 2003

[13] Tullio Jappelli and Luigi Pistaferri, *The Consumption Response to Income Changes*, The Annual Review of Economics, 2010

[14] Mark Karazosian, *Precautionary Savings - A Panel Study*, The Review of Economics and Statistics Vol. 79, 1997

[15] Peter Kennedy *A Guide to Econometrics, 3rd Edition*, The MIT Press, 1996

[16] John Maynard Keynes, *The General Theory of Employment, Interest and Money*

[17] Annamaria Lusardi, *Precautionary saving and subjective earnings variance*, Department of Economics, Dartmouth College, Hanover, 1997

[18] Jonathan A. Parker, *The Reaction of Household Consumption to Predictable Changes in Social Security Taxes*, The American Economic Review, 1999

[19] J. D. Sargan, *The Estimation of Economic Relationships using Instrumental Variables*, Econometrica, Vol. 26, 1958

[20] John Shea, *Union Contracts and the Life-Cycle/Permanent-Income Hypothesis* The American Economic Review, Vol. 85, 1995

[21] Hipólito Simón, *The Gender Pay Gap in Europe: An International Comparison with Matched Employer-Employee Data*, Universidad de Alicante

[22] Jeffrey Wooldridge *Introductory Econometrics: A Modern Approach, 2nd Edition*, South-Western College Publishing, 2002

[23] National Bank of Slovakia,
http://www.nbs.sk/en/home

[24] Organization for Economic Co-opearation and Development
http://www.oecd.org/home/0,2987,en_2649_201185_1_1_1_1_1,00.html

[25] Statistical Office of the Slovak Republic,
http://portal.statistics.sk/showdoc.do?docid=359

[26] United Nations Economic Commission for Europe
http://www.unece.org/

[27] ABCnews, *Visa predicts divorce*
http://abcnews.go.com/Business/visa-predicts-divorce/story?id=10320638

Chapter 7

Appendix

Table 7.1: Descriptive statistics for whole sample

Variable	Mean	Std. deviation
Consumption	696.21	(438.45)
Disposable income	742.18	(426.05)
Regular income	727.57	(415.42)
Irregular income	8.04	(57.54)
Number of kids	0.64	(0.94)
Size	2.93	(1.42)
Age	50.76	(14.28)
Head has college education	0.13	
Head or hus/wife has coll. ed.	0.17	
Living in regional town	0.22	
Gender of head	0.69	
Leading job	0.02	
Public sector	0.18	
Unstable job	0.10	

Note: Values are in EUR computed from SKK in conversion rate EUR 1=30.126 SKK

Table 7.2: Descriptive statistics for different groups

Variable		Disposable income		Consumption		
		Mean	Std. deviation	Mean	Std. deviation	Obs.
Region	BA	780.1	(460.31)	739.2	(490.28)	592
	TT	742.5	(429.72)	695.0	(412.19)	588
	TN	712.6	(441.49)	639.9	(325.09)	589
	NR	742.6	(447.48)	713.1	(520.33)	588
	ZA	785.0	(437.75)	760.8	(485.26)	591
	BB	642.0	(368.80)	607.6	(384.29)	589
	PO	765.5	(378.80)	705.2	(406.55)	588
	KE	767.0	(419.38)	708.7	(399.31)	585
Size	1	332.3	(171.05)	353.5	(284.56)	763
	2	625.3	(296.90)	593.4	(338.43)	1297
	3	830.3	(428.67)	772.6	(406.79)	1007
	4	937.7	(400.08)	859.4	(418.58)	1077
	5	996.0	(428.75)	912.7	(502.62)	369
	6	1069.9	(423.45)	942.0	(351.17)	129
	7	1171.9	(632.01)	1128.6	(644.04)	44
	8	1170.6	(586.14)	1013.2	(685.77)	15
	9	1001.9	(635.18)	1624.3	(2154.51)	7
	10	1707.3	(662.82)	878.9	(539.28)	2
Number of kids	0	657.4	(414.95)	609.8	(400.52)	2873
	1	861.4	(414.08)	816.2	(444.54)	909
	2	885.0	(401.64)	825.8	(407.09)	729
	3	899.0	(418.22)	918.9	(695.96)	159
	4	1010.4	(393.34)	1004.9	(523.50)	38
	5	621.5	(327.53)	876.1	(725.54)	6
	6	608.8	(307.15)	570.1	(290.86)	4
	7	673.5	(42.58)	661.7	(266.19)	2
Head has college education	yes	931.2	(476.80)	877.6	(479.23)	628
	no	713.1	(410.10)	668.3	(425.09)	4082
Head or hus/wife has coll. ed.	yes	953.2	(484.89)	893.9	(487.31)	792
	no	699.5	(399.90)	656.3	(416.73)	3918
Living in regional town	yes	787.7	(453.28)	727.2	(467.84)	1056
	no	729.0	(416.99)	687.3	(429.24)	3654
Gender of head	Woman	550.5	(372.02)	532.9	(401.68)	1466
	Man	828.8	(420.65)	770.0	(434.46)	3244
Leading job	yes	1122.5	(508.54)	1088.5	(585.28)	98
	no	734.1	(420.48)	687.9	(431.02)	4612
Public sector	yes	884.4	(392.83)	835.6	(474.03)	846
	no	711.0	(426.76)	665.7	(424.27)	3864

Note: Values are in EUR computed from SKK in conversion rate EUR 1=30.126 SKK

Table 7.3: Regional descriptive statistics from SUSR and our sample computed for family member

	Disposable income		Consumption			Avg. family
Region	Our estimate	SUSR	Nondurable est.	Total	SUSR	size
BA	329.4	326.2	312.1	317.0	313.7	2.4
TT	257.6	253.6	241.1	243.0	242.0	2.9
TN	244.4	242.2	219.5	223.9	223.1	2.9
NR	258.7	254.0	248.4	251.5	251.8	2.9
ZA	240.9	241.2	233.5	243.1	248.6	3.3
BB	249.1	242.0	235.7	237.4	232.4	2.6
PO	223.0	222.8	205.5	209.1	210.2	3.4
KE	242.3	243.7	223.9	248.1	246.7	3.2
SK	253.0	251.2	237.3	244.3	244.6	2.9

Note: Values are in EUR computed from SKK in conversion rate EUR 1=30.126 SKK. Total consumption is nondurable with added consumption on vehicles and big household appliances.

Table 7.4: Variables

Consumption	Logarithm of household consumption subtracted by expenditures on durable goods (vehicles and big household appliances).
Disposable income	Logarithm of disposable income.
Regular income	Logarithm of disposable income for each household subtracted by Irregular Income.
Irregular income	Logarithm of household irregular income, e.g. irregular income from employment, bonuses, royalties.
Number of kids	Number of persons younger than 18 years.
Size	Variable which measures number of persons in household.
Public sector	Dummy variable which takes value one if the head of a family or husband/wife is employed in public sector.
Leading job	Dummy variable which takes value one if the head of a family or husband/wife is employed as a manager or in a leading position.
Age	Age of head of the household.
Age sq	Squared age of head of the household.
Gender	Dummy variable which takes value one if the head of a family is male.
Single	Dummy variable which equals one if household consists of one member.
Big town	Dummy variable which takes value one if the household is situated in one of 8 regional cities (Bratislava, Trnava, Trenčín, Nitra, Banská Bystrica, Žilina, Prešov, Košice).
Education	Dummy variable which takes value one if the head of a family or husband/wife has college education.
Unstable job	Dummy variable which takes value one if the head of a family has short term job or job without contract.
Self-employed	Dummy variable which takes value one if the head of a family is self-employed.
Employee	Dummy variable which takes value one if the head of a family is an employee.

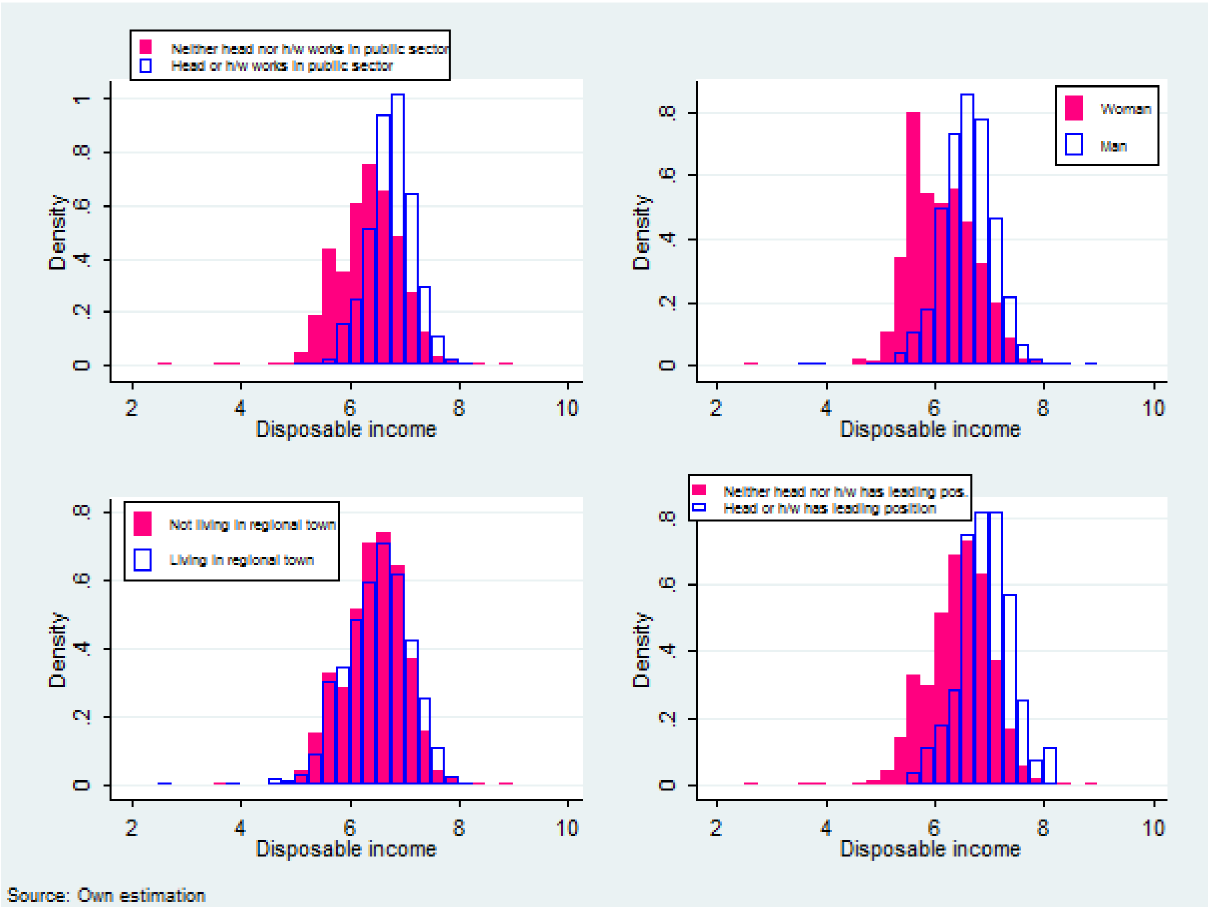


Figure 7.1: Histograms of instrumental variables used for disposable income.

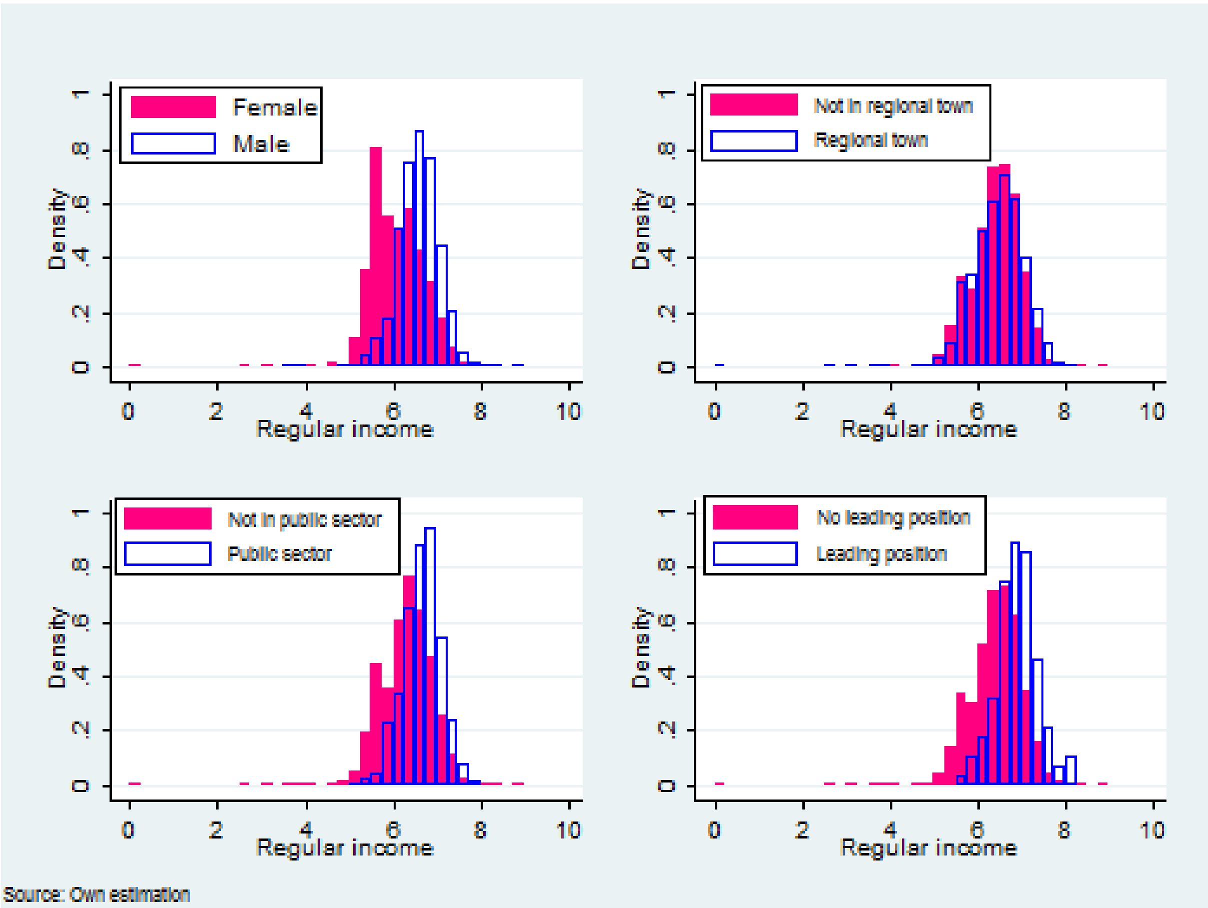


Figure 7.2: Histograms of instrumental variables used for regular income.

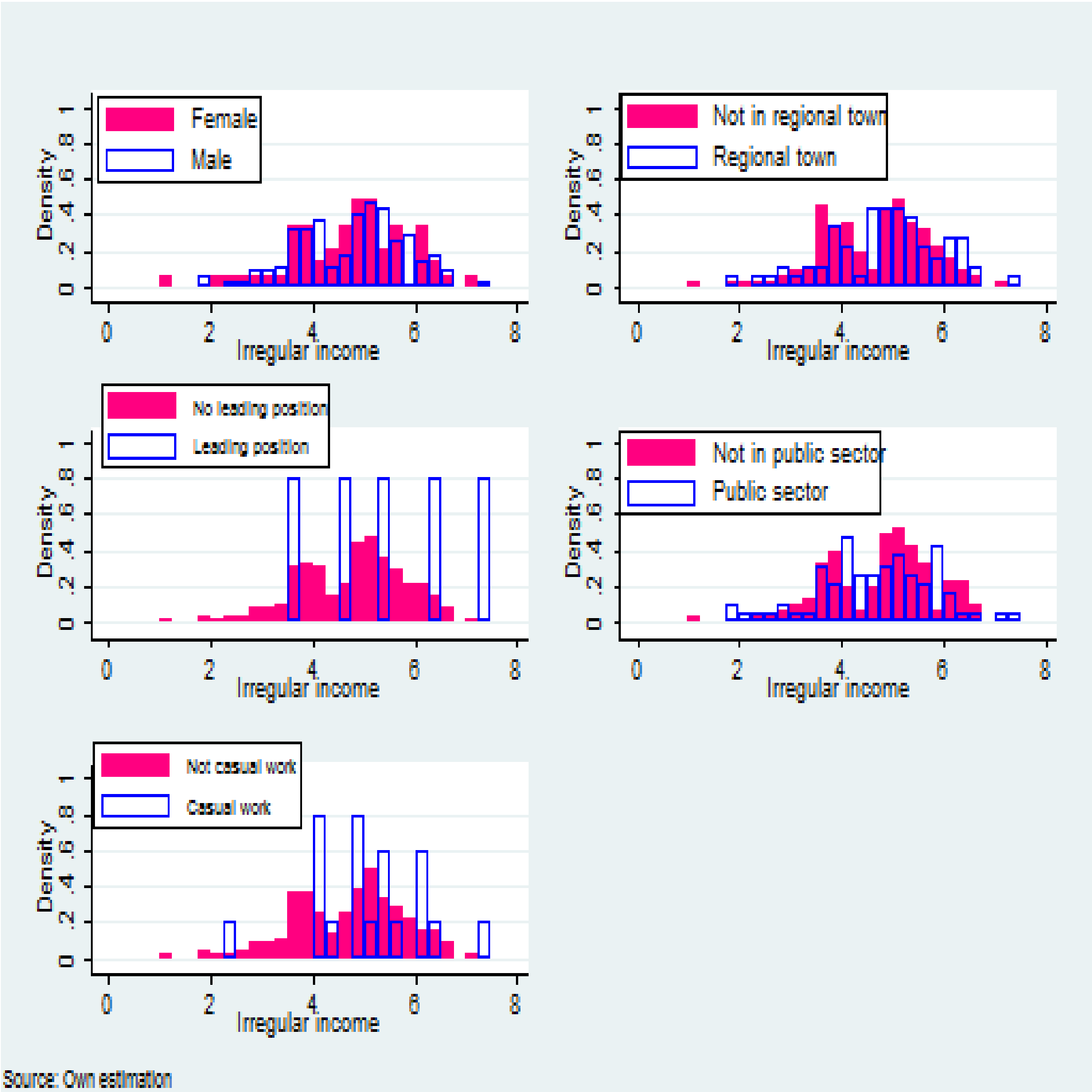


Figure 7.3: Histograms of instrumental variables used for irregular income.

Table 7.5: Basic consumption function for different age groups and single households

est: Consumption	(1)	(2)	(3)	(4)	(5)	(6)
Sample- age/single:	<35	35-44	45-54	55-64	>64	Single
Disposable income	1.039 ***	0.769 ***	0.882 ***	0.822 ***	1.081 ***	1.020 ***
	(0.080) e	(0.056) e	(0.060) e	(0.045) e	(0.053) e	(0.106) e
Results of the first stage equation for instruments for endogenous regressor - disposable income						
Leading job	0.309 ***	0.176 **	0.274 ***	0.303 ***		0.191
	(0.094) iv	(0.088) iv	(0.084) iv	(0.108) iv		(0.160) iv
Public sector	0.222 ***	0.203 ***	0.193 ***	0.375 ***		0.453 ***
	(0.035) iv	(0.029) iv	(0.028) iv	(0.035) iv		(0.049) iv
Gender	0.373 ***	0.353 ***	0.352 ***	0.451 ***	0.493 ***	0.079 **
	(0.049) iv	(0.041) iv	(0.035) iv	(0.033) iv	(0.025) iv	(0.039) iv
Big town	0.024	0.171 ***	0.100 ***	0.036	0.103 ***	0.147 ***
	(0.042) iv	(0.041) iv	(0.032) iv	(0.040) iv	(0.001) iv	(0.034) iv
Number of observations	694	984	1187	961	884	763
F-statistics	167.38	225.30	217.80	338.47	411.43	92.79
p-value of Hansen J stat.	0.55	0.44	0.71	0.65	0.81	0.60
p-value of DWH statistics	0.00	0.46	0.01	0.02	0.00	0.05
F test of excluded instr.:	29.09	36.75	43.84	96.1	199.13	46.93

Note: Coefficients report the average marginal probability effects. ***,**, and * denote significance at 1 percent, 5 percent, 10 percent level, respectively. e-Variable taken as endogenous in regression estimated by instrumental variable. iv- Instrumental variable. [1] - First stage OLS for the endogenous regressor- Disposable income. [2] - Second stage for endogenous variable- consumption. [3] -F-statistics for 2SLS, Wald for GMM. Used Robust option in STATA command to report SEs and statistics that are robust to the presence of heteroskedasticity in the error terms.