universität wien

DIPLOMARBEIT

Titel der Diplomarbeit

Item Nonresponse in the OeNB EURO SURVEY Its Determinants and Impacts on Currency Substitution Measures

Verfasser

Andreas Gulyás

angestrebter akademischer Grad

Magister der Sozial- und Wirtschaftswissenschaften (Mag. rer. soc. oec.)

Wien, im April 2009

Studienkennzahl lt. Studienblatt:	A 140
Studienrichtung lt. Studienblatt:	Diplomstudium Volkswirtschaft
Betreuerin:	Privatdoz. Dipl. Ing. Dr. Christine Zulehner

Acknowledgment

I especially want to thank Helmut Stix and the Austrian National Bank, who not only provided me the data, but also support for my diploma thesis. I owe the greatest debt to my mother, who enabled me with her everlasting support and commitment to prosecute my studies, I dedicate this work to her. I am grateful to Julia Panny for her support throughout the busy time and her restless proofreading. I also appreciated the help of all those unmentioned persons. Appraisals also to Christine Zulehner, who was an excellent supervisor.

Any opinions, conclusions, and results expressed are solely those of the author.

Table of Contents

1 Introduction

In the recent past surveys have become a more and more important source of information. Politicians put great emphasize on being informed about the public's opinion, enterprises base their investment and product decisions on market surveys and also within the framework of economics survey data plays an important role. As more and more models are micro-founded, individual data – which is in many cases obtained by surveys – becomes essential to test these models. Furthermore we know that many economic indicators depend on individual perceptions, assessments and expectations such as inflation expectation and consumer confidence – which are just obtainable through polls. In addition, surveys may provide insights in areas, which are otherwise hard to measure because the transactions are unrecorded, such as activity in the shadow economy or foreign currency cash holdings.

With increasing dependence on survey data, the question of reliability arises naturally. Surveys have many possible shortcomings which are manageable at different degrees. One danger is that the polled population is not a random sub sample of the whole population. Persons in different life situations may be harder to reach than others. For example it is more probable to contact a housewife or an unemployed person at midday than a worker. On the other hand many persons are reluctant to cooperate with pollsters. This is referred to as unit nonresponse in the literature. Because of these facts, it is not probable that the respondents are a random sub sample. In order to achieve representativeness, persons who are underrepresented with respect to certain socio-demographic characteristics get a higher weight than individuals that are overrepresented in the sample.

The next problem is that even when persons are reached and willing to take part in the survey they tend to withhold information on particular questions. This is usually called item nonresponse. If the nonresponse does not only depend on randomness, but also on certain personal characteristics, all analysis that ignore the item nonresponse mechanism will be biased.

The problem which is maybe the most difficult to handle is if the respondents' answers deviate from the truth. The respondents may want to present themselves in a better light or simply forget to mention something.

Since weighting is an effective and proper way of dealing with representativeness problems and it is very hard to correct for false answers without external data, the issue of item nonresponse is the most discussed of these three in the literature and this diploma thesis will also contribute to it. The underlying data of my investigation of item nonresponse will

be a household survey aiming to give insights in the extent and determinants of Euro holdings in Central and Southeastern Europe. Since it concerns personal wealth – from which we know that persons are reluctant to disclose information – the survey is particularly exposed to item nonresponse. Several analyses were based on this survey and although they reported item nonresponse as a possible source of uncertainty it was not taken into account. This diploma thesis will therefore investigate the determinants of item nonresponse on the key questions in the survey and its effect on currency substitution measures. I will analyze the determinants of item nonresponse with a binary choice model, which will allow us to check whether the respondents are systematically different in socio-demographic characteristics from the nonrespondents and will give us rough intuitions about the possible biases of ignoring the refusals.

Using an imputation procedure which replaces the missing data will enable us to quantify the biases which arise due to observed differences in the respondents' and nonrespondents' characteristics. I will report the observed dissemination of Euro holdings against the "unobserved" ones, the median holdings and apply this investigation also to three commonly used currency substitution measures. This overview will allow us to conclude whether the usual ignorance of item nonresponse is arguable respectively a proper way of dealing with it.

My diploma thesis is structured as follows: Consisting of three parts, the first one will introduce the reader to the topic of item nonresponse, the second one will explain the motivations of monitoring foreign currency holdings and the third one will report the results of the item nonresponse analysis for the underlying dataset. The first part – namely "Item Nonresponse" – will introduce the reader to the terminology of item nonresponse and the commonly used theoretical frameworks and considerations. I will give an overview on the possible treatment and correction methods discussed in the literature and also of previous empirical evidence on the topic of item nonresponse.

In the second part – "Implications of Currency Substitution" – I will discuss the motivations and reasons of monitoring foreign currency holdings. The reader will be informed about the extent of currency substitution in Central and Southeastern Europe and its impacts on policy issues. In the last subchapter I will summarize a paper of Stix (2008) on which my imputation method will be based.

In the third part, namely "Item Nonresponse in the OeNB Euro Survey" I will pass on to the applied part of my diploma thesis. I will investigate the determinants of item

nonresponse in the survey's key questions and quantify its impacts on currency substitution measures in Serbia in order to assess whether simply ignoring item nonresponse is an appropriate method of dealing with it.

2 Item Non Response

In this chapter the reader will be introduced to the terminology of item nonresponse, the most commonly used treatment methods of item nonresponse and will be given a summary of the empirical literature on the topic of item nonresponse.

2.1 Theoretical Background

There are many possible shortcomings of data. Aggregate data may be exposed to measurement errors, or is simply impossible to capture (like the true extent of the shadow economy) and must therefore be estimated based on correlated variables. The case of censoring occurs when data is recorded in a certain time range. The time of occurrence (e.g. death of a proband) is only observed at the units where it occurs until the date of termination of the experiment and is censored for the others. Especially survey data is exposed to a wide range of possible imperfections. First of all, the randomly or not chosen respondent may entirely refuse to take part in a survey or may not be reached by the interviewer. If the person is interviewed, he or she may refuse to answer specific questions or his or her answer may deviate from the truth in order to be seen in a better light.

With the beginning of the 1970's the problem of missing data became increasingly discussed in the literature (e.g. Hartley and Hocking (1971), Rubin (1972, 1974), Little (1976), Kalton (1983) and Griliches (1986)).

As briefly introduced in the first paragraph, we can distinguish between two types of nonresponse in surveys: Unit nonresponse, where the respondents entirely refuses to take part in the survey and item nonresponse, where just some items, viz. some entries of the data matrix are missing and not the entire case.

Before considering the methodology of dealing with missing data, we must first analyze the different missing data mechanism first formalized by Rubin (1976). Modern literature on the topic of missing data (Little and Rubin 2002) distinguishes three types of missing data mechanisms: missing completely at random (MCAR), missing at random (MAR) and not missing at random (NMAR). I will present an intuitive definition of these, readers who are interested in more technical definitions and their proofs should see Rubin (1976).

The missing data is said to be MCAR if the probability of missingness is unrelated to the values of the variable itself and all other variables, e.g. the probability of reporting the income is the same for all individuals regardless of their income itself and all other factors such as age, sex, etc. Then the observed data is observed at random and can therefore be regarded as a random subsample of the original sample.

If the probability of response is unrelated to the variable itself but to other observed factors, then the data is labeled MAR. Sticking to our example with the income of a respondent, the missing data mechanism would be MAR if the probability of refusal is related to some socio-economic factors, but within each socio-economic group, the probability of missingness is independent of the income.

Missing data is not missing at random if the probability of response depends on values of unobserved factors. For instance, the refusal of reporting the income may depend on the level of income itself, which is obviously not observed. In the case of wealth questions the nonrespondents accumulate typically in the lower and upper tails of the distribution. (see e.g. Frick and Grabka, 2005 or Biewen, 2001)

The two most prominent theoretical frameworks dealing with the response behavior in interviews and surveys are the cognitive model and the rational choice theory. The cognitive model concept distinguishes several stages of the response behavior (Tourangeau 1984). First the respondent has to interpret the question in order to understand its meaning, secondly he or she has to retrieve all relevant information from his or her memory and then integrate this information into a judgment. In the last stage the respondent has to format and edit the response based on considerations of social desirability and self presentation (for a review of studies about possible problems in the several cognitive stages and its effects on response behavior see Moore et al., 2000 and for an in-depth analysis of cognitive processes in interviews Trometer, 1996). Rational choice theory focuses on the last stage.

Rational choice theory assumes that respondents are rational decision makers who maximize their expected utility in any given situation. The individual assesses and evaluates the situation faced with and chooses the best alternative depending on the expectations of the consequences of a certain behavior. For instance in questions concerning socially undesirable issues such as sexual preferences, drug use, approval of political parties, etc. the respondent might decide to refuse or simply to lie and alter his answer to a more socially desirable one. In fear of disclosure of answers to third parties the respondents may refuse to provide information on income earn in the underground sector. Schräpler (2004) reports five alternatives for an individual faced with an interview situation: “(1) the respondent

participates in the survey; (2) the respondent participates and answers the income question truthfully; (3) the respondent participates, decides to lie, and reports false facts that may be more socially desirable; (4) the respondent participates but refuses to answer or has retrieval problems and does not know the answer; and (5) the respondent refuses to participate." Respondents assess and evaluate the costs and benefits of each option and choose the alternative with the highest utility. Schräpler (2004) considers the following potential costs and benefits for each alternative: Benefits of participating are if the respondents perceive the survey as serving a meaningful purpose or endorse a scientific or public sponsor. Further benefits are if the interviewed person finds it an interesting entertainment. But these benefits stand in contrast to the opportunity costs of time invested in answering the questionnaire. If the survey is perceived as not confidential or the topic addressed by it is too sensitive or out of interest for the respondents, the polling will deliver disutility. The consequences of these findings are that the confidence building between interviewer as the direct representative of the survey and the respondents is very important. The interviewer must try to spark interest in the survey and to enthuse the respondents for the topic. Many surveys try to reduce unit nonresponse by highlighting the importance of the respondents' input for policy decision makers, scientists or research.

The same benefits of participating in the poll obviously apply to the response of particular questions. The costs of it are the possible losses due to (1) disclosure to third parties, (2) social undesirability and/or (3) invasion of privacy. Possibilities to persuade the interviewed person to respond are again highlighting the importance of the respondents' input and try to reduce the concerns about data abuse. Stating the commissioning institution may help to build confidence in the survey.

The benefits of refusing to answer particular questions are the opposites of the costs of response: The respondents do not have to bother about social disapproval. They keep particular information private and confidential. The costs of refusal are on the other hand justification costs. Moore et al. (2000) argue that in telephone surveys the social barrier of refusing may be lower than in face-to-face interviews. Following this idea the justification costs are the greater the higher the social barriers are. They vary from very low in mail or self-completed questionnaires to relatively high in face-to-face interviews with an interviewer asking the respondent directly. The justification costs may also vary between respondents. For self-confident and self-determined persons the justification costs might be lower than for persons with opposite characteristics. On the end of the decision process in which the respondents assessed and evaluated the benefits and costs of each action, we

observe participation if the net utility (utility of the benefits minus the disutility of costs) is positive and a refusal if it is negative.

2.2 Treatment of Missing Data

After having discussed the theoretical background of item nonresponse we can now pass on to the treatment of it. In the following section I will present the most common practices in dealing with missing data. I will briefly discuss the advantages and disadvantages of the proposed methods and their practicability for applied researchers. In doing so, I will focus on the treatment of missing data in the case of survey data.

The probably most important objective in dealing with missing data is to have an idea of the missing data patterns. From the subchapter "Theoretical Background" we are already familiar with the different types of missing values which can be distinguished in MCAR, MAR and NMAR. Every treatment method has its own assumption with respect to the missing data type. Unfortunately, there are no possibilities to test whether the underlying missing data belongs to a certain category. Thus it is all the more important to consider the missing data patterns and think about their possible determinants. If for example the nonresponse rate for a specific question is not too high and there are no reasons to classify the question as sensitive, one may assume the missing data as MCAR and choose the appropriate method. Nevertheless nonresponse can have many different determinants. One can easily disregard effects which are hided from oneself. Even in questions concerning non sensitive matters there are good reasons to believe that the nonrespondents differ systematically from the responding group. An example may be a formulation of a question which is without any difficulty understandable for persons familiar with the topic, but may represent difficulties for an uneducated respondent. As an example questions about exchange rate expectations shall be mentioned. Many people are mixing an appreciation of a currency with low inflation. Not well informed persons may refuse to answer this question to conceal their nescience. A brief analysis of the missing data pattern does not hurt. A probit or logit model can reveal the observed determinants of nonresponse and might indicate effects of unobserved variables (in later chapters I will analyze in-depth the determinants of nonresponse in a household survey using a probit type framework). Hence it is very important that the applied researcher tries to empathize with "his" or "her" nonrespondents to get an idea about the possible underlying type of missing data. Some

theoretical considerations of the motives of nonrespondents were already discussed in the previous subchapter. The following chapter "Empirical Evidence" shall give the reader an intuition about the reasons and determinants of nonresponse.

I will discuss four common practices and methods to account for missing data: Complete case analysis – often also referred to as by casewise deletion, simple imputation, multiple imputation and the Heckman procedure.

The first one – complete case analysis – is the simplest and most commonly used one, but has also the strongest assumptions. Most statistical software packages use this procedure by default when missing values are found. This method simply omits every case that does contain any missing data for any of the variables selected for the analysis. This implies that the requirement for an unbiased result using this procedure is that the missing values must be MCAR, which is typically not the case. But even in the case of missing values being MCAR using casewise deletion must not be the best approach. Especially by performing multivariate analysis a considerable part of the observations may be neglected. In a dataset comprising 20 variables, a random missingness of five percent of the observations for each variable will lead to a loss of 64 percent of the sample[1].

A very common approach in treating missing values are sample selection models. I will discuss the probably most commonly used one proposed by Heckman (1976 and 1979), often also referred to its abbreviation Heckit. It is a very powerful procedure to account for missingness in the explained variable. It enables us to correct for missing data even if it is not missing at random, viz. if the missingness depends on variables which are not observed. But on the other hand it has also quite strong assumptions that might easily be not fulfilled. The next paragraphs deal with the theoretical background of the procedure as well as with a discussion of its drawbacks.

Let the sample selection determining equation be

$$r_i^* = \beta_1' w_i + u_i$$

and the equation of interest be

$$y_i = \beta_2' x_i + \varepsilon_i$$

where u_i and ε_i have a bivariate normal distribution with zero mean, correlation ρ and standard deviations σ_u and σ_ε, respectively and w_i and x_i the covariates of the sample selection and of y_i, respectively. Further we observe response if z_i^* is greater than zero and nonresponse otherwise. Formally:

$$E[y_i \mid y_i \text{ is observed}] = E[y_i \mid r_i^* > 0]$$

[1] Source Rässler and Riphahn (2006)

substituting for r_i* and rearranging yields

$$E[y_i \mid u_i > - \beta_1' w_i]$$

using the equation of interest we can rewrite the equation as follows:

$$E[\beta_2' x_i + \varepsilon_i \mid u_i > -\beta_1' w_i]$$

applying the calculus of the expectation operator we obtain:

$$E[\beta_2' x_i \mid u_i > -\beta_1' w_i] + E[\varepsilon_i \mid u_i > -\beta_1' w_i]$$

since $\beta_2' x_i$ is independent of u_i the equation simplifies to

$$\beta_2' x_i + E[\varepsilon_i \mid u_i > -\beta_1' w_i]$$

using the fact that u_i and ε_i have a bivariate normal distribution with the above explained parameters we can calculate the expectation of ε_i given that u_i is greater than $-\beta_1' w_i$. That yields:

$$\beta_2' x_i + \rho\sigma_\varepsilon \frac{\phi(\beta_1' w_i / \sigma_u)}{\Phi(\beta_1' w_i / \sigma_u)} = E[y_i \mid y_i \text{ is observed}]$$

We can consider z_i* as a latent variable where only its sign is observed. Hence we observe $z_i = 1$ if z_i* > 0 and 0 otherwise. The variable of interest y_i is only observed if z_i equals one. Applying the above calculations we obtain the equation

$$E[y_i \mid z_i = 1] = \beta_2' x_i + \rho\sigma_\varepsilon \frac{\phi(\beta_1' w_i / \sigma_u)}{\Phi(\beta_1' w_i / \sigma_u)}$$

The only term which deters us from estimating this equation is $\frac{\phi(\beta_1' w_i / \sigma_u)}{\Phi(\beta_1' w_i / \sigma_u)}$ with its unknown vector of β_1'. But β_1' can be estimated using a Probit model for the response equation where $\Pr(z_i = 1) = \Phi(\beta_1' w_i)$ and $\Pr(z_i = 0) = 1 - \Phi(\beta_1' w_i)$.

In order to make this method work in practice good predictors of response are needed. But these have to be independent of the variable of interest *y*. In earnings equations variables have to be at disposal that explain the response behavior, but are not associated with income. Since most variables might affect both, or at least act as proxy for income predictors not included as covariates in the earnings equation it seems hard to find such variables. Little and Rubin (2002) argue that without such variables the Heckit method may be "quite spurious". Biewen (2001) estimated the earnings regressions by applying the Heckman procedure. In addition to the variables used in the earnings equation he included church attendance, volunteer work and local politics as covariates for the response equation. He remarked that it is not implausible that these variables proxy unobserved personal characteristics like good will or honesty which may also be correlated with income. Lillard

et al (1982 and 1986) found considerable instability of the Heckman procedure for different specifications and different sets of exclusion restrictions. Their estimations, using data from the current Population Survey, predict implausibly higher income for nonrespondents than income figures obtained by the imputation method used by the Census Bureau. David et al (1986) showed that the Census imputations are in line with the true income amounts.

A second problem of the Heckman model is that it assumes the conditional distribution of the total population (meaning respondents and nonrespondents together) being normal. This assumption is obviously untestable since the distribution of nonrespondents is by definition unobserved.

To sum up, the Heckman procedure is a powerful method of dealing with missing data which is associated with unobserved factors. But the applied researcher has to consider very well, whether variables that can fulfill the requirements of the model are at disposal. Only with a reliable response model it is recommendable to apply sample selection models.

The next two methods I will deal with are Single and Multiple Imputation. Good reference books on this topic are the eminent book of Little and Rubin (1987, 2002), Rubin (2004) and Schafer (1997). Readers preferring a less formal description of this topic are referred to Allison (2002).

Single imputation, that is – as indicated by its name – filling in one specific value for each missing one. Two major advantages promote this method of handling nonresponse. First, standard complete data methods of analysis can be applied on the filled-in data set. Some sophisticated mathematical and statistical approaches require specialized computer programs and it is not obvious that the effort and money invested in these are really worthwhile to for example an increased sample size. Unsophisticated users may more likely come to reasonable conclusions by using familiar statistical procedures rather than complicated methods beyond their understanding.

The second major advantage of single imputation is that information from the data collector can be incorporated in the imputation procedure. The data collector, for example a Census Bureau, has typically more information than the consumer of the data set. Rubin (2004) stated**:** "Consequently, it is possible that data analysts, even those with a full arsenal of modern statistical tools, might reach better inferences by trusting the data collector's imputations than by applying sophisticated statistical models to a less rich data base".[2]

On the other hand, single imputation also has considerable disadvantages. The obvious problem is that with analyzing the filled-in data set with standard statistical

[2] Rubin (2004) page 12

procedures we treat the imputed data as real known values disregarding the implied uncertainty. Even if we perfectly understand the nonresponse mechanism, the obtained inferences by imputation will be too sharp because we do not take into account the variability due to the unknown missing values. When nonresponse is not really understood, which may often be the case, we have to take into consideration the additional uncertainty from not knowing whether the applied imputation model is appropriate. To account for these drawbacks the procedure of multiple imputation was developed. As described by its name, multiple imputation imputes several values instead of one specific value for missing observations. It involves some form of randomness in the imputation procedure to account for the uncertainty arising from the unknown missing values. If the driving mechanism behind the nonresponse is not perfectly understood or if we are unconfident whether the applied model is appropriate, several models of imputation can be used to incorporate this uncertainty. If we impute several values –regardless with which procedure – say m times, we obtain m completed data sets which are then analyzed separately with standard complete-data procedures. In order to obtain the multiple imputation estimates, the separate estimations are combined with the following procedure: The point estimate of multiple imputation is simply the averaged point estimations over these iterative imputations

$$\hat{\theta}_{MI} = \frac{1}{m}\sum_{i=1}^{m}\hat{\theta}_i$$

The variance-covariance matrix of the multiple imputation estimate $\hat{\theta}_{MI}$ adjusted for the variability due to imputation is the averaged variance-covariance matrix "within imputation" plus the variance-covariance matrix "between imputation" multiplied with a correction term. Formally

$$V_{MI} = \frac{1}{m}\sum_{i}^{m}V_i + \frac{m+1}{m}B$$

where V_i is the ordinary estimate of the variance-covariance matrix obtained from the imputed data set i and B is the between imputation sample variance-covariance matrix.

The obvious disadvantages of multiple imputation are the additional effort needed to produce multiple rather than single imputation, the more space that is needed to store the data sets and finally the extra time needed to analyze multiply-imputed data sets. In the light of an increased validity and accuracy these disadvantages seem to be quite modest.

The remaining question up to this point is which imputation algorithms are available to impute for missing values. An in-depth analysis of the methods is clearly beyond the

scope of this diploma thesis and thus I will just briefly report some available options without claiming exhaustivity.

A common method of imputation is single conditional mean imputation. The missing values are filled up with the mean or median of the respective sample subgroup. For example the nonresponse on an income question is substituted by the mean or median of the reported incomes of the group with the same characteristics and attributes.

Matching methods can be applied to impute for missing values. For each missing case a valid answer of a respondent with "similar" characteristics is searched and substituted for the missing one. Under "similar" in this context I denote cases found to be "near" by a distance-matrix or with the same characteristics. These methods are summarized in the term hot-deck, which does not have a well-defined common usage. Matching procedures that replace missing values by observation from external sources like comparable surveys or previous waves are denoted as cold deck imputation.

Regression imputation fills up the missing data with predicted values from a regression on the observed values.

Procedures can also be combined, which is then termed as composite methods. The regression approach may be combined with hot deck to find first the predicted value by a regression and then substituting an observed value which is found to be near to the predicted one. A similar method – but in the framework of multiple imputation – is applied later in this diploma thesis to analyze the impacts of item nonresponse in a household survey.

Multiple imputation methods use the same imputation procedures as single imputation methods, but incorporate some form of randomness to account for the uncertainty implied by the imputation of unknown missing values. This might be only adding random terms to predicted values or drawing randomly from a conditional distribution.

A simple simulation study by Rässler and Riphahn (2006) highlights the behavior of some of the above explained imputation procedures with respect to different missing data mechanism in a simple and informative way. They draw a sample of 2000 observations from bivariate normal distribution with

$$(Age, INC) \sim N\left(\begin{pmatrix} 40 \\ 1500 \end{pmatrix}, \begin{pmatrix} 10^2 & 0.8 \cdot 3000 \\ 0.8 \cdot 3000 & 300^2 \end{pmatrix}\right)$$

30 percent of the observations for income were deleted according to the missingness mechanisms MCAR, MAR and NMAR. The for each type of missingness simple single mean imputation (SI), single mean imputation within classes or conditional mean imputation

(SI CM), complete case analysis (CC) and finally multiple imputation using a composite method are applied to substitute for missings.

For the case where missing data is MCAR all procedures reflects – not surprisingly – the real mean income of 1500 very well (the obtained means range between 1500.14 and 1500.24). Both single mean and single conditional mean imputation understate the real standard errors by 17 and 6 percent respectively. Complete case analysis estimates the standard errors by about 20 percent above its real value. Multiple imputation also overshoots the real value, but far not by the same amount than CC.

If missing data is MAR or NMAR, multiple imputation can play out its strength. With respect to the mean, multiple imputation yields the highest accuracy of all methods. SI CM and MI have the lowest bias in estimating the standard errors. But since SI CM understates the standard error, with multiple imputation one is on the safe side.

But when is the additional effort demanded by multiple imputation really worth it? Harrell (2001) suggested some rough guidelines when to use the different approaches of dealing with missing data. He considered three methods, namely complete case analysis, single conditional mean imputation (no utilization of Y) and multiple random-draw imputation which should be applied depending on the proportion of missings. The term proportion refers to the proportion of observation with missing values in any variables.

Proportion of missings ≤ 0.05: Complete case analysis may be an option. He suggests single imputation for the missings by the median of nonmissing values for continuous variables and by the most frequent category for categorical variable since "it doesn't matter very much how you impute missings or whether you adjust variance of regression coefficient estimates for having imputed data in this case."[3]

Proportion of missings 0.05 to 0.15: If a predictor is unrelated to other covariates then Harrell suggests imputing reasonable constant values. A customized model should be applied if the predictor is correlated with other variables. He concludes that "single imputation is probably OK here, but multiple imputation doesn't hurt."[4]

If the proportion of missings is greater than 0.15, multiple imputation is recommended. In the case of frequent missing in some predictors also a sensitivity analysis with a subset of observations for which predictor is not missing is suggested.

[3] Harrell (2001), page 49
[4] Harrell (2001), page 49

2.3 Empirical Evidence

This chapter will give an overview on the item nonresponse literature of the last two decades. I will mainly focus on papers that investigated the determinants of item nonresponse and its influence on the quality of analyses. The reader should get an overview on the empirical evidence of determinants of item nonresponse and also on the commonly applied methods of its investigation.

Zweimüller (1992) investigated the survey nonresponse biases in wage regression using Austrian Survey data. He estimated selectivity corrected wage equations taking into account the missing data due to non-participation and refusals. Zweimüller concludes that for the data set used it turned out that: "(i) sample selectivity due to survey non-response is a significant problem; (ii) the bias mainly concerns the constant term and leaves the slope coefficient unaffected, and (iii) selection due to survey non-response is of larger importance than the usually addressed selectivity bias due to non-participation."[5]

Sousa-Poza and Henneberger (2000) analyzed the nonresponse behavior and its impact on wage function coefficients. They used data obtained from the Swiss Labour Force Survey (SLFS), an annual, computer aided telephone interview polling a number of labor-market related topics. On the one hand they estimated a probit style equation in order to investigate the determinants of nonresponse inclination and on the other hand they analyzed the existence of a nonresponse bias by comparing a sample selection model with an OLS estimation ignoring the item nonresponse. In the probit model explaining the nonresponse behavior 4 categories of variables were included: (i) characteristics of the interview, (ii) characteristics of the interviewer, (iii) characteristics of the respondent, (iv) mis-matches between certain characteristics of the interviewer and the respondent. Variables describing the interview situations were (i) number of attempts that were undertaken in order to contact the household, (ii) interviews taking place late in the evening and (iii) interviews taking place on the weekend. As possible factors explaining the interviewer effects age, education, whether the interviewer has a child or not, the interviewer's experience and the number of interviews conducted by the interviewer in the 1998 survey were used. As respondent's characteristics age, martial status, education, managerial position, home ownership, self employment, working time and its squared value were included. To capture the potential

[5] Source Zweimüller 1992, page 109

influence of mis-match in characteristics between the interviewer and respondent on the nonresponse behavior the absolute difference in age and the difference in education were used.[6] They estimated the equations separately for men and women. 16 percent of the females refused to provide information on their income in contrast to the male's item nonresponse rate of 13 percent. The probit model of nonresponse identified significant variables in all of the four groups of potential covariates. With regard to the interview situation only the number of contacts needed to reach the respondent has a significant negative effect on the response inclination. In contrast to the female's equation contacting after 8pm significantly increases the nonresponse probability. The results for the characteristics of interviewers yield that the more interviews were conducted by the interviewer the higher is the probability of a refusal. Low education of the interviewer has a positive influence on the response inclination in comparison to the reference group of medium educated interviewers. The age of the interviewer seems to have different effects on females and males. Elder interviewers reduce the nonresponse rate of males but increase it for females significantly. Concerning the characteristics of respondents for both females' and males' equation, age, high education in contrast to medium education and being self-employed have a significant effect on the response probability with the signs negative, positive and negative, respectively. The working time of the respondent has a significant effect on the responding propensity only for females in form of a parabolic curve with its maximum at about 26 hours. The results furthermore reveal that respondents are more likely to refuse in the case of female interviewers. No more significant effects were obtained for mismatch in characteristics.

Unfortunately omitted by most of the studies concerning item nonresponse Sousa-Poza and Henneberger also investigated the power of variation explanation of the four groups of independent variables. Their results yield that the respondent's characteristics explain the largest part of the variation and the three other groups of variables – interviewer's characteristics, interview situation and mismatches between characteristics of respondent and interviewer – "do not explain very much"[7]. On the other hand they also reported the predicting power of their model. Although 85 percent of the observations were classified correctly, only a very small fraction of the nonrespondents are identified correctly (14 percent in the female and two percent in the male samples). From this result and the fact that the pseudo-R^2 values of the two models were 0.107 and 0.065 the authors reason: "[…]

[6] Readers interested in the motives of the authors for this model are referred to Sousa and Henneberger (2000), page 81-83.
[7] Sousa-Poza and Henneberger (2000), page 88

our models are not able to capture the factors affecting an individual's nonresponse decision very well. Considering the numerous variables used in this study, one can conclude that the response inclination is, to a large degree, random".[8] This assumption is confirmed by their analysis of nonresponse bias using a sample selection model. They found no substantially variation between the OLS coefficients and the ones obtained by correcting them for selectivity bias. Similar to Zweimüller (1992) the main impact of nonresponse is on the constant term. The other comparisons between the two estimation results, namely comparing the confidence intervals and testing whether the estimated coefficients are significantly different underpin their previous outcome. As good news for the applied researcher – at least who used and will use wage data of the SLFS – Sousa-Poza and Henneberger state "[…] that the response inclination is, to large extent, randomly distributed amongst the underlying population".[9]

Biewen (2001) investigated the determinants of earnings nonresponse and its impacts on inequality measurement using data from the German Socio-Economic Panel (SOEP). In his paper he used three methods to obtain imputation. First he filled in the missing data by estimating an earnings regression on the sample of respondents. Secondly, he corrected for the possible nonresponse bias by running a sample selection model proposed by Heckman (1979). The third approach imputes values for the nonrespondents by matching them with sample members with similar personal characteristics.

He estimated the earnings equations separately for men and women with commonly used independent variables. In the sample selection equation he used in addition to these church attendance, participation in volunteer work and participation in local politics as these "variables were thought to reflect attitudes predicting the decision to answer the survey question independently from the level of earnings (which were represented by the other variables)."[10] In the selection equation for men only the occupational dummies were found to be significant with manufacturing workers, managers and workers in office jobs refusing less likely than scientist, workers in business jobs and workers in the service or agricultural sector. No other significant correlation of observed factors was found by Biewen for the response behavior. Together with an insignificant Wald test of independent equations the results point to the conclusion that no unobserved factors correlate with the nonresponse probability. On the other hand the result of the woman earnings equation provides evidence

[8] Sousa-Poza and Henneberger (2000), page 88
[9] Sousa-Poza and Henneberger (2000), page 94
[10] Source Biewen (2001), page 414-415

of unobserved variables. The incorporation of the sample selection leads to high positive corrections of the predicted incomes. Just the correction of the constant yields 60 percent higher predicted incomes than those estimated from the classical earnings regression. This finding is in line with the previously cited results of Zweimüller (1992).

Biewen obtained the imputed values by using the two models and adding a randomly drawn residual from the respective earnings equations to account for the otherwise underestimated variance. The distribution obtained by the classical regression imputation method is similar to the observed data with respect to the shape, but with much more mass in the lower tail which indicates that persons with lower incomes are more likely to refuse to provide information on their financial status. In particular, individuals with earnings up to 4000 Deutsche Mark were found to have twice as high probability of nonresponse than persons with incomes ranging between 4000 and 8000.

The imputed incomes using the Heckman model on the contrary skews the distribution to the right as this method imputes some very high earnings. Despite these differences between the observed and the estimated distribution neither the classical regression approach nor Heckman's show discernible differences between inequality measures for the observed and completed distributions. But Biewen warns, that the "predictions from the Heckman model seem rather erratic"[11] due to the drawbacks of sample selection models. (The advantages and drawbacks of them will be discussed in chapter "Treatment of Missing Data").

Very similar results were found by applying the so-called metric-matching hot-deck method. As in the case of regression imputed incomes, persons with earnings up to 4000 Deutsche Mark and around 8000 Mark have a higher probability of nonresponse, whereas individuals within these bounds are associated with low item nonresponse rates.

Riphahn and Serfling (2002) also used the SOEP data for an analysis of the determinants of item nonresponse on several income and wealth questions with focus on possible influences of matching of interviewers and respondents characteristics. For individual and household measures they estimated bivariate logit models. Taking into account that for wealth measures the dependent variables contain the additional answer category "don't know" they applied multinomial logits on these. The dependent variables

[11] Source Biewen (2001), page 419

were a range of individual and household financial variables[12]. The item nonresponse rate for these variables ranges between three and 15 percent for income questions and skyrockets up to 36.8 percent for household wealth indicators. As explanatory variables Riphahn and Serfling used the characteristics of respondent and interviewer like age, sex, education, home town size, employment, household size and the interaction effects of some of these factors. Further they controlled whether the interviewer has changed since the last wave or the respondent answered the questionnaire partly by him- or herself. The estimations reveal that in front of female interviewers respondents are more likely to refuse than in front of male interviewers. The findings indicate a weak tendency for respondents who are not full time employed to refuse. No evidence is found for matching effects with respect to employment status. Concerning the education variables neither respondent nor interviewer education level nor the matching effects of these are significant. Older respondents have a higher likelihood of nonresponse and almost all marginal effects of the age differential between respondent and interviewer point at decreasing nonresponse although only some are significant. No consistent patterns were found for the household size, interviewer change, the size of home town and presence of an interviewer. Further the authors investigated whether the "don't know" answers were relevant and independent to informative responses and to nonresponses by applying a Hausman test of the IIA property in a multinomial logit framework. They finally conclude that "[…] the group of respondents who refuse to answer the survey questions is certainly not a random draw from the population, that the group differs depending on the question they look at, that those answering "don't know" differ from non-respondents and that simply omitting these individuals from analysis may well bias results".[13]

Schräpler (2004) investigated the item nonresponse behavior related to gross income using the SOEP. Item nonresponse is factored in refusals and "don't know" responses. To account for possible differences in the driving factors behind refusals and "don't know" answers he estimated a univariate probit model for the indicator income-nonresponse and a multivariate probit model with three response variables-refuse, "don't know", and "unit response in the next wave" for each of the three waves taken. As regressors he used three groups of variables: (1) demographic and household variables for the respondent, (2)

[12] The dependent variables in of the item nonresponse rates are:
(i) for the bivariate model: income from self employment, Gross earnings last month, Net earnings last month, Vacation benefits, Income Household interest and dividend income, Household net income
(ii) for the multinomial model: Stocks/bonds, Owned home, Home loan, Savings, Total wealth.

[13] Riphahn and Serfling (2002), page 24

demographic variables for the interviewer and (3) variables that describe the interview situation. The first group consists of variables like age, the square of age, sex, occupational status, a variable indicating that the respondent does not work regularly, the presence of children in the household and whether the respondent changed residence in the prior 12 months. As interviewer variables Schräpler used the sex, age and the square of age. The variables whether the first interview of a respondent-interviewer pairing was successfully completed, whether the respondent choses to self complete the questionnaire in presence of the interviewer and whether an interviewer change occurred shall capture possible influences of the interview situation on the response behavior.

In the univariate probit model with item nonresponse as dependent variable Schräpler found age, age squared, unsteady work and the sex of the interviewer being significant factors in each of the three waves. Irregular working patterns have a positive effect on the probability of item nonresponse. Female interviewers are more likely to receive nonresponses. The higher their age is, the lower the probability of item nonresponse.

The multivariate probit model provides evidence that different attributes affect refusals and "don't know" responses. He found that the likelihood of refusals rise with increasing position and the probabilities of "don't know" answers rise with decreasing vocation position. "Don't know" responses increase significantly for persons with irregular working patterns and presence of children in the household. With regard to the interviewer characteristics only the sex has a significant effect on both, the refusal and the "don't know" responses. Female interviewers increase the probability of both. Finally the multivariate probit models reveal that refusing the income question is weakly but significantly negatively correlated with survey participation in the following wave.

Frick and Grabka (2005) also used the SOEP as underlying data for an analysis of item nonresponse on income questions and longitudinal imputation as a useful technique for coping with it. They compared the row-and-column-imputation technique described by Little and Su(1989) (hereafter L&S) with the cross sectional regression-based approach. The L&S method combines row (unit) and column (period/trend) information of the same individual. Their results reveal that this longitudinal imputation "yield more reliable results with respect to variance than the applied cross-sectional imputations methods"[14]. The findings further underpin the empirical evidence that item non-respondents are accentuated in the tails of the income distribution and thus an analysis with "complete cases" is biased.

[14] Frick and Grabka (2005), page 59

In particular they found a downward bias with respect to income levels, inequality and income variability in a society. Finally they found a positive inter-temporal correlation between item nonresponse and any kind (item and unit) of nonresponse in future waves.

3 Implications of Currency Substitution

The analyzed item nonresponse concerns a household survey aiming to provide insights in the degree of currency substitution in Eastern and Southeastern Europe. Therefore, I will explain in this chapter the motivations and reasons of national banks to monitor both, the level of its currency circulating abroad and the level of foreign currency held at home. In the first section, by presenting several papers concerning the empirical estimations of foreign currency holdings mainly in transition countries, the reader will get an insight into the degree and dimension of foreign currency holdings, either cash or deposits. The second section will deal with the impacts of currency substitution on several economic issues such as financial stability, economic activity, central bank earnings, etc. In the last section I will present a paper investigating the determinants of foreign currency cash holdings and foreign currency deposit holdings based on previous waves of the same household survey – namely the Euro Survey conducted by the Austrian National Bank (OeNB) – which is also the underlying data for my analysis of item nonresponse. The results of this paper are then used in the last chapter to set up an appropriate imputation model for nonresponse on questions concerning Euro cash and Euro savings deposit holdings.

3.1 Degree of Currency Substitution

Despite having used the term currency substitution in the introduction of this chapter, I did not yet provide a definition of it. The term refers to the phenomenon that the domestic currency is replaced to some extent by currencies with the reputation of being more stable and trustworthy and thus maintaining their purchasing power. The substitution of the domestic currency by a foreign one can serve two different purposes: Fulfill either the medium of exchange function of money – usually referred to as currency substitution – or the store of value function of it – called asset substitution. The literature often does not clearly distinguish between them and the term currency substitution is often used independently of its purpose.

The terms “euroization” and “dollarization” – synonyms often used for currency substitution, which I will also use – take into account that the two most common currencies which are used as substitutes for the domestic currencies are the Dollar, which is used

mainly in Southern American states, and the Euro mainly in the Central-, Eastern and South-eastern European countries.

Currency substitution occurs in two different ways: If a country pegs its currency to a foreign one or introduces a foreign currency unilaterally, i.e. designates a foreign currency as the sole legal tender, it is denoted by official or "de jure" euroization or dollarization. Montenegro and Kosovo have unilaterally introduced the Deutsche Mark in 1999 as legal tender as a reaction to the hyperinflation period and use the Euro since 2002. As examples for countries which peg their currencies to the Euro, Latvia, Lithuania, Estonia and Bulgaria shall also be mentioned.

On the other hand, one speaks of unofficial or "de facto" euroization or dollarization if the currency substitution is unofficial and driven by individuals' decisions to either hold foreign currencies to hedge themselves against uncertainties concerning the domestic currency – since it protects its user against the failure of a domestic bank, inflation and devaluation of the domestic currency – or to use the foreign currency as a unit of exchange. "De facto" euroization and dollarization may also be driven by the underground or unrecorded economy, since cash usage preserves anonymity and well-known, trustworthy currencies which are well protected against counterfeiting are the preferred medium of exchange.

The fundamental problem with analyzing the extent of currency substitution is that we cannot observe the underlying phenomena. Monetary authorities cannot perfectly monitor the capital in- and outflows since transactions in the underground and illegal sector are typically made via cash in a "black briefcase" and not via transaction accounts to avoid the paper trail of such transactions. To estimate the amounts of foreign currency in circulation one can apply econometric models using national accounts and monetary data or conduct surveys polling variables we are interested in. Both have their advantages and disadvantages: the drawbacks of the survey's result's are that they may be exposed to underreporting, since foreign currency holdings and foreign currency deposits holdings are related to personal wealth, which is – due to reasons explained in previous chapters – hard to obtain. Survey fieldworks are very cost-intensive and therefore there are little comprehensive polls concerning the amounts of foreign currency cash and deposits holdings. If conducted in regular periods, they give us insights into the trend in foreign currency holdings.

Estimates using national accounts and monetary data may be biased because the data neither captures the transactions by the shadow and underground economy nor the foreign

currency amounts which are “hidden under the mattress” since large foreign currency holdings may occur in countries which not only lack confidence in the domestic currency, but also in the whole financial system. Thus, currency substitution and the phenomenon of people preferring to hold their savings at home rather than in bank deposits might occur very often jointly.

A bulk of studies suggests that between 40-60% of US dollars are held abroad (Feige 1994, 1997, Porter and Judson 1996).

Feige (1994) proposed a method which measures the amount of US dollars abroad using the total amount of $100 notes put into circulation each month by the two Federal Reserve offices of New York and Los Angeles which are traditionally the main supplier of money to currency exportation. Currency specialists engaged in international currency transaction report that virtually all shipments of dollars abroad consist of $100 bills. Although the amounts being shipped abroad were not published since 1941, Feige proposed to use the amount of $100 bills which are put into circulation by these two Federal Reserve offices as a proxy measure for the US dollar outflow and vice versa the amount of $100 bills received from circulation as a proxy for the US dollar inflow. The obvious shortcomings of this proxy are on the one hand that it overstates net outflows due to some fraction of the $100 bills put in circulation by the two Federal Reserve offices satisfies the domestic currency demand. On the other hand the proxy will understate the outflows because only two of the thirty seven Federal Reserve banks and only $100 bills are taken into account and therefore smaller denomination injections are omitted. Despite these obvious defects it is a useful proxy as it is highly correlated with confidential estimates of net flows abroad calculated by the New York Federal Reserve. Additionally, Feige (1994) compared this proxy measure with two alternative approaches of US Dollar foreign holding estimations. The first one estimates the demand for US dollar treating the foreign demand as missing. If the currency demand equation is estimated during a period with low foreign currency demand (outflows minus inflows), then the forecast errors obtained by using this model in later periods are related to the net inflows or outflows of the US currency. Feige found a strong relationship between this measure and his proposed proxy with a correlation coefficient of 0.79. The second approach developed by Porter (1993) uses relative variations in cross-country seasonal patterns of currency holdings. The correlation coefficient of this method and the currency demand function residuals is 0.62.

Feige (2003) presented another proxy for inflows and outflows of US dollars. The United Customs Service is mandated to monitor systematic information on cross border

flows of US currency. Persons or institutions importing or exporting dollars exceeding sums of $10,000 are forced by the "Currency and Foreign Transactions Reporting Act" to file a "Report of International Transportation of Currency or Monetary Instruments" (CMIR). This database does not only contain information on the aggregated filed inflows and outflows, but also on their destinations and origins. Feige (2003) depicts in Figure 1 and Figure 2 the two currency inflow and outflow measures described above. CTI and CTO denote the CMIR estimates – adjusted for direct Federal Reserve net receipts from abroad that are not required to be reported on CMIR. The Inflow and Outflow Proxy curves describe the development of the proxy measure based on the $100 bills withdrawals. The two figures reveal that the two different measures evolve roughly the same over time. Taking into account that in 2000 the overall amount of dollars in circulation was $563.9 billion**,** according to the Federal Reserve[15], we observe that a considerable fraction of the US dollars in circulation move across the US borders.

For currency substitution measures in Europe, the "OeNB Euro Survey" provides the maybe most comprehensive data. This survey collects data for eleven Central, Eastern and South Eastern European countries on personal Euro cash and Euro deposit holdings and poses a range of other questions concerning the Euro in the possible determinants of its widespread use in this region. Scheiber and Stix (2009) present evidence on euroization and determinants of cross country difference in the above mentioned eleven countries. Their analysis is based on the two waves of the OeNB Euro Survey conducted in October/November 2007 and May/June 2008.

The results show substantial differences in the dissemination of Euro cash. The share of respondents declared to hold Euro cash ranges between nine percent in Hungary and 43 percent in Macedonia. The highest shares are obtained in Balkan states, namely Macedonia (43 percent), Serbia (39%) and Albania (32%). The Dollar does not seem to play a significant role with shares ranging between 7 percent in Albania and virtually zero in Hungary.

Not only the dissemination differs widely across countries, but the motivation of holding Euro cash is driven by different factors. In Central and Eastern European (CEE) countries making payments in the Euro area one of the main motives behind holdings Euro cash, while its store of value function is less important. On the contrary, in South Eastern European (SEE) countries the store value function is more important than the motive to undertake transaction in the Euro area.

[15] Source: Homepage of the Federal Reserve (March 25th 2009): http://www.federalreserve.gov/paymentsystems/coin/currcircvalue.htm

The median amounts for persons holding Euro cash confirm the results that euroization is more widespread in SEE than in CEE. The median for the former group is with 550 Euro considerably higher than the median of 190 Euro obtained in the group of CEE countries.

Scheiber and Stix calculated per capita holdings by projecting the obtained Euro cash amounts of the Euro cash holders on the whole population. The purchasing power adjusted per capita holdings reveal qualitatively the same picture obtained before. The EU member states have with amounts below 200 Euro lower per capita Euro holdings than the Balkan states. The figure for the Balkan states ranges between 259 in Bosnia and Herzegovina to 906 in Serbia.

The authors further calculated a currency substitution index (CSI), a deposit substitution index (DSI) and an overall euroization index (EI). These indices are defined according to Feige (2003), but were adjusted for the fact that the authors focused only on private households. CSI relates the foreign currency cash holdings (FCC) to the currency in circulation, which consists of local currency cash (LCC) and FCC. The deposit substitution index expresses the foreign currency deposit holdings (FCD) in terms of the total savings on savings deposits denominated in local currency (LCD) plus the FCD holdings. The overall euroization index reports the foreign currency holdings in relation to the total amount of currency in circulation plus the total amounts of deposit holdings. The used definitions are:

$$CSI = FCC / (LCC + FCC)$$

$$DSI = FCD / (LCD + FCD)$$

$$EI = (FCC + FCD) / (LCC + FCC + LCD + FCD)$$

Scheiber and Stix had taken the data for FCC from the survey while for deposits and local currency the official aggregate national figures were used. The results reveal that the projected foreign currency circulation makes up between two and 22 percent of the total currency in circulation. In Balkan countries this figure ranges from 27 to 76 percent, whereas FYR Macedonia and Serbia are clustered near the upper bound and Albania, Bosnia and Herzegovina and Croatia have shares of around 30 percent.

The deposit euroization index yields that in Serbia and Croatia with shares of 80 and 90 percent nearly all the households' deposits are held denominated in Euro. For the remaining SEE countries the share is between 40 and 60 percent and for the CEE countries between 4 and 15 percent.

The overall euroization index reveals that while the euroization level in CEE countries is only modest, the share of financial assets held in foreign currency makes up to 90 percent of all financial assets in SEE countries.

3.2 Impacts of Currency Substitution

In the following section I will discuss the policy implications of currency substitution. As explained before we distinguish between "de jure" or official euroization which means that a country either pegs its currency to a foreign one or introduces a foreign currency as the sole legal tender, and "de facto" or unofficial euroization when currency substitution occurs without legal approval of the monetary authorities. In the following, I will discuss the policy implications under these circumstances.

The widespread use of foreign currency reduces the demand for the domestic one and thus reduces the profits of the domestic national bank earned by issuing currency – known as seigniorage. Currency substitution also influences the profits of the foreign currency issuing national bank. On the one hand it boosts the seigniorage income, but on the other side such banks face higher costs for cash logistics, mainly the costs of withdrawing old banknotes. Readers interested in an in-depth discussion about seigniorage should see e.g. Fischer (1982) or Schobert (2002). In low inflation countries seigniorage income plays only a minor role in the states' revenues. Fischer (1982) shows that in many countries seigniorage income makes up for more than 10 percent of total revenues. In contrast to these findings are the seigniorage calculations of Schobert (2002) for Eastern and South-Eastern European countries summarized in table 1 on page 10 of his paper. Schobert applied two different measure methods on data of the time period of 1995-1999. Depending on the calculation method only one to three inflationary countries out of 18 receive seigniorage above 10 percent of their fiscal revenues. Seigniorage in relation to GDP ranges from close to zero to a maximum of about 3.5 percent of GDP. Schobert (2002) concludes that "Overall, the fiscal argument against euroisation in more advanced EU accession countries is quite weak[…]".

Another drawback of currency substitution for monetary authorities is that with a high degree of euroization or dollarization and a floating exchange rate regime it becomes difficult – if not impossible – to control the money supply in terms of the domestic currency. With every exchange rate fluctuation the money supply is changed and thus inflation targeting becomes very difficult.

In many transition economies foreign banks predominate the banking sector. As argued in the literature (e.g. in Luca and Petrova, 2008 or in Ize and Levy-Yeyati, 2003) banks exposed to currency risk through foreign currency liabilities try to shift this risk onto borrowers, either firms or households. Many households or small and medium-sized businesses often have no foreign currency income and lack of the possibility to hedge themselves against their currency exposure. This increases the default risk and thus bears the risk of destabilizing the financial system. In such situations currency exchange rate stabilization becomes the overwhelming policy objective of the respective national bank and reduces its ability to focus on over-issues. The close past offers a bulk of evidence for this exposure. In many Eastern European countries a considerable part of the mortgages and loans are denominated in foreign currency. In Hungary foreign currency loans account for 90 percent of new mortgages since 2006 and 20 percent of GDP[16] and in Latvia some 85 percent of loans to households and firms are denominated in foreign currencies[17]. As the financial crisis spread around in the summer of 2008 investors lost confidence in many Eastern European countries in which the cash poured in during the last boom. In an environment of a gloomy outlook on the economic performance the Hungarian National Bank saw itself forced to raise the interest rates about 300 basis points to 11.5 percent[18] on October 22nd as the Hungarian Forint lost 20 percent of its value during a few days.

An alternative to circumvent the exposure of the domestic economy to exchange rate fluctuations is to peg the domestic currency to a foreign one or a basket of foreign currencies. Many Eastern European countries like Bulgaria, Latvia, Lithuania and Estonia have chosen this monetary policy. As mentioned in the previous paragraph, the drawbacks of this policy are that the national bank ties its hand. The success of such commitments largely depends on the credibility of the monetary institutions. The case of Latvia and Russia during the financial crisis of 2008 highlights the dangers of a hard peg in turbulent times. During the financial crisis Latvia's central bank burned a fifth of its reserves – namely 1 billion Euro in 2 months (from Mid October to Mid December 2008) to defend the national currency, the lat[19]. The Russian Ruble is pegged to a currency basket consisting of Euro and Dollars. The foreign currency reserves of the Russian national bank sunk in December 2008

[16] Source: "Who's next?" The Economist October 25th 2008, Online Version (March 25th, 2009) http://www.economist.com/world/europe/displaystory.cfm?story_id=12465279
[17] Source: "Baltic brink" The Economist December 20th 2008, Online Version (March 25th 2009) http://www.economist.com/world/europe/displaystory.cfm?story_id=12815635
[18] See "Who's next?" The Economist October 25th 2008, Online Version (March 25th, 2009) http://www.economist.com/world/europe/displaystory.cfm?story_id=12465279
[19] Source: "Baltic brink" The Economist December 20th 2008, Online Version (March 25th 2009) http://www.economist.com/world/europe/displaystory.cfm?story_id=12815635

to 435 billion Dollar – about 27 percent lower than in mid August[20] – due to open market actions trying to defend the Russian Ruble. Despite these efforts the central bank had to devaluate its currency in relation to the Dollar about 20 percent from mid July to mid December 2008. It is not hard to imagine that if the confidence in such monetary authorities evaporates, the national banks would rapidly run out of cash.

If a country adopts a foreign currency unilaterally – often also referred to as full euroization or dollarization – it faces the same problems, but to a different extent, as countries with high "unofficial" euroization or dollarization rates. An additional drawback of full euroization discussed in the currency substitution literature (cf. Calvo 1992) is the loss of the "lender of last resort". The monetary authority would purchase the local currency in circulation with the National Bank's reserves. Kraft (2003) calculated, that the Croatian National Bank CNB could have financed the bailouts of the financial crisis of 1998-1999 even after purchasing all the Croatian currency, Kuna, in circulation. A bigger financial crisis might certainly be another story. But it remains undisputed that after a fully introduction of a foreign currency, the function of a lender of last resort would be constrained.

There is no clear answer when to apply the different monetary policies. The financial crisis of 2008 and 2009 also highlights the difficulties of monetary policy under euroization. The countries that pegged their currencies to the Euro were plagued with diminishing competitiveness due to the strong Euro and shrinking reserves to defend their currency pegs. On the other hand are the countries with floating exchange rate regimes, which saw their currencies strongly devaluating, which placed a huge burden on their debt service and increased the default rates on foreign currency credits.

3.3 Determinants of Euroization

Based on survey results, Stix (2008) investigated the determinants of euroization in Croatia, Slovakia and Slovenia on individual decisions level. He used household data obtained from three waves of the Euro Survey commissioned by the Austrian National Bank (OeNB) during 2004 and 2005. He followed a probit type approach to investigate the factors which

[20] Schulz, M.: "Euro und Pfund kurz vor Parität", Financial Times Deutschland Online Version (March 25th, 2009)
http://www.ftd.de/boersen_maerkte/aktien/anleihen_devisen/:Renten-und-Devisen- Euro-und-Pfund-kurz-vor-Parit%E4t/455425.html

drive the likelihood (1) that someone holds foreign currency cash and/or foreign currency deposits, (2) that someone possesses foreign currency deposits and (3) of the choice between foreign currency cash and foreign currency deposits. As explanatory variables served a range of socio-demographic variables like income, age, education, occupation, sex, household size and the size of the home town on the one hand and the following polled variables on the other hand (I'm just presenting a brief summary of the explanatory variables used by Stix in order that the reader can easier read and understand the summary of Stix' results. Readers who are interested in the in-depth description of the reasons for the choice of his variables should see Stix, 2008): The binary variable "NOTICED PAYMENTS" – which takes on the value of one if the respondent observed people in his country making payments in Euro and zero otherwise – shall capture the importance of network externalities – although being just an approximation of it. To account for the differences in expected returns between financial assets denominated in foreign and domestic currencies a set of variables concerning the expected inflation and exchange rate in one and two years ahead, respectively were polled and used as dummy variables in Stix' equations ("INFLATION WILL RISE", "INFLATION WILL FALL", "INFLATION SAME", "EXCHANGE R. WILL DEPREC"; "EXHANGE R. WILL APPREC", "EXCHANGE R. SAME"). To capture the importance of the perceived quality and credibility of the domestic banking system in the decision making the respondents were questioned how they perceive the safety of bank deposits in four categories ranging from very safe to very unsafe. Out of these, Stix constructed four dummy variables ("DEPOSITS VERY UNSAFE", ETC…). Further he used the dummy variable "CLOSE RELATIVE" denoting if a close relative is working in the Euro area to estimate the influence of remittances on the extent of euroization. Finally the three dummies (VISIT 0, VISITS 1-5 VISITS >5) control for the possible effect of the number of visits in the Euro area on the benefits of euroization by reducing transaction costs.

Stix (2008) estimated the equations separately for all three countries. He summarized the marginal effects on the probability that a person holds foreign currency cash or deposits obtained by the estimations in Table 6 on page 28 of his paper. Only coefficients at a significance level of at least 10 percent are shown. Income has a positive and significant effect on the probability of having FCC or FCD with increasing marginal returns in Croatia and Slovakia, but not in Slovenia. Being in the highest income quintile relative to the lowest one increases the likelihood of FCC/FCD possession by 26 and 35 percentage points for Croatia and Slovakia, respectively. Similar to income, significant, positive and increasing

marginal effects are obtained for education in all three countries. Furthermore having a close relative working in the Euro area, being an owner of a business relative to being an employee – with exception of Slovakia – visiting the Euro area relative to no visits and some age categories have a significant effect. Surprisingly, inflation expectations and exchange rate expectations do not play any role at all, at least in the statistical sense. The findings of Stix provide evidence that the experience with economic crises or with periods of instability influences the decision of possessing FCC or FCD. Younger generations have a lower probability of having FCC or FCD ceteris paribus than the reference group of over 55 year olds. The significant coefficients range from -5 to -16 percentage points subject to the age group, but not all of them are significant.

The results for the equation with foreign currency deposit possession as the dependent variable are quite similar to the previous ones and are summarized the same way as before in table 7 on page 29. Again, role of education, occupation, age, number of visits in the Euro area and income is found to have a significant effect with the expected sign. In contrast to the literature, the network effect variable has a significant influence on the probability of having a FCD. Stix argues that the findings are nevertheless plausible because "deposits are close substitutes for cash, in particular in the countries analyzed which have well developed financial systems". Concerning the effects of inflation and exchange rate expectations only little evidence is found. The confidence in the banking sector, measured by the perception of deposit safety, plays an important and significant role in all three countries.

Stix further modeled the choice between FCC and FCD. The dependent variable is calibrated that positive marginal effects imply that the likelihood of FCC ownership increases while the probability of FCD possession decreases. The estimation reveals, that a set of variables affects the degree of euroization, but not the choice between FCC and FCD. For income, education, age and having a close relative working abroad no convincingly significant results are found. As obtained by the previous equations, occupation – with owners of businesses having a lower likelihood of FCC ownership than blue collar workers or retirees – has a significant effect on the choice. As expected, the higher the mistrust in the deposit safety, the higher the probability of holding FCC rather than FCD, with the coefficient ranging between 23 and 27 percentage points. Furthermore, inhabitants of small villages have a substantially higher probability of holding FCC – varying between 15 and 22 percentage points – than residents of the respective capital. Assuming that the provision of banking services in rural areas is worse than in the capital (not quite a strong assumption)

the result points at the argument that a development of the banking system might induce a shift from FCC to FCD. Stix mentions secrecy as another explanation since the bank clerks in rural areas often know the customer.

Stix concludes that "With a well justified degree of caution concerning this latter result, [the insignificance of inflation and exchange rate expectations] the findings, overall, suggest that the degree of euroization is driven mainly by factors that are related to the economic record and not by expectations about the economic future." [21]

The results underpin – with providing insights in the individual's decision making – the empirical observation that despite being stable for many years, euroization and dollarization are a widespread phenomena in many transition countries. Economic policy decision makers face the problem that the degree of euroization seems to depend on issues like network externalities, experience of banking crises, etc. with long lasting persistence which are difficult to alter in a short period of time.

[21] Stix (2008), page 20

4 Item Nonresponse in the OeNB Euro Survey

Now, after being familiar with the theoretical concepts and empirical findings of item nonresponse, I will pass on to the empirical part of my diploma thesis. In the following section I will investigate the determinants of item nonresponse in the OeNB Euro Survey and quantify its impacts on the analysis of foreign currency holdings in Serbia.

4.1 Data and Descriptive Statistics

The analysis of item nonresponse in this diploma thesis is based on two waves of a survey commissioned by the Oesterreichische Nationalbank (OeNB). The survey – named OeNB Euro Survey – is conducted semiannually[22] and polls several questions about foreign currency cash and deposit holdings and their possible determinants such as inflation and exchange rate expectations, assessment of the current economic situations, confidence in the domestic currency, banking sector, etc. The data used was conducted in two waves by the respective Gallup organizations in eleven Eastern and South-Eastern European countries, namely in Albania, Bosnia and Herzegovina, Bulgaria, Croatia, the Czech Republic, Hungary, the Republic of Macedonia, Poland, Romania, Serbia and Slovakia, six of them members of the European Union. The fieldwork for the two waves was in October/November and April/May 2008 respectively, where approximately 1000 persons over the age of 14 were interviewed face-to-face per country and wave. The joint sample therefore comprises around 22000 observations.

As previously mentioned, the main aim of the survey is to give an insight into the sentiment of Eastern and Southeastern Europeans concerning the economic and financial system situation of their countries, inflation and exchange-rate expectations, dissemination of Euro cash and Euro savings deposits holdings and the motivations and probable determinants of it. In the next paragraph I will briefly consider the questions for which descriptive analysis will be undertaken. In order to provide a better understanding for the reader and to enable him or her to put him or herself in the respondents' position I will report the exact wording and the answer categories of these questions. The following five

[22] Actual information about the OeNB Euro Survey can be found on the homepage of the Austrian National Bank (25th March 2009): http://oenb.at/de/geldp_volksw/zentral_osteuropa/survey/the_euro_in_central_eastern_and_southeastern_europe.jsp#tcm:14-88216

are the key questions for which also the determinants will be investigated in the next chapter in depth: Question 5.1, 9, 15.1, 17 and the socio-economic question concerning the net monthly household income.

In question 5.1 the respondents were asked: "Sometimes people hold foreign cash e.g. for shopping, as a reserve or because they have retained foreign cash after trips abroad. Do you at present hold any foreign cash, that is banknotes and coins and not deposits at a bank account? Refer only to those cash holdings you hold personally or together with your partner." and had four answers to choose: "Yes", "No", "Don't Know" and "No Answer".

Like in many surveys concerning wealth a considerable part of the respondents refuses to provide information on income and wealth questions. In the overwhelming part of the questions two answer possibilities are provided for those respondents not willing to give an answer: "don't know" and "no answer". Over all items an average of around nine percent answered with either "don't know" or "no answer". Table 1 lists the item nonresponse rate for all questions. Not only the item nonresponse rate differs quite a lot over the questions, but also its composition. While questions concerning financial issues have an average nonresponse rate of approximately fourteen percent – where the answer category "no answer" accounted for 88 percent of the nonresponse – non financial questions have a substantial lower item non response rate of eight percent.

→ INSERT TABLE 1 ABOUT HERE ←

The "don't know"s account for 57 percent of all nonresponse in these questions. This descriptive result shows that in regard to senstive questions nonrespondents overwhelmingly choose the "no answer" category while in questions which are not concerning financial issues the nonrespondents distribute roughly equally over the two "refuse" categories. The highest item nonresponse rates are found on the one hand not very surprisingly at the questions related to the amounts of Euro and Dollar cash holdings (16 and 19 percent, respectively), the amounts of Euro and Dollar deposits holdings (23 and 21 percent, respectively) and the origin of the Euro holdings (33 percent). On the other hand questions about savings deposits (whether the respondent agrees or disagrees to the following statements: (1) interest rates on foreign currency savings deposits are higher than on local currency savings deposits, (2) savings deposits in foreign currency are better to safeguard the value of my money than savings deposits in local currency, (3) In my country it is very common to hold foreign currency deposits and (4) Savings deposits at foreign banks are much safer than those at domestic banks) have also very high nonresponse rate ranging

between 27 and 36 percent, which is a bit surprising to me since according to my perception this questions are not very sensible. But a closer look at the composition of the nonresponse yields that around 85 percent of the nonrespondents answered with "don't know". This result points at the conclusion that persons are not well informed and thus uncertain in regard to this topic.

Of particular interest are the nonresponse rates of the questions 5.1 (Euro cash possession), 9 (amount of Euro holdings), 15.1 (possession of a Euro savings deposit) and 17 (amount of Euro savings deposits holdings) since these are the key variables measuring the degree of currency substitution. By interpreting the nonresponse rates of these questions we have to take into account that question 9 and 17 are only asked from respondents, who reported to have Euro cash or Euro savings deposits, respectively. Table 2 reports the nonresponse rates and their composition out of "don't know" and "no answer" for each country.

→ INSERT TABLE 2 ABOUT HERE ←

With the exception of Poland, Romania and Serbia all other nonresponse rates are almost negligible. The overall rate is with 2.5 percent very moderate. That changes with question 9 concerning the amount of Euro cash holdings. The overall refusal rate jumps to 15.8 percent. In Bosnia, Bulgaria, Romania and Serbia it hovers around 33 percent. In comparison with these the 11.4 percent of Macedonia or the 15 percent of Croatia can be perceived as "low". Only Slovakia and the Czech Republic have relatively low rates around five percent. In Albania not one respondent refused at all. Almost the whole refusal rates are due to the "no answers". We have to consider that only respondents who declared to hold Euro cash (question 5) were asked about the amount being held. Thus persons, who regarded the question whether they hold Euro cash as too sensitive and refused to answer, are not in the base of the question concerning the Euro cash amounts (question 9) (see Table 3 - Question map for a better understanding). It might be the case that all respondents reluctant to answer wealth questions were already "dropped out" at the stage of question 5 and only those respondents who are willing to provide information on their wealth get to the question 9. Therefore, the low nonresponse rates in question 9 might be a result of high nonresponse rates at question 5. However, this seems not to be the case. First, the correlation of the nonresponse rates of question 5 to the ones of question 9 is 0.64 and secondly in particular the countries with high nonresponse rates in question 5 have also high ones in question 9.

→ INSERT TABLE 3 ABOUT HERE ←

To get a better grip of the nonresponse rate of question 9, I will construct a variable which takes into account that persons not asked about their Euro cash holdings (question 9) already gave an informative answer, namely that their amounts held are equal to zero. All persons having declared in question 5 not to have any Euro cash are classified also as respondents to question 9. Obviously also all respondents not having refused question 9 are respondents. The open point cases concern the respondents that already refused to answer question 5. Here I have to impose an assumption in order to classify these cases: I assume that persons perceiving the question whether they hold Euro cash or not as too sensitive and hence refusing to answer it, would also refuse to answer the more specific question about the exact amounts held. In my opinion this assumption is more than plausible since it is reasonable to assume that respondents not willing to provide information on a more general question like question 5 will turn about face their opinion and provide much more and specific information on their wealth. Thus the constructed question takes on the value

0 if the respondent answered neither question 5.1 nor question 9 with "Don't Know" or with "No Answer"

1 if the respondent answered any of the questions 5.1 or 9 with either "Don't Know" or "No Answer"

Table 4 reports the nonresponse rates obtained for the constructed question 9. Some of the skyrocketing nonresponse rates of the original question 9 relativize themselves.

Table 4: Nonresponse rates for the constructed variable of question 9

Countries	Nonresponse rates
Albania	0.19
Bosnia	4.99
FYR Macedonia	5.23
Bulgaria	6.55
Croatia	6.26
Poland	8.68
Romania	11.94
Serbia	18.91
Slovakia	2.42
The Czech Republic	2.4
Hungary	2.3
Whole sample	6.3

Source: OeNB Euro Survey

Most of the countries have a refusal rate of up to six percent. Only Romania and Serbia have double digit rates with approximately twelve and 19 percent, respectively. The average rate is nearly three times higher than in question 5. It seems that only a small fraction of the respondents have concerns to report whether they hold Euro cash or not, but a considerably higher share is not willing to provide insight in the exact amounts being held.

The next key question of interest is number 15.1, polling whether the respondents have a Euro savings deposit or not. The nonresponse rates range from zero in Slovakia and the Czech Republic to 7.1 percent in Poland. Again, the nonrespondents are almost solely "no answers". But similar to question 9, this question was only posed to a part of the whole sample. Only respondents who first stated in question 12b to have a savings deposit and secondly to have a savings deposit denominated in foreign currency (question 13) were asked whether they hold Euro savings deposits or not (see Table 3 for a better understanding).

The nonresponse rate of question 15.1 is biased because the nonrespondents are related only to the number of persons being asked this question. Persons who declared in question 12b not to have any savings deposit gave us - without being explicitly asked - also a meaningful answer to the question whether they possess a Euro denominated savings deposits, namely, no he or she does not have any. On the other hand we cannot ignore the individuals who refused to answer the questions which are a condition for being asked the question about Euro savings deposit, since probably many of the nonrespondents would also have refused to answer this question. Since one of the aims of this survey is to provide an insight in the degree of euroization and deposit euroization is one important factor if it, it is important to know how reliable the outcome of this question is, i.e. how many respondents have refused to answer these questions. Thus I will construct a variable which measures the refusal rate of the whole question block comprising questions 12b, 13 and 15 in a similar way to the question 9. In the cases where question 15.1 was asked from the respondent the classification between nonresponse and response is trivial. If the respondent declared not to have any savings deposit in question 12b or not to have any savings deposit denominated in foreign currency in question 13 we know without really obtaining his or her answer in question 15.1 that he or she does not have any Euro denominated savings deposit. Thus we assign the respondent in this case to the non-refusing group. The last case, namely if a respondent refused to answer question 12b or 13, we need to impose an assumption for the classification. By assigning these cases to the nonresponse group I assume implicitly that the respondents who refused to answer the question 12b or 13 would also refuse to answer

question 15.1. As already explained in the case of question 9, in my opinion this assumption is not very strong because it is plausible that a person who perceives a question as too sensible will also refuse to provide more detailed information on the same topic. Applying this to our case implies that I assume that respondents who considered the question concerning their savings deposit possessions as too sensible will also refuse to answer more detailed question whether they have savings deposits denominated in foreign currencies or not. Thus I constructed a variable taking on the value of one (response) if the respondent refused to answer any of the questions and zero (nonresponse) if the respondent did not refuse to all of the questions which were posed to him or her. In detail, the variable takes on the value

0 if the respondent answered either question 15.1 with "Yes" or with "No" or question 12b with "No" or question 13 with "No"

1 if the respondent answered any of the questions 12b, 13 or 15.1 with either "Don't Know" or "No Answer"

Table 5 shows the obtained nonresponse rate for question 15 considering the above explained "sample selection". Six percent of the whole sample refused to provide information about their Euro savings deposit holdings. Albania, Romania and Serbia have with low double digit nonresponse rates between 10.8 and 14.5 percent the highest ones. Croatia, the Czech Republic, Hungary and Slovakia are the best soils surveying Euro savings deposit holdings with negligible nonresponse rates close to zero. In the other countries the rate varies between 3.5 and 8.7 percent.

Table 5: Nonresponse rates for the constructed variable of question 15.1 and 17

	Q15.1	Q17
Albania	14.5	15.94
Bosnia	5.3	6.99
Macedonia	6.5	7.85
Bulgaria	3.5	6.85
Croatia	1.0	5.39
Poland	8.7	9.26
Romania	10.8	12.19
Serbia	11.9	13.43
Slovakia	1.3	1.9
Czech Rep.	0.7	1.08
Hungary	1.1	2.11
Whole Sample	6.0	7.56

Source: OeNB Euro Survey

The OeNB Eurosurvey collected also information about the amounts being held in Euro denominated savings deposits. Again, this question was posed only to respondents having a Euro denominated savings deposit. Thus I applied the above described procedure to obtain the refusal rate of the whole sample. But due to the fact that only a minor part of the respondents have a savings deposit denominated in Euro the results will be quite similar to the ones obtained before. The high nonresponse rates of the "crude" question 17– between 11.1 and 57.3 percent – in every country reveal that persons possessing a savings deposit denominated in Euro are particularly reluctant to provide information on the exact amount of their holdings. As expected, only minor changes are obtained by tabulating the constructed variable of question 17. The average nonresponse rate is 7.56, only slightly higher than the one of question 15.1. The nonresponse rates in Bulgaria and Croatia jump twofold and fivefold respectively, which reflects a high share of respondents holding Euro savings deposits and at the same time high refusal rates in question 17.

Since the net monthly income of the surveyed household will play an important role in almost all of the research based on this dataset, its nonresponse rate is also of particular interest. The refusal rate over the whole sample is 10.3. Like in the previously considered questions, there are sizeable differences in the nonresponse rates between the countries. In Poland, the Czech Republic and Slovakia only a very minor part of the respondents refused to answer. In contrast, in Bosnia, Bulgaria and Romania between a fifth and over a quarter of the respondent did not provide any information on their monthly income. In the remaining countries a refusal rate of around the overall rate of ten percent was obtained.

The aim of this section is to provide an insight in the degree of item nonresponse and to become an idea how the item nonresponse might bias the results. So far, I have discussed the first point. In the next section I will outline the nonresponse rates itemized by some socio-demographic variables. This will give us insights into the determinants and hence a first indication how the results will be biased if we ignore the nonresponse. Afterwards I will tabulate the distribution of Euro cash and Euro savings deposits amounts among income respondents and nonrespondents. This will reveal whether persons reluctant to provide information on their income hold different amounts than respondents.

Table 4 reports the nonresponse rates of question 5.1, 9 (constructed) 15.1, 17 (constructed) and income apportioned by the socio-demographic variables education, age, gender, income and employment status. In all five questions concerned the nonresponse rate increases with the level of education with exception of the low education group in question 15.1.

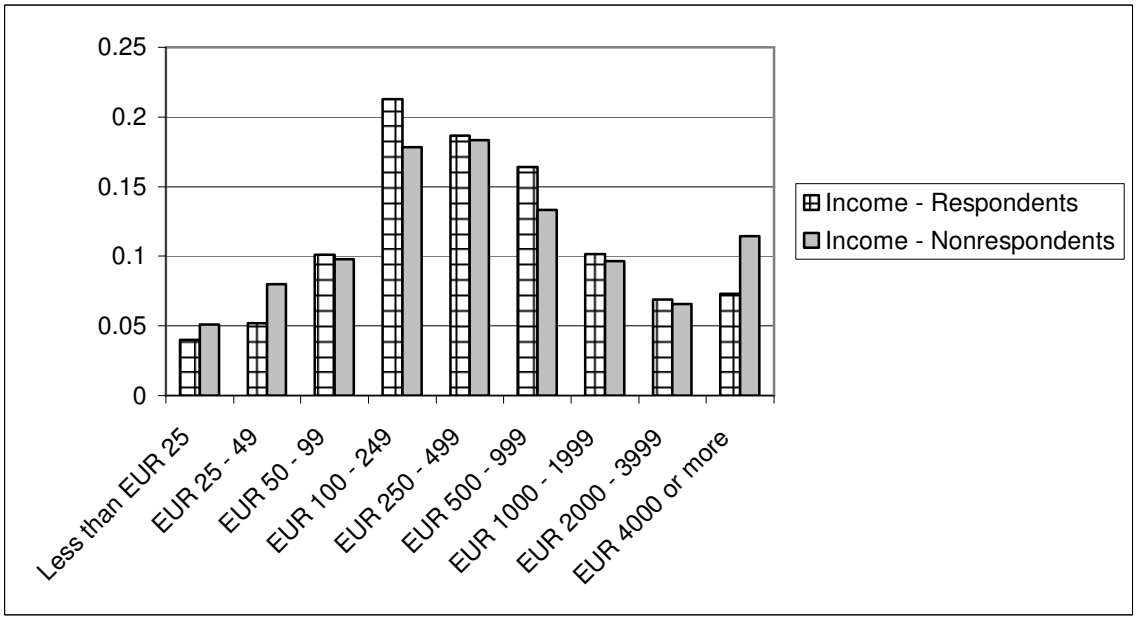
→ INSERT TABLE 4 ABOUT HERE ←

The nonresponse rates are between constant and slightly increasing over the income quantiles. For the nonresponding group with respect to income we obtain a considerably higher refusal rate than for the responding group. The refusal rate in the income nonresponse group is between 60 percent and 666! percent higher than in the income reporting group. It seems that persons who are reluctant to report information on their monthly household income are also rather unwilling to reveal information in regard to their cash and savings deposit holdings. With except of income reporting – where the nonresponse rates for males and females are nearly the same – we obtain a higher refusal rate for females than males. We observe no consistent pattern for refusal rate with respect to the employment status of the respondents. Students have an above average nonresponse rate in questions 15.1, 17 and income, but not in questions 5.1, 9. With exception of question 5.1 retired respondents have a below average nonresponse rate. Both employed and unemployed have refusal rates approximately around the average with always higher rates for employed persons. The relation between age and nonresponse seems to be nonlinear with a shape of an inverted U function. The nonresponse rate obtained for the youth is lower than for persons in working age between 25 and 55. After 55 the nonresponse rate again starts to decrease. This pattern is observed for all five variables.

Figure 1 and 2 depict the distribution of the Euro cash and Euro savings deposit amounts for income nonrespondents and respondents.

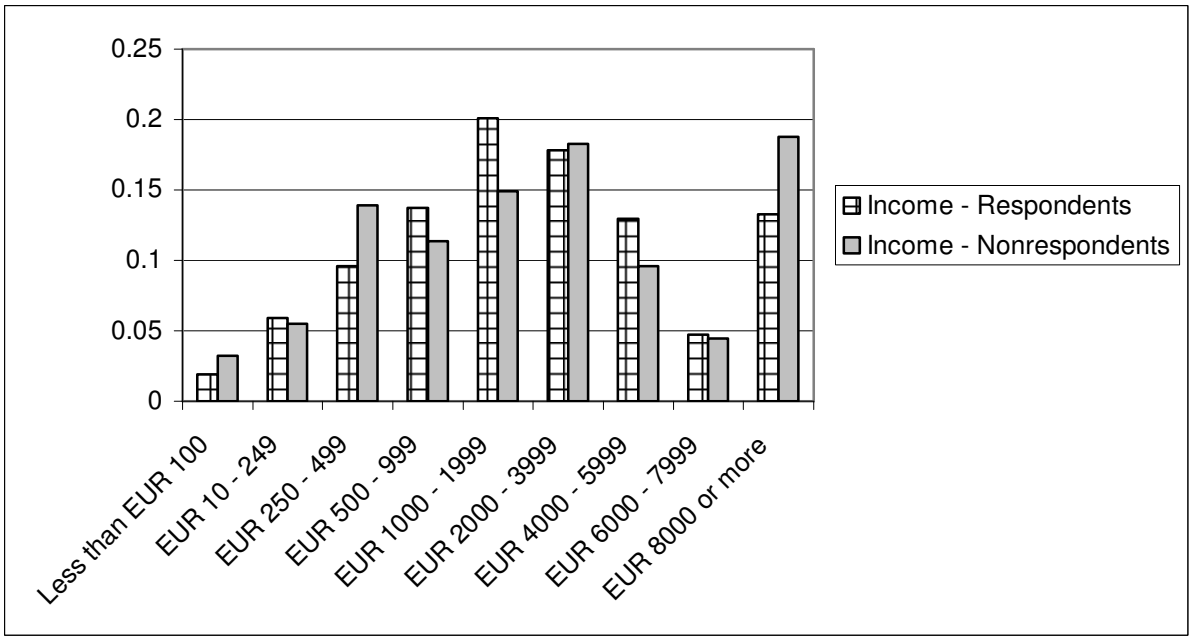
The obtained figures indicate that the amount distribution for persons who tend to withhold information on their wealth (in this particular case monthly income) have fatter tails, meaning that income nonrespondents are more likely to hold below or above average amounts than respondents having answered the income question. While seven and thirteen percent of the income respondents have more than 4000 Euro cash and more than 8000 Euro on Euro denominated saving deposits, respectively, these figures are eleven and eighteen percent for income nonrespondents. In chapter Imputation I will verify whether this first assessment holds.

Figure 1: Distribution of Euro Cash Amounts for Income Nonrespondents and Respondents



Source: OeNB Euro Survey

Figure 2: Distribution of Euro Savings Deposit Amounts for Income Nonrespondents and Respondents



Source: OeNB Euro Survey

Now, with the information of the descriptives, let us pass on to the analysis of the determinants of nonresponse.

4.2 Model Description and Variable Definitions

In order to identify the determinants of item nonresponse I will use the rational choice framework described in chapter 2 "Item nonresponse – Theoretical background". Rational choice theory assumes utility maximizing individuals who base their decision of refusing or responding on the ground of expected gains and costs. I will model this decision process by the following probit-type specification

$$NA_i^y * = \beta' x_i + u_i$$

where $NA_i^y *$ denotes the latent variable measuring the utility or disutility of respondent i refusing question y, x_i a vector of explanatory variables capturing the associated costs and benefits of item nonresponse, β' the coefficients to be estimated and u_i a normally distributed error term. Since $NA_i^y *$ is unobserved, I assume that an individual refuses if his/her utility does not exceed a threshold value which we normalize to zero. Thus we observe a person answering, i.e. $NA_i^y = 0$ if $NA_i^y * < 0$ and refusing, i.e. $NA_i^y = 1$ if $NA_i^y * > 0$.

The underlying dataset is described in "Data and Descriptive Statistics" and will not be discussed here. The model is estimated for all three dependent variables using the whole sample comprising the eleven surveyed countries and separately for each of the countries itself. In the model comprising the whole sample we are able to identify whether the considerable differences in nonresponse rates across countries reported in chapter "Data and Descriptive Statistics" are due to variations of respondent characteristics between countries or due to different "country mentalities" with respect to surveys.

4.2.1 Dependent Variables

The five questions, where the item nonresponse will be analyzed, are already familiar from "Data and Descriptive Statistics"**:** (1) Question 5.1 concerning Euro cash holdings, (2) question 9 polling the amounts of Euro cash held, (3) question 15.1 surveying the dissemination of Euro denominated savings deposits (4) question 17 collecting information on the amounts being held on the Euro savings deposits and (5) the socio-demographic question about the net monthly income of the household. The dependent variables of my model q5.1_na, q9_na, q15.1_na, q17_na and income_na take on the value of one if the

respondent refused the answer and zero if the respondent provided a meaningful answer. In the case of question 5.1 and income the classification is trivial. Nonrespondents are persons which answered question 5.1 with either "Don't Know" or "No answer" or refused to provide information on their monthly household incomes. As already mentioned in "Data and Descriptive Statistics" the cases of question 9, 15.1 and 17 are somewhat non trivial and need one assumption. As a reminder, question 15.1 is only posed to respondents having a savings deposit denominated in foreign currency which implies that these persons answered in question 12b that they a savings deposit and in question 13 that they have a savings deposit denominated in foreign currency. Persons who do neither have a savings deposit nor a denominated in foreign currency deposit obviously provided us already with the necessary information and are therefore classified in the respondent group. The only assumption we have to impose is that respondents having refused to answer question 12b or question 13 also would have refused question 15.1. The cases of question 9 and 17 are very similar. Summarizing the dependent variables take on the value of

Q5.1
- 0 if the respondent answered either question 15.1 with "Yes" or with "No" or question 12b with "No" or question 13 with "No"
- 1 if the respondent answered any of the questions 12b, 13 or 15.1 with either "Don't Know" or "No Answer"

Q9
- 0 if the respondent answered neither question 5.1 nor question 9 with "Don't Know" or with "No Answer"
- 1 if the respondent answered any of the questions 5.1 or 9 with either "Don't Know" or "No Answer"

Q17
- O if the respondents answered neither of the questions with "Don't Know" or "No Answer"
- 1 if the respondents answered any of the question 12b, 13, 15.1 or 17 with "Don't Know" or "No Answer"

4.2.2 Explanatory Variables

The choice of the explanatory variables is driven by two factors: First, theoretical considerations and secondly unfortunately by data constraints. The data set contains no information on interviewer level, therefore no effect of interviewer characteristics on the response propensity could be analyzed. By examining missing data it is needless to say that the true values of the question considered can also be determinant of nonresponse but is

omitted because of inobservances. These two points form the Achilles' heels of my model. A discussion of limitations and robustness is postponed to the results, but thus much shall be mentioned that there will be no indication of considerable biases based on both, empirical evidence and robustness tests.

Beside of several socio-demographic variables measuring the opportunity costs and cognitive abilities of the respondent some variables are used to capture network effects. I assume a higher response inclination if it is very common to hold Euro cash or Euro savings deposits. Two variables polled the perceived dissemination of both. In question 1.11 and 11.4 the participants were surveyed whether they agree or disagree on a scale from "strongly agree" to "strongly disagree" to the following statements: "In my country it is very common to hold euro cash" and "In my country it is very common to hold foreign currency deposits". If the respondent perceives a high dissemination of Euro cash holdings or Euro savings deposits I assume him or her ceteris paribus to be more likely declaring his or her Euro cash or savings deposit holdings since it is more socially desirable. Put bluntly, if everybody holds Euro Cash or Euro denominated savings deposits, why not declare it? One could argue that this is only true for respondents possessing Euro cash and that variable might have the reverse effect on the probability of refusal if the person does not have Euro cash. But taking into account that the refusals probably accumulate mainly in the group of Euro cash and deposit holders the reverse effect will only play a minor role. Thus I included these two variables split up in the dummies "Cash Common agree", "Cash common disagree", "Cash common don't know", "Deposit common agree", "Deposit common disagree" and "Deposit common don't know" in the respective equations. The answers "strongly agree", "agree" and "somewhat agree" are assigned to the agree group and all form of disagree answers to disagree. Persons which answered with "don't know" were matched to the "don't know" group.

In contrast to the previous two subjective network variables the next one is objective. It measures the nonresponse rates of the respective underlying question of the dependent variables in the inhabited region of the respondents (the respondents' own declaration is disregarded to circumvent endogenous concerns). The lower the item nonresponse rate in the region the higher is the barrier to refuse due to the lower the social desirability of nonresponse. Therefore I expect that these variables have a significant positive effect on the probability on refusal.

The effect of the following variable will be of particular interest. The respondents were asked whether they agree or disagree on a scale of strongly agree to strongly disagree

that they prefer to hold cash rather than a savings account. The dummy variables "prefer cash - agree", "prefer cash - disagree" and "prefer cash - don't know" are constructed the same way as in the previous cases of question about the perceived dissemination of Euro cash and Euro savings deposits. This variable may influence the propensity of nonresponse in two ways. Persons preferring to hold their savings in cash "under their mattresses" rather than on savings accounts may perceive this question especially sensible. These people may fear a disclosure to third parties because their savings are particularly exposed to burglaries and thefts. But this is not the reason for the special interest. It is due to the fact that individuals preferring to save in cash will also hold Euros with a higher probability, since the Euro is the preferred currency in the region for storing value. Thus this variable might be strongly correlated with the unobserved, true Euro cash variable. A significant effect of the attitudes towards holding cash rather than preferring savings deposit on the response inclination will therefore be an indication for the missing data being NMAR.

Several socio-demographic characteristics of the respondents control for the associated costs, either interpreted as the cognitive effort being demanded or as opportunity costs and potential benefits. Thus gender, age, size of household, whether the respondent is the household head or not, number of children, employment status, education level, size of the inhabited town and net monthly income of the household were included in the equations as explanatory variables. The following paragraphs deal with the detailed descriptions and definitions of the included socio-demographic variables and theoretical considerations of them. That may be said right away: The signs of the effects of almost any socio-demographic variables are very difficult to anticipate, since they typically have antipodal influences on the response inclination.

The gender dummies "Male" and "Female" as reference group were included to control for possible gender differences in the response inclination. On the one hand men are typically more decisive and assertive than females and women are on the other hand more consensus orientated in their behavior and might therefore be more likely to fulfill the interviewer's requests. Due to these two points I assume men to have a lower social barrier to refuse and thus higher probabilities of nonresponse.

The variable "age" is coded continuously. Possible nonlinear effects are captured by including the square of age, "age2". The anticipated effect of this variable is not clear. Older people have in general more leisure time than persons in working-age and thus lower opportunity costs. On the other hand older people may be more reclusive and suspicious of strangers and therefore be more reluctant to provide sensible information.

The household size is summarized in three categories and is included as three dummies in the nonresponse equations: (1) “hhsize2” if the household is smaller or equal to two persons, (2) “hhsize34” if the household comprises between three or four persons and (3) “hhsize5+” if the household is equal or larger than five. Ceteris paribus, the higher the household size, the higher might be on the one hand the effort of doing the housework and thus the opportunity costs of the time spent on the certain question and the lower on the other hand the patience of the respondents. Moreover in the question concerning the monthly income of the household it may become more difficult with increasing household size to keep an overview over the earnings of each household member. With increasing household size – especially in the case of respondents and their parents and/or grandparents living under the same roof – it is assumable that the financial wealth becomes more interweaved between the household members and thus more difficult to distinguish.

The next variable “hh head” measures whether the respondent is the household head with taking on the value of one if yes and of zero otherwise. Similar considerations as in household size are applied but with little different conclusions. The household head has to bear the main part of managing the household. This might reduce the response inclination because of the above explained reasons. On the other hand, the household head also has better overview and information about the personal wealth status and income status and thus the lower efforts demanded might increase the response likelihood. This argument seems to be the strongest in the case of the income question since income information about all household members is needed to answer this question. It is a priori not possible to judge which of the two opposite effects will dominate.

The next variable considered is the number of children in the household. It is coded in three dummy variables “no child”, “one child” and “Two or more Children” taking on the value of one if the respondent has no, one and two or more children, respectively and zero otherwise. I expect a negative correlation between the number of children and response inclination, since the opportunity costs and impatience are assumed to be the higher the more children the respondents have.

The employment status of the respondent is harmonized as follows: A person may either be employed, unemployed, retired or student/pupil. Respondents declared to be a housewife or househusband are considered as employed. This categorical variable is included as the dummy variables “employed”, “unemployed”, “retired” and “student/pupil” in the nonresponse equations. This variable should capture the opportunity costs of the answers’ duration. A person employed has typically less leisure time than a person

belonging to the other categories and may a higher valuation of the time consumed by the question.

The respondent's education level is measured by the three dummy variables "high education", "medium education" and "low education". If he or she has a university or college degree, he or she grouped in the high education category. Respondents having finished a secondary, vocational or high school are classified as members of the medium education group. Finally, a person without formal schooling or elementary or primary school is considered as low education respondents. Ex ante it is not clear in which direction the education level affects the response inclination. Different mechanisms influence the probability of nonresponse with opposite signs. The interview time's opportunity costs are obviously the higher the higher the education level is. But simultaneously the cognitive effort associated with the answering will be the lower the higher the education is. Higher educated persons might be easier to enthuse to report their income or wealth since they may easier identify themselves with a scientific work than individuals never been in touch with research. It is hard to guess which effect is stronger. The empirical literature suggests that higher education is linked to a higher response probability.

I included the size of the respondent's town of residence as an explanatory variable for nonresponse behavior because it is a potential indicator of an attitude of openness and trustworthiness. Respondents may refuse because of fear of crimes and large cities entail a sense of anonymity, which may be better guarded by its inhabitants than in rural areas.[23] The variable was surveyed in categories. Unfortunately every institute used somewhat different intervals for the town sizes. Therefore a sharp definition of the harmonized variable is not possible. The town size is harmonized in the three categories small, medium and large town, which try to summarize the categories consistent across countries as far as possible. It is included in the equations as the three dummy variables "small town", "medium town" and "large town". In Poland the survey has been only carried out in the ten largest towns and thus it has been just differentiated between very large and large towns. A classification in the sense of small medium and large towns is therefore impossible. Thus, in the equations using the pooled sample of every country, the entries for Poland are disregarded. The next paragraph deals with the exact definition of the harmonized variable, readers with no particular interest in its exact definition may skip it.

In Albania, Bosnia and Herzegovina and Macedonia settlements with up to 5000 inhabitants are coded as small towns. Cities with more than 100000 inhabitants are coded as

[23] Cf. for example De Maio (1980). She found a significant higher response inclination among rural people than among urban dwellers.

large cities and the towns in-between are defined as medium towns. In Bulgaria all towns with more than 100000 inhabitants are large towns. Settlements in-between 15000 and 100000 inhabitants are denoted as medium towns. Small towns are regarded as municipals up to a population of 15000 people. In Croatia small, medium and large towns are defined as towns up to 10000, in-between 10001 and 100000 and more than 100001 inhabitants, respectively. As already mentioned in Poland this question was polled only in two categories namely more than 500000 or in-between 300000 and 500000 inhabitants. In equations describing the response inclination with town size as an explanatory variable, the Polish observation are left out. In the tables summarizing the estimates and results for each countries the two categories are reported under "medium town" for towns with a population of between 300000 and 500000 and "large town" for towns with more than 500000 inhabitants. (See e.g. table 6). In Romania rural towns are considered as small towns. Non rural cities with up to 200000 inhabitants are coded as medium towns and Bucharest and all over towns with a population size of over 200000 people are large towns. The interval bounds for small, medium and large towns in Serbia are 0-5000, 5001-99999 and more than 100000 inhabitants, respectively. Small towns in Slovakia and the Czech Republic are considered as municipalities with up to 4999 inhabitants. Medium towns have a population of in-between 5000 and 99999 whereas large towns have more than 100000 inhabitants. Finally, Hungary has almost the same classification with cut-boundaries of 5000 and 100000 inhabitants.

A quite similar problem arises in the harmonization of the net monthly income of the household. All countries polled this variable with different currencies and boundaries. Since it would be impossible to construct any somewhat consistent groups over all countries, I had to apply a relative measure of income. In every country the respondents were classified in three groups belonging to the high income group if having an income of over the 66 percent quantile of the respective countries' income distribution, to the medium income class if the respondent's earnings is in-between the 33 percent percentile and 66 percent quantile and to the lower income group if the income is under the 33 percent quantile. Respondents had also the possibility to refuse. These categories are split up in four dummy variables labeled "low income", "medium income", "high income" and "income NA".

High income is correlated with Euro cash and Euro savings deposit possession (see Stix, 2008), hence the estimated effects of the variable income on the nonresponse probability provide valuable insights in the pattern of missing data and provide incidence whether the missing data is MCAR, MAR or NMAR. Significant coefficients would not

only violate the assumption of MCAR but would also due to the high correlation of income with Euro cash and deposit holdings provide incidence (but not evidence) on missing data being NMAR. Respondents with very low incomes and hence probably low Euro cash and savings deposit holdings may refuse to answer because of the fear that the interviewer could disapprove of their low wealth. The fear of social desirability may also be an incentive for persons with high income to refuse, but in my opinion the fear of disclosure to third parties is a stronger motivation for nonresponse in the high income group. My hypothesis is that persons in the lower and upper part of the income distribution have a higher probability of nonresponse than individuals in the medium income group (cf. Biewen, 2001). Uncertainty remains due to the fact that around ten percent of the respondents refused to answer the income question itself. It is possible that respondents reluctant to provide information due to the above explained reasons are concentrated in the income nonresponse group and thus the variable "income na" captures the effects mentioned above. Under the assumption that all persons who provided information on their income have no concerns about trustworthiness, disclosure to third parties and social desirability we might observe no influence of the level of income on the probability of response but a significant effect of the refusing income group. Thus only if we obtain no significant income variable at all indicates that income has no effect on the response inclination.

4.3 Estimation Results

In the following sections the regression results for the five dependent variables will be described. The equations were estimated with the whole sample. As the impacts of nonresponse on currency substitution measures will be later investigated in the case of Serbia, I will also estimate the equations for Serbia only. The reasons for choosing Serbia are that it is the country with the highest currency substitution rate out of the eleven surveyed countries and has also one of the highest nonresponse rates in the sample.

Since it is not necessary to use harmonized definitions of variables in the case of Serbia, some explanatory variables have slightly changed. The income variable is not longer defined as a relative measure, it is now coded in six salary categories "up to 6000 RSD", "6001-12000 RSD", "12001 – 18000 RSD", "18001 – 30000 RSD", "over 30001 RSD" and "Inc No Answer" where RSD denotes the Serbian Dinar. The region dummies "Belgrade", "Vojvodina", "Central-west Serbia" and "South-east Serbia" are used instead of the region's

nonresponse rate variable. All other covariates have the same definitions as previously explained.

In order to preserve the clarity of the results and to not overload the reader with information I decided to discuss only significant variables in detail. Hence, in most of the cases I will not report the sign and the marginal effects of insignificant variables. Readers interested in these figures can take a look at the tables, where all results are reported in detail.

4.3.1 Nonresponse on Question 5.1 – Whether the Respondents Has Euro Cash or Not

In the following paragraphs I will report the estimation results for the determinants of nonresponse on the question 5.1 surveying whether the respondents hold Euro cash or not. The results for the whole sample, sometimes also referred to as Central and Southeastern Europe, are summarized in Table 7 and the Serbian results in Table 13.

Because the observations in Poland could not be classified within the variable "town size", all individuals in Poland must be left out in equations comprising the variable "town size" and vice versa. I thus estimated two equations for the whole sample, on the one hand with town size as an explanatory but observations for Poland excluded and on the other with Poland included and townsize omitted. In addition, I imputed values for the income nonrespondents and reran the estimations to check for the sensitivity of the results. In Table 7 column (1) refers to the equation with "town size", column (2) reports the estimation with the Polish observations included and in column (3) refers to the estimate with the imputed income values. In the Serbian Table 13 column (1) represents the estimation results with the income nonrespondents included as an additional income category and (2) the results with the imputed income values. It must be pointed out that the columns with the imputed values are ordinary probit coefficients while the others represent marginal effects.

→ INSERT TABLE 7 and 13 ABOUT HERE ←

Neither significant gender nor age effects on the response behavior were found in both, the whole sample and in Serbia. In both cases by performing a Wald test for the hypothesis that both coefficients "age" and "age squared" are jointly being equal to zero we cannot reject the null hypothesis.

A series of country dummies could be found to have a significant impact on the response probability. Respondents from Albania and Macedonia have a lower likelihood of refusal in comparison to the reference group of Hungary. On the other hand I obtained a higher probability of nonresponse for Poland (if included), Romania and Serbia. The magnitudes of the revealed effects are nevertheless between small and moderate with marginal effects between 0.006 and 0.03.

If these country effects reflect only differences in the response mentality of the people we can ignore the nonrespondents. But if the probability of nonresponse is related to the unobserved Euro cash holdings, then projections of the gross Euro cash holdings in Central and South-Eastern Europe will be downward biased, since respondents in countries with higher Euro cash holdings will be less likely to reveal their amounts and thus be underrepresented in the sample. To check whether this is the case, I correlated the country dummy coefficients with the projected average Euro cash amounts from Scheiber and Stix (2009). The correlation coefficients is 0.16. In this question it seems that the relation between the probability of nonresponse and the average Euro cash holdings is negligible.

The results for the income coefficients are of particular interest since it is probable that income is positively correlated with the likelihood of Euro cash holdings (cf. Stix 2008). The concerns about the nonrespondents group were proved to be true. All income categories are insignificant with exception of the highest income group in the Serbian sample and the income nonresponding group in both. By interpreting the marginal effects we have to take into account the low observed probabilities which range between 2.5 and eight percent in the whole sample and 5.5 and 19 percent in Serbia. Nonrespondents have an about one percentage point higher likelihood in the case of the whole sample and a five percentage points higher probability in the case of Serbia of refusing than the reference group of persons with low income. The question is whether the level of income has really no effect on the response inclination or persons with certain income characteristics are more likely to refuse and hence the bias the coefficients of the income dummies. To account for this, I imputed income figures for the income nonrespondents and reran the estimation. The results yield that income has really no significant influence on the response probability, neither in the whole sample nor in Serbia.

With regard to the household size no convincingly significant effects are obtained. With only some of the variables being significant at the ten percent level but not robust against different specifications, it is seems that household size is not a determinant.

Being household head or having children does not significantly (with one exception at the ten percent level) affect the probability of nonresponse in all samples and specifications.

With exception of students and pupils in Serbia, who have a lower probability of nonresponse than unemployed persons, all other employment, education and town size coefficients throughout all samples and specifications are not found to be influential on the response inclination.

The next variable measures the influence of whether a respondent prefers holding cash rather than savings deposits on the response behavior. This variable was not found to have a significant influence on the response likelihood in the sample with all observations. In Serbia, respondents preferring cash rather than savings deposits have a significantly lower probability of nonresponse than persons not preferring to save with cash. On the other hand individuals who answered the questions about their preference with "Don't Know" have a lower likelihood of nonresponse. These effects are hard to interpret since the signs stand in contrast to my expectations. It seems not to be the case that persons who prefer to save with cash perceive the questions whether they hold Euro cash or not as more sensitive than others.

The last two variables are concerning somewhat like "network effects" on the nonresponse probability. The results of the variables measuring the respondents' subjective assessment of the dissemination of Euro cash holdings show that persons disagreeing with the statement that holding Euro cash is common in their country have no significant differences in the likelihood of nonresponse than respondents agreeing with the mentioned statement. Nevertheless an interesting pattern is revealed. We obtain a higher probability of nonresponse for persons who answered this question with "don't know". It seems that persons not familiar with the topic of Euro cash holdings are also more reluctant to provide information in whether they hold Euro cash or not. If the reason for answering the assessment question with "don't know" is that they simply do not know and are not familiar with the topic, the resulting higher nonresponse probability on the dependent variable can be explained by too much cognitive afford being demanded for answering this question (cf. cognitive theory in chapter "Item Nonresponse").

The second network variable measuring the respective nonresponse in the region of the respondent is objective. In the case of Serbia the regional dummies fulfill this task. The highly significant regional dummies and region's nonresponse rate suggest strong regional effects on the response inclination.

4.3.2 Nonresponse on Question 9 – The Euro Cash Amounts Being Held

In the next paragraphs I will present the results obtained by the equations describing the nonresponse determinants for the question 9 – surveying the amounts of Euro cash holdings.

Neither significant gender nor age effects are found. Most of the country dummies are significant. Except for Albania, all country dummies have positive signs. Again I correlated the coefficients with the Euro cash amounts from Scheiber and Stix (2009) to test whether the probability is related to the amounts. The correlation coefficient is again rather small with 0.27. With a well certified degree of cautiousness it may be said that the results for Euro cash indicate mentality differences across countries.

→ INSERT TABLE 8 ABOUT HERE ←

An interesting result was obtained for income. In contrast to the equation whether a respondent holds Euro cash or not, not only the income nonresponse group has a significant effect. For the whole sample, persons in the third income quantile have a higher likelihood of nonresponse than persons in the low income class. In Serbia the probability of nonresponse also increases with the income level.

Neither the household size nor the number of children in the household seems to play a determining role in the response inclination. The household head dummies also reveal no significant effects.

Concerning possible employment status effects no convincing results were obtained. While the respondent's education does not have a significant impact on the response likelihood in Serbia, it has an elevating effect on the probability of nonresponse in equations comprising the whole sample.

The obtained results for the variables measuring the preference towards cash rather than savings deposits are contradictory. The results for the whole sample confirm my expectation that respondents not preferring to hold cash rather than savings accounts have a lower probability of refusal than persons preferring cash. But in Serbia no such effect is found. Combined with the results in the question whether the respondent holds Euro cash or not such an effect is disputable.

In the whole sample none of variables measuring the perception of Euro cash dissemination is found to be significant. For the Serbian observations again, the "Don't Know" answers increase the probability of nonresponse.

Similar to the previous equations strong and significant regional effects were found.

4.3.3 Nonresponse on Question 15.1 – Whether the Respondents Holds Euro Denominated Savings Deposits or Not

The equation for the nonresponse of question 15 is estimated the same way as for question 5.1. Again two specifications were estimated, one with the variable town size and without Poland and one with the opposite configuration which are reported two columns in Table 9. The Serbian results are reported in Table 15

No significant gender and age effects are found. The Wald test of jointly testing the hypothesis that age and age2 equal to zero yields also no significance.

→ INSERT TABLE 9 and 15 ABOUT HERE ←

The estimations confirm the results of the descriptive analysis of the nonresponse rate across countries. For those living in Albania, Bosnia, Macedonia, Poland, Romania or Serbia we find a significant higher likelihood of refusal than for persons living in Hungary, the reference group. The marginal effects range between three percentage points (pp) and seven pp. The probabilities of nonresponse are moderately correlated with the projected amounts with a correlation coefficient of 0.52.

A quite different pattern to that obtained in the Euro cash questions is observed for income. Estimated for all observations, nonresponse on the income question again increases the nonresponse probability. In Serbia, the income nonrespondents group has no significant different probability of nonresponse as the reference group with an income up to 6000 SRD. Interestingly and contrary to my expectations, the likelihood of refusing to answer the question whether the respondent holds Euro savings deposits or not decreases with increasing income. Almost all coefficients of the income categories are significant. It is surprising to obtain such different response patterns for Euro cash and Euro savings deposits.

No significant effects are found for household size, household head, number of children and the employment status.

The results for the education dummies reveal that the higher the education of the respondent, the higher the probability of nonresponse. The high education coefficient is highly significant in both samples. The “medium education” level dummy was found to be only compellingly significant in Serbia at the 1 percent level.

Medium town and large town inhabitants have a higher probability of nonresponse than residents of small towns in both samples. But only in Serbia the coefficients are compellingly significant.

Respondents disagreeing to the statement that holding Euro deposits is very common in their country have a lower likelihood of nonresponse than individuals agreeing to the statement, but only in the whole sample the coefficients are statistically different from zero. These findings do not correspond with my expectations. I expected persons perceiving a higher dissemination to be less likely to refuse. The only explanation I can provide is that persons not possessing Euro savings deposits and perceiving that it is common to hold these have refused to answer the question 15.1 in fear of social desirability. However, the previous insignificant and contradictory findings suggest that this variable is not a reliable predictor of nonresponse.

In sample comprising all observations the nonresponse rate in the respondents' region reveals that the probability of refuse increases with the nonresponse rate. In contrast to the strong significant regional differences in the response probability in the previous questions we observe almost no significant differences in Serbia.

4.3.4 Nonresponse on question 17 - The Amounts Being Held on Euro Savings Deposits

In the next few paragraphs I will report the outcome of the analysis of nonresponse on the constructed question 17. The question surveys the amounts being held on savings deposits denominated in Euro. Table 10 reports the estimation results for the whole sample and Table 17 for the Serbian counterparts. Since in many countries the share of euro savings deposit possessors is very low, the sample is very similar to the previous one. Hence, no huge differences are to be expected.

Neither significant age nor significant gender effect were found. With the exception of Slovakia all country dummies are significant. Ceteris paribus while respondents in the Czech Republic have a lower likelihood of nonresponse than in Hungary all other countries' respondents have a higher probability of refusal. The correlation between the obtained probabilities and Euro savings deposits amounts is 0.48.

→ INSERT TABLE 10 and 16 ABOUT HERE ←

We observe a similar pattern of income effects as in the previous equation. The level of income seems only to play a determining role in the response behavior in the Serbian sample. On the other hand, only in the equation with all observations nonrespondents on the income question have a significant higher probability of nonresponse on question 17.

In neither one of the two samples household characteristics influence the response inclination significantly.

The evidence of employment status effects is anything but convincing. With only few dummies being significant at the 10 percent level, it is hardly a driving factor of nonresponse in both samples.

In contrast to the whole sample where high educated persons have a significantly higher probability of 3.5 percentage points of nonresponse than respondents with low education, the evidence for education effects in Serbia is weak. This is in line with the previous findings in the other equations. It seems that while education is a significant predictor in Central and Southeastern Europe, it does not affect the response inclination in Serbia.

In Serbia the results reveal that the nonresponse increases with increasing town size. Respondents in large town are 5.7 percentage points more likely to refuse the answer than rural dwellers. It is interesting that if the respondents were asked about their Euro cash holdings, the town size plays no significant role, but if it comes to the Euro savings deposits it turns significant. It is hard to interpret why there are such differences in the patterns of determinants in such similar topics. In the equations for Central and Southeastern Europe the significance is not robust against the income imputation. Without income imputation, the dummies are hardly significant.

Similar results to the previous question about whether the respondents hold Euro savings deposits or not were obtained for the variable measuring the perceived dissemination of Euro savings deposits. As previously mentioned the results on this variable are generally puzzling and contradicting, thus I must conclude that my expectations turned out not to be true.

In the whole sample, the nonresponse rate of question 17 in the region of the respondents is again highly significant, notable with by far the lowest p-value of all coefficients. In Serbia we observe almost no significant differences in the response inclination among the regions.

4.3.5 Nonresponse on the Income Question

In the next paragraphs the determinants of nonresponse on the household income question will be investigated. The estimation results for Central and Southeastern Europe are reported in Table 11 and the Serbian equations in Table 17**.** Again, in the sample comprising all

observations, two specifications are being estimated, one without the observations of Poland but with town size as explanatory and one with the reverse combination. The model is robust against these two specification, we obtain only slight differences between them.

While in Serbia no gender effects were found, in the whole sample we observe the first time significant differences in the response inclination between males and women. Male respondents have a higher probability of nonresponse than females. Like in the previous equations, the age of the respondents does not seem to influence the response inclination.

→ INSERT TABLE 11 and 17 ABOUT HERE ←

Almost all country dummies show significant coefficients. Although the reference country Hungary does not have the highest nonresponse rate, all coefficients of the country dummies are negative which indicates that ceteris paribus, respondents not living in Hungary have a lower likelihood to refuse than Hungarian persons.

The results further reveal that the household characteristics have a significant influence on the probability of nonresponse in both samples considered. The probability of nonresponse increases with the household size. Two possible effects may drive this finding. With increasing household size it is assumable that the respondents have less spare time and therefore are less willing and less patient in answering the questions. Secondly, since the question surveys the monthly household income, with increasing household members it may become more difficult to have an overview on the sum of monthly income of each household member. The last hypothesis is confirmed by the significant negative coefficient of the household head dummy that indicates lower income nonresponse likelihood for household heads. Since a household head is typically better informed about the incomes of household members than non household heads, he or she will have a higher probability of nonresponse than non households heads.

The effects of the employment variables are qualitatively similar across the samples, but not their significance levels. While in the sample with all observations the coefficients for retired persons and students/pupils are highly significant, we observe in Serbia only a weak significance for employed persons. Retirees have also a lower likelihood of nonresponse than the reference group of unemployed. An interesting result yields the coefficient of the student/pupil dummy. It reveals that students or pupils have a lower response inclination than unemployed persons. This might reflect the low or nonexistent incomes or if the respective person does not administrate its own household simply the lack of information about the incomes of the other household members.

The evidence for education effects in the sample with all observations is weak. The high education dummy is significant at the 10 percent level. In contrast to the previous findings, we observe significant education effects in Serbia. Surprisingly, with increasing education the probability of nonresponse decreases.

In contrast to the previous equations no income variable could be included in the estimations since we are analyzing the nonresponse of the income question. If the nonresponse would be related to income I would expect variables which are assumably correlated with the unobserved income like age, education and employment status to have significant effects on the response inclination. With only minor evidence of an effect of these, that does not seem to be the case.

The likelihood of nonresponse is not monotonically increasing or decreasing with the town size. The evidence for significant town size effects is rather weak in the sample with all observations. In Serbia only the medium town size coefficient is significant.

Such as in all previous equations the higher the nonresponse in the region, the higher is the obtained probability of refusals. In Serbia, we observe in all regions a lower nonresponse probability than in Belgrade.

4.3.6 Discussion and Summary of the Results

Before beginning with the discussion of the estimation results I will consider the reliability of the estimated equations. A bulk of goodness of fit measures are available, I will stick to the most common one, namely McFadden's likelihood ration index. Additionally I report a 2 x 2 table of the predicted versus the actual values by using the prediction rule $\hat{y} = 1$ if $\hat{F} > F^*$ and 0 otherwise where $\hat{F}$ denotes the predicted probability and F^* a user specified threshold value most commonly set to 0.5. Although being a good summary of the predictive ability it has its flaw in the case of unbalanced samples, e.g. the proportion of 1s is much bigger than the proportion of 0s or vice versa.[24] From the descriptive analysis we know that this is the case for all of our five dependent variables. The item nonresponse rate ranges between 2.5 and 7.5 percent thus the samples are especially unbalanced. In such circumstances most models will classify all observations into the group with the highest a priori probability since it needs an extreme configuration of characteristics to achieve a $\hat{F}$ greater than 0.5. The obvious adjustment of reducing the threshold value F^* comes at

[24] See Amemiya (1981)

costs. Any prediction rule of this form makes two errors: Classifying 1s wrongly as 0s and vice versa. Reducing the threshold value increases the correct classification of 1s, but reduces the incorrect prediction of 0s and vice versa.[25] Since the aim of this study is not to develop a model that perfectly predicts nonresponse and of the above mentioned flaws not too much emphasis is given on the prediction tables. I have choose F^* being equal to 0.1 for the questions concerning Euro holdings and 0.3 for the income equations. This seems arbitrary, but there is no correct value to specify and thus I have chosen values near (2-2.5 times) the observed probability of nonresponse.

The McFadden R^2 is in all equations around 0.13 which is quite well in equations with so many observations (around 20000). The joint Wald test of all coefficients being equal to zero is in all equations highly significant.

Table 12 shows that between 81 and 96 percent of the observations are correctly classified. We would obtain an even higher percentage if the threshold value would be set higher.

→ INSERT TABLE 12 ABOUT HERE ←

Not only the overall prediction rate is of interest, but also the one of the 1s and 0s. Between 84 and 98 percent of responses are correctly predicted. The percentage of correctly classified nonresponses is substantially lower. In the equation of question 5.1 only around 15 percent of the nonresponses are correctly predicted. In the other probit equations predict between 40 and 52 percent of the nonresponses correctly.

The above discussed figures are very similar in the Serbian equations and hence their discussion will be omitted here. Nevertheless, table 18 reports all classification tables in detail and interested readers are referred to it.

The reader has to take into account two main concerns of this model. If the missing values are NMAR, i.e. the probability of nonresponse depends on unobserved factors, then these variables are per definition omitted from the model. Two not included variables due to data constraints are considered to potentially have an influence on the response inclination. First and obviously, the true amount of Euro cash or deposits held or earned income is due to the nonresponse not completely collected and hence omitted in the equations. Secondly, the data set contains no interviewer information. The literature suggests on the one hand that interviewer variables have not much explanation power (cf. Sousa-Poza and Henneberger, 2000) and on the other hand provides no consistent evidence on interviewer characteristics

[25] See Greene (2008)

except one: All studies considered by this diploma thesis (cf. chapter "Empirical Evidence") found that interviewer's gender have an influence on response probability, notably that respondents are more likely to refuse in front of female interviewers. In the OLS case, if the omitted and included variables are orthogonal, then the coefficients are unbiased. Yatchew and Griliches (1985) showed that in the probit case even if the variables are uncorrelated with each other, the coefficients are biased. In a perfect world, all relevant information is at disposal to the applied researcher. Unfortunately, especially when working with microeconomic data, it is almost impossible to capture all relevant data which influences the individual decision making. Being aware of the limitations of the reported goodness of fit measures R^2 and prediction tables, I conclude with a well certified degree of cautiousness that they do not indicate flaws in the models and thus I assume the estimation result being trustworthy. But uncertainty will always remain up to a specific degree and it is important to be aware of it.

The second concern is the nonresponse of income. Around 10 percent of the respondents refused to answer the income question. Thus especially where the nonresponse group in the estimations is significant, the income effect may be unreliable. To check the sensitivity of the results, I reran the estimations with imputed values for income. The results do not change qualitatively. It seems that income nonrespondents are not different enough from respondents to alter the estimation results.

After having discussed the limitations and robustness of the estimations we can pass on to the interpretation and summary of the results. By examining the impacts of nonresponse on currency substitution measure we do not have to bother of all significant variables. The nonrespondents and respondents can be systematically different and the results will nonetheless be unbiased. The important point is that the refusing group must not differ systematically in characteristics which are known (or unknown) to be related to the variable of interest, in our case the Euro cash and deposit holdings. For example, if respondents living in bigger households are more likely to refuse but not more likely to hold greater amounts of Euro cash, then our calculations of Euro cash holdings will be unbiased. But great cautiousness is advisable because many variables are possibly correlated with unobserved, which may be related to the variable of interest. In our sample only the household size may be a respondent's characteristic that may not be related to Euro cash or deposit holdings.

It is surprising that the determinants for the relatively similar topics of Euro cash and Euro savings deposit holdings are so different. Not only that different variables are

identified as determinants, but some effects also changed their signs. While with increasing income the probability of refusing to answer the questions about the Euro cash amounts increases, it decreases for Euro savings deposits. This confirms the suggestion that all questions are different and it is not trivial to guess the determination patterns of item nonresponse.

The results further show that the assumption for applying complete case analysis is in all cases clearly violated. The probability of nonresponse depends at least on observed factors. The likelihood of nonresponse does also depend on variables that might be correlated with the Euro cash or Euro savings deposit holdings. In the next section I will prove whether these observed differences in the probabilities of nonresponse have a quantitative impact on currency substitution measures in the case of Serbia.

4.4 Impacts of Item Nonresponse on Currency Substitution Measures

In the previous chapter we concluded that the missing data seems not to fulfill the assumption of MCAR for applying complete-case analysis. One of the aims of the OeNB Euro Survey is to present a picture of the extent of currency substitution in the surveyed countries. Hence, it is essential to know the potential bias of item nonresponse on currency substitution measures. I will focus my analysis on Serbia, the country with one of the highest currency substitution rates in Europe and with one of the highest nonresponse rates in the sample. I will impute the missing values due to item nonresponse by applying multiple imputation. This notion was already introduced in chapter “Treatment of Missing Data”. To account for the uncertainty implied by imputation, multiple values are substituted for the missing ones rather than one single value. This will enable us not only to quantify the impact of item nonresponse on currency substitution measure, but also to investigate whether nonrespondents and respondents hold statistically significantly different amounts of Euro cash and on savings deposits. Several methods for imputation were briefly introduced in subchapter “Treatment of Missing Data”. I have chosen the one introduced in Van Burren et al (1999). Royston implemented this procedure in STATA, named as ice (Royston 2004, 2005a 2005b). Ice imputes missing values by using an iterative multivariable regression technique, namely switching regressions.

The regression models are reported in Table 19. Van Burren et al. (1999) recommend including both, variables that predict the variable to be imputed and covariates of the response equation. Hence, the selection of prediction variables are based on the model of Stix (2008), who investigated the determinants of Euro cash and Euro savings deposit possession and on my results of the nonresponse equations.

→ INSERT TABLE 19 ABOUT HERE ←

In all prediction equations all available socio-demographic variables were included. In addition, in the models predicting Euro cash and Euro savings deposits possessions, respectively the amounts being held, the inflation expectations of the respondent for the next twelve months, the EUR/RSB exchange rate expectations for the next five years, the number of his or her visits in the Euro zone and whether he or she noticed big or small payments in Serbia were included as predictors.

Table 20 reports the obtained dissemination of Euro cash and Euro savings deposits. For complete case analysis – which is denoted in the table as observed cases – we obtain that 41.45 percent of the non refusing respondents hold Euro cash. For the imputed values in the table denoted as "unobserved" – we observe a slightly higher share of Euro cash holders, namely almost 43 percent. If we regard the full data set, i.e. the observed and imputed values together, the differential nearly vanishes due to the low nonresponse rate.

Table 20: Dissemination of Euro Cash

Q5.1: Do you hold Euro cash?

	Yes	No
Observed cases	41.45	58.55
Unobserved cases	42.94	57.06
All cases	41.59	58.41

Test statistic for zero differences between respondents and nonrespondents	
z	-0.24

Note: Test statistic z is approximately standard normal distributed. The critical value at a significance level of 5% is 1.96
Source: OeNB Euro Survey

But are the differences between respondents and nonrespondents statistically significant, or can they be explained by potential variation in the data? A test of zero proportion difference is applied to answer this question. The reported test already accounts

for the uncertainty implied by the imputation. We cannot reject the null hypothesis of the proportion difference being zero. Although we observe slightly higher Euro dissemination for the nonresponding group, the difference is not statistically significant. Nor is it economically significant since the complete case analysis reveals almost the same share of Euro cash holders than the multiple imputation analysis.

The results for the Euro deposits dissemination are summarized in table 21. The share of Euro savings deposit holders is about 2 percentage points higher for nonrespondents than for respondents. This effect may be perceived as small, but taking into account that the share is only around nine percent for the observed cases, the proportion jumps about 20 percent.

Again, the difference is statistically not significant. The computed share for the observed and unobserved cases together is only slightly different (0.08 percentage points) from the complete case proportion. Taking into account that the underlying data was obtained by a survey, the difference is not worth mentioning. With such narrow differences in the results it is hard to see that the extra effort of imputation is really worth it.

Table 21: Dissemination of Euro denominated Savings Deposits

Q15.1: Do you hold Euro savings deposits?

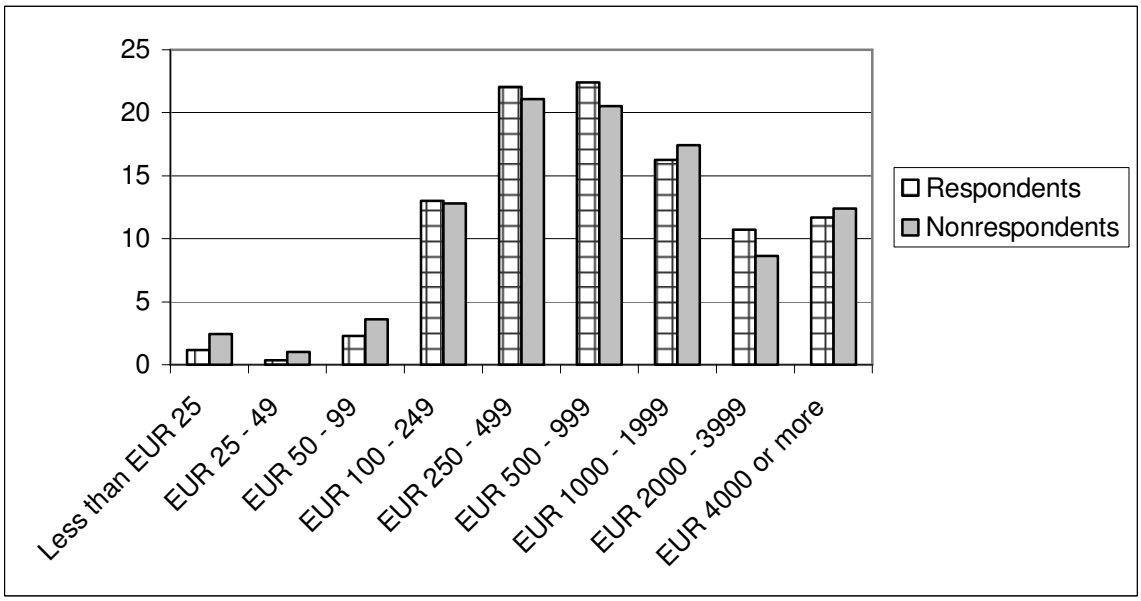
	Yes	No
Observed cases	8.86	91.14
Unobserved cases	11.01	88.99
All cases	8.94	91.06

Test statistic for zero differences between respondents and nonrespondents	
z	-0.86

Note: Test statistic z is approximately standard normal distributed. The critical value at a significance level of 5% is 1.96
Source: OeNB Euro Survey

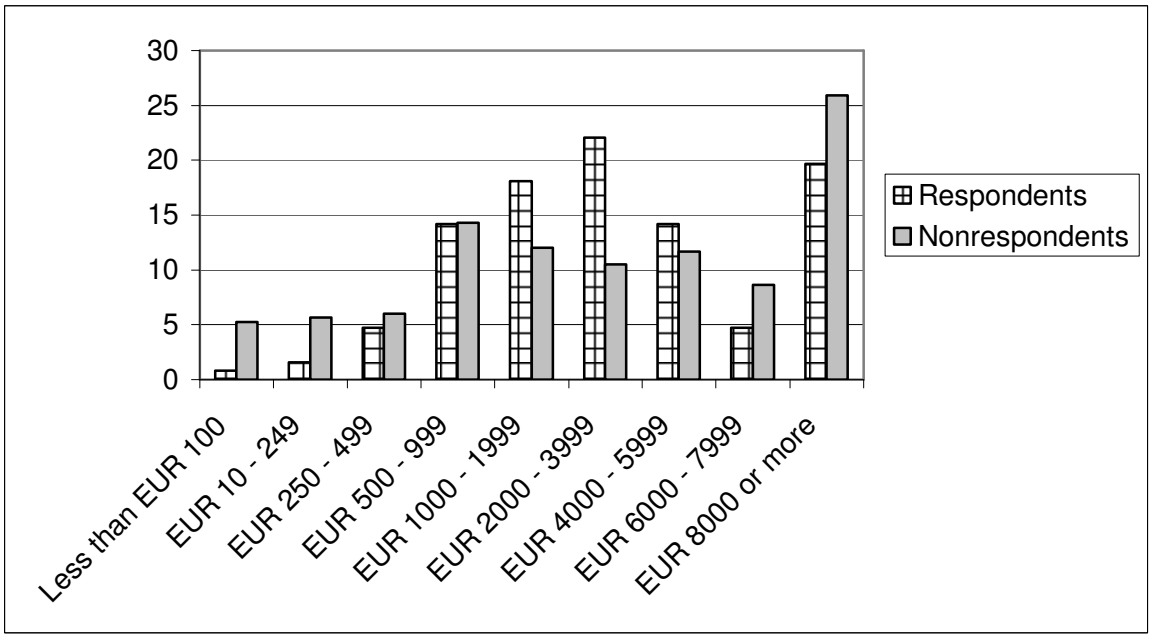
In the next few paragraphs I will investigate the differences in the distribution of Euro cash and Euro savings deposit amounts between respondents and nonrespondents, depicted in Figure 3 and Figure 4. The imputation reveals the expected results, namely that nonrespondents are accentuated in the upper and lower tails of Euro cash and Euro deposits distribution.

Figure 3: Distribution of Euro Cash Amounts for Respondents and Nonrespondents



Source: OeNB Euro Survey

Figure 4: Distribution of Euro Savings Deposit Amounts for Respondents and Nonrespondents

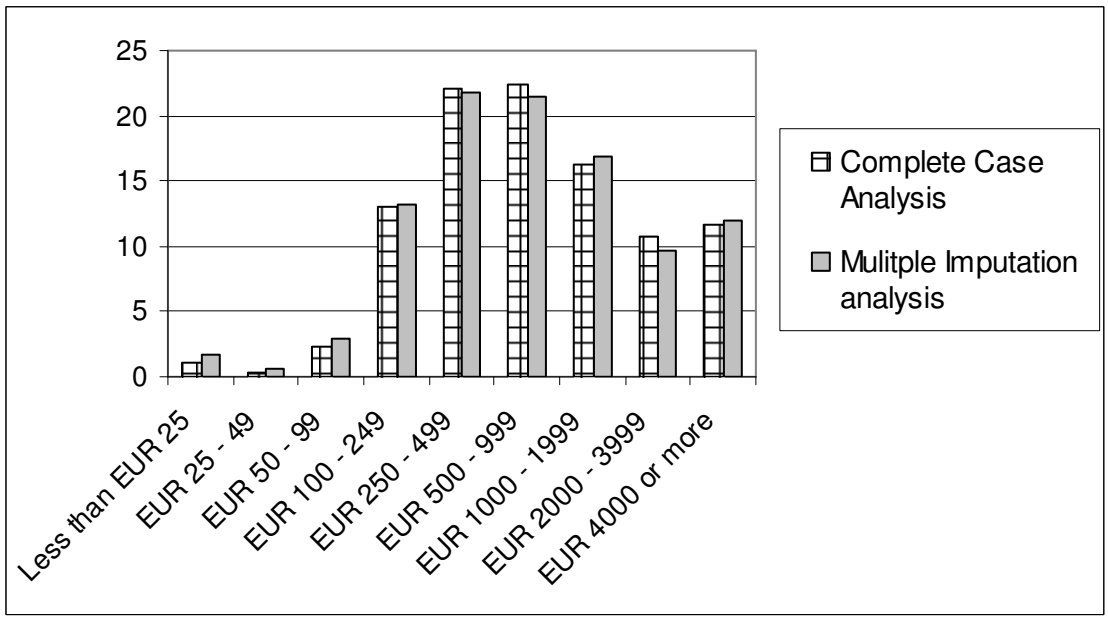


Source: OeNB Euro Survey

This effect is somewhat stronger for Euro deposit amounts than for Euro cash. In the cash amount distribution of nonrespondents more weight is put on the lower cash intervals. While 17 percent of the respondents hold less than 249 Euro, about 20 percent of the nonrespondents possess Euros up to this figure. There is almost no difference in the relative weight of the two highest cash amount category between respondents and nonrespondents. The shape of savings deposit distribution for respondents differs to a greater extent from the respondents' one. The responses are much more accentuated in the tails of the distribution. While 21 percent of the respondents hold less than 249 Euro and 24 percent hold more than 2000 Euro, these figures are 31 percent and 35 percent for the nonrespondents.

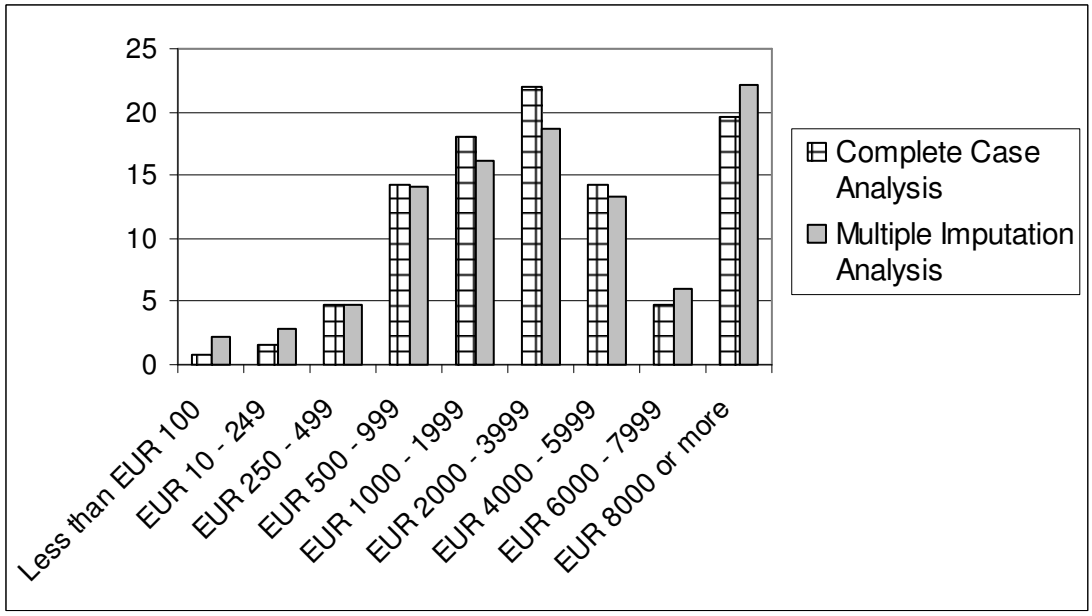
The difference between complete case analysis and multiple imputation analysis is quite moderate in the case of cash amounts. The highest difference in relative frequencies is about 1 percentage point. For Euro savings deposits the differences ranges between nearly zero and around 3.5 percentage points. Figure 5 and Figure 6 depicts the distribution of cash and savings deposits for complete case and multiple imputation analysis.

Figure 5: Distribution of Euro Cash Amounts for Complete Case and Multiple Imputation Analysis



Source: OeNB Euro Survey

Figure 6: Distribution of Euro Savings Depost Amounts for Complete Case and Multiple Imputation Analysis



Source: OeNB Euro Survey

The above obtained results are also mirrored by the median amounts which are computed by linear interpolation between classes. The complete case medians for respondents holding Euro cash is 733 Euro and for savings deposit 3188 Euro. The median amounts for nonrespondents is 702 and 3545 Euro for cash and savings deposits respectively or in relative figures, 4 percent lower respectively 11 percent higher than for observed cases.

Table 22: Median Amounts of Euro Cash and Euro Savings Deposits

	Euro Cash	Euro Savings Deposits
Complete Case Analysis	733.5	3187.8
Multiple Imputation	715.9	3257.6
Nonrespondent	702.0	3545.0

Note: The table reports the median amounts of Euro cash and Euro denominated savings deposit holdings for Euro cash and Euro deposits holders, respectively.
Source: OeNB Euro Survey

To test whether the obtained figures for the nonrespondents are statistically significantly different from the respondents' ones, the Wilcoxon rank-sum test is applied. Unfortunately we cannot incorporate the uncertainty from imputation into nonparametric

tests. To account for the uncertainty all the same, I apply the Wilcoxon test on all five imputations. All five tests for Euro cash are insignificant. This leads us to the conclusion that the medians for respondents and nonrespondents are statistically not significantly different.[26] In the case of Euro savings deposits, except of one, all tests are at least at the 10 percent level significant. In particular, one of them is significant at the 10 percent, two are significant at the 5 percent and one at the 1 percent level.

Scheiber and Stix (2009) investigated the extent of currency substitution using amongst others the three currency substitution measures CSI, DSI and EI. The definitions were already provided in chapter Currency Substitution.

The obtained figures for Euro cash and Euro savings deposit holdings are first projected on the whole population and then related to monetary aggregates according to the respective formulas. Table 23 reports the obtained figures for complete case analysis and multiple imputation analysis. The potential bias of complete case analysis may arise from uncertainty of three different questions. First, the amounts being stated may be biased. Secondly, if the proportion of Euro cash or Euro savings deposit holders is different for respondents and nonrespondents, then the complete case projections may over or understate the real amounts. Scheiber and Stix (2009) further corrected the projections for jointly held Euro cash and savings deposit amounts. Thirdly, the respondents were asked whether the stated amounts are being held personally or jointly with their partners. This potential cause of bias is disregarded in my calculations because of two reasons. First of all, the question is in my opinion not sensitive and hence may not be exposed to under or over reporting. Secondly, only a small fraction of around 3 percent of the respondents withholds information. Because I do not correct for jointly holdings, my obtained projections differ from the ones calculated by Scheiber and Stix (2009).

Table 23 reports different currency substitution measures obtained by complete case and multiple imputation analysis. The differences of the per capita holdings are quite small. The obtained figure for per capita Euro cash holdings is with multiple imputation analysis about five Euros or one percent lower than with complete case analysis. The per capita amounts on savings deposits denominated in Euros are about 12 Euros or 3.5 percent higher with the imputed values than with complete case analysis.

[26] It is not strictly true, that the Wilcoxon rank sum test tests for differences in medians. The null hypothesis is that the two samples are drawn from a single population.

Table 23: Currency Substitution Measures for Complete Case and Multiple Imputation Analysis.

	Complete Case Analysis	Multiple Imputation Analysis
Euro Cash Holdings per Capita	581.4	576.2
Euro Savings Deposits Holdings per Capita	352.6	364.8
Currency Substitution Index (CSI)	82%	81%
Deposit Substitution Index (DSI)	94%	95%
Euroization Index (EI)	86%	86%

Source: OeNB Euro Survey and Statistical Bulletin June 2008 of the Serbian National Bank

The three currency substitution indexes were already introduced in the chapter "Implications of Currency Substitution". As a reminder, the CSI relates the foreign currency in circulation to the overall currency in circulation (local and foreign), the DSI compares the Euro savings deposits holdings to the sum of local and Euro savings deposit holdings. The EI is the combination of the two indexes.

The minor differences between complete case and imputation analysis are also mirrored in the currency substitution indexes. They are in all three cases just about one percentage point. Taking into account the high figures, they are not worth mentioning – to put it bluntly. This picture could have been already guessed from the low differences in the dissemination rates of Euro cash and Euro savings deposit holder and per capita holdings.

This leads us to the conclusion, that although the MCAR assumption is clearly violated and nonrespondents differ from respondents – in some cases significantly – the number of nonrespondents is too low and the differences too narrow to bias the results even slightly. The extra effort of multiple imputation seems to be not worth it.

Nevertheless, rash conclusions on data reliability are unfortunately misplaced. The applied imputation method corrects only for observable differences. Statistics of the Serbian National Bank show that the per capita foreign currency savings on savings deposits – there are no such figures for foreign currency cash available – are about twice the figure obtained for per capita Euro savings deposit holdings by the survey. Since the Euro is the predominant currency in Serbia, it is likely that the true share of per capita Euro savings deposit holdings on foreign currency savings deposits is higher than the obtained 50 percent. Hence it is plausible that the nonresponse depends on unobservable factors, which means that the type of missing data is NMAR. The probability of refusal may depend on the Euro

cash or Euro savings deposit amount. After controlling for all other observed factors, it seems that persons with higher amounts tend to withhold information more often than participants with relatively low amounts. One other important factor may be that the given answers deviate from the truth. The respondents must not even have an iniquitous motivation (in the researcher's point of view) for the deviation, but it may be the case that some persons simply forget to take account of one or more long term savings deposit.

The maybe disappointing conclusion is that imputation does not achieve the desired correction for underreporting and we must interpret the results with a well certified degree of caution.

5 Conclusion

Despite the lively discussion of item nonresponse in the literature in the last two decades, the most common method of dealing with it is still simply to ignore it. Complete case analysis is the default treatment method in almost all statistical packages. Despite its convenience it is important to take into account the strong assumptions of it. In all cases where the nonresponse depends not only on randomness, the obtained results will be biased.

This diploma thesis aims at contributing to the item nonresponse literature in the following ways: Besides introducing the reader in the topic of item nonresponse, it provides an applied example of the item nonresponse problem and an investigation on the question whether the commonly applied complete case analysis is appropriate this case. It is known from previous studies that all surveys concerning the personal wealth of the respondents are particularly exposed to refusals. From national authority data it becomes clear that the OeNB Euro Survey is also affected by underreporting. In a first step I tried to identify the determinants of item nonresponse for the key questions in the questionnaire.

The mechanism of nonresponse is very difficult and many opposing effects are determining the response inclination. It is hard – if not say impossible – to guess or expect which effect will dominate. Let me pin this down with an example: A higher educated person may more endorse scientific work or research than a respondent with less academic background. On the other hand, the opportunity costs rise with higher education. A priori, it is hard to guess which effect outweighs the other.

The results suggest that nonresponse behavior is not driven by any general pattern or characteristics which would be valid in all cases. Consequently, questions and countries cannot be lumped together. Strong country and regional effects suggest that the likelihood of answering might depend on mentality. The response equations reveal also that – as expected – questions concerning whether the respondents hold Euro or not are not as sensitive than questions surveying the exact amounts being held. Not even the response equations of Euro cash and Euro savings deposits are comparable. In every question, not only other determinants seem to drive the response inclination, but often the effects for one and the same variable changes. In the case of Serbia the response inclination for the Euro cash amounts decreases with higher income but increases with it for the questions surveying the Euro savings deposit amounts. Combined with previous contradictory empirical findings, the results reveal that one cannot derive general conclusions about the response inclination for specific respondent characteristics**.** The applied researcher must always separately

investigate the determinants to get a grip of the nonresponse mechanism and nonresponse characteristics.

From the response equations it became clear that the assumption of missing values being MCAR for applying complete case analysis is clearly violated. In order to quantify the bias of complete case analysis, I performed additionally the multiple imputation analysis that corrects for observable differences between the respondents and nonrespondents and allows us also to investigate whether these differences are statistically significant or just a possible random variation. The obtained figures yield in most cases no statistically significant differences between respondents and nonrespondents. But even in the case of statistical differences – the amounts of Euro holdings on foreign currency savings deposits – these are too narrow and there are too few nonrespondents to have a significant impact – in an economical sense – on the investigated currency substitution measures. The differences in the calculated measures (per capita Euro cash holdings, per capita Euro savings deposit holdings, CSI, DSI, EI) between complete case analysis and imputation analysis do not justify the extra effort of imputation. The differences in per capita holdings are only 1 and 3.5 percent and in the case of the indexes never above 1 percentage points.

Thus, the observed differences in the characteristics of respondents and nonrespondents cannot explain the underreporting. It is very probable that the nonresponse not only depends on observed, but also on unobserved factors such as the true amounts of Euro holdings. A further possible factor is that the answers may simply deviate from the truth. Respondents may not want to tell their true holdings or just forget to take into account some long term deposits they are not frequently confronted with.

As previously mentioned, a general conclusion that is also valid for other surveys is difficult to draw, but at least within the investigated framework this diploma thesis provides good news for persons hoping to find confirmation for complete case analysis, but may be disappointing for those researchers who put hope into imputation analysis to correct for some of the underreporting.

6 References

Allison, Paul David (2002): "Missing Data" Sage university papers: Quantitative applications in the social sciences 136, Thousand Oaks, California.

"Baltic brink" The Economist December 20th 2008, Online Version (March 25th 2009) http://www.economist.com/world/europe/displaystory.cfm?story_id=12815635

Biewen, M. (2001): "Item nonresponse and inequality measurement: Evidence from the German earnings distribution." Allgemeines Statistisches Archiv, 85: 409–425.

De Maio, Theresa J., (1980): "Refusals: Who, where and why", Public Opinion Quarterly 44, 223-233.

Feige, Edgar L. (1994): "The Underground Economy and the Currency Enigma." Public Finance, 1994, 49, pp. 119-39. Reprinted in Fiorentini, Gianluca and Zamagni, Stefano, eds. The Economics of Corruption and Illegal Markets, International Library of Critical Writings in Economics, Edward Elgar, 1999.

Feige, Edgar L. (1997) "Revised Estimates of the Size of the U.S. Underground Economy: The Implications of US Currency Held Abroad." in Lippert, Owen and Walker, Michael, eds. The Underground Economy: Global Evidence of its Size and Impact, Fraser Institute, 1997.

Feige, E. L. (2003). "Dynamics of Currency Substitution, Asset Substitution and De Facto Dollarisation and Euroisation in Transition Countries", Comparative Economic Studies 45(3): 358–83.

Fischer, S. (1982): "Seignorage and the case for a national money". Journal of Political Economy 90(2): 295–313.

Frick, J. R. and Grabka, M. M. (2005): "Item non-response on income questions in panel surveys: incidence, imputation and the impact on inequality and mobility." Allgemeines Statistisches Archiv, 89: 49-61

Greene, William H. (2008): "Econometric Analysis", 6th edition, Pearson Prentice Hall, Upper Saddle River, New Jersey

Griliches, Zvi (1986) "Economic Data Issues", in: Z. Griliches and M.D. Intriligator (eds.), Handbook of Econometrics-Volume III, Elsevier Science Publishers, North Holland, 1465-1514.

Harrell, Frank E. (2001): "Regression Modeling Strategies: With Applications to Linear Models, Logistic Regressions, and Survival Analysis" New York: Springer-Verlag; 2001.

Hartley, H.O. and R.R. Hocking (1971), "The Analysis of Incomplete Data", Biometrics 27, 783-808.

Heckman, J (1976): "The Common Structure of Statistical Models of Truncation, sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models." Annals of Economic and social Measurement, 5, 1976 pp. 475-492

Heckman, J. (1979): "Sample selection bias as a specification error". Econometrica, 47, 153 – 61.

Ize, A., Levy-Yeyati, E., (2003): "Financial dollarization". Journal of International Economics 59, 323–347.

Kalton, Graham (1983): "Compensating for missing survey data" Research Report Series, Institute for Social Research, University of Michigan, Ann Arbor.

Kraft, E. (2003), "Monetary Policy under Dollarisation: The Case of Croatia", Comparative Economic Studies, 45(3): 256-277.

Lillard, Lee & Smith, James P. & Welch, Finis (1982): "What Do We Really Know About Wages: The Importance of Nonreporting and Census Imputation," The Rand Corporation, Santa Monica, CA.

Lillard, Lee & Smith, James P & Welch, Finis, (1986). “What Do We Really Know about Wages? The Importance of Nonreporting and Census Imputation,” Journal of Political Economy, University of Chicago Press, vol. 94(3), 489-506.

Little, R.J.A., Su, H.L. (1989): “Item nonresponse in panel surveys. In D. Kasprzyk, G. Duncan, & M. P. Singh (Eds.), Panel surveys(pp. 400-425). John Wiley, New York

Little, R.J.A. (1976), Inference about means from incomplete multivariate data, Biometrika 63, 593-604.

Little, R.J.A. and Donald B. Rubin (1987, 2002), Statistical analysis with missing data, 1. / 2. edition, John Wiley & Sons Inc., Hoboken, NJ.

Luca, A., Petrova, I., (2008): What drives credit dollarization in transition economies?, Journal of Banking & Finance 32, 858-869

Moore, Jeffrey C., Linda L. Stinson, and Edward J. Welniak Jr. (2000). “Income Measurement Error in Surveys: A Review.” Journal of Official Statistics 16(4):331–62.

Porter, R (1993). Appendix to G. Hess, et al., „Nominal Income Targeting with the Monetary Base“, in M. Goodfriend and D.Small (Eds.) Operating Procedures and the Conduct of Monetary Policy; Conference Proceedings in Finance and Economics Discussion Series, Working Studies 1, Part 2, Federal Reserve Board.

Porter, Richard, and Judson, Ruth (1996) “The Location of US Currency: How Much is Abroad?” Federal Reserve Bulletin, 1996, 82, pp. 883 – 903.

Rässler, Susanne & Riphahn, Regina (2006): "Survey item nonresponse and its treatment," AStA Advances in Statistical Analysis, Springer, vol. 90(1), pages 217-232, March.

Riphahn, Regina T. and Oliver Sering (2002), Item non-response on income and wealth questions, IZA Discussion Paper No. 573, IZA Bonn.

Royston, P. (2004): “Multiple imputation of missing values” Stata Journal 4: 227-241.

Royston, P. (2005a): “Multiple imputation of missing values: update” Stata Journal 5: 188-201.

Royston, P. (2005b): “Multiple imputation of missing values: update of ice. Stata Journal 5: 527-536.

Rubin, Donald, B. (1972), A Non-iterative Algorithm for Least Squares Estimation of Missing Values in Any Analysis of Variance Design, The Journal of the Royal Statistical Society, Series C (Applied Statistics) 21, 136-141.

Rubin, Donald, B. (1974), Characterizing the estimation of parameters in incomplete data problems, Journal of the American Statistical Association 69, 467-474.

Rubin, Donald, B. (1976), Inference and missing data, Biometrika, 63, 581-592.

Rubin, Donald b. (2004): “Multiple Imputation for Nonresponse in Surveys” John Wiley Interscience, Hoboken, NJ.

Schafer, Joseph L. (1997): “Analysis of incomplete multivariate data”, 1. Edition Chapman & Hall, London

Scheiber, Thomas; Stix, Helmut (2009): “Euroization in Central, Eastern and Southeastern Europe – New Evidence on Its Extent and Its Causes” mimeo

Schobert, F. (2002): Seigniorage: An Argument for a National Currency? CEPS Research Report No. 28, August 2002

Schräpler, Jörg-Peter (2004), Respondent Behavior in Panel Studies. A Case Study for Income Nonresponse by Means of the Germany Socio-Economic Panel (SOEP), Sociological Methods and Research 33(1), 118-156.

Schulz, M.: “Euro und Pfund kurz vor Parität”, Financial Times Deutschland Online Version (March 25th, 2009) http://www.ftd.de/boersen_maerkte/aktien/anleihen_devisen/:Renten-und-Devisen-Euro-und-Pfund-kurz-vor-Parit%E4t/455425.html

Sousa-Poza, Alfonso and Fred Henneberger (2000): Wage data collected by telephone interviews: an empirical analysis of the item nonresponse problem and its implications for the estimation of wage functions, Schweizerische Zeitschrift für Volkswirtschaft und Statistik Vol. 136(1), 79-98.

Stix, H. (2008): “Euroization: What Factors Drive its Persistence? Household Data Evidence for Croatia, Slovenia and Slovakia”. OeNB Working Paper 140.

Statistical Bulletin June 2008, The National Bank of Serbia, Belgrade 2008

Stocké, Volker & Stark, Tobias, (2006) "Trust in Surveys and the Respondents’ Susceptibility to Item Nonresponse," Sonderforschungsbereich 504 Publications 06-06, Sonderforschungsbereich 504, Universität Mannheim & Sonderforschungsbereich 504, University of Mannheim.

Tourangeau, R., (1984): “Cognitive Science and Survey Methods.” Pp. 73–100 in Cognitive Aspects of Survey Methodology: Building a Bridge Between Disciplines, edited by T. Jabine, M. Straf, J. Tanur, and R. Tourangeau. Washington, DC: National Academic Press.

Trometer, Reiner, 1996, Warum sind Befragte "meinungslos"? Kognitive und kommunikative Prozesse im Interview, Inauguraldissertation an der Universität Mannheim, mimeo.

van Buuren, S., H. C. Boshuizen, and D. L. Knook. (1999): “Multiple imputation of missing blood pressure covariates in survival analysis” Statistics in Medicine 18: 681-694.

Webpage of the Federal Reserve (March 25th 2009): http://www.federalreserve.gov/paymentsystems/coin/currcircvalue.htm

"Who's next?" The Economist October 25th 2008, Online Version (March 25th, 2009) http://www.economist.com/world/europe/displaystory.cfm?story_id=12465279

Yatchew, Adonis & Griliches, Zvi, (1985): "Specification Error in Probit Models" The Review of Economics and Statistics, MIT Press, vol. 67(1), pages 134-39, February.

Zweimüller, Josef, 1992, Survey non-response and biases in wage regressions, *Economics Letters* 39, 105-109.

7 Appendix

Table 1: Nonresponse Rates for All Questions

Qestion Number	Topic of the question	Don't Know	No Answer
q1_01	Questions concerning the assessment of the economic and financial situation of the country and question about the Euro	2.2	0.9
q1_02		8.2	1
q1_03		5	1.2
q1_04		12.9	1.5
q1_05		14.3	2.8
q1_06		8.2	1.4
q1_07		18.8	3.1
q1_08		5.8	3.1
q1_09		6.7	1.4
q1_10		12.1	1.5
q1_11		12	1.6
q1_12		10.3	1.4
q1_13		5	2
q1_14		15	2
q2	Perceived inflation	0	1.3
q3	Expected inflation	3.7	1.1
q4	Expected exchangerate development	0	18.4
q5_1	Holding Euro cash	0.4	2.1
q5_2	holding US dollar cash	0.5	3.3
q5_3	Holding Swiss franc cash	0.5	3.4
q5_4	Holding other chash	0.5	5
q6a	Personal or joint holdings	0.8	1.2
q7_1	Motivations for holding Euro cash	0.6	2.1
q7_2		1	2.3
q7_3		1.6	2.8
q7_4		1.1	2.5
q7_5		4	2.6
q7_6		4.4	2.5
q7_7		2.1	2.9
q7_8		4.7	3.1
q7_9		1.4	4
q8_1	Motivations for holding US dollar cash	2	4.6
q8_2		2	4.6
q8_3		1.5	4.5
q8_4		2.3	5.2
q8_5		3.8	4.9
q8_6		2.5	5
q8_7		3.5	5
q8_8		2.2	7.2
q9	Amounts of Euro and Dollar	2.1	13.7
q10		1.5	17.5
q11_1	Questions concerning savings deposits	31.2	4.7
q11_2		22.9	4.8
q11_3		20.1	4.3
q11_4		22.9	4.4

Note: The table summarizes the share of respondents answered with "Don't Know" and "No Answer"

Table 1: (cont'd)

q12a	Questions concerning savings deposits denominated in local and foreign currency	0.4	4.2
q12b		0.5	5.1
q13		0.1	1.2
q14		4.5	9.3
q14a		0.6	5.8
q15_1	Holding savings deposits in serval currency	0.1	2.1
q15_2		0.6	9.9
q15_3		0.6	11.1
q15_4		0.7	12.5
q16_1	Motivations for holding Euro savings deposits	0.2	1.8
q16_2		1.6	2.3
q16_3		1.5	2.4
q16_4		0.9	2.7
q16_5		1.9	2.5
q16_6		2.5	4.5
q16_7		2.2	4.6
q16_8		5.3	3.7
q16_9		2	4.7
q17	Amounts held on Euro or Dollar savings deposits	1.5	21.4
q18		1.3	19.5
q19_1	Source of Euro holdings	4	28.7
q19_2		4	28.7
q19_3		4	28.7
q23_1	Visits	0	7
q23_2		0	10.2
q23_3		0	8.5
q24	Expected Euro introduction	25	4
q25a	Noticed payments	8.5	1.8
q25b		9	2.6
q26	Make any payments	1	2.5
s01	Gender	0	0
s02	Age	0	0
s03	Head of Household	0	0.1
s04	Size of Household	0	0
s05	Number of Children	0	0
s06	Number of Children Age 1	0	0
s07	Number of Children Age 2	0	0
s08	Number of Children Age 3	0	0
s09	Number of Children Age 4	0	0
s10	Number of Children Age 5	0	0
s11	Employment status	0	0.1
s12	Profession	0	16.1
s13	Region	0	0
s14	Size of town/village	0	0.3
s15	Education	0	0.1
s16	Net income of household (monthly)	0.1	10.2

Source: OeNB Euro Survey

Table 2: Nonresponse rates for selected questions

	Albania	Bosnia	FYR Macedonia	Bulgaria	Croatia	Poland
Q5) Do you at present hold any foreign cash, that is banknotes and coins and not deposits at a bank account. - Euro						
Do not know	0.0	0.3	0.0	0.3	0.2	1.2
No answer	0.2	0.8	0.2	2.4	1.3	4.3
Nonrepsonse	0.2	1.2	0.2	2.7	1.5	5.5
Q9) I am now showing you a card with different amounts. Could you choose the range that best fits the amount of euro cash that you hold? Please refer to banknotes and coins only, not deposits at a bank account.						
Do not know	0.0	4.7	1.1	1.7	2.4	3.5
No answer	0.0	27.8	10.3	30.6	13.2	17.5
Nonresponse	0.0	32.5	11.4	32.2	15.6	21.0
Q15) You mentioned that you have foreign currency savings deposits. In which currency are these savings deposits denominated? - Euro						
Do not know	0.0	0.0	0.0	0.0	0.0	1.8
No answer	1.7	2.2	3.6	2.2	0.2	5.4
Nonresponse	1.7	2.2	3.6	2.2	0.2	7.1
Q17) I am now showing you a card with different amounts. Could you choose the range that best fits the amount of savings deposits that you have in euro?						
Do not know	0.4	3.7	1.1	0.9	1.1	2.6
No Aanswer	11.5	37.0	10.0	56.4	22.7	25.6
Nonresponse	11.9	40.7	11.1	57.3	23.7	28.2
S16) Net income of household (monthly)						
No Answer	13.2	28.3	8.7	19.1	8.3	0.0

Note: Nonresponse denotes the sum of “Don’t Know” and “No Answers”
Source: OeNB Euro Survey

Table 2: (cont'd)

	Romania	Serbia	Slovakia	The Czech Republic	Hungary	Whole Sample
Q5) Do you at present hold any foreign cash, that is banknotes and coins and not deposits at a bank account. - Euro						
Do not know	1.2	0.4	0.2	0.3	0.0	0.4
No answer	6.6	5.1	1.2	0.5	0.8	2.1
Nonrepsonse	7.8	5.5	1.4	0.8	0.8	2.5
Q9) I am now showing you a card with different amounts. Could you choose the range that best fits the amount of euro cash that you hold? Please refer to banknotes and coins only, not deposits at a bank account.						
Do not know	1.4	7.3	0.2	0.4	0.5	2.1
No answer	29.2	26.9	3.9	5.3	15.5	13.7
Nonresponse	30.7	34.2	4.1	5.6	16.1	15.8
Q15) You mentioned that you have foreign currency savings deposits. In which currency are these savings deposits denominated? - Euro						
Do not know	0.0	0.6	0.0	0.0	0.0	0.1
No answer	5.4	3.0	0.0	0.0	2.0	2.1
Nonresponse	5.4	3.6	0.0	0.0	2.0	2.2
Q17) I am now showing you a card with different amounts. Could you choose the range that best fits the amount of savings deposits that you have in euro?						
Do not know	1.0	5.1			2.3	1.5
No Aanswer	27.0	14.6	15.5	12.1	46.5	21.4
Nonresponse	28.0	19.6	15.5	12.1	48.8	22.9
S16) Net income of household (monthly)						
No Answer	22.8	12.3	0.6	1.9	9.5	10.3

Table 3: Questionnaire

Q5.1: Do you at present hold any foreign cash, that is banknotes and coins and not deposits at a bank account. Refer only to those cash holdings you hold personally or together with your partner - **Euro**

No question block is skipped
DK question block is skipped
NA question block is skipped
Yes proceed

Q9: I am now showing you a card with different amounts. Could you choose the range that best fits the amount of euro cash that you hold?

Less than EUR 25
EUR 25 - 49
EUR 50 - 99
EUR 100 - 249
EUR 250 - 499
EUR 500 - 999
EUR 1000 - 1999
EUR 2000 - 3999
EUR 4000 or more
Do not know
No answer

Q12b: Do you have a savings deposits?

No question block is skipped
DK question block is skipped
NA question block is skipped
Yes proceed

Q13: Are these savings deposits denominated in foreign currency?

No question block is skipped
DK question block is skipped
NA question block is skipped
Yes proceed

Q15.1: You mentioned that you have foreign currency savings deposits. In which currency are these savings deposits denominated? - EURO

No question block is skipped
DK question block is skipped
NA question block is skipped
Yes proceed

Q17: I am now showing you a card with different amounts. Could you choose the range that best fits the amount of savings deposits that you have in euro?

Less than EUR 100
EUR 10 - 249
EUR 250 - 499
EUR 500 - 999
EUR 1000 - 1999
EUR 2000 - 3999
EUR 4000 - 5999
EUR 6000 - 7999
EUR 8000 or more
Do not know
No answer

Table 6: Nonresponse rates tabulated by selected socio-demographics

	Question 5.1	Question 9	Question 15.1	Question 17	Income
Education Level					
Low	2.1	4.8	6.0	6.7	14.9
Medium	2.3	6.0	5.5	6.9	15.1
High	3.4	8.8	7.5	10.8	17.5
Age					
under 24	2.6	5.4	6.9	7.3	22.5
25 - 34	2.2	7.1	5.9	7.7	16.3
35 - 44	2.5	7.1	6.9	8.7	14.2
45 - 54	3.0	7.0	6.2	8.3	14.5
55 - 64	2.3	5.9	5.2	6.9	13.6
over 65	2.1	4.5	3.9	5.2	10.0
Net Household income					
low	2.2	4.5	5.8	6.6	
med	2.1	5.2	5.1	6.1	
high	1.9	6.3	5.3	7.5	
No Answer	4.6	25.0	8.9	50.0	
Gender					
Male	2.4	6.7	5.6	7.4	15.5
Female	2.6	6.0	6.3	7.7	15.5
Employment status					
Employed	2.7	7.1	6.4	8.3	15.0
Unemployed	1.7	5.3	5.8	6.9	15.6
Retired	2.5	5.1	4.3	5.9	11.3
Student/Pupil	1.9	4.9	6.5	7.0	25.7
Overall nonrepsonse rate					
	2.5	6.3	6.0	7.56	10.3

Source: OeNB Euro Survey

Table 7: Estimation results for question 5.1 – Do you have Euro cash

VARIABLES	(1)	(2)	(3)
Female	Reference group		
Male	-0.001	-0.001	-0.022
	(0.001)	(0.001)	(0.042)
Age	-0.000	-0.000	-0.002
	(0.000)	(0.000)	(0.015)
Age2	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)
Hungary	Reference group		
Albania	-0.007**	-0.007**	-0.377*
	(0.003)	(0.003)	(0.201)
Bosnia	0.002	0.002	0.0799
	(0.003)	(0.004)	(0.112)
Macedonia	-0.008***	-0.008***	-0.487**
	(0.002)	(0.003)	(0.204)
Bulgaria	0.005	0.006	0.180
	(0.004)	(0.005)	(0.124)
Croatia	0.007	0.009	0.197
	(0.005)	(0.006)	(0.125)
Poland		0.025***	
		(0.008)	
Romania	0.016**	0.022***	0.410***
	(0.007)	(0.008)	(0.119)
Serbia	0.022***	0.029***	0.490***
	(0.008)	(0.010)	(0.110)
Slovakia	0.005	0.006	0.086
	(0.005)	(0.006)	(0.147)
Czech Republic	0.002	0.002	-0.001
	(0.005)	(0.005)	(0.156)
Income low	Reference group		
Income medium	-0.001	-0.001	-0.055
	(0.002)	(0.001)	(0.064)
Income high	-0.002	-0.002	-0.088
	(0.002)	(0.002)	(0.077)
Income NA	0.008**	0.011***	
	(0.003)	(0.004)	
hhsize1-2	Reference group		
hhsize3-4	-0.003*	-0.002	-0.096
	(0.002)	(0.002)	(0.060)
hhsize5+	-0.002	-0.003	-0.061
	(0.002)	(0.002)	(0.080)
hh head	0.002	0.000	0.054
	(0.002)	(0.002)	(0.063)

Notes: Robust standard errors adjusted for clustering in parentheses, "***" significance at 1%; "**" significance at 5%; "*" significance at 10%; See continuation.

Table 7 (cont'd)

No child	Reference group		
One Child	0.001	0.002	0.018
	(0.002)	(0.002)	(0.073)
Two or more Children	0.000	0.001	-0.001
	(0.00)	(0.003)	(0.096)
Unemployed	Reference group		
Employed	0.000	0.002	0.013
	(0.003)	(0.003)	(0.099)
Retired	-0.002	0.000	-0.088
	(0.003)	(0.003)	(0.122)
Student/Pupil	-0.002	-0.003	-0.032
	(0.003)	(0.003)	(0.125)
Low education	Reference group		
Medium education	0.001	0.002	0.055
	(0.002)	(0.002)	(0.075)
High education	0.003	0.006*	0.123
	(0.003)	(0.003)	(0.087)
Small town	Reference group		
Medium town	0.001		0.034
	(0.002)		(0.088)
Large town	0.002		0.115
	(0.003)		(0.088)
Prefer cash - agree	Reference group		
Prefer cash - disagree	-0.001	-0.001	-0.047
	(0.002)	(0.002)	(0.073)
Prefer cash - dont know	-0.002	-0.002	-0.068
	(0.003)	(0.003)	(0.128)
Euro cash - agree	Reference group		
Euro cash common - disagree	-0.000	0.002	-0.002
	(0.002)	(0.002)	(0.066)
Euro common - dont know	0.008***	0.008***	0.246***
	(0.003)	(0.003)	(0.067)
Region's nonresponse rate	0.170***	0.182***	6.782***
	(0.028)	(0.025)	(0.846)
Observations	19683	21707	
R-squared	0.135	0.128	
Wald test	0.000	0.000	
Obs. Prob.	0.018	0.021	

Note: Constant is included but not shown; "R-squared" refers to McFadden R^2, "Wald test" to Wald test of jointly zero coefficients and "Obs. Prob" to observed Probability. Model (1): Without observations in Poland, Model (2): With observations in Poland, Model (3): Model with imputed values for income nonresponse. For (1) and (2) marginal effects are reported and for (3) the ordinary probit coefficients.
Source: OeNB Euro Survey

Table 8: Estimation results for question 9 – Euro cash amounts

VARIABLES	(1)	(2)	(3)
Female	Reference group		
Male	0.001	0.002	0.028
	(0.002)	(0.002)	(0.030)
Age	0.001	0.001	0.009
	(0.001)	(0.001)	(0.009)
Age2	-0.000	-0.000	-0.000
	(0.000)	(0.000)	(0.000)
Hungary	Reference group		
Albania	-0.04***	-0.037***	-0.832***
	(0.004)	(0.004)	(0.198)
Bosnia	0.014*	0.014*	0.196**
	(0.001)	(0.008)	(0.078)
Macedonia	0.025**	0.026***	0.173**
	(0.010)	(0.010)	(0.084)
Bulgaria	0.014	0.015	0.184**
	(0.009)	(0.009)	(0.082)
Croatia	0.030***	0.031***	0.284***
	(0.010)	(0.010)	(0.077)
Poland		0.042***	
		(0.012)	
Romania	0.038***	0.037***	0.359***
	(0.012)	(0.012)	(0.089)
Serbia	0.052***	0.050***	0.407***
	(0.018)	(0.017)	(0.103)
Slovakia	0.005	0.004	-0.064
	(0.009)	(0.009)	(0.100)
Czech Republic	0.008	0.007	0.001
	(0.007)	(0.007)	(0.076)
Income low	Reference group		
Income medium	0.007	0.004	0.057
	(0.005)	(0.005)	(0.058)
Income high	0.014***	0.009*	0.123**
	(0.005)	(0.005)	(0.052)
Income NA	0.057***	0.055***	
	(0.008)	(0.008)	
hhsize1-2	Reference group		
hhsize3-4	-0.004	-0.003	-0.032
	(0.004)	(0.004)	(0.055)
hhsize5+	-0.003	-0.003	-0.012
	(0.005)	(0.005)	(0.073)
hh head	0.005	0.003	0.047
	(0.004)	(0.004)	(0.049)

Notes: Robust standard errors adjusted for clustering in parentheses, "***" significance at 1%; "**" significance at 5%; "*" significance at 10%; See continuation.

Table 8 (cont'd)

No child	Reference group		
One child	0.002	0.001	0.014
	(0.004)	(0.004)	(0.054)
Two or more Children	0.002	0.002	0.0148
	(0.004)	(0.004)	(0.053)
Unemployed	Reference group		
Employed	0.005	0.009*	0.064
	(0.005)	(0.005)	(0.060)
Retired	-0.003	-0.001	-0.062
	(0.006)	(0.006)	(0.077)
Student/Pupil	-0.007	-0.005	-0.053
	(0.007)	(0.007)	(0.094)
Low education	Reference group		
Medium education	0.007*	0.009**	0.089
	(0.004)	(0.004)	(0.058)
High education	0.018***	0.024***	0.214***
	(0.007)	(0.007)	(0.072)
Small town	Reference group		
Medium town	-0.000		-0.008
	(0.004)		(0.055)
Large town	0.004		0.089*
	(0.004)		(0.050)
Prefer cash - agree	Reference group		
Prefer cash - disagree	-0.010***	-0.010***	-0.131***
	(0.003)	(0.003)	(0.046)
Prefer cash – don't know	-0.010	-0.009	-0.101
	(0.007)	(0.007)	(0.111)
Euro cash common - agree	Reference group		
Euro cash common - disagree	-0.006	-0.005	-0.075
	(0.004)	(0.004)	(0.059)
Euro common – don't know	-0.001	-0.004	-0.017
	(0.005)	(0.004)	(0.063)
Region's nonresponse rate	0.282***	0.319***	3.896***
	(0.038)	(0.035)	(0.387)
Observations	19683	21707	
R-squared	0.139	0.127	
Wald test	0.000	0.000	
Obs. Prob.	0.056	0.059	

Note: Constant is included but not shown; "R-squared" refers to McFadden R^2, "Wald test" to Wald test of jointly zero coefficients and "Obs. Prob" to observed Probability. Model (1): Without observations in Poland, Model (2): With observations in Poland, Model (3): Model with imputed values for income nonresponse. For (1) and (2) marginal effects are reported and for (3) the ordinary probit coefficients.
Source: OeNB Euro Survey

Table 9: Estimation results for question 15.1 – Do you have Euro savings deposits

VARIABLES	(1)	(2)	(3)
Female	Reference group		
Male	-0.004	-0.003	-0.057
	(0.003)	(0.003)	(0.043)
Age	-0.000	-0.000	-0.005
	(0.000)	(0.000)	(0.008)
Age2	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Hungary	reference group		
Albania	0.042**	0.045***	0.431***
	(0.016)	(0.016)	(0.129)
Bosnia	0.041**	0.042**	0.455***
	(0.017)	(0.018)	(0.140)
Macedonia	0.052***	0.061***	0.510***
	(0.013)	(0.014)	(0.093)
Bulgaria	0.026***	0.026***	0.342***
	(0.010)	(0.009)	(0.092)
Croatia	0.002	0.001	0.011
	(0.008)	(0.009)	(0.126)
Poland		0.061***	
		(0.014)	
Romania	0.029**	0.030**	0.346***
	(0.012)	(0.012)	(0.111)
Serbia	0.066***	0.070***	0.615***
	(0.015)	(0.015)	(0.093)
Slovakia	0.007	0.009	0.048
	(0.009)	(0.011)	(0.126)
Czech Republic	-0.010	-0.010	-0.229
	(0.007)	(0.007)	(0.140)
Income low	Reference group		
Income medium	-0.004	-0.003	-0.037
	(0.003)	(0.003)	(0.057)
Income high	-0.004	-0.003	-0.031
	(0.004)	(0.004)	(0.063)
Income NA	0.013**	0.022***	
	(0.006)	(0.007)	
hhsize1-2	Reference group		
hhsize3-4	-0.002	-0.002	-0.024
	(0.003)	(0.003)	(0.050)
hhsize5+	-0.002	-0.004	-0.011
	(0.004)	(0.004)	(0.068)
hh head	0.004	0.002	0.060
	(0.003)	(0.003)	(0.044)

Notes: Robust standard errors adjusted for clustering in parentheses, "***" significance at 1%; "**" significance at 5%; "*" significance at 10%; See continuation.

Table 9 (cont'd)

No child	Reference group		
One child	-0.001	0.000	-0.019
	(0.003)	(0.003)	(0.048)
Two or more children	-0.002	-0.003	-0.039
	(0.004)	(0.004)	(0.065)
Unemployed	Reference group		
Employed	0.001	0.006	0.019
	(0.004)	(0.004)	(0.068)
Retired	-0.005	-0.005	-0.104
	(0.005)	(0.005)	(0.089)
Student/Pupil	0.012	0.011	0.182*
	(0.008)	(0.007)	(0.102)
Low education	Reference group		
Medium education status	0.005	0.007*	0.083
	(0.004)	(0.004)	(0.058)
High education status	0.018***	0.019***	0.249***
	(0.006)	(0.006)	(0.074)
Small town	Reference group		
Medium town	0.010*		0.148**
	(0.005)		(0.076)
Large town	0.006		0.098*
	(0.004)		(0.058)
Euro deposits common - agree	Reference group		
Euro deposits common - disagree	-0.006**	-0.005*	-0.094**
	(0.003)	(0.003)	(0.045)
Euro deposits common – don't know	0.003	0.003	0.062
	(0.005)	(0.004)	(0.066)
Region's nonresponse rate	0.275***	0.302***	4.382***
	(0.036)	(0.036)	(0.540)
Observations	19572	21588	
R-squared	0.1416	0.1341	
Wald test	0.000	0.000	
Obs. Prob.	0.045	0.048	

Note: Constant is included but not shown; "R-squared" refers to McFadden R^2, "Wald test" to Wald test of jointly zero coefficients and "Obs. Prob" to observed Probability. Model (1): Without observations in Poland, Model (2): With observations in Poland, Model (3): Model with imputed values for income nonresponse. For (1) and (2) marginal effects are reported and for (3) the ordinary probit coefficients.
Source: OeNB Euro Survey

Table 10: Estimation results for question 17 – Euro savings deposits amounts

VARIABLES	(1)	(2)	(3)
Female	Reference group		
Male	-0.004	-0.003	-0.029
	(0.003)	(0.003)	(0.039)
Age	-0.000	-0.000	0.005
	(0.000)	(0.000)	(0.007)
Age2	0.000	0.000	0.000
	(0.000)	(0.000)	(0.000)
Hungary	Reference group		
Albania	0.042**	0.045***	0.201*
	(0.016)	(0.016)	(0.120
Bosnia	0.041**	0.042**	0.305***
	(0.017)	(0.018)	(0.107)
Macedonia	0.052***	0.061***	0.308***
	(0.013)	(0.014)	(0.070)
Bulgaria	0.026***	0.026***	0.278***
	(0.009)	(0.010)	(0.068)
Croatia	0.002	0.001	0.276***
	(0.008)	(0.009)	(0.064)
Poland		0.061***	
		(0.014)	
Romania	0.029**	0.030**	0.173**
	(0.012)	(0.012)	(0.086)
Serbia	0.066***	0.070***	0.392***
	(0.015)	(0.015)	(0.070
Slovakia	0.007	0.009	-0.053
	(0.009)	(0.011)	(0.095)
Czech Republic	-0.010	-0.010	-0.289***
	(0.007)	(0.007)	(0.100)
Income low	Reference group		
Income medium	-0.004	-0.003	-0.006
	(0.003)	(0.003)	(0.048)
Income high	-0.004	-0.003	0.072
	(0.004)	(0.004)	(0.053)
Income NA	0.013**	0.022***	
	(0.006)	(0.007)	
hhsize1-2	Reference group		
hhsize3-4	-0.002	-0.002	0.005
	(0.003)	(0.003)	(0.047)
hhsize5+	-0.002	-0.004	-0.003
	(0.004)	(0.004)	(0.061)
hh head	0.004	0.002	0.065
	(0.003)	(0.003)	(0.040)

Notes: Robust standard errors adjusted for clustering in parentheses, "***" significance at 1%; "**" significance at 5%; "*" significance at 10%; See continuation.

Table 10 (cont'd)

No child	Reference group		
One child	-0.001	0.000	-0.024
	(0.003)	(0.003)	(0.043)
Two or more children	-0.002	-0.003	-0.023
	(0.004)	(0.004)	(0.060)
Unemployed	Reference group		
Employed	0.001	0.006	0.053
	(0.004)	(0.004)	(0.057)
Retired	-0.005	-0.005	-0.008
	(0.005)	(0.005)	(0.074)
Student/Pupil	0.012	0.011	0.157*
	(0.008)	(0.007)	(0.087)
Low education	Reference group		
Medium education status	0.005	0.007*	0.101**
	(0.004)	(0.004)	(0.051)
High education status	0.018***	0.019***	0.327***
	(0.006)	(0.006)	(0.070)
Small town	Reference group		
Medium town	0.010*		0.115*
	(0.005)		(0.068)
Large town	0.006		0.059
	(0.004)		(0.052)
Euro deposits common - agree	Reference group		
Euro deposits common - disagree	-0.006**	-0.005*	-0.160***
	(0.003)	(0.003)	(0.047)
Euro deposits common – don't know	0.003	0.003	-0.098*
	(0.005)	(0.004)	(0.059)
Region's nonresponse rate	0.275***	0.302***	4.170***
	(0.036)	(0.036)	(0.399)
Observations	19572	21588	
R-squared	0.142	0.134	
Wald test	0.000	0.000	
Obs. Prob.	0.045	0.048	

Note: Constant is included but not shown; "R-squared" refers to McFadden R^2, "Wald test" to Wald test of jointly zero coefficients and "Obs. Prob" to observed Probability. Model (1): Without observations in Poland, Model (2): With observations in Poland, Model (3): Model with imputed values for income nonresponse. For (1) and (2) marginal effects are reported and for (3) the ordinary probit coefficients.
Source: OeNB Euro Survey

Table 11: Estimation results for the income question

VARIABLES	(1)	(2)
Female	Reference group	
Male	0.013**	0.010*
	(0.006)	(0.005)
Age	-0.002	-0.002
	(0.002)	(0.001)
Age2	0.000	0.000
	(0.000)	(0.000)
Hungary	Reference group	
Albania	-0.040***	-0.037***
	(0.007)	(0.007)
Bosnia	-0.017**	-0.023***
	(0.008)	(0.006)
Macedonia	-0.056***	-0.050***
	(0.010)	(0.011)
Bulgaria	-0.022***	-0.025***
	(0.007)	(0.007)
Croatia	-0.009	-0.011**
	(0.006)	(0.005)
Poland		-0.049***
		(0.011)
Romania	-0.003	-0.005
	(0.008)	(0.007)
Serbia	-0.028***	-0.023***
	(0.009)	(0.008)
Slovakia	-0.124***	-0.116***
	(0.008)	(0.008)
Czech Republic	-0.068***	-0.063***
	(0.016)	(0.014)
hhsize1-2	Reference group	
hhsize3-4	0.027***	0.027***
	(0.010)	(0.009)
hhsize5+	0.058***	0.053***
	(0.014)	(0.012)
hh head	-0.017**	-0.017**
	(0.009)	(0.008)
No child	Reference group	
One child	-0.017**	-0.017**
	(0.007)	(0.007)
Two or more Children	-0.012	-0.013*
	(0.008)	(0.008)

Notes: Robust standard errors adjusted for clustering in parentheses, "***" significance at 1%; "**" significance at 5%; "*" significance at 10%; See continuation.

Table 11 (cont'd)

Unemployed	Reference group	
Employed	-0.006	-0.008
	(0.009)	(0.009)
Retired	-0.034***	-0.032***
	(0.012)	(0.011)
Student/Pupil	0.063***	0.057***
	(0.014)	(0.013)
Low education	Reference group	
Medium education	0.006	0.008
	(0.009)	(0.009)
High education	0.027*	0.027*
	(0.014)	(0.014)
Small town	Reference group	
Medium town	-0.004	
	(0.011)	
Large town	0.021*	
	(0.012)	
Region's nonresponse rate	0.649***	0.678***
	(0.043)	(0.038)
Observations	20469	22567
R-squared	0.133	0.132
Wald test	0.000	0.000
Obs. Prob.	0.162	0.155

Note: Constant is included but not shown; "R-squared" refers to McFadden R^2, "Wald test" to Wald test of jointly zero coefficients and "Obs. Prob" to observed Probability. Model (1): Without observations in Poland, Model (2): With observations in Poland
Source: OeNB Euro Survey

Table 12 Prediction Tables for the respective equations

Q5.1 equation

		Actual 0	Actual 1	Total
Predicted	0	52	383	435
	1	303	18945	19248
	Total	355	19328	

correctly predicted 0s	15%
correctly predicted 1s	98%
over prediction rate	97%

Q9 equation

		Actual 0	Actual 1	Total
Predicted	0	577	2670	3247
	1	530	15906	16436
	Total	1107	18576	

correctly predicted 0s	52%
correctly predicted 1s	86%
over prediction rate	84%

Q15.1 equation

		Actual 0	Actual 1	Total
Predicted	0	338	1698	2036
	1	541	16995	17536
	Total	879	18693	

correctly predicted 0s	38%
correctly predicted 1s	91%
over prediction rate	89%

Notes: 0 refers to response and 1 to nonresponse.
Source: OeNB Euro Survey

Table 12 (cont'd)

Q17 equation

		Actual 0	Actual 1	Total
Predicted	0	575	2874	3449
	1	641	15482	16123
	Total	1216	18356	
	correctly predicted 0s		47%	
	correctly predicted 1s		84%	
	over prediction rate		82%	

Income equation

		Actual 0	Actual 1	Total
Predicted	0	1031	1574	2605
	1	2286	15578	17864
	Total	3317	17152	
	correctly predicted 0s		31%	
	correctly predicted 1s		91%	
	over prediction rate		81%	

Notes: 0 refers to response and 1 to nonresponse.
Source: OeNB Euro Survey

Table 13: Estimations results for question 5.1 in Serbia

	(1)	(2)
Female	Reference group	
Male	-0.000	0.003
	(0.008)	(0.108)
Age	-0.001	-0.010
	(0.002)	(0.026)
Age2	0.000	0.000
	(0.000)	(0.000)
Up to 6000 SRD	Reference group	
6001 - 12000 SRD	-0.005	0.014
	(0.016)	(0.225)
12001 - 18000	0.007	0.160
	(0.019)	(0.209)
180001 - 30000	0.021	0.215
	(0.023)	(0.216)
over 30001	-0.022*	-0.104
	(0.013)	(0.263)
Inc No Answer	0.050*	
	(0.028)	
hhsize1-2	Reference group	
hhsize3-4	-0.019	-0.179
	(0.012)	(0.141)
hhsize5+	-0.021*	-0.222
	(0.011)	(0.183)
hh head	-0.002	-0.062
	(0.010)	(0.129)
No child	Reference group	
One Child	-0.017*	-0.246
	(0.010)	(0.161)
Two or more Children	0.006	0.043
	(0.013)	(0.157)
Unemployed	Reference group	
Employed	0.018	0.211
	(0.012)	(0.165)
Retired	0.013	0.135
	(0.023)	(0.250)
Student/Pupil	-0.028***	-0.467
	(0.010)	(0.293)

Notes: Robust standard errors adjusted for clustering in parentheses, "***" significance at 1%; "**" significance at 5%; "*" significance at 10%; See continuation.

Table 13 (cont'd)

Low education	Reference group	
Medium education	0.009	0.108
	(0.012)	(0.155)
High education	-0.002	-0.098
	(0.016)	(0.207)
Small town	Reference group	
Medium town	0.014	0.115
	(0.012)	(0.140)
Large town	0.020	0.245*
	(0.014)	(0.147)
Prefer cash - agree	Reference group	
Prefer cash - disagree	0.032**	0.344***
	(0.013)	(0.123)
Prefer cash – don't know	-0.028***	-0.490*
	(0.009)	(0.262)
Euro cash - agree	Reference group	
Euro cash common - disagree	-0.008	-0.085
	(0.010)	(0.151)
Euro common – don't know	0.062**	0.546***
	(0.030)	(0.189)
Belgrade	Reference group	
Vojvodina	-0.025***	-0.388**
	(0.009)	(0.159)
Central-west Serbia	-0.039***	-0.656***
	(0.008)	(0.190)
South-east Serbia	-0.020*	-0.375**
	(0.010)	(0.154)
Observations	1829	
R-squared	0.118	
Wald test	0.000	
Obs. Prob	0.052	

Note: Constant is included but not shown; "R-squared" refers to McFadden R^2, "Wald test" to Wald test of jointly zero coefficients and "Obs. Prob" to observed Probability. Model (1): Without income imputation, Model (2): With imputed values for missing income. Marginal effects are reported for (1). The ordinary probit coefficients are reported for Model (2)

Table 14: Estimations results for question 9 in Serbia

	(1)	(2)
Female	Reference group	
Male	0.018	0.073
	(0.018)	(0.074)
Age	0.004	0.016
	(0.005)	(0.018)
Age2	-0.000	-0.000
	(0.000)	(0.000)
Up to 6000 SRD	Reference group	
6001 - 12000 SRD	0.076	0.247*
	(0.048)	(0.149)
12001 - 18000	0.191***	0.562***
	(0.055)	(0.148)
180001 - 30000	0.194***	0.574***
	(0.057)	(0.149)
over 30001	0.126**	0.446***
	(0.051)	(0.151)
Inc No Answer	0.262***	
	(0.057)	
hhsize1-2	Reference group	
hhsize3-4	-0.043	-0.141
	(0.026)	(0.103)
hhsize5+	-0.053*	-0.178
	(0.028)	(0.127)
hh head	-0.001	-0.025
	(0.022)	(0.089)
No child	Reference group	
One Child	0.026	0.097
	(0.026)	(0.100)
Two or more Children	0.048*	0.167
	(0.029)	(0.106)
Unemployed	Reference group	
Employed	0.044	0.139
	(0.027)	(0.113)
Retired	0.013	0.026
	(0.045)	(0.174)
Student/Pupil	0.049	0.183
	(0.046)	(0.162)

Notes: Robust standard errors adjusted for clustering in parentheses, "***" significance at 1%; "**" significance at 5%; "*" significance at 10%; See continuation.

Table 14 (cont'd)

Low education	Reference group	
Medium education	0.045*	0.160
	(0.025)	(0.105)
High education	0.054	0.145
	(0.039)	(0.138)
Small town	Reference group	
Medium town	-0.018	-0.104
	(0.022)	(0.090)
Large town	0.017	0.067
	(0.025)	(0.100)
Prefer cash - agree	Reference group	
Prefer cash - disagree	0.009	0.031
	(0.022)	(0.088)
Prefer cash - dont know	-0.141***	-0.766***
	(0.020)	(0.185)
Euro cash common - agree	Reference group	
Euro cash common - disagree	0.032	0.127
	(0.027)	(0.102)
Euro common - dont know	0.090*	0.326**
	(0.047)	(0.150)
Belgrade	Reference group	
Vojvodina	0.058*	0.191*
	(0.031)	(0.111)
Central-west Serbia	-0.074***	-0.323**
	(0.026)	(0.129)
South-east Serbia	0.001	-0.087
	(0.028)	(0.111)
Observations	1829	
R-squared	0.076	
Wald test	0.000	
Obs. Prob	0.188	

Note: Constant is included but not shown; "R-squared" refers to McFadden R^2, "Wald test" to Wald test of jointly zero coefficients and "Obs. Prob" to observed Probability. Model (1): Without income imputation, Model (2): With imputed values for missing income. . Marginal effects are reported for (1). The ordinary probit coefficients are reported for Model (2)

Table 15: Estimations results for question 15.1 in Serbia

	(1)	(2)
Female	Reference group	
Male	-0.002	-0.013
	(0.015)	(0.086)
Age	0.000	0.001
	(0.004)	(0.021)
Age2	-0.000	-0.000
	(0.000)	(0.000)
Up to 6000 SRD	Reference group	
6001 - 12000 SRD	-0.056***	-0.289*
	(0.019)	(0.150)
12001 - 18000	-0.061***	-0.372**
	(0.018)	(0.159)
180001 - 30000	-0.015	-0.082
	(0.024)	(0.150)
over 30001	-0.063***	-0.353**
	(0.019)	(0.153)
Inc No Answer	-0.028	
	(0.022)	
hhsize1-2	Reference group	
hhsize3-4	-0.007	-0.031
	(0.021)	(0.121)
hhsize5+	-0.023	-0.121
	(0.023)	(0.149)
hh head	-0.002	-0.019
	(0.018)	(0.104)
No child	Reference group	
One Child	-0.031*	-0.202
	(0.018)	(0.123)
Two or more Children	-0.009	-0.059
	(0.020)	(0.124)
Unemployed	Reference group	
Employed	0.029	0.162
	(0.020)	(0.125)
Retired	0.041	0.219
	(0.039)	(0.193)
Student/Pupil	0.034	0.198
	(0.037)	(0.184)

Notes: Robust standard errors adjusted for clustering in parentheses, "***" significance at 1%; "**" significance at 5%; "*" significance at 10%; See continuation.

Table 15 (cont'd)

Low education	Reference group	
Medium education	0.039**	0.243**
	(0.019)	(0.123)
High education	0.023	0.117
	(0.031)	(0.165)
Small town	Reference group	
Medium town	0.053***	0.286***
	(0.021)	(0.108)
Large town	0.064***	0.342***
	(0.025)	(0.122)
Euro deposits common - agree	Reference group	
Euro deposits common - disagree	-0.019	-0.116
	(0.016)	(0.098)
Euro deposits common - dont know	-0.024	-0.141
	(0.017)	(0.116)
Belgrade	Reference group	
Vojvodina	-0.019	-0.117
	(0.021)	(0.133)
Central-west Serbia	-0.038*	-0.252*
	(0.021)	(0.150)
South-east Serbia	-0.004	-0.041
	(0.022)	(0.132)
Observations	1737	
R-squared	0.041	
Wald test	0.004	
Obs. Prob.	0.104	

Note: Constant is included but not shown; "R-squared" refers to McFadden R^2, "Wald test" to Wald test of jointly zero coefficients and "Obs. Prob" to observed Probability. Model (1): Without income imputation, Model (2): With imputed values for missing income. . Marginal effects are reported for (1). The ordinary probit coefficients are reported for Model (2)

Table 16: Estimations results for question 17 in Serbia

	(1)	(2)
Female	Reference group	
Male	-0.007	-0.038
	(0.016)	(0.083)
Age	0.003	0.016
	(0.004)	(0.021)
Age2	-0.000	-0.000
	(0.000)	(0.000)
Up to 6000 SRD	Reference group	
6001 - 12000 SRD	-0.058***	-0.262*
	(0.022)	(0.146)
12001 - 18000	-0.062***	-0.320**
	(0.021)	(0.159)
180001 - 30000	-0.011	-0.045
	(0.027)	(0.146)
over 30001	-0.045*	-0.207
	(0.023)	(0.152)
Inc No Answer	-0.022	
	(0.026)	
hhsize1-2	Reference group	
hhsize3-4	-0.014	-0.067
	(0.022)	(0.116)
hhsize5+	-0.024	-0.112
	(0.025)	(0.143)
hh head	0.000	-0.008
	(0.019)	(0.100)
No child	Reference group	
One Child	-0.030	-0.173
	(0.019)	(0.117)
Two or more Children	-0.012	-0.069
	(0.021)	(0.119)
Unemployed	Reference group	
Employed	0.032	0.159
	(0.022)	(0.122)
Retired	0.069	0.313*
	(0.044)	(0.186)
Student/Pupil	0.044	0.215
	(0.041)	(0.181)

Notes: Robust standard errors adjusted for clustering in parentheses, "***" significance at 1%; "**" significance at 5%; "*" significance at 10%; See continuation.

Table 16 (cont'd)

Low education	Reference group	
Medium education	0.040*	0.216*
	(0.021)	(0.118)
High education	0.040	0.185
	(0.034)	(0.155)
Small town	Reference group	
Medium town	0.054**	0.264**
	(0.021)	(0.103)
Large town	0.057**	0.277**
	(0.025)	(0.117)
Euro deposits common - agree	Reference group	
Euro deposits common - disagree	-0.010	-0.059
	(0.017)	(0.093)
Euro deposits common – don't know	-0.030	-0.161
	(0.019)	(0.113)
Belgrade	Reference group	
Vojvodina	-0.028	-0.163
	(0.022)	(0.128)
Central-west Serbia	-0.046**	-0.269*
	(0.022)	(0.144)
South-east Serbia	-0.011	-0.080
	(0.023)	(0.127)
Observations	1737	
R-squared	0.036	
Wald test	0.008	
Obs. Prob	0.119	

Note: Constant is included but not shown; "R-squared" refers to McFadden R^2, "Wald test" to Wald test of jointly zero coefficients and "Obs. Prob" to observed Probability. Model (1): Without income imputation, Model (2): With imputed values for missing income. . Marginal effects are reported for (1). The ordinary probit coefficients are reported for Model (2)

Table 17: Estimations results for the income question in Serbia

Female	Reference group
Male	0.007
	(0.018)
Age	-0.003
	(0.004)
Age2	0.000
	(0.000)
hhsize1-2	Reference group
hhsize3-4	0.066***
	(0.025)
hhsize5+	0.117***
	(0.037)
hh head	-0.071***
	(0.022)
No child	Reference group
One Child	-0.003
	(0.024)
Two or more Children	-0.014
	(0.024)
Unemployed	Reference group
Employed	-0.042*
	(0.025)
Retired	-0.057*
	(0.032)
Student/Pupil	0.056
	(0.039)
Low education	Reference group
Medium education	-0.007
	(0.023)
High education	-0.064**
	(0.027)
Small town	Reference group
Medium town	-0.061***
	(0.020)
Large town	-0.003
	(0.023)

Notes: Robust standard errors adjusted for clustering in parentheses, "***" significance at 1%; "**" significance at 5%; "*" significance at 10%; See continuation.

Table 17 (cont'd)

Belgrade	Reference group
Vojvodina	-0.058**
	(0.023)
Central-west Serbia	-0.026
	(0.026)
South-east Serbia	-0.204***
	(0.019)
Observations	1959
R-squared	0.106
Wald test	0.000
Obs. Prob.	0.192

Note: Constant is included but not shown; "R-squared" refers to McFadden R^2, "Wald test" to Wald test of jointly zero coefficients and "Obs. Prob" to observed Probability.
Source: OeNB Euro Survey

Table 18: Prediction Tables for the Serbian Estimation Results

Q5.1 equation

		Actual 0	Actual 1	Total
Predicted	0	25	91	116
	1	71	1642	1713
	Total	96	1733	
	correctly predicted 0s	26%		
	correctly predicted 1s	95%		
	over prediction rate	91%		

Q9 equation

		Actual 0	Actual 1	Total
Predicted	0	97	163	260
	1	247	1322	1569
	Total	344	1485	
	correctly predicted 0s	28%		
	correctly predicted 1s	89%		
	over prediction rate	78%		

Q15.1 equation

		Actual 0	Actual 1	Total
Predicted	0	28	74	102
	1	153	1482	1635
	Total	181	1556	
	correctly predicted 0s	15%		
	correctly predicted 1s	95%		
	over prediction rate	87%		

Notes: 0 refers to response and 1 to nonresponse.
Source: OeNB Euro Survey

Table 18 (cont'd)

Q17 equation

		Actual 0	Actual 1	Total
Predicted	0	29	103	132
	1	177	1428	1605
Total		206	1531	
correctly predicted 0s		14%		
correctly predicted 1s		93%		
Over prediction rate		84%		

Income equation

		Actual 0	Actual 1	Total
Predicted	0	150	212	362
	1	227	1370	1597
Total		377	1582	
correctly predicted 0s		40%		
correctly predicted 1s		87%		
over prediction rate		78%		

Notes: 0 refers to response and 1 to nonresponse.
Source: OeNB Euro Survey

Table 19: Imputation Models

Imputed Variables	Model	Independent Variables
Income	ologit	Employment status, Education, Region, Number of children, Household head, town size, gender and age
Q5.1	logit	Employment status, Education, Region, Number of children, Household head, town size, gender, age, income, noticed big payments, noticed small payments, exchange rate expectations inflation expectations and visits in the euro zone
Q9	ologit	
Q15.1	logit	
Q17	ologit	

5 imputaton iterations

Source: OeNB Euro Survey

Abstract

Surveys have many different sources of uncertainty. One of them is item nonresponse, namely if a respondent refuses to answer a specific question. Despite a lively discussion in the literature in the last two decades, the most common method of dealing with it is still simply to ignore it. If the nonresponse depends on other factors than solely on randomness, this approach will yield biased results. This diploma thesis provides a discussion of the empirical and theoretical item nonresponse literature and an investigation of the determinants of item nonresponse in the OeNB Euro Survey – a household survey aiming to provide insights in the extent and determinants of Euroization in eleven Eastern and Southeastern European countries. In addition, imputation will enable me to quantify the biases arising from observable differences between nonrespondents and respondents.

This diploma thesis also contains a chapter about the motivation of a national bank to monitor the foreign currency holdings at home. It shows that currency substitution – that is if the domestic currency is substituted by some extent by a foreign one – has significant impacts on monetary policy issues and it is therefore essential to have reliable data about it. This is also one aim of the OeNB Euro Survey. The key questions concern the Euro cash and Euro savings deposit holdings which are particularly exposed to item nonresponse.

The analysis of the determinants of item nonresponse yields that the assumption of applying complete case analysis – which is the default dealing method of all commonly used statistical packages – is clearly violated. The results further reveal that even questions that are quite similar have different nonresponse determinants. Variables that might be correlated with personal wealth such as education, income and occupation are not as strongly correlated with nonresponse as expected, but their influence cannot be denied.

By imputing values for the nonrespondents with a regression based approach I was also able to analyze the differences of dissemination and amounts of Euro holdings between respondents and nonrespondents respectively between complete case analysis and imputation analysis – where also the nonrespondents are counted. The results yield that there are almost no significant differences between respondents and nonrespondents. Thus, the differences between complete case analysis and imputation analysis are so narrow that they do not justify the extra effort of imputation. But uncertainty remains up to a certain degree, because the imputation procedure only corrects for observable differences between respondents and nonrespondents.

Zusammenfassung

Umfragen haben viele verschiedene Unsicherheitsquellen. Eine von ihnen ist Item Nonresponse, d.h. wenn ein Befragter verweigert, auf bestimmte Fragen zu antworten. Trotz einer regen Diskussion in der Item Nonresponse Literatur der letzten zwanzig Jahre ist das Ignorieren der Verweigerungen noch immer die üblichste Art mit ihm umzugehen. Wenn die Antwortverweigerung nicht ausschließlich vom Zufall abhängt, dann liefert diese Herangehensweise verzerrte Resultate. Diese Diplomarbeit beinhaltet neben einer Diskussion der theoretischen und empirischen Fachliteratur eine Analyse der Item Nonresponse Determinanten in der OeNB Euro Survey – eine Haushaltsumfrage mit dem Ziel, Einsicht in Umfang und Motivation der Eurohaltung in Ost- und Südosteuropa zu gewähren. Zusätzlich quantifizierte ich mittels Imputation die Verzerrung der Resultate aufgrund beobachtbarer Unterschiede zwischen Verweigerern und Antwortenden.

Diese Diplomarbeit beinhaltet außerdem noch ein Kapital über die Motivation von Nationalbanken zur Messung von Fremdwährungsbeständen. Es zeigt, dass Currency Substitution – i.e. wenn die Inlandswährung zu einem bestimmten Grad durch eine ausländische ersetzt wird – erhebliche Folgen auf die Monetärpolitik hat und es daher essentiell ist, verlässliche Daten darüber zu haben. Dies ist auch einer der Ziele der OeNB Euro Survey. Die Schlüsselfragen betreffen Ersparnisse in Eurobargeld und auf Eurosparbüchern, welche besonders Verweigerungen ausgesetzt sind.

Die Analyse der Item Nonresponse Determinanten führt zu dem Ergebnis, dass die Annahme der Complete Case Analyse – welche die Standardmethode ist, mit Item Nonresponse umzugehen – klar verletzt ist. Die Resultate zeigen weiters, dass sogar Fragen, die sehr ähnlich sind, verschiedene Determinanten aufweisen. Variablen wie Einkommen, Bildung und Beschäftigung, die üblicherweise mit den Vermögen korreliert sind, sind nicht so stark mit der Verweigerung korreliert, wie es erwartet worden ist, jedoch ist ihr Einfluss nicht von der Hand zu weisen.

Durch Imputation mittels einer regressionsbasierten Methode war ich in der Lage, die Differenzen in der Verbreitung und Beträge der Eurohaltung zwischen Verweigerern und Antwortenden bzw. zwischen Complete Case Analyse und Imputation Analyse – in der auch die Verweigerer miteingerechnet werden – zu analysieren. Die Resultate zeigen, dass es kaum signifikante Unterschiede zwischen Verweigerern und Antwortenden gibt. Daher sind auch die Unterschiede zwischen Complete Case Analyse und Impuationsanalyse so gering, dass sich der zusätzliche Aufwand der Imputation nicht zu rechtfertigen ist. Bis zu einem

gewissen Grad verbleiben aber Unsicherheiten, da die Imputationprozedur nur für beobachtete Unterschiede zwischen Verweigerern und Antwortenden korrigiert.

Curriculum Vitae Andreas Gulyás

Citizenship:	Austrian
Born:	1985 in Vienna

Education

June 2003	High School Diploma (Matura) in GRg XII Erlgasse
2004 – 2009	Studies in Economics and Statistics at the University Vienna
September 2007 – February 2008	Studies in Economics at the Corvinus University of Budapest

Employment

2003 – 2004	Civilian Service at Caritas Socialis
2006 – until now	Research Assistant in the Austrian National Bank (Department for Economic Studies)

Teaching

October 2008 – January 2009	Tutor for "Grundzüge der Statistik für Statistiker"
March 2009 – June 2009	Tutor for "Quantitative Methoden der Volkwirtschaftslehre"

Languages

German	Mother tongue
Hungarian	Second mother tongue
English	Very good knowledge
French	Basic skills