SOEPpapers
on Multidisciplinary
Panel Data Research

DIW Berlin

German Institute
for Economic Research

The German
Socio-Economic
Panel Study

70

Anika Rasner
Ralf K. Himmelreicher
Markus G. Grabka
Joachim R. Frick

Best of Both Worlds – Preparatory Steps in Matching Survey Data with Administrative Pension Records.

The Case of the German Socio-Economic Panel and the Scientific Use File Completed Insurance Biographies 2004

Berlin, December 2007

SOEPpapers on Multidisciplinary Panel Data Research
at DIW Berlin

This series presents research findings based either directly on data from the German Socio-Economic Panel Study (SOEP) or using SOEP data as part of an internationally comparable data set (e.g. CNEF, ECHP, LIS, LWS, CHER/PACO). SOEP is a truly multidisciplinary household panel study covering a wide range of social and behavioral sciences: economics, sociology, psychology, survey methodology, econometrics and applied statistics, educational science, political science, public health, behavioral genetics, demography, geography, and sport science.

The decision to publish a submission in SOEPpapers is made by a board of editors chosen by the DIW Berlin to represent the wide range of disciplines covered by SOEP. There is no external referee process and papers are either accepted or rejected without revision. Papers appear in this series as works in progress and may also appear elsewhere. They often represent preliminary studies and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be requested from the author directly.

Any opinions expressed in this series are those of the author(s) and not those of DIW Berlin. Research disseminated by DIW Berlin may include views on public policy issues, but the institute itself takes no institutional policy positions.

The SOEPpapers are available at
http://www.diw.de/soeppapers

ISSN: 1864-6689

German Socio-Economic Panel Study (SOEP)
DIW Berlin
Mohrenstrasse 58
10117 Berlin, Germany

Contact: Uta Rahmann | urahmann@diw.de

Best of Both Worlds

Preparatory Steps in Matching Survey Data with Administrative Pension Records

The Case of the German Socio-Economic Panel and the Scientific Use File

Completed Insurance Biographies 2004

Anika Rasner
FDZ-RV Berlin & Max Planck Institute for Demographic Research

Ralf K. Himmelreicher
FDZ-RV Berlin

Markus G. Grabka
DIW Berlin

&

Joachim R. Frick
DIW Berlin and Berlin University of Technology (TUB)

December 2007

Corresponding Author

Anika Rasner
Forschungsdatenzentrum der Rentenversicherung
Deutsche Rentenversicherung Bund
Hallesche Strasse 1
10963 Berlin
Fon +49 30 865 89548
Mail anika.rasner@drv-bund.de

TABLE OF CONTENTS

1 Introduction

The confluence of recent reforms of the pension system and changes in employment histories, paired with various demographic trends, is expected to have a strong impact on the distribution of old-age income and the evolution of old-age poverty in Germany. Over the last decade, the public pension system has undergone a sequence of reforms, prompted by the quest for financial sustainability and demographic challenges ahead. These reforms aimed at halting the trends in early retirement, decelerating the growth in public pension benefits, and changing the public-private mix in the provision of income for the elderly. Simultaneously, the persistent unemployment that followed German reunification led to labor market reforms that promoted atypical and marginal forms of employment (e.g. so-called *Mini-Jobs*) and changed the unemployment benefit scheme. As a result, employment patterns become much more heterogeneous and deviated from the employment history of the typical German worker who works full-time, year in and year out, until retirement.

The normal employment career is well-embedded in the German welfare state, whereas atypical employment forms are less well-protected. As a consequence, it may be expected that the reforms will alter the level and composition of the retirement incomes of future retirees. The issue therefore arises as to how the confluence of changing employment patterns and public pension reforms, paired with demographic changes, will affect the old-age income of current and future cohorts of retirees. The goal of the research project reported herein was to trace the consequences of work and family choices through to outcomes in old age. In particular, we investigated whether the changes in employment patterns that interact with the effects of public pension reforms will undo the successes that Germany has had in alleviating poverty amongst the elderly, and amongst elderly women in particular.

However, given the available data, we were unable to quantify the impact of the interplay of pension reforms, changes in employment patterns, and demographic trends on the economic

situation at higher ages. Survey data usually suffers from small numbers of observations and – even in the case of a large number of cases – missing lifecycle earnings and employment information. Meanwhile, administrative data lacks other important covariates, in particular in the German case, variables pertaining to the household context that allow researchers to draw conclusions about the economic well-being of the elderly. The lack of adequate data motivated the elaboration of a statistical matching procedure that links administrative pension records with survey data on a completely anonymous basis.

This paper presents preparatory steps that were carried out in order to prepare two longitudinal micro datasets for a statistical matching procedure, namely the Scientific Use File "Completed Insurance Biographies 2004" (*SUF Vollendete Versichertenleben 2004 – SUF VVL 2004*) provided by the Research Data Center of the Federal German Pension Insurance, and household panel data from the "German Socio-Economic Panel" (*Sozio-Oekonomisches Panel - SOEP*). The SUF VVL 2004 provides detailed information that is relevant for the calculation of an individual's public pension benefit, as well as monthly information about an individual's earnings[1], whereas the SOEP gives information about the household context and other relevant components of income. A successful matching of the two datasets would allow us to bring together the "best of both worlds" by combining their respective benefits and circumventing their drawbacks.

Statistical matching does not aim at finding the exact same person in both datasets. This is impossible, because, due to the measures instituted to protect the confidentiality of personal information, no common identifiers are available. Hence, the two datasets cannot be merged in the strict sense. However, through statistical matching, cases that are similar in terms of the observed characteristics of a person can be identified and linked. By combining information from

[1] This is true for all earnings that are subject to social insurance contributions. Certain occupational groups are systematically excluded from the public pension insurance, such as farmers, civil servants, or the self-employed. The SUF VVL 2004 does not provide information about the earnings of these occupational groups.

different sources, one can obtain a much more comprehensive dataset for the study of the topic of interest (Van der Puttan et al. 2002, p. 2).

Statistical matching is becoming increasingly popular in economics and social sciences. It is proving to be a useful tool in the evaluation of public policies. For example, Hujer et al. and Caliendo have applied statistical matching methods in the evaluation of the effects of job creation schemes on success in the labor market (Caliendo 2006; Hujer et al. 2004).

The dataset that would result from matching the SOEP and the SUF VVL 2004 would serve two purposes. First, it would allow us to simulate the old-age income of actual and future cohorts of retirees. On the basis of the available information on household context, we would be able to make qualified statements about the distribution of old-age income and quantify the prevalence of old-age poverty among the population of interest. Second, the dataset would help us to approximate the social security wealth of individuals who have not yet retired. Research that addresses the distribution of wealth and income needs to take this wealth component into account. Up to now, these accumulated pension rights have not been considered adequately in distributional analyses, even though it is essential for obtaining unbiased wealth estimates. For example, this becomes relevant when comparing the wealth of individuals who are insured in the public pension insurance scheme with the wealth of those groups who are excluded from public pension insurance (e.g. the self-employed or civil servants). Furthermore, the longitudinal dataset would allow us to evaluate the behavioral effects of recent policy reforms.

The goal of this paper is to present the preparatory steps that we have carried out in the statistical matching of administrative pension records with survey data. The paper will not focus on distributional analyses and does not intend to present any results. It is structured as follows. In Section II, issues of the confidentiality of data are presented: the German data protection law and its implications for social science research in general and for the statistical matching in particular are discussed. In Section III a short description of both datasets will be provided. In Section IV

follows an outline why the two data sources complement one another and pinpoint the potential pitfalls that may be encountered when matching the two datasets. In Section V, the population of interest will be specified, key variables used in the matching approach will be presented, and the distributions of the respective core variables will be compared in both datasets.[2] In Section VI, several regression models for different demographic groups are estimated and the predictive quality of the model assessed. Section VII presents the out-of-sample predictions, which show whether the regression results estimated on the basis of one dataset can be replicated, applying the estimated coefficients to the other dataset.

2 Issues of Data Confidentiality

Data from the Federal German Pension Insurance are social security data that are protected by the Social Security Data Protection Act, which is part of the Social Code (*Sozialgesetzbuch*). The Social Code establishes rules for the collection, processing, and use of highly sensitive personal and privacy data in the branches of the social insurance system, such as the Federal German Pension Insurance (Bundesministerium für Arbeit und Soziales 2006). Some uses of the data are regarded as an infringement of the individual's personal rights, in particular, the right of informal self-determination (*informationelle Selbstbestimmung*). Laws that safeguard the use of social security data are laid down in the provisions on the confidentiality of social security data in § 35 Book I of the Social Code (SGB I), on the protection of social security data in § 67 - 85a, Chapter 2, Book X of the Social Code (SGB X) and supplementary provisions for the protection of data in other sections of the Social Code (Bundesministerium für Arbeit und Soziales 2006).

The Articles of the Social Code do not apply if the data have been anonymized, in which case the disclosure of persons is no longer possible. The process of anonymization therefore allows the Research Data Center of the Federal German Pension Insurance to provide Scientific Use Files

[2] The distinction between the "SUF Completed Insurance Biographies 2004" and the dataset "Completed Insurance Biographies 2004" is important. The dataset "Completed Insurance Biographies 2004" is the total population of first-time pensioners in 2004, whereas the "SUF Completed Insurance Biographies 2004" is only a sample of the total population.

to researchers who are interested in the empirical analysis of retirement and disability. According to the legal definition of § 67 of SGB X, social security data are anonymized if they have been altered in such a way that the identity of the individuals can only be inferred by expending an unreasonable effort in terms of time, money, and manpower. This type of anonymization is called *de facto* anonymization. In contrast, if it is impossible in principle to infer the identity of the individual from the data, then we speak of absolute anonymization.[3] The high opportunity costs of absolute anonymization outweigh its benefits and furthermore, compromise the research value of the data. Anonymization is a trade-off between the risk of personal information being disclosed and the usability of data for research. *De facto* anonymization makes it almost impossible to re-identify individuals and is providing analytically valid micro-data to researchers (Hawala et al. 2005).

In order to analyze the factually anonymized Scientific-Use Files provided by the FDZ-RV, researchers have to sign a data use contract. The data transfer from the FDZ-RV to the researcher adheres to the principles of safe harbor.

The use of the SOEP data is bound by the strict requirements in Germany for the protection of the confidentiality of data (see *Bundesdatenschutzgesetz*). In order to work with the anonymized micro-data, researchers have to sign a data transfer contract. Further technical and organizational requirements have to be met before access is granted to the data so that the data is protected from unauthorized access. These requirements involve a personal computer or a computer network that is password-protected. Furthermore, persons who work with the data are obliged to protect its confidentiality. The data transfer contract explicitly prohibits any attempt to de-anonymize the data or to re-identify individual respondents in the data.

Despite the above restrictions which are of a technical nature only, and do not limit research, the statistical matching of the two datasets, the SUF VVL 2004 and the SOEP, is allowed. However,

[3] For further details in German, see Heese 2004.

according to the data transfer contract of the SOEP group and the data protection representative of the Federal German Pension Insurance, statistical matching is allowed only if the matched datasets are both anonymized. Consequently, statistical matching is not allowed if an anonymized dataset is to be matched with non-anonymized micro-data. Whereas the statistical matching of two anonymized micro-datasets is allowed, because in this case the matched file is still factually anomymized. However, in order to protect confidentiality, the new and unique dataset resulting out of the statistical matching can only be used on the safe-harbor computers in the Research Data Center of the Federal German Pension Insurance.

3 The Data

3.1 Completed Insurance Biographies 2004 (SUF VVL 2004)

The Scientific Use File Completed Insurance Biographies 2004, provided by the Research Data Center of the Federal German Pension Insurance, is based on administrative records or pension accounts of individuals, who are entitled to receive public pension benefits.[4] It is the first longitudinal dataset that the FDZ-RV provided to researchers who are interested in retirement and disability (Stegmann 2006).[5]

The SUF VVL 2004 is a systematic random sample of individuals who received public pension benefits for the first time in 2004.[6] A two-stage sampling procedure was applied. In the first step, a 20% sample was drawn from the pool of first-time retirees in 2004. In the second step, a subsample of 25% was drawn for selected age groups. The final data product, the SUF VVL 2004, is a 5% sample of first time pensioners that contains a total of 39,331 cases (Stegmann

[4] In the remainder of the paper, we will use the abbreviation SUF VVL 2004 when speaking of the "Scientific Use File Completed Insurance Biographies 2004" (*Scientific Use File Vollendete Versichertenleben 2004)*. The abbreviation "VVL 2004" refers to the total population of first-time retirees.

[5] The data, as well as more detailed information, can be found at www.fdz-rv.de or in the special issue of Deutsche Rentenversicherung Volume 61, Issue 9-10, which deals exclusively with the SUF VVL 2004 and empirical applications based on the data.

[6] The sample of completed insurance biographies comprises first-time old-age pensioners as well as first-time disability pensioners. The analysis will be confined to old-age pensioners.

2006, p. 550).[7] The sample is selective for several reasons. First, only persons eligible for public pension benefits were considered. Certain subgroups of the population were therefore systematically excluded, e.g. the self-employed, or civil servants in the case that they never accumulated any entitlements within the social security system.[8] Second, only two types of benefit were considered: old-age pensions and disability pensions. Beneficiaries who only receive other benefit types, such as educational pensions, or survivor's pensions (i.e. no personal pension entitlements) were excluded from the sample. Third, persons were excluded if they were eligible for public pension benefits in a foreign country and Germany has a social security agreement with the respective country.

As a result of these selection criteria, the sample is representative neither of the population as a whole, nor of the group of the elderly. The lack of representativeness is due to the fact that access to public pension benefits depends greatly on criteria of eligibility. This peculiarity of the data makes inter-cohort analysis strictly speaking, impossible (Fachinger and Himmelreicher 2006, p. 568).

The SUF VVL 2004 consists of two main components. The first part contains technical variables (e.g. person ID, year of first-time receipt of pension, etc.) and demographic information (e.g. sex, year of birth, nationality, etc.), as well as aggregated data related to the calculation of the individual's public pension benefit. The second component is subdivided into several longitudinal files. Ideally, the longitudinal information is available for a maximum of 624 months, starting in January in the year the person turned 14 years up to December in the year the person turned 65 years. A missing value appears in the data if a person was not employed in a job that is subject to social insurance contributions or if no other situation applied that is relevant to pension entitlements. For our purposes, the individual's earnings point history and the information about

[7] A precondition for being part of the sample is that the individual's Statutory Pension Insurance account has been clarified (*Versicherungskontenklärung*).

[8] Other groups are farmers, lawyers, medical doctors, and certain craftsmen, because they are covered by their respective profession-based pension scheme, such as the farmers' pension scheme.

the social employment situation is the most relevant. The sum of earning points, which is the central outcome variable of the study, is explained below (see also Himmelreicher/Frommert):

Earning Points (EP_i): The individual earning points describe the earnings position of an individual relative to the average earnings of all the individuals that pay contributions into the public pension system:

$$EP_i^t = \frac{Y_i^t}{\overline{Y}^t}$$

Y stands for the i^{th} individual's earnings in a given year t. For any year t, the earning point (EP) equals 1 if the i^{th} individual earns as much as the average of contributors ($\overline{Y}$) in time period t. The earning points are summed up over the entire working life of an individual and determine the final pension benefit. The total sum of the earning points, where n is the number of years of employment or equivalent periods of pension credits, is then used for the calculation of the final pension benefit:

$$\sum_{t-1}^{n} EP_i$$

Source: Rasner 2005, own illustration

The earnings point information is available for 624 months; hence, earnings dynamics and mobility can, in principle, be followed over time.[9] In addition, it is possible to analyze how certain demographic events (e.g. the birth of a child) affect a person's earnings. The longitudinal information on the employment situation enables us to analyze the effect of the duration of different activities over the life-course (e.g. schooling or unemployment) on the level of public pension benefits.

Overall, the administrative micro-data provided by the Research Data Center of the Federal Public Pension Insurance Institutes is of exceptionally high quality with respect to all the important details related to the calculation of public pension benefits. Other variables, which are not relevant for the calculation of the public pension benefits, such as educational attainment or occupational status, have a high number of missing values. The high precision is due to the data

[9] Earnings information in the data provided by the Statutory Pension Insurance only refers to earnings that are subject to social insurance contributions up to the contribution ceiling (*Beitragsbemessungsgrenze*). Amounts that are earned above the maximum contribution ceiling cannot be detected in the data, which implies that the earnings data is right-censored. The analysis of earnings dynamics and mobility is therefore restricted.

being process produced,[10] in that the Federal German Pension Insurance receives daily information about the earnings and employment situation of the individual from the employer, which are then converted into monthly information in the Scientific Use File. In contrast to survey data, administrative data therefore do not suffer from recall errors or non-response. Furthermore, panel attrition is not an issue for administrative data (Himmelreicher et al. 2006, p. 5). The main advantage of using administrative data is the large sample size. The SUF VVL 2004 comprises nearly 40,000 cases. However, a major drawback is the lack of relevant covariates necessary for any kind of multivariate analyses, such as information on the household context or other sources of income and assets. The attempt made in this paper to develop a procedure for the statistical matching of the SOEP and SUF VVL 2004 is intended to overcome these drawbacks.

3.2 The German Socio-Economic Panel (SOEP)

The German Socio-Economic Panel (SOEP) is a household panel study that started in 1984. The SOEP is a broad interdisciplinary survey that covers a representative sample of the total population living in private households in Germany.[11] To date, 23 waves of data for West Germany and 17 waves for East Germany are available. The most recent accessible data was collected in 2006, when about 12,499 households and 22,665 individuals (among those 5,143 children) were interviewed. Detailed information about the SOEP can be found at http://www.diw.de/gsoep and in further readings (e.g. Haisken-DeNew and Frick 2005; SOEP Group 2001; Wagner et al. 2006).

The micro-data provide information on individuals, households and families, and enable researchers to measure stability and change in living conditions over time. The survey measures a

[10] For an opposing account of the precision and quality of administrative data, see Kapteyn and Ypma 2006.

[11] This implies that certain segments of the population, which may be relevant for the analysis at hand, are at least partly excluded from the survey; namely, the institutionalized population, the homeless, emigrants, and potential immigrants (Wagner, Frick und Schupp 2006).

broad variety of objective indicators that cover such topics as demography and population, labor market and occupation or income, taxes and social security. It also contains a large choice of subjective indicators that aim at investigating the individual's perceptions, tastes and preferences, as well as (in more recent years) cognitive abilities and personality traits. The standard components are surveyed year by year, whereas certain special topic modules (e.g. Social Security and Poverty in 2002 or Use of Time and Preferences in 2005) are asked every few years. The richness of the data and continuous extensions attract researchers from various academic disciplines, for example, economics, sociology, statistics, demography, psychology, and geography.

Ideally, information is collected by asking (i) every person in the household above age 16 to complete an individual questionnaire, and (ii) one person, usually the head of the household, to complete a household questionnaire. Most relevant for our purpose is the biographical information surveyed, which contains the individual's complete employment history, starting at age 15. The information in the PBIOSPE file is gathered through a special biographical questionnaire that is administered only once, in order to obtain information for the time prior to the first interview. The PBIOSPE file stores information about the employment history, categorized into different types of activities. The biographical data are then updated year by year on the basis of the ongoing survey. The annual individual questionnaire collects information about the person's occupational status in the previous calendar year and is then aggregated into yearly values (Pischner 2006, p. 24).[12] The major advantage of the SOEP data is that all income components, apart from the individual's pension entitlements, are collected in order to obtain a comprehensive measure of the economic well-being of the household.

[12] For more detailed information on this file, see http://www.diw.de/documents/dokumentenarchiv/17/60061/bio2006.pdf#page=28).

3.3 Perfect Complements? The Best of Both Worlds

We want to develop a statistical matching procedure in order to obtain a dataset that combines the best of both worlds. The two datasets complement each other perfectly, for several reasons. As outlined above, the dataset SUF VVL 2004 provides high-quality work histories with information about monthly earnings and the employment situation, as well as reliable data for the calculation of the individual's monthly pension benefit. However, other important covariates are missing. First and foremost, the data lacks information about the household context, as well as benefits and transfers from other pension schemes. This information is necessary for investigating issues related to inequality or the distribution of old-age income. Without additional information about income, definite statements about the development of old-age poverty are highly speculative, if not impossible.[13] Statistical matching with the SOEP will enable us to address this shortcoming of the SUF VVL 2004; namely, the lack of contextual information. The SOEP provides very detailed information about income, not only for the individual respondent, but also for the household in which the person lives. The income information ranges from wage and salary income, and private and government transfers, to asset income (for further details, see Grabka 2006; Himmelreicher 2001).

The data also provides comprehensive demographic information about the birth of children, marital status, and changes in status over the entire life span. The information is stored in the BIOMARSY file. This file is set up accordingly to the PBIOSPE file. Information in the BIOMARSY file is much more differentiated than the marital status variable in the administrative pension data. The SUF VVL 2004 distinguishes only two status categories ("married" and "not married")[14] and is measured at the time a person retires.[15] By contrast, the SOEP data measures

---

[13] The old-age poverty rate of women would be highly overestimated if we did not consider additional income information. In the majority of cases, it is the public pension benefit of the husband that lifts women above the poverty threshold or, in the case that the husband dies, the survivor's benefit. The importance of survivor's benefits for the economic well-being of widows is stressed in (Deutsche Rentenversicherung Bund 2006a; Hagen, Himmelreicher und Hoffmann 2007).

[14] The category "married" includes married and remarried persons. The category "not married" covers widowed, divorced, and never married persons.

five different marital status categories (single, married, widowed, divorced, no longer married), which are surveyed year by year.

One shortcoming of the SOEP data is the lack of earnings information for the years prior to the first interview. The SOEP surveys the respondent's occupational status retrospectively, but not the individual's earnings history. This reduces the response burden, but it is also motivated by the lack of reliability and accuracy of earnings information that is collected retrospectively (Ferber and Birnbaum 1979, p. 112). If the SOEP and SUF VVL 2004 data are matched statistically, this shortcoming can be circumvented, at least with respect to earnings that are subject to social insurance contributions, which are available over the entire lifecycle. However, no lifecycle information is available for other components of income. Therefore, the statistical matching will also enable earnings information to be taken into account, thus yielding a more comprehensive measure of social security wealth as a share of the total household wealth.

The combination of administrative and survey data will significantly expand research opportunities beyond those provided by the SOEP and the SUF VVL 2004 data alone. The survey data provides very detailed contextual information (e.g. demographic and income information) that is usually missing in administrative data, whereas the administrative data provides very accurate longitudinal information about earnings and the social employment situation. The unique dataset that will result from the statistical matching is well-suited to trace the consequences of lifecycle work and family choices through to outcomes in old age.

3.4 Potential Pitfalls: "When Worlds Collide"

Despite the fact that the two datasets complement each other, there are certain pitfalls that need to be taken into consideration in both the preparation and implementation of the matching procedure. Three major pitfalls have been identified: 1) population sample versus inflow sample;

---

[15] It may be expected that changes in the marital status over the life course will explain much more variance in public pension benefits than the marital status at the point of retirement.

2) differences in sampling probabilities; and 3) differences in sample sizes. We now consider each of these potential pitfalls.

3.4.1 Population Sample versus Inflow Sample

The SOEP is a population sample, a quite large representative sample of the total population living in German households. Hence, it is possible to generalize from the sample data to the total population. However, we cannot use the entire sample population, because in this analysis we are interested in first-time pensioners only. Therefore, the sample population must be reduced considerably in order to specify the population of interest. Yet this reduced sample still needs to be large enough to allow legitimate generalization from this small segment of the sample population; namely, first-time pensioners.

The SUF VVL 2004, on the other hand, is a so-called inflow sample (Fitzenberger and Speckesser 2005). We use the inflows into retirement in the year 2004, more specifically inflows into old-age pensions. Being part of the sample is therefore conditional on the first-time receipt of old-age pension benefits. This entails that a person must have accumulated some sort of pension entitlements throughout his/her working life. Certain segments of the population can, by definition, never be part of the SUF VVL 2004 sample population (e.g. persons who were employed as civil servants or the self-employed for a large proportion of their working lives).

These differences between a population sample and an inflow sample need to be considered when specifying the sample population. Persons might be part of the SOEP sample population but not be part of the SUF VVL 2004 sample population. The correct specification of the population needs to yield two sample populations that resemble each other in the key dimensions. The sample population is specified in Chapter 5.

3.4.2 Differences in Sampling Probabilities

In a representative sample, the probability for each person that they will be part of the survey population is theoretically the same. However, the sampling probability in the SOEP is only theoretically the same. There are two reasons for this. First, the institutionalized population was not representatively included in the first wave.[16] Second, certain groups are oversampled deliberately. Oversampling means that the sampling probability for some groups is higher than for others. The purpose of oversampling is to obtain high enough numbers of observations for the analysis of certain subgroups of the population. For example, East Germans and foreigners have a higher sampling probability than West Germans.[17] Hence, in the SOEP, the probability of being part of the sampled population is not the same for every person.[18]

In the SUF VVL 2004, being part of the sample is conditional on the first-time receipt of public pension benefits. As noted above, this entails that certain segments of the population are systematically excluded from the VVL 2004 sample population. If the condition of first-time benefit receipt holds true, the sampling probability is the same for every person. The effects of oversampling and different sampling probabilities in the SOEP for the statistical matching with the SUF VVL 2004 are further illustrated in Chapter 6. We show how the sampling probabilities contribute to differences in the distribution of certain core variables. These differences need to be taken into consideration when developing the matching procedure by applying analytic weights.

3.4.3 Different Sample Sizes

Differences in sample sizes come into play when comparing the distribution of certain variables in both datasets. If sample sizes are small, the distribution is much more susceptible to outliers,

---

[16] However, persons of the initial sample population who lived permanently or temporarily in institutions were followed in later waves (Haisken-DeNew und Frick 2005). Individuals who moved from private households to institutional housing will be followed. Nevertheless, the SOEP does not aim at being representative for this population.

[17] The sampling probability for East Germans is 0.0004 and for foreigners it is 0.0008, compared to a sampling probability of 0.0002 for West Germans (Haisken-DeNew und Frick 2005, p. 19).

[18] However, these differences are corrected for by appropriate weighting factors that explicitly control for the underlying differences in sample design.

which in turn impairs the comparability of the two datasets. Section 7.2 illustrates the outlier problem when comparing the variable *monthly public pension benefit* in the two datasets. The differences in sample size will be addressed in the implementation of the matching procedure, but not in this paper.

4 Specification of the Sample Population

For the matching procedure to be successful, the sample population must be specified correctly. It is important to understand the structure of the two sample populations and to know the summary statistics and the distribution of certain core variables (e.g. *gender, age, marital status,* etc.). A statistical matching requires two populations that resemble each other as closely as possible in relevant ways, especially in some key dimensions. Otherwise, unequal populations will be matched to each other, which will impair the reliability of the results.

First, it is necessary to identify the population of interest in both datasets. In our case, the population of interest is first-time old-age pensioners. It is much easier to identify the population of interest in the SUF VVL 2004 because the dataset consists only of such individuals who retired in 2004 (inflow sample). However, in the SOEP, we have to isolate those individuals who retired recently and identify recipients of old-age public pension benefits, which is slightly more complicated. Once the sample populations of the two datasets have been identified, all individuals must be subject to the same pension rules. This is an important precondition, because if pension rules differed for the populations, these differences might affect the labor supply and the retirement behavior of the individuals, which would, in turn, complicate the matching procedure. Although plenty of social security reforms were passed between 2000 and 2005, they were directed principally towards future cohorts and only partially affect the public pension benefits and retirement behavior of this recent cohort of retirees. Hence, pension rules may be considered to be constant. In Sections 5.1 and 5.2, we explain in detail how the two sample populations were identified.

4.1 Specification of the Sample Population within the SOEP

Despite a relatively large total sample size of 11,400 households and 21,000 individual respondents in 2005, the sample population has to be specified in accordance with the respective research question. In our analysis, we focused on the financial well-being of first-time retirees. Therefore, the analysis was confined only to a very small segment of the total SOEP population.

In the first step of the analysis, we did not use the panel structure of the SOEP. We based the analysis solely on data for the 2005 wave, which comprises 21,097 cases. We used the data for 2005 instead of for 2004 because the crucial information is collected retrospectively, and the majority of questions in the 2005 questionnaire, especially those related to the income situation, refer to the year 2004. Figure 1 shows the original question 103 from the 2005 Individual SOEP questionnaire.

Figure 1 Original Question from the Individual Questionnaire in the SOEP

103. Who pays your pension and how high were the monthly payments in 2004?

☞ *Please state the gross amount, **excluding** taxes. If you receive more than one pension, please check each that applies.*
If you do not know the exact amount, please estimate.

	Own pension		Widower's pension, orphan's pension	
Gesetzliche Rentenversicherung (LVA, BfA, Knappschaft)		euros		euros
Beamtenversorgung		euros		euros
Kriegsopferversorgung		euros		euros
Unfallversicherung (z.B. der Berufsgenossenschaft)		euros		euros
Zusatzversorgung des öffentlichen Dienstes (z.B. VBL)		euros		euros
Betriebliche Altersversorgung (z.B. Werkspension)		euros		euros
Private pension scheme		euros		euros
Other:		euros		euros

Source: (TNS Infratest 2005, p. 25)

Question 103 (variable name: *vp10301*) also helped us to distinguish retirees (from the public pension insurance) from non-retirees in the data. Every person who reports a monthly public pension benefit from the public pension insurance is coded as a "retiree in 2005". A total of 4,518 persons receive public pension benefits in 2004.

Since the population of interest is "first-time old-age pensioners", the population has to be specified further. Persons who reported having received public pension benefits, but who were below age 60 in the year 2005, were coded as disability pensioners.[19] The group of disability pensioners cannot be identified by a specific variable in the SOEP questionnaire. Therefore, we had to work around this difficulty by using plausible assumptions. Current pension rules do not allow the receipt of old-age pension benefits before age 60. Hence, by definition, any public pension benefits paid before age 60 are disability benefits. Using the PBIOSPE data as a basis, we identified those individuals who retired between 2000 and 2004.[20] If a person who received public pension benefits reported that he/she had retired (*spelltype = 8*) and that this period started later than 1999 (*beginy > 1999*), the person was coded as a "first-time old-age pensioner between 2000 and 2004".[21] Altogether, 949 persons were identified as belonging to the population of interest. Through the statistical matching of the two datasets, we will be able to obtain household information for each of the 949 individuals and information about all other members living in the respective household, by using the unique household identifier (variable name: *$hhnr*).

---

[19] A total of 447 persons received public pension benefits and retired prior to age 60 and were therefore coded as disability/invalidity pensioners in the data. Due to the young age of some respondents coded as disability pensioners, we assume that some might also have received orphan's pensions; however, this is very difficult to ascertain.

[20] It is impossible to base the analysis solely on first-time retirees in 2004 because of small case numbers. Therefore, we prolonged the timeframe and consider first-time pensioners who retired in the years from 2000 to 2004.

[21] Additional plausibility checks have shown that some respondents, who reported being retired, did not report any public pension benefits. We double-checked whether these people receive public pension benefits from other pension schemes. If this was not the case, the individuals were excluded from the population of first-time old-age pensioners from 2000 to 2004.

4.2 Specification of the Sample Population within the SUF VVL 2004

The specification of the population of interest for the VVL data is less complicated than for the SOEP data. The original dataset consists of 39,331 cases. From the outset, so-called "Vertragsrentner" were excluded from the Scientific Use File VVL 2004 (Stegmann 2006, p. 538).[22]

In the SUF VVL 2004, only two different types of public pension benefit are distinguished: old-age pensions and disability pensions. Given that old-age pensioners are the focus of our research question, we excluded all recipients of disability pensions.[23] We considered the following benefit types of old-age pensions in the analysis: the regular old-age pension, old-age pensions due to unemployment or partial employment in old age, old-age pension for women, old-age pension for persons with disabilities, and the old-age public pension benefit for persons with long insurance periods.[24] A total of 7,730 persons receive other public pension benefits and were therefore excluded from the sample.

Furthermore, we excluded retirees who receive German public pension benefits while living in a foreign country. This group has to be excluded from the VVL because they are not part of the SOEP sample. In the SOEP, a person drops out of the sample if he or she is no longer living in Germany. Therefore, we also had to exclude from the VVL sample persons who are living in a foreign country but receive benefits from the German public pension insurance. The same

---

[22] So-called "Vertragsrentner" are persons who have spent time working in two different countries and hence have accumulated pension entitlements within the Federal German Pension Insurance and some other social security system (Himmelreicher 2005). Persons qualify for the payment of a so-called *Vertragsrente* if the two countries the person worked in have a bilateral social security agreement, also called a "totalization agreement". A totalization agreement governs the payment of benefits between the two countries (Social Security Administration 2007). The monthly public pension benefits of *Vertragsrentner* depend on the rules of the totalization agreement and therefore need to be interpreted in the light of these rules. For *Vertragsrentner,* a straightforward interpretation of the impact of the employment history on the level of public pension benefits is no longer possible. These persons cannot be identified in the SOEP.

[23] Old-age pensioners were identified over the variable *leat*, which classifies the individuals according to the type of public pension benefit they receive.

[24] Originally, these public pension benefits differed in terms of the eligibility criteria and the retirement age. The eligibility criteria (e.g. statutory retirement age & earliest possible age limit for the receipt of public pension benefits) were harmonized in the course of several reforms. For all benefit types, except for the old-age pension for person with disabilities, the statutory retirement age was raised to 65. Early retirement is penalized by permanent benefit reductions.

applies to persons who fall under the regulations of the Foreign Pension Law (*Fremdrentengesetz*).[25] A total of 446 persons fall under the regulations of the Foreign Pension Law. It was necessary to exclude this group of individuals, because we do not have any information about their employment in areas outside Germany. If these persons have been employed abroad, the SUF VVL 2004 data will not contain information about these periods, but the SOEP data does contain information about these periods. Due to this discrepancy in the two datasets, we have to exclude this group of persons.

In addition, we excluded beneficiaries of partial public pension benefits (*Teilrente)* (n=67). In the SOEP, we specified the population on the basis of whether a person reports being retired in a given year and receives a monthly public pension benefit. If both conditions applied, the person was considered to be "retired". It is not possible to control whether a person receives only partial public pension benefits. Therefore, we excluded the group of partial social security recipients from the SUF VVL 2004. After the specification, the total sample population consisted of 30,829 individuals.

5 Finding Matching Variables

For the statistical matching procedure to be successful, the datasets need to share a set of common variables measured in comparable ways. It is useful to choose the set of common variables on the basis of theoretical considerations and the research question that is addressed. In our analysis, we focused on the impact of the individual's employment history on the level of public pension benefits. The individual's public pension benefit is our dependent variable.

[25] The Foreign Pension Law was enacted in 1960. Public pension benefits were paid to individuals of German ancestry who lived in areas outside of Germany and who were forced to flee their homelands due to adverse political conditions. For individuals who fall under the regulations of the Foreign Pension Law, public pension entitlements earned in Eastern Europe are taken into account when calculating the German public pension benefit (Himmelreicher 2005).

5.1 Monthly Public Pension Benefit

SOEP: In the SOEP data, the monthly public pension benefit is easy to identify. Question 103 in the 2005 questionnaire asks: "Who pays your pension? How high were the monthly payments you received in 2004?" (see Figure 1). Persons are supposed to report the gross social security payment they receive each month from the Statutory Pension Insurance. Hence, for the statistical matching of the two datasets, we will simply use the value reported by each respondent.

We want to mention one important detail pertaining to the interplay of public pension benefits and health insurance contributions and how the interplay affects the accuracy of our dependent variable. Depending on the individual's earnings before retirement, the recipients of public pension benefit can either be insured in the statutory health insurance or hold a private health insurance plan.[26] The type of health insurance coverage determines the amount of the monthly public pension benefit payment. Health insurance contributions of persons covered by the statutory health insurance are deducted from the public pension benefit before it is paid out to the individual. By contrast, persons covered by private health insurance or persons insured voluntarily in the statutory health insurance receive a higher social security payment, but are obligated to pay their health care premiums out of the effective social security payment. For illustration, let us assume that a person covered by the statutory health insurance has the same gross public pension benefit as a person who is privately or voluntarily insured (e.g. both persons receive 980 Euro), then for the person covered by the statutory health insurance one half of the health and long-term care contributions is paid directly from the gross public pension benefit into the statutory health insurance. Hence, the amount paid for this individual is smaller than the gross pension benefit; namely. 955 Euro. For a privately and voluntarily insured person, the health-care and long-term care contributions are not paid directly to the private health insurance

[26] Persons with earnings below the maximum contribution ceiling are automatically insured in the compulsory health insurance scheme, whereas persons with earnings above this margin can opt for a private health care provider.

carrier, but are paid out to the individual. Hence, the amount paid out to the individual is higher than the gross pension benefit; namely, 1120 Euro (Deutsche Rentenversicherung 2007).

When calculating the monthly public pension benefit on the basis of SUF VVL 2004 data, we did not consider the distinction between persons covered by the statutory or private health insurance. We assumed that the calculated benefit is the disposable social security income of the person. We think this assumption is valid, because it is likely that respondents in the SOEP tend to report the public pension benefit that is transferred to their account every month. Even though respondents are explicitly asked to report the gross public pension benefit, it needs to be asked whether respondents are able to distinguish between their gross and net public pension benefit in the interview situation. For income from the statutory pension insurance, the comparison of income aggregates in the SOEP with official statistics shows that respondents in the SOEP tend to report a slightly higher public pension benefit, relative to the benefit they actually receive according to the official statistics (see Grabka 2004, p. 189).

Table 1 presents the summary statistics for the dependent variable (the monthly public pension benefit) for the population of "first-time pensions between 2000 and 2004" in the SOEP data. The table also shows the sample size of the four main demographic groups; namely, men and women in East and West Germany, their respective average public pension benefits, and both the median and standard deviation.

Table 1 Average Public Pension Benefits for First-Time Pensioners, 2000 & 2004 [27]

	WEST	**EAST**
MEN	Mean: 1,268 Euro Standard Deviation: 487 Median: 1,300 Euro n=304	Mean: 1,048 Euro Standard Deviation: 267 Median: 1,000 Euro n=139
WOMEN	Mean: 537 Euro Standard Deviation: 366 Median: 429 Euro n= 358	Mean: 732 Euro Standard Deviation: 306 Median: 687 Euro n=148

Source: SOEP 2005, own calculations

As expected, West German men receive the highest average public pension benefit (1,268 Euro) followed by East German men (1,048 Euro). East German women have a considerably higher average pension (732 Euro) than West German women, whose average public pension benefit is 537 Euro.

We defined another subsample in the SOEP data, namely, first-time pensioners from 2003 to 2004, in order to approach the SUF VVL 2004 sample as closely as possible. We identified a total of 351 first-time pensioners from 2003 to 2004. Table 2 displays the results.

Table 2 shows that the average public pension benefits have fallen for East German men (minus 35 Euro) and even more so for East German women (minus 129 Euro), whereas they have increased slightly for West German men (plus 22 Euro) and women (plus 29 Euro). However, the apparent changes in average monthly public pension benefits obtained from comparing the group of first-time pensioners between 2000 and 2004 with the group of first-time pensioners between 2003 and 2004 might be an indication of the negative impact of longer periods of unemployment as a result of the worsening economic situation in East Germany.

[27] The variable "monthly public pension benefit" was topcoded at 2,500 Euro, because some implausible cases were detected in the SOEP data. The reason for the topcoding is stated in more detail in Section 6.2.1.

Furthermore, Table 2 illustrates that the number of cases is quite small when this specification of the sample population is chosen. For these two reasons, we decided that the group of first-time pensioners between 2000 and 2004 is a sample population of reasonable size.

Table 2 Average Public pension benefits for First-Time Pensioners, 2003 and 2004 [28]

	WEST	**EAST**
MEN	Mean: 1,290 Euro Standard Deviation: 518 Median: 1,280 Euro n=102	Mean: 1,013 Euro Standard Deviation: 237 Median: 990 Euro n=55
WOMEN	Mean: 567 Euro Standard Deviation: 397 Median: 469 Euro n= 134	Mean: 603 Euro Standard Deviation: 211 Median: 600 Euro n=60

Source: SOEP 2005, own calculations

Table 2 shows that the average public pension benefits have fallen for East German men (minus 35 Euro) and even more so for East German women (minus 129 Euro), whereas they have increased slightly for West German men (plus 22 Euro) and women (plus 29 Euro). However, the apparent changes in average monthly public pension benefits obtained from comparing the group of first-time pensioners between 2000 and 2004 with the group of first-time pensioners between 2003 and 2004 might be an indication of the negative impact of longer periods of unemployment as a result of the worsening economic situation in East Germany. Furthermore, Table 2 illustrates that the number of cases is quite small when this specification of the sample population is chosen. For these two reasons, we decided that the group of first-time pensioners between 2000 and 2004 is a sample population of reasonable size.

SUF VVL 2004: The SUF VVL 2004 lacks explicit information about the individual's public pension benefit. However, all variables necessary for calculating the public pension benefit are

[28] The variable "monthly public pension benefit" was topcoded at 2,500 Euro, because some implausible cases were detected in the SOEP data. The reason for the topcoding is further illustrated in Section 6.2.1.

included in the data. The data only contains information for the independent public pension benefits, which are benefits based on the individual's own entitlements as opposed to derived pension benefits, such as survivor's or orphan's pensions. Explicit information about the individual's public pension benefit was not included in the SUF VVL 2004, because it was identified as a potential source for the re-identification of persons in the sample.[29] The calculation of the benefit is based on the variable "sum of individual earning points" (*PSEGPT90*). Roughly speaking, these are primarily all full contribution periods, reduced contribution periods, and non-contributory periods (Himmelreicher and Mai 2006).[30] In addition to these contribution periods, the variable *PSEGPT90* takes into account the pension type factor and the actuarial adjustment in the case of early or late retirement. The pension type factor varies with the type of pension a person receives and lies between 1 (for old-age pensions) and 0.25 (for an orphan's pension). Given that our analysis is bound to old-age pensioners, the pension type factor equals 1 for the entire sample population. In contrast, the actuarial adjustment factor varies from person to person. The actuarial adjustment factor depends on the retirement age of the individual. If the person retires at the statutory retirement age, the factor equals 1. In the case of early retirement, the factor is reduced by 0.3% per month up to a maximum of 18% (Börsch-Supan 2000, p. 30). Late retirement increases the factor accordingly.

Despite the consideration of the pension type factor and the actuarial adjustment, it is not possible to derive the individual's monthly public pension benefit directly from the sum of individual earning points. Due to the different actual pension values in East and West Germany, it is necessary to consider the share of earning points that a person accumulated in East and West Germany, respectively. For 2004, the actual pension value for West Germany amounted to 26.13 Euro and for East Germany to 22.97 Euro (Deutsche Rentenversicherung Bund 2005b). In the

---

[29] The decision to exclude the variable "individual's monthly public pension benefit" is worth reconsidering, because it is *the* variable of interest in the data for most of the researchers. For the matching, the variable is particularly useful because it plays such a central role in the matching procedure. According to information from the Research Data Center, the variable will be included in future Scientific Use Files.

[30] Additional components go into the variable "sum of earning points". However, their relative importance is negligible (Himmelreicher und Mai 2006).

SUF VVL 2004, it is possible to adjust for the share of earning points accumulated in each region by using the variable *anteilos*, which describes the share of earning points accumulated in East Germany. Table 3 illustrates the calculation of the individual's monthly pension benefit in the SUF VVL 2004 data:

Table 3 Calculation of Individuals' Public Pension Benefit in the SUF VVL 2004 Data

Pension $_{EAST}$ = PSEGPT90 * ANTEILOS * Pension Value $_{EAST}$
+
Pension $_{WEST}$ = PSEGPT90 * (1 – ANTEILOS) * Pension Value $_{WEST}$
=
Pension sum
, where
PSEGPT90 = sum of individual earning points
ANTEILOS = share of earning points accumulated in East Germany
(1 – ANTEILOS) = share of earning points accumulated in West Germany
Pension Value $_{EAST}$ = 22.97 Euro in the year 2004 for East Germany
Pension Value $_{WEST}$ = 26.13 Euro in the year 2004 for West Germany

Source: Own illustration

Table 4 provides the summary statistics for the monthly public pension benefit in the SUF VVL 2004. The case numbers for the four demographic groups are significantly higher than in the SOEP data.

Table 4 Average Public Pension Benefits for First-Time Pensioners in SUF VVL 2004

	WEST	**EAST**
MEN	Mean: 1,064 Euro Standard Deviation: 498 Median: 1,136 Euro n=10,463	Mean: 1,000 Euro Standard Deviation: 307 Median: 966 Euro n=3,520
WOMEN	Mean: 474 Euro Standard Deviation: 331 Median: 384 Euro n= 13,193	Mean: 723 Euro Standard Deviation: 276 Median: 689 Euro n=3,653

Source: FDZ-RV - SUFVVL2004, own calculation

The comparison of the summary statistics for East and West German men and women in the SOEP and VVL data shows that the distribution of public pension benefits is quite similar in the two datasets, with the exception of West German men. Furthermore, it is noticeable that for all four demographic groups, the average public pension benefits in the SOEP are higher than in the SUF VVL 2004. Potential explanations for this might be either over-reporting of earnings or rounding effects. Hence, earnings tend to cluster at 50 Euro or 100 Euro steps. The over-reporting in survey data is systematic in such a way that respondents tend to report earnings of either 1,500 Euro or 1,450 Euro, rather than earnings of 1,435 Euro, whereas administrative data supposedly provides exact data (Hanisch and Rendtel 2002; Wolff and Augustin 2000).[31]

For East German men and women, the fit between SOEP and VVL data is exceptionally good. In the SUF VVL 2004, East German men receive an average public pension benefit of 1,000 Euro compared to 1,048 Euro in the SOEP (a difference of 48 Euro). For East German women, the fit is even better. In the SUF VVL 2004, East German women receive an average public pension benefit of 723 Euro compared to 732 Euro in the SOEP (a difference of 9 Euro). The standard deviation for the public pension benefits of East German women confirms the similarity of the distribution of public pension benefits (SUF VVL 2004: 277; SOEP: 306). The results for West German women also lie within a tolerable margin. In the SUF VVL 2004, West German women receive an average public pension benefit of 474 Euro compared to 537 Euro in the SOEP (a difference of 64 Euro).

The largest discrepancy between the two datasets is found for the group of West German men. In the SUF VVL 2004, West German men receive an average public pension benefit of 1,064 Euro, whereas in the SOEP they receive an average benefit of 1,268 Euro (a difference of 205 Euro). One explanation for the large discrepancy might be that West German men are a very

[31] Administrative data is generally expected to represent the truth, whereas survey data is assumed to be prone to over- or underreporting (Kapteyn und Ypma 2006). However, Kapteyn and Ypma show in their comparison of administrative data and survey data that measurement error is also an issue in administrative data.

heterogeneous group (standard deviation of 487). Compared to the other groups, they are much more often self-employed or work as civil servants. Hence, they receive public pension benefits from different pension schemes (e.g. private or civil servant pensions). It is therefore possible that men simply report their total retirement income when they are asked to state their social security benefit from the statutory pension insurance. We will state how we intend to address this problem in Section 7.2.

5.2 Time Spent in Different Types of employment

5.2.1 Preparation of the Data

SOEP: In our analysis, we focused on the effect of the employment history on the level of old-age income. We therefore needed to aggregate the information from PBIOSPE by adding up the time a person spent in each type of employment. PBIOSPE distinguishes the nine types of employment/activities listed in Table 5, plus the category missing if none of the nine types of employment applies:[32]

Table 5 Activities Distinguished in the SOEP Data

	ACTIVITY
A 1	School/University
A 2	Training/Apprenticeship
A 3	Military/Civilian Service
A 4	Full-Time Employment
A 5	Part-Time Employment
A 6	Unemployment
A 7	Homeproduction
A 8	Retirement
A 9	Other Activities
A 10	Missing

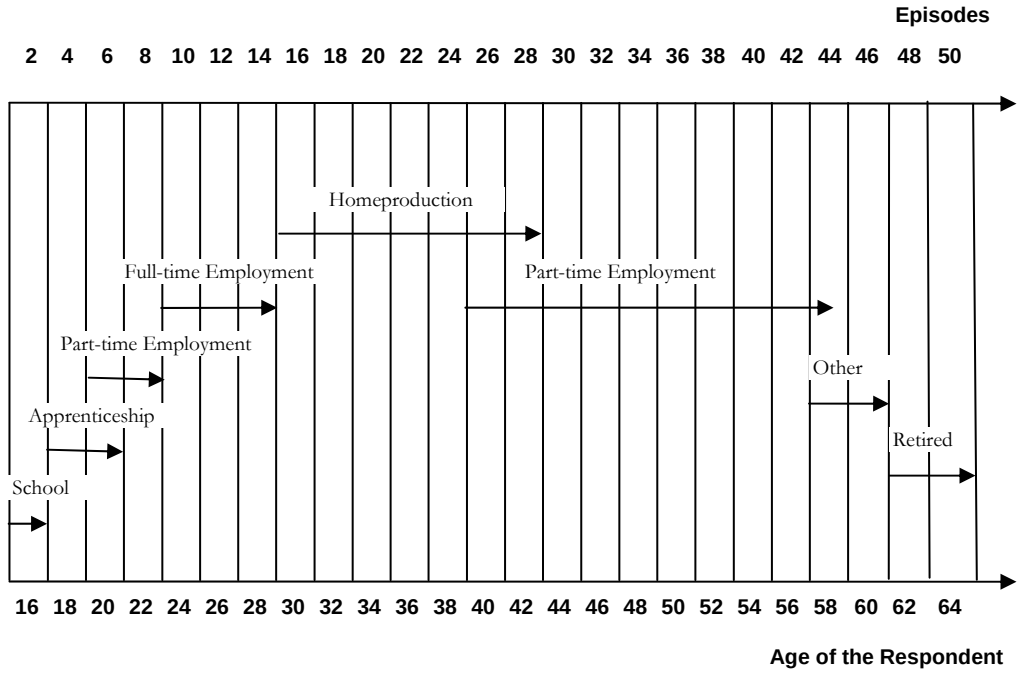
Source: (Pischner 2006, p. 24)

In the ideal case, we have information for 51 years. Between ages 15 to 65, the individual *i* spends his/her time in different activities *a*. Activities can overlap, which means that a person can report

[32] In the remainder of the paper the terms "types of employment" or "types of activities" will be used as equivalents.

more than one activity in a given year y. Figure 2 illustrates the fictitious employment history of a person *i* between ages 15 to 65.

Figure 2 Fictitious Employment History of Person *i*



Source:(compare to Himmelreicher und Viebrok 2002), own illustration

In the example above, periods of apprenticeship and part-time employment overlap in the age group 19 and 20, homeproduction and part-time employment overlap in the age group 39 to 43, and part-time employment and other activities overlap in the age group 57 and 58. In the case of overlapping periods, activities were weighted according to the number of activities reported in a given year. We applied an "equal distribution assumption", which means that every full year is divided by the number of activities reported in that year. Our example employment history reports two activities at age 19; namely, "apprenticeship" and "part-time employment". According to the "equal distribution assumption", the year is divided by two activities. Hence, six months were credited towards each category ("apprenticeship/training" & "part-time

employment"). We need to use this simplifying assumption because information is only available on an annual basis.[33]

If a person reported the type of activity "homeproduction", we deviate from the "equal distribution assumption". "Homeproduction" is not counted in a given year if other types of employment are reported simultaneously. This is because some women are likely to report that they are in homeproduction while they are working full-time, whereas others are not likely to report being in homeproduction. In these cases, we do not apply the "equal distribution assumption", because otherwise the time women spend in homeproduction would be overstated relative to the time spent in other types of employment. In our example employment history, part-time employment and homeproduction overlap in the age group 39 to 43. In this situation, we count four years in part-time employment and ignore the time spent in homeproduction. Homeproduction is only considered if no other activity is reported.

Additional problems are caused by short spells of employment and other activities. It can be assumed that these short spells are not reported in a yearly based activity calendar. This might yield a slight underestimation of certain activities. Furthermore, it is not possible to distinguish between different forms of employment in the SOEP data. We are unable to say whether a person was self-employed or was an employee whose earnings are subject to social insurance contributions.[34] We do not face this situation in the SUF VVL 2004, because all periods considered in the dataset are relevant for the calculation of the public pension benefit. In turn, we can't distinguish between part-time and full-time employment in the SUF VVL 2004.

[33] Monthly information is available for the years an interview was given. In the ideal case, we have monthly information on the occupational status if a person participated in all 22 waves of the SOEP. For the time prior to the first interview, information is only collected on an annual basis in the employment history questionnaire.

[34] We tried to control for the possibility of self-employment by considering the variable *stib*, which reflects the occupational status of a person in a given year.

For each person, the time spent in the nine different types of employment is summed up over the years 15 to 65. If the person reports no type of employment in a given year, the year is coded as missing. Even if there are gaps in the employment history, the number of years should add up to 51 years for every retired person. Table 6 shows how we translated information from the example employment history for our purposes.

Table 6 Translating an Hypothetical Employment History in the SOEP

EPISODE	# OF ACTIVITIES	COUNTED
15 to 17 years school	1	2 years school
17 to 19 years apprenticeship	1	2 years apprenticeship
19 to 21 years apprenticeship and part-time employment	2	1 year apprenticeship & 1 year part-time employment
21 to 23 years part-time employment	1	2 years in part-time employment
23 to 29 years full-time employment	1	6 years full-time employment
29 to 39 years homeproduction	1	10 years homeproduction
39 to 43 years homeproduction and part-time employment	2	0 years homeproduction & 4 years part-time employment
43 to 57 years part-time employment	1	14 years part-time employment
57 to 58 years part-time employment and other	2	6 months part-time employment & 6 months other activity
58 to 61 years other	1	3 years other
61 to 65 years retired	1	5 years retired
Total		51 years

Source: Own Illustration

VVL: The SES-file in the SUF VVL 2004 data is the equivalent of the PBIOSPE file in the SOEP data. Unlike PBIOSPE, the SES file distinguishes between thirteen different types of employment, which are listed in Table 7.

Table 7 Activities in the SUF VVL 2004[35]

	ACTIVITY
SES 1	School/University
SES 2	Apprenticeship/Training
SES 3	Homeproduction
SES 4	Unemployment
SES 5	Military/Civilian Service
SES 6	Other Activities
SES 7	Care Giving
SES 8	Invalidity/Sickness
SES 9	Employment subject to social insurance contributions
SES 10	Marginal Employment
SES 11	Self Employment
SES 12	Invalidity Pension
SES 13	Old-Age Pension

Source: (Stegmann 2006)

An employment situation is only defined if a certain period is relevant for a person's pension entitlements. For example, the self-employed can opt to pay social insurance contributions on a voluntary basis. Under these circumstances, the employment situation "self-employed" applies. However, if a self-employed person does not pay voluntary contributions in the social security system but instead invests in a private pension scheme, this type of employment does not fall under the social employment situation "self-employed". If none of the above types of employment applies in a given month, a missing value appears.

In the SUF VVL 2004, information is available on a monthly basis. Hence, the time a person spent in each employment situation can be summed up more precisely in the SUF VVL 2004 than in the SOEP. The SES file starts in January the year a person turned 14 and ends in December the year a person turns 65 (Stegmann 2006). In the ideal case, information is available for 624 months (52 years times 12 months), which is illustrated in a simplified way in Table 8.

35 For more detailed information, consult Volume 9/10, 2006 of Deutsche Rentenversicherung, User Guide provided by the Research Data Center (Forschungsdatenzentrum der Deutschen Rentenversicherung 2006) or www.fdz-rv.de.

Table 8 Structure of the Longitudinal File Social Employment Situation

	SES001	…	…	SES312	…	…	SES624
Activity	School			Employment Subject to Social Insurance Contributions			Retired

Source: Stegmann (2006), p. 549, modified for own purposes

In the case of the SES file, we did not use the simplifying "equal distribution assumption". Even if types of employment overlap, only one type of employment is recorded. In the case of overlapping types of employment, the decision as to which type of employment to record depends on a set of priority rules. The priority rules are already applied when the data is being prepared and serve the purpose of anonymization (Stegmann 2006, p. 545). The rules are related to the type of contributions that are paid into the system. Employment that is subject to social insurance contributions is prioritized against all other types of employment. Then follow voluntary contributions (*freiwillige Beitragszeiten*), creditable periods (*Anrechnungszeiten*), credited substituted periods (*Ersatzzeiten*), receipt of public pension benefits (*Rentenbezug*), childcare credits and the raising of several children (*Kindererziehungszeit und Erziehung mehrerer Kinder*), as well as childcare periods and credits (*Kinderberücksichtigungszeit und Gutschrift*) (for further details, see Stegmann 2006, p. 542). Due to these priority rules, the time spent in the different types of employment can easily be summed up over the respective time span. Another useful set of files provides longitudinal information in the form of flag variables that indicate whether a certain pension-relevant situation applied at some point in time.

SOEP: In order to get a better understanding of the data, we first calculated the average time spent in various types of employment in the age group 15 to 65 for three different populations: all retirees in 2005 (1), first-time pensioners from 2000 to 2004 (2), and first-time pensioners in 2003 and 2004 (3). The average time spent can be calculated in two different ways. In one approach, all persons are considered in the denominator, independent of whether or not they have spent time in a certain type of employment. If a person spent no time at all on homeproduction, he/she will still be counted in the denominator. An average value of five years

spent on homeproduction therefore needs to be interpreted as follows: for all persons in the defined subsample, the average duration spent on homeproduction amounts to five years. In the alternative approach, only non-zero observations are considered, which means only those individuals that have spent time in a certain type of employment. If a person did not spend any time in homeproduction, the case is not considered in the denominator. A person that spent five years in homeproduction is considered in the denominator. An average value of 12 years in homeproduction therefore needs to be interpreted as follows: for those persons who have spent time in homeproduction, the average duration spent in homeproduction amounts to 12 years.

We distinguished between different demographic groups when calculating the average time spent in the nine types of employment. In the first set of calculations, we only distinguished between men and women. In the next step, we distinguished between men and women in East and West Germany. The East-West distinction is based on the variable *vbula* in the ppfad-File.[36] Furthermore, we distinguished between Germans and persons with a history of migration.[37] The average time spent on different types of employment were first calculated for Germans and persons with a history of migration together and then calculated separately for the two groups. It is necessary to distinguish between Germans and persons with a history of migration because we need to determine how the group of migrants differs from Germans. This step helped us to understand how the group of persons with a history of migration should be handled in the multivariate analysis and the matching procedure.[38]

In the results, we distinguished two different categories for homeproduction. The first category sums up all periods of homeproduction, independently of whether they overlap with other types

[36] The variable *vbula* distinguishes the 16 different states ("Länder") of the Federal Republic of Germany. The variable "East" captures the following five states: Brandenburg, Mecklenburg-Vorpommern, Sachsen, Sachsen-Anhalt, and Thüringen. The variable "West" captures the following 11 states: Baden-Württemberg, Bayern, Berlin, Bremen, Hamburg, Hessen, Niedersachsen, Nordrhein-Westfalen, Rheinland-Pfalz, Saarland, and Schleswig-Holstein. Given that it is not possible to distinguish between East and West Berlin in the VVL data, we subsume Berlin under the "West" category.

[37] For a more detailed description of the definition of persons with a history of migration, see Section 6.2.9.

[38] For the calculation of the average time spent in various activities, we apply the analytic weights attached to each observation in the SOEP to control for the different sampling probabilities.

of employment. The second category considers periods of homeproduction only if no other types of employment were reported. As mentioned above, the spelltype "missing" applies if no activity was reported in a given year. The spelltype "sum" sums up the time spent over all types of employment.

VVL: In principle, the calculation of the average time spent on the different VVL types of employment follows the same rules. Two different sets of means were calculated, one based on the total population independent of whether persons have spent any time in the respective activity and the other based only on non-zero values. The same demographic groups were distinguished in the calculations: first, men and women; and in a second step, men and women in East and West Germany. The classification into "East" and "West" was carried out according to the same rules as for the SOEP data. We also distinguished Germans and persons with a history of migration.[39] The calculation of the average time spent in the different VVL types of employment differed in some respects from the calculations based on SOEP data. First, the average time spent on the VVL types of employment was only calculated for the population of first-time old age pensioners in 2004. Furthermore, no analytic weights were considered in the calculations for the VVL, because no such weights exist in the VVL.

(Insert Appendix A & B)

5.2.2 Making Results Comparable

After aggregating the time that individuals spent in different types of employment in both datasets, we want to be able to use the variables when we match the datasets. However, the two datasets contain a different number, and different kinds of employment. Therefore, the types of

[39] Due to differences in the definition of "persons with a migration history" (see Section 5.8), the comparison of the time spent in various types of employment is hardly comparable in the VVL and the SOEP.

employment have to be aligned according to plausible assumptions. Table 9 illustrates how we proceeded.

The types of employment were aligned in two steps. First, the 14 VVL 2004 categories (Column 1) were aligned with the 10 SOEP categories (Column 2). We expected the VVL categories "employment subject to social insurance contributions", "marginal employment" and "self-employment" to capture the same types of employment as the SOEP categories "full-time employment" and "part-time employment", respectively. The VVL categories "other", "care" and "invalidity & sickness" were subsumed under the SOEP category "other". For the purposes of implementing the matching procedure, the SOEP categories "full-time" and "part-time" were subsumed under the category "employment". The third column lists the final nine categories that are relevant for the statistical matching procedure.

Table 9 Streamlining Types of Employment from VVL 2004 & SOEP

COLUMN 1	COLUMN 2	COLUMN 3
VVL 2004 CATEGORIES	**SOEP CATEGORIES**	**FINAL CATEGORIES**
School/ University	School/University	School/University
Apprenticeship/ Training	Apprenticeship/ Training	Apprenticeship/ Training
Homeproduction	Homeproduction (only years in which person does not report any other activities).	Homeproduction
Unemployment	Unemployment	Unemployment
Military/ Civilian Service	Military/Civilian Service	Military/ Civilian Service
Other Caregiving Invalidity and Sickness	Other (which can be periods of maternity leave)	Other
Employment subject to social insurance contributions Marginal employment Self-employed	Full-time employed (including self-employment) Part-time employed	Employment
Invalidity Pension Old-Age Pension	Retirement	Retirement
Years Missing	Years Missing	Years Missing

Source: Own Illustration

In Appendix C, we compare the mean time spent in different types of employment after aligning the categories in the SOEP and the SUF VVL 2004. We compare the results of the SUF VVL 2004 with the results of the group of first-time pensioners from 2000 to 2004 and with the results of the group of first-time pensioners in 2003 and 2004. Again, we provide two sets of tables. One set shows the calculation of mean values only for those individuals who actually have spent some time on a certain activity, while the other set shows the calculation of mean values for the total population, independently of whether or not the individuals spent time on the respective activity.

(Insert Appendix C)

5.2.3 Gender

In addition to the time each individual spent in the different types of employment, further variables are needed if we are to be able to match the datasets. One of the most important of these is "gender". The employment histories of women differ to a great extent from those of men, with corresponding consequences for the public pension benefits. The German public pension system is employment-centered. Individuals who have a continuous employment history and (above) average earnings throughout their working lives receive a final public pension benefit that is high enough to maintain their standard of living even after they have retired. However, the majority of West German women do not have such an employment history. The reasons for the discontinuity in the employment histories of West German women are manifold.[40] Women enter the labor market in jobs below their qualification levels. Women earn lower wages in comparable jobs in companies of comparable size. Women are more likely to work in part-time or marginal part-time jobs, where their earnings are below average. Women are more likely to interrupt employment when they give birth to a child and exit the labor market for the child-rearing years while the children are small. Further, they are more affected by the problem of reconciling work and family duties than their male counterparts. In addition, generous social policies, as well as joint income taxation, present substantial disincentives that inhibit women, particularly married women, from entering the labor market or encourage them to only work part-time (Rasner 2006a, 2006b). For these reasons, gender is one of the most important variables for the matching procedure. Table 10 shows how the variable "gender" is distributed in the different sample populations.

[40] Most of the above reasons for less continuous employment histories apply to West German women. Due to the dual-earner policy promoted by the former GDR-regime, East German women tend to have career paths that are more similar to those of men.

Table 10 Distribution of Variable "Gender" in SOEP & SUF VVL 2004[41]

GENDER	SOEP 2000 – 2004		SUF VVL 2004	
	n	Percent	n	Percent
Male	443	44.0	13,983	45.4
Female	506	56.0	16,846	54.6
Total	949	100.0	30,829	100.0

Source: FDZ-RV - SUFVVL2004 & SOEP 2005, own calculations

Table 10 shows that the distribution of the variable "gender" is quite similar in the SOEP and the SUF VVL 2004. The share of females is higher in both datasets, with a male/female ratio of 1 to 1.17.

5.2.4 Region

As mentioned above, the variable "region" distinguishes East and West. Under East, we subsumed all federal states of the former German Democratic Republic. Under West, we subsumed all federal states of the former Federal Republic of Germany, including the entire city of Berlin.[42] We decided to use the East-West distinction rather than less aggregated state dummies, because of the greater explanatory power of the distinction between East and West Germany. This variable best captures the geopolitical, institutional, and economic differences between the former German Democratic Republic (GDR) and the Federal Republic of Germany (FDR). The distinction between these two parts of Germany is necessary, despite the reunification of Germany in 1990. The cohort of retirees we are interested in, at least for now, spent most of its working life under one or the other regime, which in turn strongly affected their respective employment histories. For example, the average employment history of an East German woman was more similar to the employment history of a West German man than to the employment history of a West German woman. Table 11 shows how the variable "region" is distributed in the different sample populations.

[41] For the following cross-tabulations, analytic weights were applied.

[42] In the VVL data, it is not possible to distinguish between East and West Berlin. We therefore subsumed Berlin under the *West* category.

Table 11 Distribution of Variable "Region" in SOEP & SUF VVL 2004

REGION	SOEP 2000 - 2004		SUF VVL 2004	
	n	Percent	n	Percent
West	662	74.6	23,656	76.7
East	287	25.4	7,173	23.3
Total	949	100.0	30,829	100.0

Source: FDZ-RV - SUFVVL2004 & SOEP 2005, own calculations

5.2.5 Marital Status

Another relevant variable for the matching procedure is "marital status". Information about marital status can be found in both datasets, but the information differs in two respects. First, the VVL data measures marital status (variable: *fmsd*) only at the point of retirement. Hence, there is no information about changes in marital status over the life-course. Second, the VVL marital status category distinguishes only between two categories: "married" and "not married". The category "married" indicates that a person is either married or remarried. The category "not married" comprises persons who are widowed, divorced, or were never married. In contrast, in the SOEP, marital status is measured longitudinally; hence, changes in status can be followed over the life-course. Furthermore, the SOEP distinguishes five different categories of marital status: married and living together, married but living apart, never married, divorced, or widowed. For the matching procedure, the SOEP data has to be aligned in accordance with the VVL data. For each person, we used the marital status information at retirement. The two SOEP marital status categories (married and living together, married but living apart) were subsumed under the new marital status category "married". The three other categories (never married, divorced, and widowed) were subsumed under the new marital status category "not married". After the matching procedure has been completed successfully, we can return to the more detailed information on marital status that the SOEP contains (see BIOMARSY). The differences in the distribution of the variable "marital status" between the SOEP and the SUF VVL 2004, as summarized in Table 12, can be explained by the differences in measurement.

Table 12 Distribution of Variable "Marital Status" in SOEP & SUF VVL 2004

MARITAL STATUS	SOEP 2000 – 2004		SUF VVL 2004	
	n	Percent	n	Percent
Not Married	169	23.4	7,173	22.4
Married	780	76.6	23,656	77.6
Total	949	100.0	30,829	100.0

Source: FDZ-RV - SUFVVL2004 & SOEP 2005, own calculations

The frequencies of the variable "marital status" in the SOEP and the SUF VVL 2004 show that the two variables are distributed similarly in both datasets.

5.2.6 Number of Children

In both datasets, information about the number of children is only available for women. In the SOEP, information about the birth history of female respondents can be found in the data file BIOBIRTH (Frick and Schmitt 2006). In the first interview, the birth history is reconstructed from the biographical questionnaire and then updated each year on the basis of the data collected in the individual questionnaire. In the biography questionnaire, women are asked about the number of children, the sex, and the year of birth of each child. Through this procedure, the file BIOBIRTH captures the complete birth history of all female respondents in the SOEP. For the purpose of matching the datasets, we were interested in the variable *sumkids*, which indicates the total number of children born.

In the VVL data, information about the number of children is usually assigned to the mother. Exceptions to the rule occur either if the mother of the children died or if the mother works as a civil servant. Consider a situation in which the mother works as a civil servant and the spouse is gainfully employed and obligated to pay contributions into the statutory pension insurance. In this situation, childcare credits are credited to the account of the spouse.[43] In the SUF VVL 2004, only 1% of the children are assigned to the male records (n=180), whereas 99% or in 15,178 of

[43] According to Paragraph §56, (4) SGB VI civil servants are not eligible for childcare credits from the Federal Statutory Pension Insurance.

the cases children are assigned to the mother's pension accounts (Himmelreicher and Mai 2006, p. 38 f.). Differences in the distribution of the variable "number of children" in SOEP and VVL, as illustrated in Table 13, can be attributed to these exceptions.

Table 13 Distribution of the Variable "Number of Children" in SOEP & SUF VVL 2004

NUMBER OF CHILDREN	SOEP 2000 – 2004		SUF VVL 2004	
	n	Percent	n	Percent
No children	50	13.0	2,003	11.9
One Child	109	26.1	4,000	23.7
Two Children	205	37.6	6,311	37.5
Three Children	90	14.0	2,918	17.3
Four Children	35	7.0	1,010	6.0
Five+ Children	17	2.3	604	3.6
Total	506	100.0	16,846	100.0

Source: FDZ-RV – SUFVVL2004 & SOEP 2005, own calculations

In Table 13, we can see that the congruence between the two datasets is better for high-parity mothers with four and more children and for mothers with two children, whereas there are small differences of ~ 2% for women with no children and one or three children.

5.2.7 Retirement Age

Information about the age of retirement is provided in both datasets.[44] We rounded the retirement age in the SUF VVL 2004 to integer numbers so that the results are comparable with the information in the SOEP. Given that we excluded disability pensioners from our sample population, the earliest possible retirement age was 60 years.[45] There is also an upper retirement age limit of 65, which is due to the sample design.[46]

[44] In the VVL data this information is captured in the variable ZTPTR1, which indicates the age of retirement of a person.

[45] This is in line with our expectations. According to current pension rules, it is impossible to receive any kind of old-age public pension benefit (e.g. old-age pensions for women, old-age pensions due to unemployment, etc.) before age 60.

[46] After the variable *ZTPTR1* was rounded, we found 22 persons with a rounded retirement age of 66 years (0.07%).

The SOEP questionnaire does not include a question about a person's retirement age. Information about an individual's retirement age can be reconstructed using multiple variables in the PBIOSPE file.[47] Due to the way in which the variable "retirement age" is operationalized in the SOEP, it is not possible to have a retirement age higher than 65, because the biographical information in the PBIOSPE ends at age 65. The same is true for the SUF VVL 2004. Persons with a retirement age of 66 are only the result of rounding. We therefore decided to topcode the variable "retirement age" in the SUF VVL 2004 at age 65. Table 14 summarizes the distribution of the retirement age in SOEP and VVL, with the mean retirement age of the sample given at the bottom of the table.

Table 14 Distribution of the Variable "Retirement Age" in the SOEP and SUF VVL 2004

IN %	**(1A)**	**(1B)**	**(3A)**	**(3B)**
Retirement Age	**SOEP 2000 – 2004**		**SUF VVL 2004**	
	n	Percent	n	Percent
Age 60	299	31.3	7,998	25.9
Age 61	169	17.8	2,287	7.4
Age 62	86	9.4	3,152	10.2
Age 63	119	13.0	3,911	12.7
Age 64	82	9.8	1,979	6.4
Age 65	194	18.7	11,502	37.3
Total	949	100.0	30,829	100.0
Mean Retirement Age	62.10		62.77	

Source: FDZ-RV – SUFVVL2004 & SOEP 2005, own calculations

There are clear differences in the distribution of the retirement age in the SOEP and the VVL, in particular at age 61. In the group of first-time old-age pensioners from 2000 to 2004, 17.8% of retirees retired at age 61, compared to only 7.4% in the VVL data. The results in the SUF VVL 2004 are supported by the official statistics of the Federal German Pension Insurance. We see

[47] There is one problem with using this approach. Some persons report repeated periods of "retirement", some of which starting before age 60. Persons with repeated periods of "retirement" might have received disability pension benefits before age 60. We solved the problem by taking the maximum starting age of the period "retirement". For a person who reported the beginning of retirement for the first time at age 40 and again at age 63, we record a retirement age of 63. To double-check, we control whether the person receives a public pension benefit from the social security system.

spikes in the distribution at ages 60, 63, and 65 (Deutsche Rentenversicherung Bund 2006a). Comparing Table 14, Column 3B with Table 15, Column 5 (see page 45), we see that despite small deviations, the SUF VVL 2004 data corresponds to the data from the official statistics of the Federal German Pension Insurance. In Column 6, aggregate data from the official statistics for the retirement cohorts 2000 to 2004 were pooled so that we have a measure that allows comparison with the SOEP data. Apart from the large deviation concerning the share of individuals who retired at age 61, the distribution of the retirement age for the group of first-time pensioners from 2000 to 2004 in the SOEP corresponds roughly with the official statistics of the Federal German Pension Insurance. Despite these deviations, the mean retirement age is almost exactly the same in both datasets, about 62 years.

Table 15 Distribution of Retirement Age for First Time Old-Age Pensioners 2000-2004

IN %	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Retirement Age	**Retired in 2000**	**Retired in 2001**	**Retired in 2002**	**Retired in 2003**	**Retired in 2004**	**Retired 2000-2004 (pooled)**
Age 60	46.10	40.09	30.93	26.24	25.90	33.85
Age 61	5.64	7.08	10.92	10.80	6.42	8.17
Age 62	3.43	5.58	6.12	6.48	9.10	6.14
Age 63	10.90	11.21	12.15	13.51	13.19	12.19
Age 64	1.48	1.77	2.74	2.78	2.80	2.31
Age 65	30.19	32.04	34.88	37.33	39.87	34.86
Age 66+	2.26	2.23	2.26	2.87	2.72	2.46

Source: Deutsche Rentenversicherung Bund: Rentenzugang 2000-2004, own calculations

A potential explanation for the differences between the SOEP and SUF VVL 2004 might be the interaction of age, cohort, and period effects that result from the pooling of first-time retirees in the SOEP in the years from 2000 to 2004 (Fachinger and Himmelreicher 2006). In addition, the small number of observations in the SOEP contributes to the differences in the retirement age in both datasets.

5.2.8 Migration History

At the beginning of the study, the indicator "migration history" was defined quite broadly in the SOEP. This broad indicator was used when calculating the mean time spent on different types of employment. The aim was to determine how persons with a history of migration differ from Germans and whether these differences might affect the statistical matching. First, we checked whether a person had German citizenship in the year 2005 (*nation05*). Then we checked whether a person has had German citizenship since birth, or whether it was obtained later (*vp137*). The variable *germborn* indicates whether a person was born in Germany or immigrated after 1948. If a person reported that he/she immigrated after 1948, then the variable *migration* was coded with 1. The construct validity of our migration variable was double-checked with the variable *immiyear*, which indicates the year of immigration. If a person reported a year of immigration, the person was expected to have a migration history; hence, the variable *migration* equals 1. However, for the purposes of matching the datasets, the migration variable in the SOEP was aligned with the migration variable in the SUF VVL.

In the SUF VVL 2004, persons with a history of migration were identified using the variable *SA* (*Staatsangehörigkeit* or citizenship). The variable *SA* only discriminates between German citizenship and citizenship of another country. Hence, the "migration" construct in the SUF VVL is less broad than the construct applied in the SOEP. The lack of additional variables makes a broader measure of the variable "migration history" infeasible.[48] Table 16 illustrates the distribution of the variable "migration history" in the SOEP and the SUF VVL 2004.

[48] In the specification of the SUF VVL 2004 sample population, we decided to forego a broader definition of the variable "migration history" by excluding persons who fall under the regulations of the Foreign Pension Law. Persons whose pension is subject to a bilateral social security agreement were excluded from the sample completely.

Table 16 Distribution of the Variable "Migration History" in the SOEP and SUF VVL

MIGRATION HISTORY	SOEP 2000 – 2004		SUF VVL 2004	
	n	Percent	n	Percent
Yes	103	4.1	660	2.2
No	846	95.9	30,169	97.8
Total	949	100.0	30,829	100.0

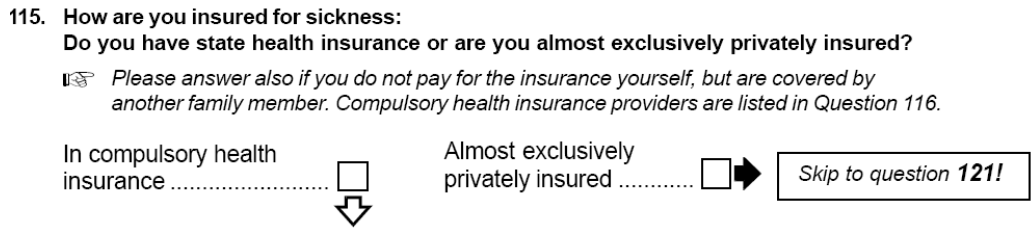
Source: FDZ-RV - SUFVVL2004 & SOEP 2005, own calculations

With more than 4%, the share of persons with a history of migration is larger in the SOEP than in the VVL (~2.2%). One explanation for the difference in the share of persons with a history of migration might be that we excluded persons who fall under the Foreign Pension Law (*Fremdrentner*) in the SUF VVL 2004. On the other hand, we were unable to exclude these persons from the SOEP, because there is no way to identify them.

5.2.9 Type of health insurance

Retirees can either be insured in statutory health insurance or hold a private health insurance plan. In the statutory health insurance, the payment of contributions by members can be either mandatory or voluntary. Question 115 in the SOEP asks for the type of health insurance the respondent holds. The original question from the individual questionnaire in the SOEP is shown in Figure 3 below:

Figure 3 Original Question from the SOEP Questionnaire "Type of Health Insurance"



Source: (TNS Infratest 2005, p. 27)

Question 117 asks for the type of member a person is in the respective health insurance. The original question from the SOEP questionnaire is shown in Figure 4 below:

Figure 4 Original Question from the SOEP Questionnaire "Type of Member"

117. Are you personally in this health insurance ...

- a mandatory paying member ☐
- a volluntary paying member ☐
- covered by the insurance of a family member ☐
- insured as pensioner, unemployed, student, or draft soldier or through your community service work ("Zivildienst")? ☐

Source: (TNS Infratest 2005, p. 27)

In the SUF VVL 2004, the variable AT provides information about the type of health insurance.

Figure 5 Original Item from the SUF VVL 2004 Codebook "Type of Health Insurance"

Stellen von - bis	Feld-länge	Feldbe-zeichnung	Erläuterung
109	1	AT	**Art des Krankenversicherungsverhältnisses** Private Versicherung oder Beiträge zur Pflichtversicherung sind vom Rentenberechtigten selbst zu zahlen. Bei den Renten mit privater Krankenversicherung wird die Rente zunächst mit AT=8 festgesetzt. Der Zuschuss wird häufig im Nachhinein gewährt. Deshalb ist ein großer Teil der Fälle, die mit AT=8 gemeldet werden, inhaltlich deckungsgleich mit AT=0. (a) freiwillige und private Versicherung 0 = Beitragszuschuss nach §§ 106, 315, 319 SGB VI, ggf. wird die Höhe einer anderen Rente bei der Berechnung des Zuschusses berücksichtigt/freiwillig versichert mit Beitragszuschuss bei einer anderen Rente, die Höhe der Rente wird aber bei der Berechnung des Zuschusses mit berücksichtigt. (b) Pflichtversicherung in der gesetzlichen Krankenversicherung 5 = pflichtversichert in der gesetzlichen Krankenversicherung (c) Renten ohne Beitragszuschuss und ohne Pflichtbeitrag zur Krankenversicherung 8 = nicht nach deutschem Recht versichert, Auslandsrenten ohne AT-Kennzeichnung (blank)

Source: (Deutsche Rentenversicherung Bund 2005a)

The SUF VVL 2004 summarizes "voluntary paying members" and "members of a private health insurance" in one category (*AT*=0). Another category is the group of "mandatory paying members" (*AT*=5). The third category is the group of "persons that are not insured according to German law" (*AT*=8). In order to harmonize the variable "type of health insurance" in the SOEP with the VVL, we combined the information from variables *vp115* and *vp117* in the SOEP. Persons who reported being privately insured in question *vp115* were grouped as "voluntary paying members or members of a private health insurance". The same applies to persons who reported being voluntary paying members in question *vp117*. All other groups were considered to

be "mandatory paying members". The category "persons not insured according to German law" is a peculiarity in the SUF VVL 2004. According to information from the Federal German Pension Insurance, most cases that fall into this category are persons whose health insurance status has not been validated at the point of data preparation. We therefore subsumed these cases under the category "mandatory paying members" in order to obtain a comparable measure for the "type of health insurance" and accept the slight inaccuracy of this procedure.

Table 17 Distribution of the Variable "Type of Health Insurance" in the SOEP and SUF VVL 2004

HEALTH INSURANCE	SOEP 2000 – 2004		SUF VVL 2004	
	n	Percent	n	Percent
Statutory	863	90.9	28,613	92.8
Private	86	9.1	2,216	7.2
Total	949	100.0	30.829	100.0

Source: FDZ-RV - SUFVVL2004 & SOEP 2005, own calculations

Inconsistencies between the SOEP and VVL can be traced to the differences in the categories of the variable "type of health insurance" in both datasets.

5.2.10 Educational Attainment

Educational attainment is a crucial variable for explaining the individual's lifetime earnings and consequently, the level of public pension benefit the person receives as he/she retires. Variables that describe the educational attainment of a person are available in both datasets. However, the ways in which these variables are measured differ considerably. In the SUF VVL 2004, educational attainment is measured by combining the highest secondary or tertiary schooling degree with information about the completion of vocational training (Fitzenberger et al. 2005).

However, the reliability of the measure needs to be called in question. This is because the information has no relevance whatsoever for the calculation of the public pension benefit.

Hence, there is no incentive for employers to invest much time and manpower in providing accurate information to the branches of the social insurance system. As a consequence, the variable in the SUF VVL 2004 has a high number of missing values.

Therefore, it needs to be determined whether the variables that measure "educational attainment" are comparable in the SOEP and the SUF VVL 2004 and hence, are useful variables for the matching procedure. To determine this, we first had to align the operationalization in both datasets. We wished to modify the SOEP information so that it would fit the information provided in the SUF VVL 2004. In a second step, we compared the distribution of the variable in both datasets and analyzed whether we could find the positive effect of higher educational attainment on the level of public pension benefits that we expected.[49] Table 17 illustrates the operationalization of the variable in the SUF VVL 2004.

[49] Educational attainment has been found to have a positive effect on the level of public pension benefits in other studies based on data from the Statutory Pension Insurance (Rehfeld, Bütefisch und Hoffmann 2007).

Table 18 Distribution of the Variable "Educational Attainment" in the SUF VVL 2004

Value Labels for Different Categories of Educational Attainment (based on TTSC3)	**Value**	**Share in % (n)**
Missing Information	-9	49.8 (15,347)
Secondary school or higher secondary school without vocational training *(Hauptschule/Realschule ohne abgeschlossene Berufsausbildung)*	1	6.4 (1,967)
Secondary school or higher secondary school with completed vocational training *(Hauptschule/Realschule ohne abgeschlossene Berufsausbildung)*	2	27.1 (8,355)
High school or technical high school without vocational training *(Abitur oder Fachhochschule ohne abgeschlossenen Berufsausbildung)*	3	0.2 (57)
High school or technical high school with completed vocational training *(Abitur oder Fachhochschule ohne abgeschlossenen Berufsausbildung)*	4	0.9 (278)
Completed degree at *Fachhochschule*	5	2.1 (647)
Completed degree at a university or technical university	6	2.4 (746)
No information available/ degree unknown	7	11.1 (3,432)

Source: FDZ-RV - SUFVVL2004, own calculation

To obtain a comparable measure in the SOEP data, we had to restructure the information. Table 19 illustrates the approach that we used. The four upper boxes present the four educational attainment variables in the SOEP. At the bottom of the table, the first four columns show how these variables needed to be combined in order to match the measure in the SUF VVL 2004. Adapting the approach of Haak (Haak 2006), we then constructed a new education variable that differentiates between low, medium, and high educational attainment. The category of school dropouts, which is a category in the SOEP, but not in the SUF VVL, was grouped under low educational attainment. [50]

[50] In contrast to Haak (2006) and Clemens et al. (2007), persons who have completed high school or technical high school but have not completed vocational training were categorized in the group of medium educational training.

Table 19 Educational Attainment Variables in the SOEP and how to align them to the SUF VVL 2004 Categories[51]

SOEP	
$PSBIL School Education	
Missing	-1
Secondary School	1
Higher Secondary School	2
Fachhochschulreife	3
High school	4
Other degree	5
No completed degree	6

SOEP	
$PBBIL01 Vocational Training	
Does not apply	-2
Missing	-1
Apprenticeship	1
Full-time vocational school	2
School for health care professions	3
Trade and technical school for vocational education	4
Training for public employees	5
Other training	6

SOEP	
$PBBIL02 Tertiary Education	
Does not apply	-2
Missing	-1
University of Applied Sciences	1
University, Technical University	2
University in a foreign country	3
Engineering School and School of Applied Sciences of former GDR	4
University of former GDR	5

SOEP	
$PBBIL03 Completed Degree	
Does not apply	-1
Missing	-2
No completed degree	1
College Degree	2

COMBINATION OF SOEP VARIABLES					**VVL CATEGORIES FOR EDUCATIONAL ATTAINMENT**	**NEW EDUCATION VARIABLE**
PSBIL	**PBBIL01**	**PBBIL02**	**PBBIL03**		**TTSC3**	
5					No information available/ Degree unknown	Unknown (-2)
- 1					Missing Information	Missing (-1)
6					(School Drop-Out: no degree)	Low
1 or 2			1		Secondary school or higher secondary school without vocational training	Low
1 or 2	>0			→	Secondary school or higher secondary school with completed vocational training	Medium
3 or 4			1		High school or technical high school without vocational training	Medium
3 or 4	>0				High school or technical high school with completed vocational training	Medium
3 or 4		1 or 4			Completed degree at university of applied science	High
3 or 4		2 or 3 or 5			Completed degree at a university or technical university	High

[51] If information was missing for the variable `PSBIL`, we simply combined information from `PBBIL01-PBBIL03` in order to obtain a comparable measure in the SOEP.

The distribution of the new variable for "educational attainment" is nearly congruent in the two datasets. Table 20 illustrates the distribution in the two datasets, considering only valid values.

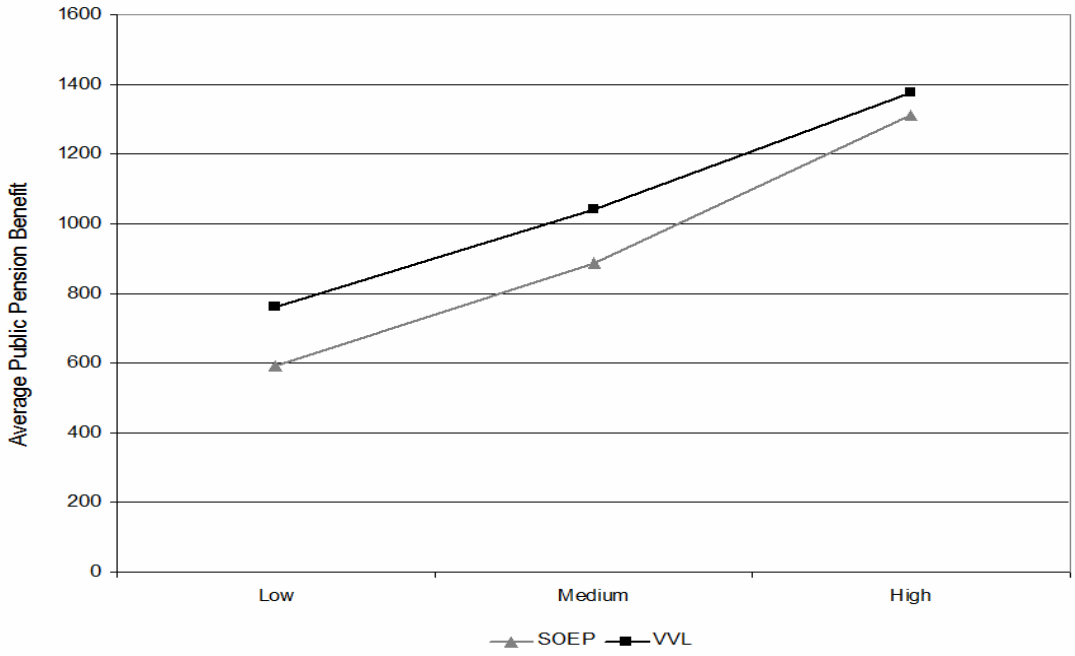
Table 20 Distribution of New Variable "Educational Attainment" in SOEP & SUF VVL

	SOEP		SUF VVL 2004	
New "Educational Attainment" Variable	n	Percent	n	Percent
Low (1)	156	17.8	1,967	16.6
Medium(2)	590	66.0	8,690	72.1
High (3)	175	16.3	1,393	11.6
	921	100.00	12,050	100.00

Source: FDZ-RV - SUFVVL2004 & SOEP 2005, own calculations

We found an education gradient in the data. Graph 1 illustrates the returns to education with respect to the average public pension benefit. The positive relationship between higher educational attainment and public pension benefits can be demonstrated in both datasets.

Graph 1 Returns to Education in the SOEP and SUF VVL 2004 based on the Total Sample Population



Source: FDZ-RV - SUFVVL2004 & SOEP 2005, own calculations

6 Estimating Regression Equations

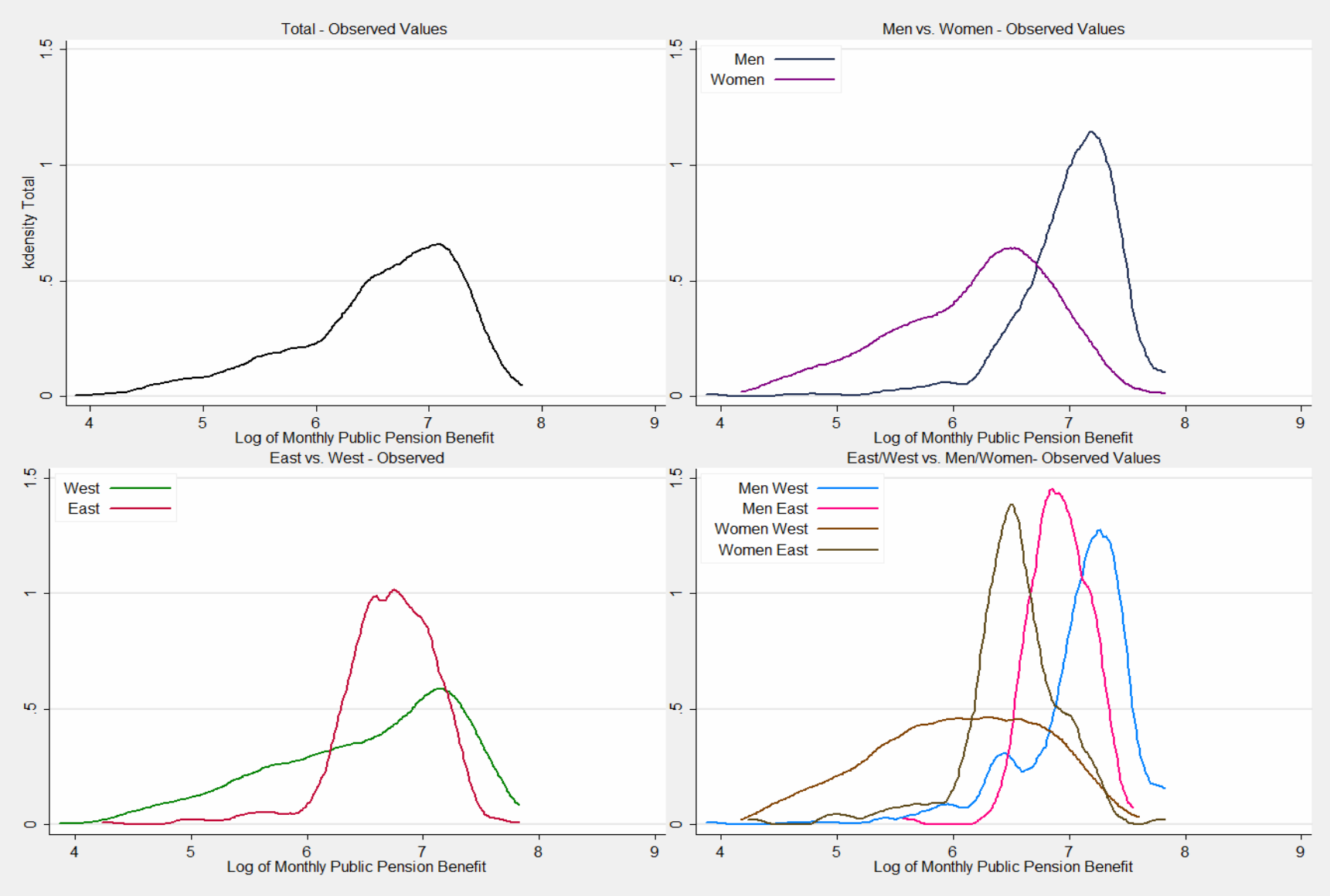
6.1 Which Variables Enter Which Model?

After specifying the two populations of interest and ensuring that certain core variables are distributed in similar ways, we needed to check whether the SOEP dataset can be compared with the VVL in a multivariate analysis. This step is necessary because the actual matching procedure will be carried out over the estimated regression coefficients. Therefore, we had to analyze whether the regression estimates in the SOEP and the VVL correspond in terms of strength and direction. A total correspondence is rather unlikely, due to differences in the measurement of certain variables and considerable differences in sample sizes.

The regression equations were estimated for our newly specified SUF VVL 2004 population and for the group of first-time pensioners from 2000 to 2004 in the SOEP. We opted for this SOEP population because it appeared to come closest to the SUF VVL 2004 population with respect to the distribution of relevant matching variables. In addition, the pooled population of first-time pensioners has a reasonable sample size, which made it possible to differentiate further into various demographic groups.

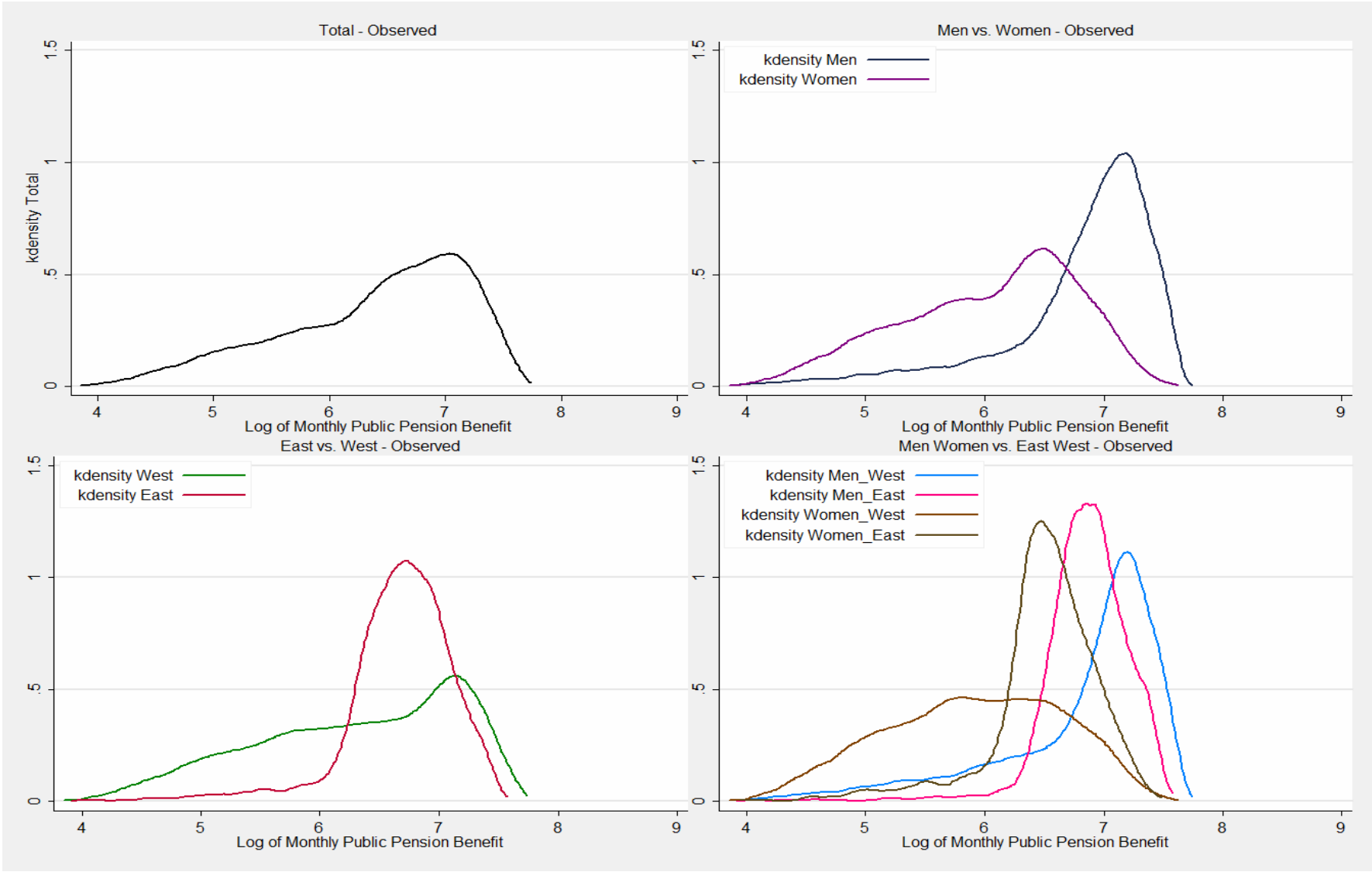
Graphs 2 and 3 illustrate why further differentiation is necessary in both datasets. The distribution of the monthly public pension benefit differs considerably between different demographic groups. In particular, the distribution of public pension benefits of West German women deviates quite clearly from the rest of the population. Furthermore, the calculation of the average time spent in different types of employment revealed considerable dissimilarities between the groups (for example East and West German women). We took these dissimilarities into account by estimating separate regressions for various subsamples. In order to assess which model is the best for our purposes, we went from a very general model that was based on the total sample population to subsamples specified by gender and region (e.g. Model VIII for West German women).

Graph 2 Comparison of Distribution for Different Demographic Groups - SOEP



Source: SOEP 2005, own calculations

Graph 3 Comparison of Distribution for Different Demographic Groups – SUF VVL 2004



Source: FDZ-RV - SUFVVL2004, own calculations

Table 21 summarizes the models estimated, with different subsamples being specified. Column 2 briefly describes each subsample. Column 3 lists the abbreviation we use for each subsample in the remainder of the paper. Columns 4 and 5 compare the case numbers per subsample in the SOEP and the SUF VVL 2004.

Table 21 Subsamples within the Sample Population and Case Numbers

(1)	(2)	(3)	(4)	(5)
	Population	**Abbreviation**	**n SOEP**	**n SUF VVL**
I	Total sample population	Total	949	31,744
II	Total West population: Only West, Men & Women	Total West	662	24,213
III	Total East population: Only East, Men & Women	Total East	289	7,261
IV	Total Male population: Only Men; East & West	Total Men	443	12,274
V	Total Female population: Only Women; East & West	Total Women	506	17,200
VI	Men-West population: Only Men, Only West	West-Men	304	10,727
VII	Men-East population: Only Men, Only East	East-Men	139	3,547
VIII	Women-West population: Only Women, Only West	West-Women	358	13,486
IX	Women-East population: Only Women, Only East	East-Women	148	3,714

Source: Own illustration

The regression was estimated for the subsamples summarized in Table 20. In a first set of regressions, we considered only the aggregated time spent in different types of employment, plus some basic controls for sex and region. This regression equation comes closest to our initial research goal of assessing the impact of the individual's employment history on the level of public pension benefits.

In a second set of regressions, we expanded the number of controls by including additional variables in the estimation, such as migration, family status, type of health insurance, retirement age, education, number of children, and educational attainment. These variables are other potential matching variables that are measured in both datasets.

Not every variable has to be included in each subsample. Therefore, the number of variables varies per model. For women, we excluded the variable "years in the military". For men, we excluded the variable "years in homeproduction". Even though some women and men have valid values in the respective types of employment, we did not consider these variables in the regression estimation. Their inclusion in the model would lead to biased estimates. In the extended models, the variable "number of children" was excluded in the male subsamples, because we only have information on the birth history for women.

6.2 Regression Diagnostics and Modifications

6.2.1 Dependent variable: Monthly Public Pension Benefit

After running the first set of regressions, we realized that the results for the SOEP and the SUF VVL 2004 were quite different. In some of the SOEP regressions, the value of the constant was highly negative, contrary to the first intuition. A closer look at the distribution of the dependent variable revealed that there were some striking outliers in the SOEP data. Column 2 in Table 21 lists the largest values in the distribution of the "monthly public pension benefit".

Table 22 Summary Statistics of "Monthly Public Pension Benefit" in the SOEP

(1)		(2)	(3)	
Percentiles	**Amount in €**	**Smallest Values**		
1%	97			
5%	160	48		
10%	242	65	**Mean**	895.9684
25%	500	69	**Std. Dev.**	611.133
		75		
50%	811			
		Largest Values	**Variance**	373483.6
75%	1,200	3,780	**Skewness**	3.413083
90%	1,540	4,500	**Kurtosis**	34.19637
95%	1,720	5,800		
99%	3,000	8,500		

Source: SOEP 2005, own calculations

The four highest values in the SOEP data range from 3780 Euro to 8500 Euro, which is far beyond anything possible within current pension legislation. Due to the maximum contribution

ceiling, a person can accumulate a maximum of two earning points per year.[52] Using a hypothetical earnings profile, we tried to determine the maximum monthly public pension benefit that it is possible for any person to reach within the rules and regulations of the Statutory Pension Insurance. We assumed that the hypothetical person accumulates two earning points per year, each point being worth the actual pension value of 26.13 Euro (Deutsche Rentenversicherung Bund 2006b).[53] Furthermore, the person was assumed to work year in and year out for 45 years. Plugging these numbers into the simplified pension benefit formula, our hypothetical person would receive a maximum monthly public pension benefit of 2,351 Euro. The value of 2,351 Euro is in line with the range of values we calculated for the SUF VVL 2004, as illustrated in Table 22.

Table 23 Summary Statistics of "Monthly Public Pension Benefit" in the SUF VVL 2004

(1)		(2)	(3)	
Percentiles	**Amount in €**	**Smallest Values**		
1%	76			
5%	125	21		
10%	178	21	**Mean**	813.04
25%	363	26	**Std. Dev.**	500.29
		29		
50%	773			
		Largest Values	**Variance**	250288.7
75%	1,184	2,077	**Skewness**	.355654
90%	1,536	2,088	**Kurtosis**	2.137249
95%	1,712	2,166		
99%	1,913	2,294		

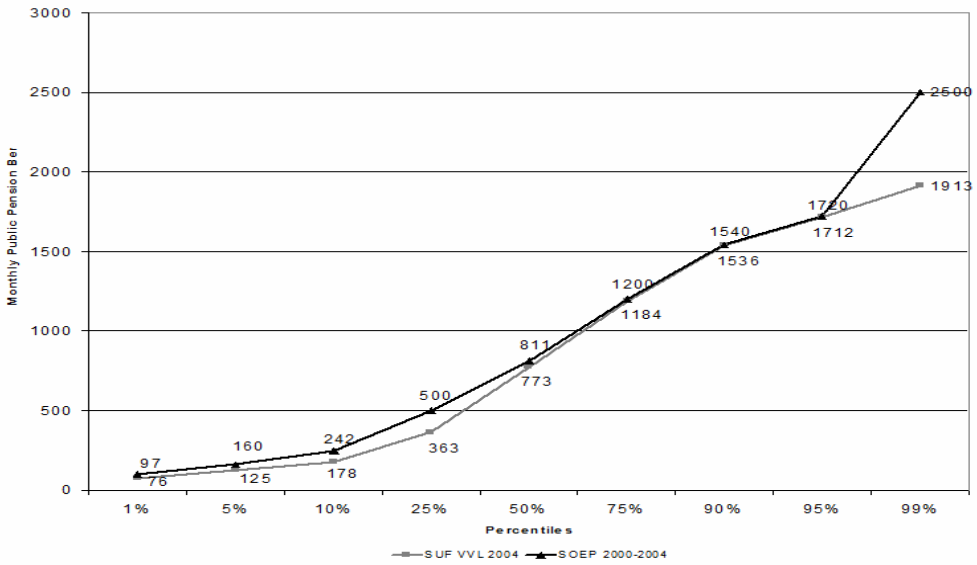
Source: FDZ-RV - SUFVVL2004, own calculations

We assumed that the outliers in the SOEP are cases of nonsampling errors. For example, the respondent might have misinterpreted the question and therefore reported the annual pension benefit instead of the monthly benefit received from the statutory pension insurance, or the respondent might have interpreted the question in such a way that he reported the total old-age

[52] In the year 2004, the ceiling was set at monthly earnings of 5,150 Euro. No social insurance contributions are paid for earnings above this ceiling. Monthly earnings of 5,150 Euro roughly correspond to two earning points per year (Deutsche Rentenversicherung Bund 2006b).
[53] This is the actual pension value for West Germany.

income adding up income from all different sources. Another explanation for nonsampling errors might be error on the part of the interviewer. Instead of noting a monthly public pension benefit of 850 Euro, the interviewer might have noted a monthly public pension benefit of 8,500 Euro. Given that we are unable to assess which kind of error applies, we decided to topcode the monthly public pension benefit at 2,500 Euro. We opted against dropping these implausible cases, because the number of cases in the SOEP was already small. In the group of first-time pensioners from 2000 to 2004, 34 cases were affected by the topcoding.[54] Graph 4 illustrates the percentile-comparison of the SOEP and the SUF VVL 2004 after the topcoding. The distribution of the dependent variable in both datasets appears to be nearly congruent in the lowest decile. Between the second and the fifth deciles the distributions disperses, but become very similar again in the further course of the distribution. The large deviation at the 99th percentile persists even after the topcoding.

Graph 4 Comparison of Percentiles in the Distribution of "Monthly Public Pension Benefit" SOEP vs. SUF VVL 2004



Source: FDZ-RV - SUFVVL2004 & SOEP 2005, own calculations

[54] The summary statistics provided in Section 5.1. already consider the topcoding of the variable "monthly public pension benefit" in the SOEP.

6.2.2 The Relationship between Public Pension Benefits and Years of Employment

At the outset of the analysis, we expected to find a strongly positive relationship between public pension benefits and years of employment. In an employment-centered system, such as the Federal German Pension Insurance, benefits are closely linked to previous periods of employment. Most forms of employment are subject to social insurance contributions.[55] With the payment of these contributions, the individual accumulates entitlements that qualify for the later receipt of public pension benefits.

Roughly speaking, individuals with long periods in employment usually receive high public pension benefits. We did find a positive relationship between public pension benefits and years of employment in the VVL data. Surprisingly, we did not find the expected relationship between public pension benefits and years of employment in the multivariate regression results that were based on the SOEP data.

The differences in the regression results between SOEP and SUF VVL 2004 are due to the differences in the two sample populations. As mentioned above, the SUF VVL 2004 contains data only for individuals who retired in the year 2004. All employment periods are pension-relevant employment periods. Hence, anything other than a strong relationship between years in employment and the monthly public pension benefit would have been implausible. We do not have information about periods of self-employment or other non pension-relevant forms of employment (e.g. illegal employment *Schwarzarbeit*) that could affect this relationship. Missing information could be an indication of these forms of employment. However, we cannot be sure about this.

In contrast, SOEP respondents report periods of employment, irrespective of whether or not these periods are pension-relevant. Hence, we do not have any way of discriminating between

[55] Certain occupational groups are exceptions to the rule in that they are not obligated to pay social insurance contributions, e.g. the self-employed who can opt to pay voluntary contributions into the public pension insurance or pay money into a private pension plan.

these periods. This explains why we did not find a clear-cut positive relationship between years of employment and public pension benefits. In the data, we might have cases who reported that they have worked for 40 years, but who receive only a very small pension. It is possible that these are cases in which the person worked for a few years in employment, during which time he/she paid social insurance contributions and then worked for many years in self-employment, during which no pension entitlements were accumulated.

To address this problem, we controlled for the occupational status of a person for the period 1995 to 2004 (variables *stib95-stib04*). We created two more dummy variables for the regression equation; namely, *selfemployment* and *civilservant*. If a person reported being self-employed or having worked as a civil servant in any of the years, we coded the respective variables with 1. The data shows that in this cohort of retirees, very few people worked as civil servants in the years prior to retirement (n=14), whereas the number of self-employed is slightly higher (n=76). When we incorporated the two dummy variables into the regression models, the coefficients appeared to be more robust in the "Total"-models (Models I-V). In these models, the coefficients have intuitive strength and direction; namely, they are strongly negative and significant. This appears plausible, because the self-employed and civil servants are, by definition, excluded from the public pension system unless they pay voluntary contributions. The coefficients are also more robust in the "Total Men" model than in the "Total Women" model. The difference is due to the fact that men are more often self-employed or civil servants than women, at least in the cohort of retirees we are interested in. If we differentiate the sample populations by region and gender (e.g. West-Men Model), the results are less robust. Due to the small number of self-employed and civil servants in our sample population, we decided to summarize the variables *selfemployment* and *civil servant* (variable *civil_self*).[56] For the reasons mentioned above, we expected the coefficient for *civil_self* to be negative. It is not possible to identify self-employed persons or civil servants in the VVL data. The modifications were therefore confined to the SOEP data.

[56] The new variable *civil_self* indicates whether a person was either self-employed or a civil servant (*civil_self* = 1).

6.2.3 Years in Schooling and Years in Training

Some additional, but minor, modifications were made in both datasets. The variables "years in school" and "years in training" were top-coded. On average, the respondents in the SOEP dataset report approximately two years of schooling[57] and 2.3 years of training.[58] In the SUF VVL 2004, respondents report an average of 0.7 years of schooling[59] and 1.35 years of training.[60] However, the distribution is distorted by some very high values. These values appear rather implausible. In the SOEP, the maximum value reported for years of schooling is 22 years and 28.5 years for training. In the SUF VVL 2004 in turn, the maximum value reported for schooling is 24.58 years and 12.75 years for training.

For our analysis, we were only interested in those times that are relevant for calculating the monthly public pension benefit. The variables "years in school" and "years in training" were therefore top-coded at a maximum of 10 years. Table 23 illustrates how many cases were affected by the topcoding in the SOEP and SUF VVL 2004 sample populations.

6.2.4 Years in Other Activities, Years Retired, and Years Missing

The distribution of the variables "years in other", "years retired" and "years missing" also reveals a large variance. However, topcoding is not an appropriate way to handle these variables. The issue is whether the variables should enter the model on a continuous scale. Coefficients should be interpreted as follows: one additional year in "other activities" increases or decreases the monthly public pension benefit by a certain amount. Given that we do not know what type of activity falls into the category "years in other", the interpretation does not necessarily make sense. As an alternative to the variables entering the model on a continuous scale, we could recode the variables into dummies. If the values of the variable "years in other" exceeded three years, the

[57] Considering only non-zero values, respondents report on average 3.5 years in schooling.
[58] Considering only non-zero values, respondents report on average 3 years of training.
[59] Considering only non-zero values, respondents report on average 2.9 years in schooling.
[60] Considering only non-zero values, respondents report on average 2.7 years of training.

new variable "other" was coded with 1. If the number of years missing exceeded three, the new variable "missing" was coded with 1. If the number of years retired exceeded four years, the new variable "retired" was coded with 1.

The striking differences between the two datasets are due to the fact that the SUF VVL 2004 only records those periods that are relevant for calculating the public pension benefit. If none of the 13 employment situations applied, the respective month was coded as a missing. In the SOEP in turn, respondents are free to report any activity they consider as relevant in the biography questionnaire. In the multivariate analysis, the coefficients of the dummy variables allude to whether persons that have high values in the three original variables are systematically different from others, everything else being kept constant. Table 23 summarizes the modifications and lists the number of cases that were affected by each modification.

Table 24 Data Modifications for Regression Analysis in SOEP and SUF VVL 2004

VARIABLE	NEW VARIABLE LABEL	MODIFICATION	NUMBER OF OBSERVATIONS AFFECTED BY MODIFICATION
Years in school	Years_school_n	Top Coding: If years in school exceed 10, then top-coding at 10 years of schooling.	SOEP: 35 VVL: 8
Years in training	Years_training_n	Top Coding: If years in training exceed 10, then top-coding at 10 years of training.	SOEP: 2 VVL: 1
Years in other activities	Other	Dummy Variable if years in other activities exceed three.	SOEP: 23 VVL: 4,016
Years with missing information	Missing	Dummy Variable if years with missing information exceed three.	SOEP: 138 VVL: 29,018
Years in retirement	Retired	Dummy Variable if years in retirement exceed four.	SOEP: 24 VVL: 119

Source: FDZ-RV - SUFVVL2004 & SOEP 2005, own calculations

6.2.5 Years in Military

In addition, we checked whether the procedure described above makes sense for the variable "years in military". The variable entered the regression model continuously and as a dummy. It did not make a difference in which form the variable entered the model, because the coefficients were not significant and weak in strength in both versions. The variable "years in military" was therefore excluded from the regression equation.

6.2.6 Migration History

Further analyses revealed that the variable "migration" is only significant in the Total-models and the West-models. The lack of statistical significance in the East-Models is due to the small number of persons in East Germany with a history of migration. Due to the heterogeneity within the group of persons with a history of migration, we further refined our measure for persons with a history of migration. For example, it can be assumed that respondents from France have much more in common with Germans than respondents from Ghana. Therefore, we decided to distinguish between EU and Non-EU migrants. The group of EU-migrants consists of persons that come from the EU-14 countries (EU-15 minus Germany). All the other persons with a history of migration were placed in the group of Non-EU migrants. Contrary to our expectations, it did not make a difference whether we included a general "migration"-measure or a further refined measure to distinguish between "EU- and Non-EU migrants". In both cases, the variables were dropped from the estimation of the East models because of the small case numbers. The strength and the direction of the coefficients (EU-migrants and Non-EU migrants) correspond to the "migration"-coefficient. Therefore, we retained the "migration" variable in its original form.

7 Regression Results

The matching procedure will be carried out based on the actual predictions of the estimated regression coefficients. We estimated a multivariate OLS regression. In the OLS regression, the dependent variable (in our case, the logged monthly public pension benefit) was assumed to be a linear function of our independent or explanatory variables (e.g. time spent in different types of employment, gender, region, etc.) that appear on the right-hand side of the equation. The variables on the right-hand side are expected to explain the variance in the monthly public pension benefit. The variance that is left unexplained by the specified model is captured in the error term, the so-called residual. All non-observables go into the error term, even though they have explanatory power with respect to the dependent variable. The basic idea behind the ordinary least squares regression is to minimize the sum of squared errors; namely, the distance between the observed and predicted values. The estimated regression coefficients indicate how a change in one of the independent variables affects the dependent variable, holding everything else constant.

We therefore estimated roughly the same regression equation in both datasets, considering all the modifications discussed in Section 7. The overarching goal was to find a model that best predicts the monthly public pension benefit.

Table 25 compares the explained variance (r^2) in each of the nine estimated models.

Table 25 Comparison of Explained Variance in SOEP and SUF VVL 2004

	(1) TOTAL	(2) TOTAL WEST	(3) TOTAL EAST	(4) TOTAL MEN	(5) TOTAL WOMEN	(6) MEN WEST	(7) MEN EAST	(8) WOMEN WEST	(9) WOMEN EAST
SOEP	0.57	0.59	0.57	0.34	0.42	0.31	0.37	0.35	0.56
SUF VVL	0.78	0.80	0.70	0.74	0.70	0.76	0.61	0.67	0.67

Source: FDZ-RV - SUFVVL2004 & SOEP 2005, own calculations

As expected, the regression models that were based on the SUF VVL 2004 data explain much more of the variance than those that were based on the SOEP data. The differences in explained variance between the VVL 2004 and the SOEP are due to the fact that the SUF VVL 2004 data only considers those periods that are actually relevant for the calculation of the individual's monthly public pension benefit, whereas the SOEP considers all periods, irrespective of whether or not these periods are pension-relevant. In both datasets, the model fit is best for the "Total West" model, with 80% of the variance explained in the SUF VVL 2004 and 59% explained in the SOEP. The "Men West" model has the least good fit in the SOEP data, with 31% of the variance explained. The "Men East" model has the least good fit in the SUF VVL 2004 data, with 61% of the variance explained.

(Insert Appendix D)

Table 26 compares the direction and significance levels of the coefficients in the two datasets. SOEP results are presented in the upper left part of the box, SUF VVL 2004 results in the lower right. Boxes are highlighted in green if the effect of the regression coefficients works in the same direction in both datasets. Boxes are highlighted in red if the effect of the regression coefficients works in different directions in both datasets. The boxes are white if the respective variable is measured in only one of the two datasets or if it was dropped due to small case numbers. The significance level does not matter in the highlighting of the boxes. For example, if the coefficients in the two datasets work in the same direction, but the coefficient in the SOEP is significant at the 10% level and the SUF VVL 2004 coefficients at the 1% level, the box is still highlighted in green. This is because significance levels are largely a matter of case numbers. Given that the SUF VVL 2004 contains so many cases, most of the coefficients are significant at the 1% level.

Table 26 Comparison of Direction and Significance Levels of Regression Coefficients in the SOEP and SUF VVL 2004

Dependent Variable	(1) Total	(2) Total West	(3) Total East	(4) Total Men	(5) Total Women	(6) Men West	(7) Men East	(8) Women West	(9) Women East
Years in School (topcode)	+/*** / +/***	+/*** / +/***	+/*** / +/***	+/*** / +/***	+/** / +/***	+/*** / +/***	+/*** / +/***	+ / +/***	+/*** / +/***
Years in Training (topcode)	+/*** / +/***	+/*** / +/***	+/*** / +/***	+/* / +/***	+/*** / +/***	+/* / +/***	+ / +/***	+/** / +/***	+/*** / +/***
Years in Employment	+ / +/***	+ / +/***	+/** / +/***	+ / +/***	- / +/***	+ / +/***	+/** / +/***	- / +/***	+ / +/***
Years in Unemployment	-/*** / +	-/*** / -	-/* / -/*	-/*** / +/***	-/*** / +/***	-/*** / +/**	-/** / -/***	-/*** / +	+ / +/***
Years in Homeproduction	-/*** / +/***	-/*** / +/***	-/* / +/***		-/*** / +/***			-/*** / +/**	- / +/***
Retired (Dummy)	- / +/***	- / +/***	+ / +/***	-/* / +/***	+ / +/***	-/* / +/***	+ / +/***	- / +/***	+ / +/***
Other (Dummy)	-/** / +/***	-/** / +/***	- / +/***	-/** / +/***	-/** / +/***	-/* / +/***	Dropped / +/***	-/** / +/***	+ / +/***
Missing (Dummy)	-/* / -/***	- / -/***	-/** / -/***	-/* / -/***	- / -/***	- / -/***	- / -/***	- / -/***	-/*** / -/***
Receives Pension for Civil Servants (Dummy)	-/*** / n/a	-/*** / n/a	- / n/a	-/*** / n/a	dropped / n/a	-/*** / n/a	- / n/a	dropped / n/a	dropped / n/a
Receives Private Pension (Dummy)	-/* / n/a	-/* / n/a	- / n/a	-/** / n/a	+ / n/a	-/** / n/a	- / n/a	+ / n/a	+ / n/a
Worked as Civil Servant/ Self Employed (Dummy)	-/*** / n/a	-/*** / n/a	-/** / n/a	-/*** / n/a	-/*** / n/a	-/*** / n/a	-/** / n/a	-/** / n/a	+ / n/a
Educational Attainment: low (Reference Category: medium)	- / -/***	- / -/***	- / -/***	-/* / -/***	- / -/***	- / -/***	- / -/***	- / -/***	- / -/***
Educational Attainment: high (Reference Category: medium)	+/** / +/***	+/** / +/***	+ / +/***	+ / +/***	+ / +/***	+/* / +/***	+ / +/***	- / +/***	+ / +/***
Educational Attainment: missing (Reference Category: medium)	+ / -/***	- / -/***	+/*** / -/***	-/ / -/***	+/** / -/***	- / -/***	+ / -/***	- / -/***	+/*** / -/***
Educational Attainment: unknown (Reference Category: medium)	-/** / -/***	- / -/***	- / -/***	-/*** / -/***	+ / -/***	-/*** / -/***	dropped / -/***	- / -/***	- / -/***
Sex (female=1)	-/*** / -/***	-/*** / -/***	-/*** / -/***						
West (west=1)	+/*** / +/***			+/*** / +/***	- / +/***				

7.1 Discussion

The majority of boxes in Table 26 are colored in green, which indicates that the independent variables work in the same directions in both datasets. The results meet our expectations and the coefficients point in the intuitive direction. This is also true for the constants in all models, which are all positive and highly significant. Pronounced differences between the two datasets can be stated for the following variables (red boxes): "years in unemployment", "years in home-production", "retired" and "other", as well as "educational attainment: missing". In what follows, we discuss the reasons for the inconsistencies in the coefficients and search for better functional equivalents in the two datasets.

7.1.1 Years in Homeproduction

The inconsistencies in the variables "years in homeproduction" and "years in unemployment" are due to the fact that the variables do not measure the same thing in both datasets. In the SUF VVL 2004, "years in homeproduction" only refers to pension-relevant periods, such as child-care periods or child-care credits (*Kinderberücksichtigungszeiten* or *Kindererziehungszeiten*). If a person opted to stay at home thereafter, this will not be captured in the variable "years in homeproduction". Instead, if no other pension-relevant circumstance applies, the respective period will be coded as a missing. Furthermore, we need to consider the priority rules that were applied when the data was prepared. If two pension-relevant types of employment overlap, the type of employment that we observe in the data depends on the priority rules. Given that child-care periods have the lowest overall priority (compare Section 6.2), we only observe them if no other pension-relevant circumstance applies.

In contrast, in the SOEP, "years in homeproduction" can cover all those periods in which a person stayed at home to manage the household or care for children, irrespective of whether or not these periods were pension-relevant. We tried to control for the fact that the SUF VVL 2004 follows priority rules by considering homeproduction in the SOEP only if a person reported no

other type of employment in a given year. This approach did not yield the desired results. We therefore needed to find a better functional equivalent in the two datasets.

A promising way to obtain functional equivalents was to combine the variables "years missing" and "years in homeproduction". Since "years in homeproduction" in the SUF VVL 2004 considers only pension-relevant periods, the same has to be true for the SOEP. We therefore had to make plausible assumptions on the basis of the applicable pension rules for the group of female first-time pensioners. Women receive one year of child-care credits for all children born before January 1st 1992 (§ 56 SGB VI). For all children born thereafter, women receive three years of child-care credits. In the SOEP sample of first-time old-age pensioners in 2004, there are no women with children born after 1992.

To solve the problem of the two datasets measuring different things, we constructed a new variable for homeproduction that depends on the number of children. A mother of three children receives three years of child-care credits.[61] Equivalently, a mother with one child receives one year of child-care credits. The difference between the actual number of years in homeproduction and the new homeproduction variable was set to missing. Given that in the SUF VVL 2004, a month is set to missing if no other pension-relevant period applies, we did the same in the SOEP to obtain a functional equivalent.[62]

7.1.2 Years in Unemployment

The regression results also reveal inconsistencies in the variable "years in unemployment". In the SUF VVL 2004, the variable only represents periods of registered unemployment (§58 Abs. 3

[61] We assume that all child-care periods are credited to the pension account of the mother. In this instance, we deviate from the SUF VVL 2004.

[62] It is not feasible to take non-contributory periods (*Berücksichtigungszeiten für Kindererziehungszeiten*) into account. These periods serve to close gaps in the insurance history but do not have an increasing effect on the monthly public pension benefit (§ 57 SGB VI) There is no straightforward solution to how many years of non-contributory periods are considered per child. The maximum is 10 years. However, these non-contributory periods only apply if there is no other pension-relevant circumstance (e.g. periods of employment that are subject to social security contributions).

SGB VI), whereas respondents in the SOEP can also report unregistered periods of unemployment that went unnoticed by the social security system.[63] It is not feasible to find a functional equivalent for the two datasets with respect to the "years in unemployment". Given that we cannot control for the problem of the "hidden labor force", we will have to accept this imperfection in the matching procedure.

7.1.3 Other

The explanation for the discrepancies in the variables "other" is not as straightforward. The inconsistencies might indicate that the variable captures completely different things in the two datasets.

As illustrated in Table 9, the newly constructed category "other" is a summary measure of three different social employment situations (SES) in the SUF VVL 2004: "care giving", "invalidity and sickness", and "other". The category "other" in the VVL refers, among other things, to voluntary contributions or creditable periods, which explains the strongly positive coefficient in all models (Stegmann 2006, p. 547). The variable "other" also captures periods of sickness and invalidity and periods of care giving. Periods in which voluntary contributions were made are tantamount to periods of employment that are subject to social insurance contributions. Self-employed persons typically pay contributions into the public pension insurance on a voluntary basis. Social security contributions are also paid during periods of invalidity and sickness. During the first six weeks, a sick person is eligible for the continuation of payment ("Lohnfortzahlung im Krankheitsfall") of his/her prior earnings if he or she worked in the position for more than four weeks (§ 3 EntgFG).[64] In this case, employers and employees continue to pay contributions into the public

[63] Persons who are unemployed but not officially registered as unemployed are often referred to as the „hidden labor force" (or *Stille Reserve*). For an encompassing overview over the phenomenon of the "hidden labor force" in the German labor market see (Holst 2000).

[64] EntgFG standing for "Gesetz über die Zahlung des Arbeitsentgelts an Feiertagen und im Krankheitsfall".

pension insurance as if the person was employed.[65] If a person is still sick after six weeks, he or she will receive a sickness allowance ("Krankengeld"). In this case, contributions are paid by the employee and the health care insurance.[66] Voluntary contributions and contributions that come from sickness and invalidity both have an increasing effect on the final public pension benefit, because they are based on either actual earnings if a person works as a self-employed, or past earnings in periods of sickness and invalidity.

The "other" category in the SOEP does not cover the same circumstances as the SUF VVL 2004. Instead, quite heterogeneous types of employment are subsumed under the category "other", such as being on maternity leave, traveling around the world, or being incarcerated. Obviously, these situations do not have an increasing effect on the level of pension benefits and therefore explain the differences in the direction of influence between the SOEP and the SUF VVL 2004.

Given that periods of "sickness and invalidity" as well as periods in "other" types of employment have an increasing effect on public pension benefits, we decided to treat them as if they are equivalent to regular employment. The categories "sickness and invalidity" and "other" are therefore classified under "employment subject to social insurance contributions" in the SUF VVL 2004. The category "care giving" remains in the "other" category.

7.1.4 Retired

The "retired" dummy variable in the VVL is consistently positive and highly significant in all models.[67] Intuitively, the variable should have a negative effect on the level of public pension benefits, because the German pay-as-you-go system is strongly employment-centered. It is highly likely that cases that fall under the "retired" dummy variable are cases that previously received

[65] The level of contributions to be paid depends on prior earnings.

[66] The sick allowance can be paid for up to 78 weeks within a period of three years. The level of contributions equals 80% of the contributions paid when the person received the continuation of payment.

[67] As a reminder, the variable "retired" is coded with 1 if a person has more than four years of retirement.

disability benefits. In German pension legislation, the time a person spends receiving disability pension benefits is counted as a creditable period (§ 58 Abs. 1 Ziff. 5 SGB VI). When a person receives the old-age public pension benefit (*Altersrente*) for the first time, these creditable periods are credited towards the pension account as if they were contribution periods (§ 71 Abs. 1 & 2 SGB VI). For this purpose, the Federal German Pension Insurance simply extrapolates from the employment history. The extrapolation is based on the previous employment history and prior earnings or the so-called total evaluation of contributions (*Gesamtleistungsbewertung*). Hence, if the employment history was continuous and earnings were high prior to being disabled, the total evaluation of contributions for a person is quite favorable. In fact, times in disability can then lead to an increase in pension benefits.[68]

In the SOEP, there are several explanations for what is captured in the variable *retired*. First, it might capture the receipt of disability benefits. Alternatively, it might reflect partial retirement agreements (*Altersteilzeit* or *Vorruhestand*). Elderly employees in partial retirement can negotiate with their employer to work only part-time after reaching a certain age and then slowly phase into retirement.[69] Ideally, the employee should spend the last five years of his career working part-time. However, most employees prefer the so-called "block-model". They spend 2.5 years working full-time and then 2.5 years in full retirement. In the official statistics, employees in partial retirement are considered to be employed. We do not know how SOEP respondents categorize periods in partial retirement. It is possible that they report being retired even though they are only partially retired and hence, employed according to the official statistics. The fact that the variable *retired* captures several different circumstances might explain the inconsistent effect of the variable in the SOEP and the SUF VVL 2004. Unfortunately, there is no apparent solution to construct functional equivalents in both datasets.

---

[68] Persons with more than four years in retirement accumulated on average 42 earning points compared to 31 earnings points for persons who spent less than four years in retirement.

[69] Employers and employees have a mutual interest in partial retirement, even though the motives differ quite clearly. For employers, partial retirement is a way to rejuvenate the workforce, whereas for employees, it is an alternative to early retirement that circumvents costly actuarial adjustments (Brenke 2007; Hoffmann 2007).

(Insert Appendix E)

Table 27 illustrates whether the modifications discussed in the previous paragraphs rendered the expected results.

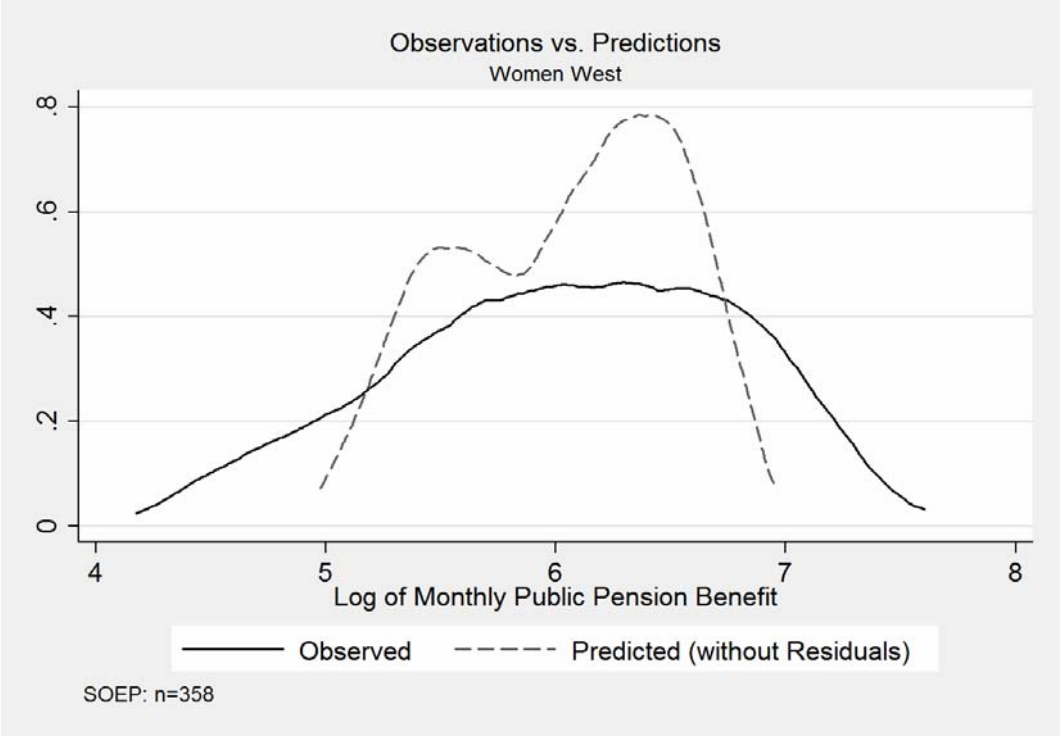
Table 27 Comparison of Direction and Significance Levels of Regression Estimates in SOEP and SUF VVL 2004

Dependent Variable Monthly Public Pension Benefit	(1) Total	(2) Total West	(3) Total East	(4) Total Men	(5) Total Women	(6) Men West	(7) Men East	(8) Women West	(9) Women East
Years in School (topcode)	+/*** / +/***	+/*** / +/***	+/*** / +/***	+/*** / +/***	+/*** / +/***	+/*** / +/***	+/*** / +/***	+*** / +/***	+/*** / +/***
Years in Training (topcode)	+/*** / +/***	+/*** / +/***	+/*** / +/***	+ / +/***	+/*** / +/***	+ / +/***	+ / +/***	+/*** / +/***	+/*** / +/***
Years in Employment	+/*** / +/***	+/*** / +/***	+/*** / +/***	+ / +/***	+/*** / +/***	+ / +/***	+/** / +/***	+/*** / +/***	+/*** / +/***
Years in Unemployment	-/*** / +***	-/*** / +/***	- / +/***	-/*** / +/***	- / +/***	-/*** / +/***	-/** / +/***	- / +/***	+ / +/***
Years in Homeproduction	-/*** / +/***	-/* / +/***	+ / +/***		-/* / +/***			-/*** / +/**	+ / +
Retired (Dummy)	+ / +/***	+ / +/***	+/** / +/***	-/* / +/***	+/** / +/***	-/* / +/***	+ / +/***	+ / +/***	+/* / +/*
Other (Dummy)	-/** / -/***	-/** / -/***	- / -/***	-/** / -	-/** / -/***	-/* / -/*	Dropped / -	-/** / -/***	+ / -/***
Missing (Dummy)	- / -/***	- / -/***	-/** / -/***	-/* / -/***	- / -/***	- / -/***	- / -/***	- / -/***	-/* / -/***
Worked as Civil Servant/ Self Employed (Dummy)	-/*** / n/a	-/*** / n/a	-/** / n/a	-/*** / n/a	-/*** / n/a	-/*** / n/a	-/** / n/a	-/** / n/a	+ / n/a
Educational Attainment: low (Reference Category: medium)	-/* / -/***	- / -/***	- / -/***	- / -/***	- / -/***	- / -/***	- / -/***	- / -/***	- / -/***
Educational Attainment: high (Reference Category: medium)	+/*** / +/***	+/*** / +/***	+/*** / +/***	+/*** / +/***	+/*** / +/***	+/** / +/***	+/*** / +/***	+/* / +/***	+/*** / +/***
Educational Attainment: missing (Reference Category: medium)	- / -/***	- / -/***	+ / -/***	- / -/***	+ / -/***	- / -/***	+ / -/***	- / -/***	Dropped / -/***
Educational Attainment: unknown (Reference Category: medium)	-/* / -/***	- / -/***	- / -/***	- / -/***	- / -/***	- / -/***	dropped / -/***	- / -/***	Dropped / -/***
Sex (female=1)	-/*** / -/***	-/*** / -/***	-/*** / -/***						
West (west=1)	+/*** / +/***			+/*** / +/***	+ / +/***				

7.2 In-Sample Predictions

In the next step, we needed to assess how well our models predict the monthly public pension benefit. It is important for the models to approach our observations as closely as possible, so that we can carry out the actual matching procedure over the predictions. The quality of the predictions would be highest if all the variance in the dependent variable was explained by our models. Since this is not the case, the density graphs that illustrate the observed distribution of the public pension benefit deviate quite clearly from the predicted public pension benefit. The in-sample predictions were made for all the specified subsamples described above. Hence, we had a total of nine in-sample predictions.[70] Graph 5 illustrates the in-sample prediction for the Women West model

Graph 5 Example for In-Sample Prediction SOEP

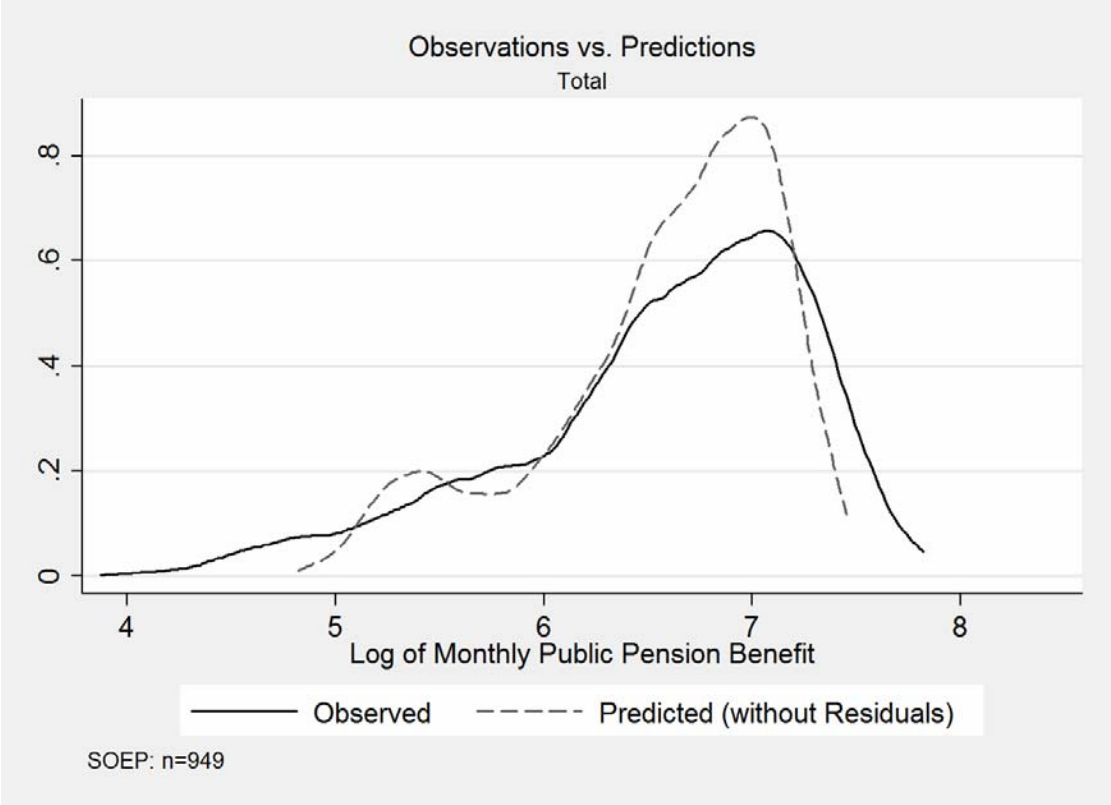


Source: SOEP 2005, own calculations

[70] We illustrate our method with the subsample of West German women. For the results of the other subsamples, please consult Appendix F.

Graph 5 shows that the distribution of the observed logged monthly public pension benefit is centered on a mean value of 6 (roughly 400 Euro). The distribution of the predicted value shows two peaks: one at the log value 5.3 (roughly 200 Euro) and the second at 6.3 (roughly 550 Euro). Our model appears not to be well-suited for predicting the smallest and largest values in the distribution of the logged monthly public pension benefit. This is due to the "regression to the mean" effect, according to which the predicted values tend to move closer to the observed sample mean than one might anticipate from the distribution of the observed values. As a result, the distribution of the predicted values is shrinking, displaying a far smaller variance than the distribution of the observed value. In this case, the variance of the observed values is 0.534, whereas the variance of the fitted values is 0.222. According to Copas, the "regression to the mean" effect is linked closely to the goodness-of-fit statistics. The better the "goodness-of-fit" of the model, the smaller the shrinkage effect in the distribution of the predictions. The problem is more pronounced in the case of small sample sizes and/or a high number of covariates, which is exactly the situation that we face (Copas 1997). In fact, the "shrinkage" and "regression to the mean" effect is not as pronounced in the larger subsamples. In order to illustrate the impact of small sample sizes and/or high numbers of covariates, Graph 6 compares the observations and predictions based on the Total SOEP sample population (n=949 compared to n= 358 in Women West sample). The smallest and largest values are also not predicted well. However, the shrinkage effect is less distinct than in the smaller subsample of West German women.

Graph 6 Example for In-Sample Prediction SOEP



Source: SOEP 2005, own calculations

7.2.1 Random Residual Assignment

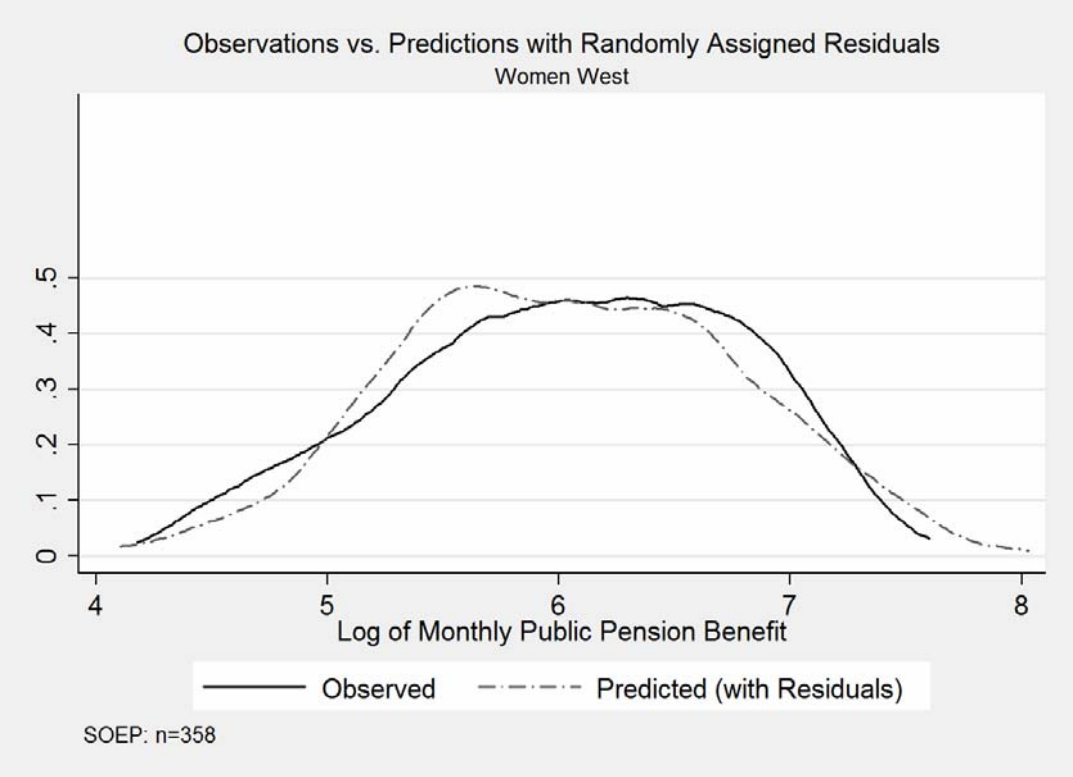
Given that we want to carry out the matching over the predicted values, it is necessary to recapture the variance in the predictions. Basically, the predictions all lie on the estimated regression line that minimizes the sum of squared errors. The residual describes the distance between the observed and the predicted values. In order to regain the variance in the prediction, we developed the following procedure.

For each pair of observations and predictions, we calculated the respective residual. If we add these residuals on top of the prediction, we obtain identical values for observations and predictions, which is not in our interest. In order to introduce some randomness, we assigned

each residual randomly to a new observation.[71] The newly assigned residual is now added on top of the prediction.

We then plotted the observations against the newly predicted values, including the randomly assigned residuals. Graph 7 illustrates whether or not the random residuals procedure was successful.

Graph 7 In-Sample Predictions with Randomly Assigned Residuals SOEP



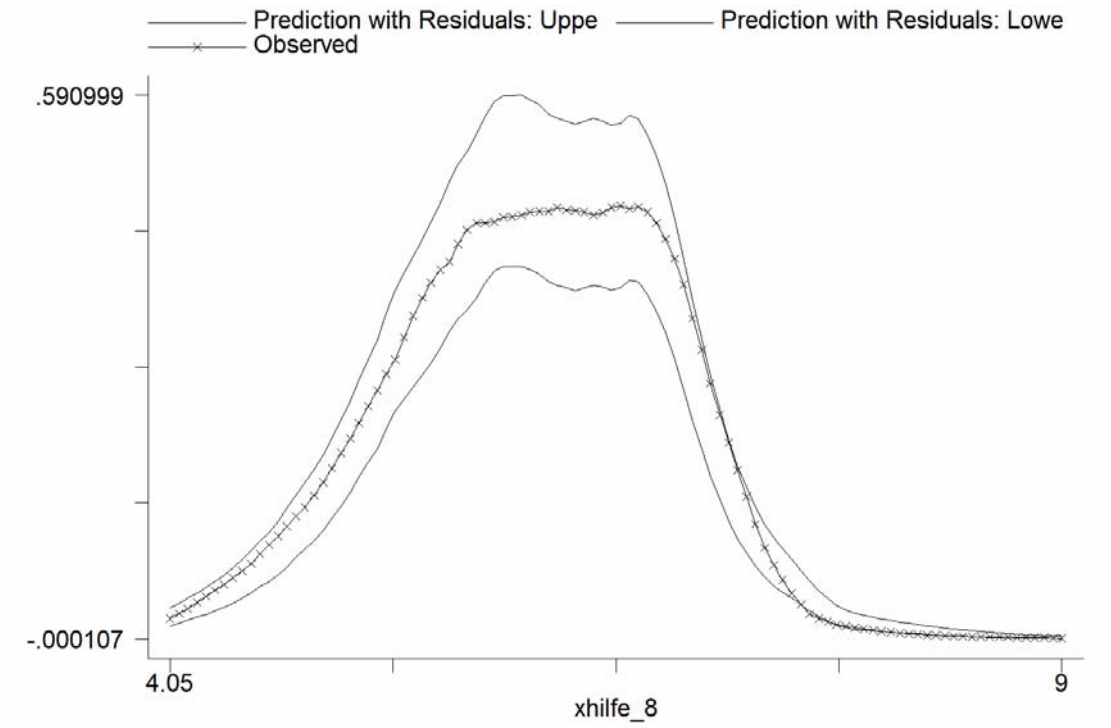
Source: SOEP 2005, own calculations

When we compare Graph 7 to Graph 5, we can see that applying the proposed procedure largely improved the accordance between the distribution of the observed and predicted values. The dashed grey line, which represents the predictions with randomly assigned residuals, no longer has two peaks and is shifted downwards to the distribution of the observed values. Furthermore, we were able to recapture the variance in the distribution. The variance of the new predictions is 0.521, compared to 0.534 for the observations.

[71] For information on the implementation of this procedure in Stata, contact the corresponding author.

Finally, it was useful to control whether the observations and the predictions with randomly assigned residuals are significantly different from each other. Therefore, we plotted the 95% confidence bands of the new predicted values around the observations. The two variables are significantly different from each other if the observed values lie outside the 95% confidence bands. Graph 8 depicts the results.

Graph 8 95% Confidence Bands for Predictions with Randomly Assigned Residuals and Observations



Source: SOEP 2005, own calculations

The solid lines represent the upper and lower limit of the 95% confidence band, whereas the dashed line represents the observed values for West German women. As we can see, the observations lie perfectly within the confidence bands, which indicate that the observations and the modified predictions do not differ significantly from each other.

7.3 Out-of-Sample Predictions

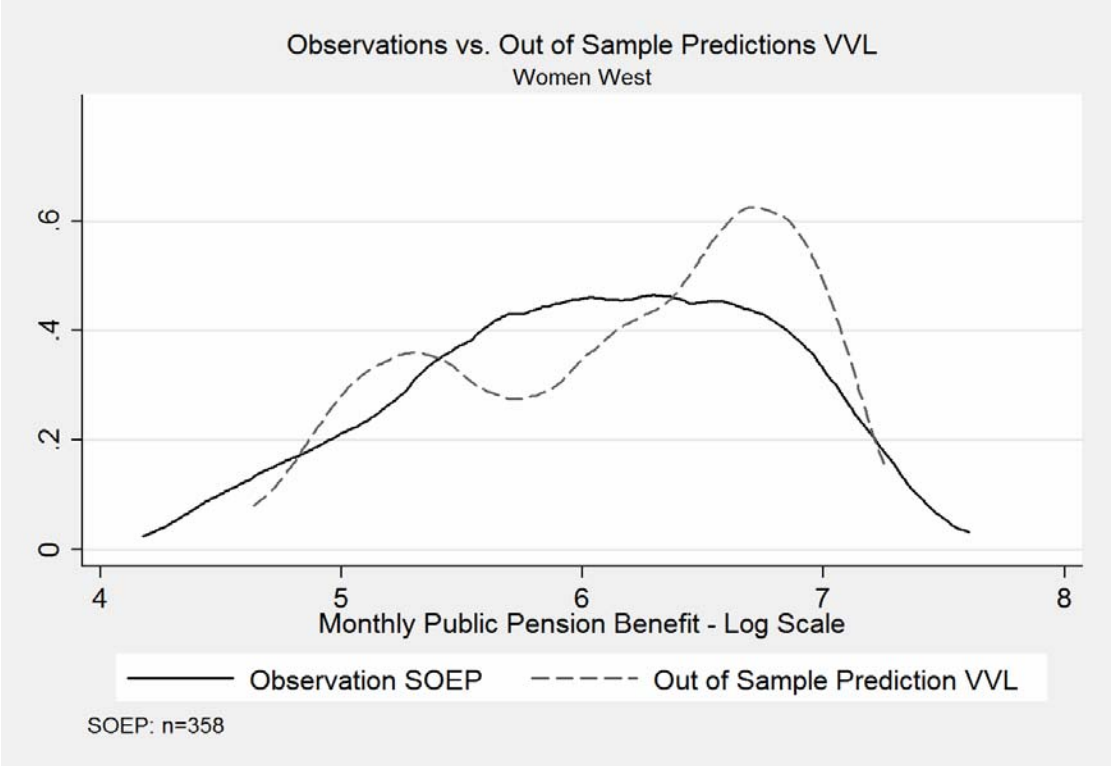
In the final step, we needed to check whether we can replicate the SOEP results with the VVL data. In the out-of-sample prediction, we plugged the coefficients that we estimated in the SUF

VVL data into the SOEP, because the coefficients estimated on the basis of the SUF VVL 2004 are much more robust than those estimated on the basis of the SOEP. The following equation shows the calculation of the out-of-sample prediction for West German women, where α is the estimated constant of the model for West German women in the SUF VVL 2004 and β_1 to β_{12}, inclusive, are the estimated coefficients for West German women model. These coefficients were then multiplied by the respective values (x_1 through x_{12}) in the SOEP.

$$\hat{y}_{\text{Out-of-.sample Women West}} = \alpha_{\text{Women West VVL}} + \beta_{1\ \text{Women West VVL}} * \text{years\_school\_n}(x_1)_{\text{SOEP}} + \beta_{2\ \text{Women West VVL}} * \text{years\_training\_n}(x_2)_{\text{SOEP}} + \beta_{3\ \text{Women West VVL}} * \text{years\_employment}_{\text{SOEP}}(x_3) + \beta_{4\ \text{Women West VVL}} * \text{years\_unemployed}_{\text{SOEP}}(x_4) + \beta_{5\ \text{Women West VVL}} * \text{homeproduction\_new}_{\text{SOEP}}(x_5) + \beta_{6\ \text{Women West VVL}} * \text{retired}_{\text{SOEP}}(x_6) + \beta_{7\ \text{Women West VVL}} * \text{other}_{\text{SOEP}}(x_7) + \beta_{8\ \text{Women West VVL}} * \text{missing\_dummy}_{\text{SOEP}}(x_8) + \beta_{9\ \text{Women West VVL}} * \text{education\_low}_{\text{SOEP}}(x_9) + \beta_{10\ \text{Women West VVL}} * \text{education\_high}_{\text{SOEP}}(x_{10}) + \beta_{11\ \text{Women West VVL}} * \text{education\_missing}_{\text{SOEP}}(x_{11}) + \beta_{12\ \text{Women West VVL}} * \text{education\_unknown}_{\text{SOEP}}(x_{12})$$

Graph 9 illustrates the out-of-sample prediction for West German women comparing the distribution of the observations in the SOEP with the predictions that were based on coefficients estimated in the SUF VVL 2004.

Graph 9 Example for Out-of-Sample Prediction

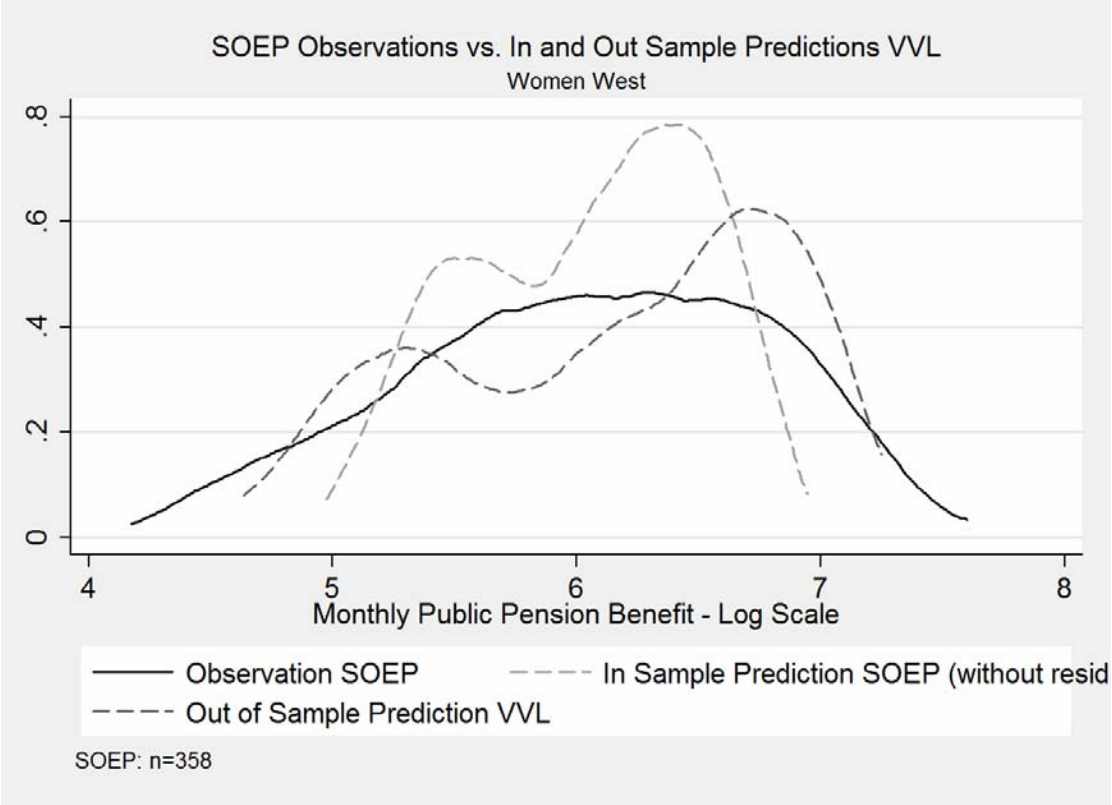


Source: FDZ-RV - SUFVVL2004 & SOEP 2005, own calculations

Similar to our first in-sample-prediction, the out-of-sample prediction for West German women deviates quite clearly from the distribution of the observations. The distribution of the out-of-sample prediction also shows two peaks: one at a log value of 5.3 (~ 200 Euro) and the other at 6.8 (~900 Euro), which is shifted slightly to the right compared to the second peak in the in-sample prediction. The "regression to the mean" effect is less pronounced in the out-of-sample prediction. Hence, the smallest and highest values in the distribution of the observations are predicted better by the out-of-sample prediction. The next graph confronts the distribution of the observations with the in- and out-of-sample predictions.[72]

[72] The out-of-sample predictions for the other demographic groups are illustrated in Appendix G.

Graph 10 Comparison of Observations and In and Out of Sample Prediction

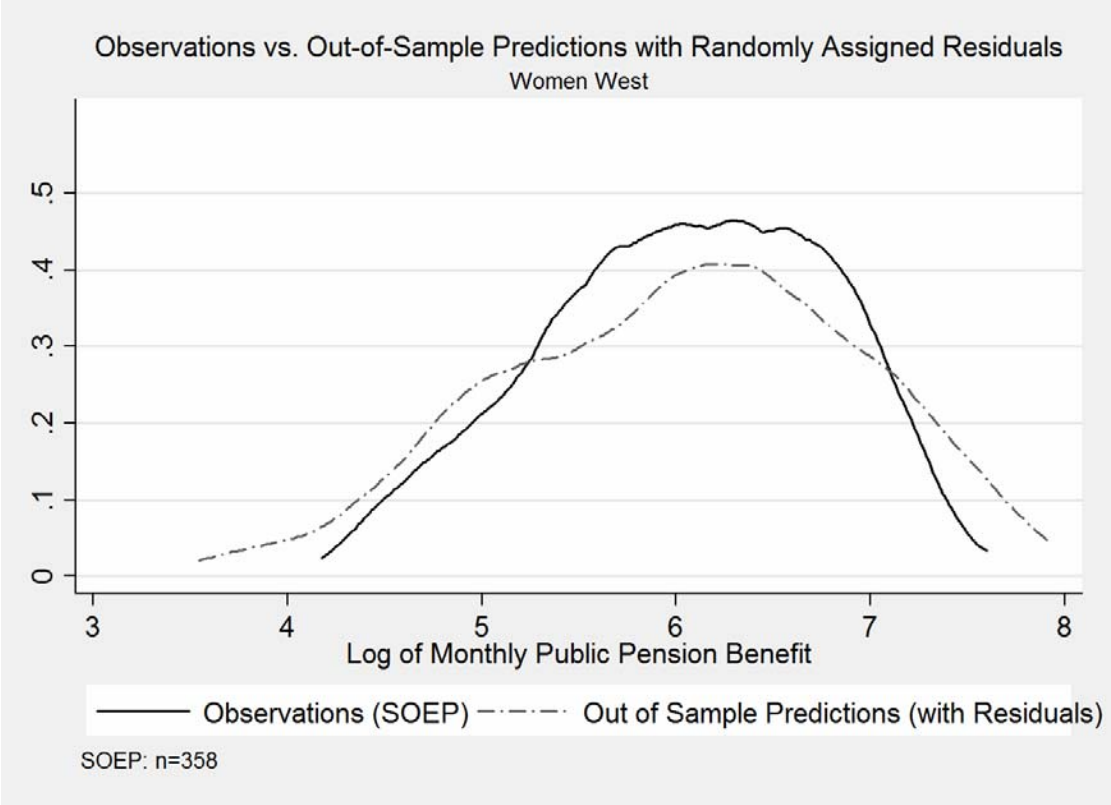


Source: FDZ-RV - SUFVVL2004 & SOEP 2005, own calculations

The graph shows that the shrinkage in the out-of-sample prediction is less distinct relative to the in-sample prediction. The variance of the out-of-sample prediction is 0.494, compared to a variance of 0.534 for the observations. As a reminder, the variance of the original in-sample prediction is 0.222. One explanation for the less distinct shrinkage effect might be the more robust SUF VVL 2004 coefficients, which we applied to the SOEP.

We apply the same procedure as before and assign the residuals randomly to our out-of-sample predictions. In this instance, the residual is the difference between the observation and the respective out-of-sample prediction. Graph 11 shows how the results improve after assigning random residuals to the out-of-sample prediction.

Graph 11 Out-of-Sample Predictions with Randomly Assigned Residuals SOEP

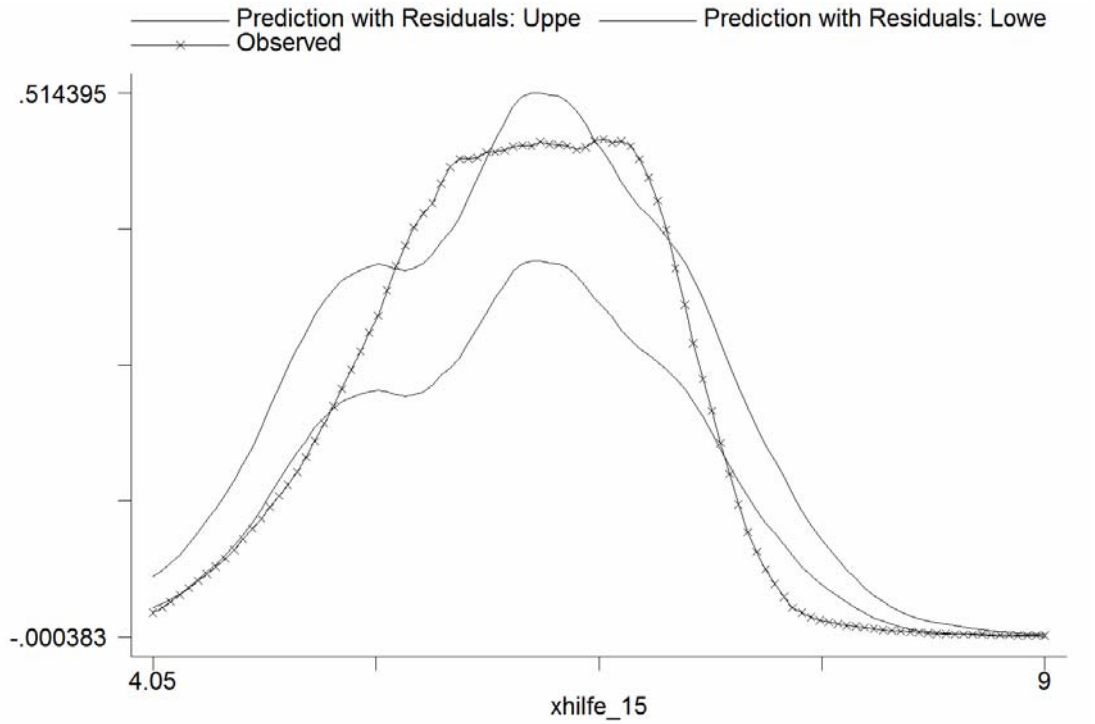


Source: FDZ-RV - SUFVVL2004 & SOEP 2005, own calculations

Again, applying the proposed procedure largely improved the accordance between the distribution of the observed and predicted values. The dashed grey line, which represents the out-of-sample predictions with randomly assigned residuals, no longer has two peaks and is shifted even below the distribution of the observed values. It has a much larger variance than the observations, because we obtain more extreme predictions for smaller values as well as for larger values.

Finally, we controlled whether the observations and the modified out-of-sample predictions differed significantly from each other. Therefore, we plotted the 95% confidence bands of the out-of-sample predictions with randomly assigned residuals around the observations. Graph 12 illustrates the results.

Graph 12 95% Confidence Bands for Out-of-Sample Predictions with Randomly Assigned Residuals and Observations



Source: FDZ-RV - SUFVVL2004 & SOEP 2005, own calculations

It can be seen that the result is not as satisfying as in the case of the Graph 8, which represents the confidence bands of observations vs. in-sample-predictions with randomly assigned residuals. In some areas, the distribution of the observations lies outside of the confidence bands. In these areas, the observations differ significantly from the out-of-sample predictions with randomly assigned residuals. However, for the most part, the observations lie within the confidence bands of the out-of-sample predictions.

The results of the out-of-sample predictions show that we are able to replicate our SOEP results with the SUF VVL 2004 out-of-sample predictions. Hence, the preparatory steps for the matching are now complete. We can now proceed and test various matching approaches in order to determine which approach is the most appropriate for our data.

8 Conclusion and Outlook

We have described the preparatory steps for a statistical matching of administrative pension data and longitudinal household survey data.

The next step will be to test different matching procedures. In order to find the best possible matching approach for our data, we need to weigh the respective merits and drawbacks of each matching procedure.

The ultimate goal is to obtain a new and unique dataset that can be used to trace the consequences of lifecycle work and family choices through to outcomes in old age. If the matching is successful, a similar matching approach will be applied to a newly released dataset. The dataset Aktive Versicherte (Sample of Active Pension Accounts) consists of individuals who are not yet retired. The Aktive Versicherte will then be matched to the latest wave of the Socio-Economic Panel.

Overall, the rich and comprehensive linked datasets resulting from the statistical matching will provide a unique opportunity to study the economic well-being of the present and future elderly in Germany, as it is determined by their work and family choices in the context of new pension legislation. The linked datasets will also allow researchers from other disciplines to tackle a multitude of research questions that are related to lifecycle choices and policy outcomes. In addition, the matching procedure is an important contribution to the methodological discussion of how to find ways to improve the infrastructure of research data in Germany.

Bibliography

Börsch-Supan, Axel. 2000. Incentive Effects of Social Security on Labor Force Participation: Evidence in Germany and Across Europe. *Journal of Public Economics* 78:25-49.

Brenke, Karl 2007. Die Bedeutung der Älteren auf dem Arbeitsmarkt nimmt deutlich zu. *DIW Wochenbericht* 74 (21/2007):337-45.

Bundesministerium für Arbeit und Soziales. 2006. Social Security Data Protection (Sozialdatenschutz) [cited January 1, 2006]. Available from http://www.bmas.bund.de/BMAS/Redaktion/Pdf/Publikationen/Soziale-Sicherung-im-Ueberblick/en/soziale-sicherung-21-sozialdatenschutz,property=pdf,bereich=bmas,sprache=de,rwb=true.pdf.

Caliendo, M. 2006. Lecture Notes in Economics and Mathematical Systems. Vol. No. 568. Berlin: Springer.

Copas, J. B. 1997. Using Regression Models for Prediction: Shrinkage and Regression to the Mean. *Statistical Methods in Medical Research* 6:167-83.

Deutsche Rentenversicherung. 2007. Rentenzugang 2006. Berlin: Deutsche Rentenversicherung.

Deutsche Rentenversicherung Bund. 2005a. Codeplan: FDZ-Biografiedatensatz - VVL 2004. Würzburg: Forschungsdatenzentrum der Rentenversicherung.

———. 2006a. Rentenbestand am 31.12.2005. Berlin: Deutsche Rentenversicherung Bund.

———, ed. 2005b. Rentenversicherung in Zeitreihen. Edited by Deutsche Rentenversicherung Bund. Vol. Band 22. Berlin.

———, ed. 2006b. Rentenversicherung in Zeitreihen. Edited by Deutsche Rentenversicherung Bund. Vol. Band 23. Berlin.

Fachinger, Uwe and Ralf K. Himmelreicher. 2006. Die Bedeutung des Scientific Use Files Vollendete Versichertenleben 2004 (SUFVVL2004) aus der Perspektive der Ökonomik. Deutsche Rentenversicherung 61 (9-10):562-82.

Ferber, Marianne A. and Bonnie G. Birnbaum. 1979. Retrospective Earnings Data: Some Solutions for Old Problems. *The Public Opinion Quarterly* 43 (1):112-8.

Fitzenberger, Bernd, Aderonke Osikominu and Robert Völter. 2005. Imputation Rules to Improve the Education Variable in the IAB Employment Subsample. Mannheim: Zentrum für Europäische Wirtschaftsforschung GmbH.

Fitzenberger, Bernd and Stefan Speckesser. 2005. Employment Effects of the Provision of Specific Skills and Techniques in Germany. Mannheim: ZEW.

Forschungsdatenzentrum der Deutschen Rentenversicherung. 2006. Benutzerhinweise FDZ-Biografiedatensatz - VVL 2004. Würzburg: Deutsche Rentenversicherung Bund.

Frick, Joachim R. and Christian Schmitt. 2006. Biobirth - A Data Set on the Birth Biography of Female Respondents. In: Biography and Life History Data in the German Socio-Economic Panel, Joachim R. Frick and Jürgen Schupp (Hrsg.). Berlin.

Grabka, Markus G. 2006. Codebook for the $PEQUIV File 1984-2005. In: SOEP Data Documentation. Berlin: Deutsches Institut für Wirtschaftsforschung.

Grabka, Markus M. 2004. Alternative Finanzierungsmodelle einer sozialen Krankenversicherung in Deutschland - Methodologische Grundlagen und exemplarische Durchführung einer Mikrosimulationsstudie. Dissertation, Fakultät VIII - Wirtschaft und Management, Technische Universität zu Berlin, Berlin.

Haak, Carroll. 2006. Von Künstlern lernen: Mehrfachbeschäftigung, Bildung und Einkommen auf den Arbeitsmärkten von Künstlern. Berlin: Wissenschaftszentrum Berlin für Sozialforschung.

Hagen, Christine, Ralf K. Himmelreicher and Hilmar Hoffmann. 2007. Typologie des Mehrfachrentenbezugs. *RV Aktuell* (3).

Haisken-DeNew, John P. and Joachim R Frick (Eds.) 2005. Desktop Companion to the German Socio-Economic Panel (SOEP). Berlin: Deutsches Institut für Wirtschaftsforschung (DIW).

Hanisch, Jens and Ulrich Rendtel. 2002. Quality of Income Data from Panel Surveys with Respect to Rounding. Paper for the ICIS Conference. Wiesbaden: Statistisches Bundesamt.

Hawala, Sam, Martha Stinson and John Abowd. 2005. Disclosure Risk Assessment Through Record Linkage. In: Joint UNECE/Eurostat Work Session on Statistical Confidentiality. Geneva Switzerland.

Heese, Claudia. 2004. Aspekte des Datenschutzes im Forschungsdatenzentrum der Rentenversicherung. In: Das Forschungsdatenzentrum der gesetzlichen Renten-versicherung (FDZ-RV) im Aufbau. Bericht vom Ersten Workshop des FDZ-RV am 28. und 29. Juni 2004 in Würzburg (Hrsg.). Bad Homburg: WDV.

Himmelreicher, Ralf K. 2005. Analysepotenzial des Scientific Use Files Versicherten-rentenzugang. In: Forschungsrelevante Daten der Rentenversicherung: Bericht vom zweiten Workshop des Forschungsdatenzentrums der Rentenversicherung (FDZ-RV) vom 27. bis 29. Juni 2005 in Würzburg, Deutsche Rentenversicherung Bund (Hrsg.). Berlin.

Himmelreicher, Ralf K. 2001. Soziodemographie, Erwerbsarbeit, Einkommen und Vermögen von westdeutschen Haushalten: Eine Längsschnitt-Kohortenanalyse auf Datenbasis des SOEP (1984-1997). Berlin: Logos Verlag.

Himmelreicher, Ralf K. and Dirk Mai. 2006. Retrospektive Querschnitte: Das Analysepotenzial des Scientific Use Files Versichertenrentenzugang 2004 unter besonderer Berücksichtigung der Berechnung von Entgeltpunkten. *DRV Schriften* Band 55:23-51.

Himmelreicher, Ralf K., Hans-Martin von Gaudecker and Rembrandt D. Scholz. 2006. Nutzungsmöglichkeiten von Daten der gesetzlichen Rentenversicherung über das Forschungsdatenzentrum der Rentenversicherung (FDZ-RV). Rostock: Max-Planck-Institute for Demographic Research.

Himmelreicher, Ralf and Holger Viebrok. 2002. Die "Riester-Rente" und einige Folgen für die Alterseinkünfte. Bremen: Zentrum für Sozialpolitik.

Hoffmann, Hilmar. 2007. "Wege in den Ruhestand." *Deutsche Rentenversicherung* 62 (4-5):298-320.

Holst, Elke. 2000. Die Stille Reserve am Arbeitsmarkt: Größe – Zusammensetzung – Verhalten. Berlin: Edition Sigma.

Hujer, R., M. Caliendo and S. Thomsen. 2004. New Evidence on the Effects of Job Creation Schemes in Germany - A Matching Approach with Threefold Heterogeneity. *Research in Economics* 58 (4):257-302.

Kapteyn, Arie and Jelmer Yeb Ypma. 2006. Measurement Error and Misclassification: A Comparison of Survey and Register Data. Santa Monica: RAND.

Pischner, Rainer 2006. Activity Biography in the File PBIOSPE. In: Biography and Life History Data in the German Socio-Economic Panel, Joachim R. Frick and Jürgen Schupp (Eds.). Berlin: Deutsches Institut für Wirtschaftsforschung.

Rasner, Anika. 2006a. Das Konzept der geschlechtsspezifischen Rentenlücke. *DRV-Schriften* 55:270-84.

———. 2006b. Mind the Gap! Einbeziehung internationaler Benchmarks in der Beurteilung der geschlechtsspezifischen Rentenlücke in Deutschland. *Deutsche Rentenversicherung* 61 (11-12):737-54.

Rehfeld, Uwe, Thomas Bütefisch and Hilmar Hoffmann. 2007. Gesundheitsbedingte Leistungen der gesetzlichen Rentenversicherung für Frauen und Männer; Indikatoren für die Morbidität. In: Fehlzeiten-Report 2007 - Chronische Krankheiten. Bernhard Badura, Henner Schellschmidt and Christian Vetter (Eds.). Berlin: Springer Verlag.

Social Security Administration. 2007. Totalization Agreement with Germany. Social Security Online, May 2, 2007 [cited May 24, 2007]. Available from http://www.ssa.gov/international/Agreement_Pamphlets/germany.html#intro.

SOEP Group. 2001. The German Socio-Economic Panel Study (GSOEP) after more than 15 Years - Overview. In: Proceedings of the Fourth International Conference of Socio-Economic Panel Study Users, E. Holst, D. R. Lillard and T. A. DiPrete (Hrsg.): Vierteljahreshefte zur Wirtschaftsforschung.

Stegmann, Michael. 2006. Aufbereitung der Sondererhebung "Vollendete Versichertenleben 2004" als Scientific Use File für das FDZ-RV. *Deutsche Rentenversicherung* 61 (9-10):537-53.

TNS Infratest. 2005. Living in Germany: Survey 2005 on the Social Situation of Households.

Wagner, Gert G., Joachim R. Frick and Jürgen Schupp. 2006. Enhancing the Power of Household Panel Studies - The Case of the German Socio-Economic Panel Study (SOEP). In: Longitudinal Social Surveys in an International Perspective. McGill University Montreal.

Wolff, Joachim und Thomas Augustin. 2000. "Heaping and Its Consequences for Duration Analysis." München: Institut für Statistik der Ludwig-Maximilians-Universität München, Sonderforschungsbereich 386.

Appendix A

Average Time Spent in Different Activities

SOEP Data

Note: The calculation of the average time spent in different activities in this first set of Tables is based on all individuals, independent of whether they have actually spent time in the respective activity or not. Hence, the denominator is always the total number of persons.

(A): Germans and Persons with Migration History together

	ALL PENSIONERS IN THE YEAR 2005[i]				FIRST TIME PENSIONERS B/W 2000-2004[ii]				FIRST TIME PENSIONERS B/W 2003-2004[iii]			
Spelltype_	**Men**		**Women**		**Men**		**Women**		**Men**		**Women**	
	Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
_1: School/University	1.75	3.03	1.31	2.13	2.70	4.21	1.37	2.11	2.05	2.86	1.38	2.13
_2: Apprenticeship/Training	2.62	2.64	1.66	1.69	2.64	2.06	2.02	1.63	2.90	1.89	2.23	1.67
_3: Military/Civilian Service	1.30	2.84	0.19	1.37	0.59	1.71	0.18	2.36	0.76	1.25	0.08	0.50
_4: Full-time employed	37.82	7.46	22.87	14.70	38.16	7.11	20.75	14.09	37.42	6.59	20.54	13.72
_5: Part-time employed	0.43	2.05	5.35	8.93	0.43	1.27	7.58	10.04	0.64	1.89	7.46	9.89
_6: Unemployed	1.10	2.30	0.84	2.10	2.20	3.34	1.70	2.89	2.24	2.53	2.13	3.31
_7: Homeproduction [iv]	0.23	1.18	16.40	16.12	0.27	1.06	16.54	15.64	0.29	1.06	16.73	14.63
_7n: Homeproduction [v]	0.04	0.50	12.27	14.28	0.01	0.16	12.48	14.30	0.01	0.16	12.77	13.57
_8: Retired	4.54	3.98	4.64	4.35	2.28	1.70	2.03	1.56	1.42	2.10	1.13	0.99
_9: Other	0.44	2.44	0.95	3.58	0.17	0.52	0.59	2.71	0.19	0.47	0.40	1.25
atyp_sum[vi]	51	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
atyp_missing[vii]	0.97	3.65	0.91	3.36	1.80	3.17	2.30	4.41	3.37	4.66	2.88	2.67
Mean Age	71.37	6.77	73.07	7.48	65.07	2.17	65.05	2.35	63.59	1.86	63.74	1.93
Sum of Weights	6436520.09		8635570.99		1376688.64		1755168.17		467585.529		725210.011	
Number of cases (n=)	1827		2113		392		465		140		180	

(B) Only Germans

	ALL PENSIONERS IN THE YEAR 2005				FIRST TIME PENSIONERS B/W 2000-2004				FIRST TIME PENSIONERS B/W 2003-2004			
Spelltype_	**Men**		**Women**		**Men**		**Women**		**Men**		**Women**	
	Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
_1: School/University	1.70	2.80	1.34	2.15	2.32	3.42	1.35	2.09	2.13	2.89	1.41	2.16
_2: Apprenticeship/Training	2.70	2.63	1.72	1.68	2.64	1.57	2.12	1.61	2.98	1.80	2.35	1.64
_3: Military/Civilian Service	1.27	2.76	0.20	1.33	0.57	1.76	0.19	2.45	0.66	1.17	0.09	0.51
_4: Full-time employed	37.97	7.20	22.63	14.67	39.03	5.59	20.55	14.08	37.41	6.42	20.46	13.78
_5: Part-time employed	0.42	2.10	5.40	8.97	0.44	1.33	7.68	10.21	0.68	1.98	7.45	10.08
_6: Unemployed	0.97	1.98	0.82	2.10	1.70	2.45	1.67	2.89	2.04	2.36	2.17	3.34
_7: House-Husband/Wife	0.25	1.23	16.61	16.15	0.28	1.08	16.48	15.59	0.27	0.98	16.73	14.84
_7n: House-Husband/Wife	0.04	0.51	12.44	14.34	0.01	0.18	12.51	14.33	0.01	0.17	12.67	13.76
_8: Retired	4.54	4.01	4.59	4.37	2.27	1.73	2.01	1.56	1.46	2.22	1.14	1.00
_9: Other	0.43	2.52	0.94	3.66	0.18	0.53	0.59	2.77	0.20	0.49	0.36	1.12
atyp_sum	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
atyp_missing	0.96	3.68	0.92	3.42	1.83	3.14	2.33	4.52	3.43	4.50	2.91	2.74
Mean Age	71.41	6.71	73.17	7.54	65.04	2.19	65.02	2.37	63.51	1.86	63.73	1.96
Sum of Weights	5761265.56		8021728.66		1194423.76		1632365.91		412960.638		681393.712	
Number of cases (n=)	1631		1933		346		423		122		165	

(C) Only Persons with a Migration History[viii]

	ALL PENSIONERS IN THE YEAR 2005				FIRST TIME PENSIONERS B/W 2000-2004				FIRST TIME PENSIONERS B/W 2003-2004			
Spelltype_	**Men**		**Women**		**Men**		**Women**		**Men**		**Women**	
	Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
_1: School/University	2.20	4.50	1.00	1.80	5.22	7.15	1.53	2.38	1.49	2.58	0.87	1.47
_2: Apprenticeship/Training	1.98	2.59	0.92	1.60	2.66	4.02	0.79	1.38	2.29	2.39	0.45	1.07
_3: Military/Civilian Service	1.59	3.43	0.15	1.78	0.69	1.31	0.00	0.00	1.50	1.59	0.00	0.00
_4: Full-time employed	36.54	9.28	25.96	14.79	32.49	11.95	23.46	14.12	37.47	7.97	21.82	13.20
_5: Part-time employed	0.43	1.55	4.70	8.43	0.40	0.84	6.21	7.32	0.36	1.02	7.58	6.44
_6: Unemployed	2.19	3.99	1.08	2.13	5.49	5.73	2.08	2.99	3.72	3.29	1.49	2.77
_7: House-Husband/Wife	0.12	0.65	13.56	15.46	0.17	0.89	17.34	16.39	0.44	1.58	16.71	11.26
_7n: House-Husband/Wife	0.02	0.30	10.04	13.27	0.00	0.00	12.08	13.99	0.00	0.00	14.29	10.65
_8: Retired	4.49	3.67	5.31	3.95	2.33	1.49	2.18	1.59	1.07	0.65	1.04	0.78
_9: Other	0.54	1.48	1.07	2.33	0.10	0.43	0.70	1.80	0.11	0.27	1.12	2.49
atyp_sum	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
atyp_missing	1.01	3.37	0.78	2.37	1.62	3.39	1.96	2.59	2.99	5.85	2.35	1.27
Mean Age	71.10	7.26	71.88	6.55	65.22	2.09	65.34	2.14	64.13	1.82	63.82	1.30
Sum of Weights	675254.532		613842.328		182264.882		122802.26		54624.8904		43816.2995	
Number of cases (n=)	196		180		46		43		18		15	

(D) Comparison East vs. West[ix]: Germans and Persons with a Migration History together

	ALL PENSIONERS IN THE YEAR 2005							
	WEST				EAST			
Spelltype_	**Men**		**Women**		**Men**		**Women**	
	Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
_1: School/University	1.73	3.08	1.35	2.17	1.80	2.86	1.19	1.99
_2: Apprenticeship/Training	2.64	2.78	1.62	1.68	2.56	2.08	1.79	1.72
_3: Military/Civilian Service	1.26	2.82	0.24	1.54	1.46	2.93	0.05	0.39
_4: Full-time employed	37.85	7.68	20.63	14.78	37.70	6.65	30.18	11.80
_5: Part-time employed	0.44	2.18	5.74	9.23	0.37	1.51	4.07	7.76
_6: Unemployed	1.12	2.45	0.71	1.99	1.00	1.71	1.25	2.37
_7: House-Husband/Wife	0.27	1.30	19.81	16.34	0.12	0.59	5.24	8.61
_7n: House-Husband/Wife	0.05	0.55	14.77	14.95	0.02	0.23	4.11	7.25
_8: Retired	4.51	4.05	4.14	3.99	4.63	3.72	6.27	5.01
_9: Other	0.47	2.50	0.99	3.84	0.34	2.20	0.82	2.53
atyp_sum	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
atyp_missing	0.92	3.72	0.80	3.36	1.11	3.40	1.27	3.33
Mean Age	71.65	6.84	73.41	7.43	70.43	6.43	71.97	7.57
Sum of weights	4981314.42		6611747.11		1455205.67		2023823.88	
Number of cases (n=)	1268		1423		559		690	

(E) Comparison West vs. East: Germans and Persons with a Migration History together

	FIRST-TIME PENSIONERS BETWEEN 2000 AND 2004							
	WEST				EAST			
Spelltype_	**Men**		**Women**		**Men**		**Women**	
	Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
_1: School/University	2.87	4.52	1.39	2.11	2.16	3.06	1.26	2.13
_2: Apprenticeship/Training	2.74	2.22	1.96	1.70	2.36	1.41	2.25	1.27
_3: Military/Civilian Service	0.48	1.81	0.22	2.65	0.91	1.29	0.00	0.00
_4: Full-time employed	38.37	7.44	18.17	13.89	37.54	6.00	30.48	10.00
_5: Part-time employed	0.45	1.33	8.37	10.42	0.37	1.07	4.59	7.76
_6: Unemployed	2.14	3.62	1.10	2.33	2.41	2.26	3.98	3.59
_7: House-Husband/Wife	0.33	1.19	20.09	15.53	0.08	0.38	3.15	5.59
_7n: House-Husband/Wife	0.01	0.19	15.13	14.78	0.00	0.00	2.51	5.11
_8: Retired	2.30	1.81	1.87	1.55	2.23	1.32	2.60	1.48
_9: Other	0.18	0.54	0.64	3.01	0.15	0.47	0.41	0.90
atyp_sum	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
atyp_missing	1.45	2.19	2.13	4.78	2.87	4.98	2.94	2.50
Mean Age	65.25	2.18	65.47	2.36	64.48	2.04	63.45	1.43
Sum of Weights	1039575.36		1387319.91		337113.279		367848.26	
Number of cases (n=)	266		328		126		137	

(F) Comparison East vs. West: Germans and Persons with a Migration History together

	FIRST-TIME PENSIONERS BETWEEN 2003 AND 2004							
	WEST				EAST			
Spelltype_	Men		Women		Men		Women	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
_1: School/University	1.89	2.88	1.48	2.24	2.48	2.78	0.98	1.55
_2: Apprenticeship/Training	3.03	2.03	2.24	1.75	2.54	1.39	2.21	1.29
_3: Military/Civilian Service	0.67	1.16	0.10	0.55	1.01	1.44	0.00	0.00
_4: Full-time employed	38.56	6.04	18.57	13.84	34.44	7.06	28.60	9.82
_5: Part-time employed	0.74	2.06	8.39	10.52	0.37	1.36	3.63	5.36
_6: Unemployed	1.88	2.47	1.30	2.47	3.17	2.48	5.49	4.11
_7: House-Husband/Wife	0.39	1.23	19.60	14.55	0.00	0.00	5.02	7.32
_7n: House-Husband/Wife	0.02	0.19	14.86	14.04	0.00	0.00	4.25	6.57
_8: Retired	1.51	2.40	1.11	1.04	1.16	0.92	1.25	0.76
_9: Other	0.20	0.50	0.45	1.37	0.16	0.35	0.23	0.53
atyp_sum	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
atyp_missing	2.50	2.85	2.51	2.40	5.67	7.12	4.37	3.20
Mean Age	63.78	1.88	64.12	1.90	63.08	1.71	62.18	1.09
Sum of Weights	338237.769		582328.701		129347.759		142881.311	
Number of cases (n=)	91		124		49		56	

(G) Comparison East vs. West: Germans only

	ALL PENSIONERS IN THE YEAR 2005							
	WEST				EAST			
Spelltype_	Men		Women		Men		Women	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
_1: School/University	1.66	2.79	1.38	2.21	1.80	2.83	1.19	1.96
_2: Apprenticeship/Training	2.75	2.80	1.68	1.67	2.55	2.07	1.83	1.73
_3: Military/Civilian Service	1.21	2.70	0.24	1.51	1.45	2.94	0.06	0.40
_4: Full-time employed	38.04	7.39	20.27	14.67	37.75	6.61	30.12	11.90
_5: Part-time employed	0.44	2.26	5.81	9.28	0.37	1.52	4.11	7.79
_6: Unemployed	0.96	2.06	0.68	1.98	1.00	1.71	1.26	2.39
_7: House-Husband/Wife	0.29	1.37	20.22	16.31	0.13	0.60	5.22	8.60
_7n: House-Husband/Wife	0.05	0.58	15.06	15.01	0.02	0.23	4.16	7.32
_8: Retired	4.52	4.12	4.07	3.98	4.61	3.68	6.23	5.09
_9: Other	0.46	2.62	1.01	3.96	0.34	2.22	0.73	2.45
atyp_sum	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
atyp_missing	0.91	3.76	0.79	3.42	1.10	3.41	1.31	3.40
Mean Age	71.72	6.78	73.54	7.47	70.45	6.43	71.98	7.66
Sum of Weights	4330301.22		6094837.39		1430964.34		1926891.27	
Number of cases (n=)	1081		1268		550		665	

(H) Comparison East vs. West: Germans only

	FIRST-TIME PENSIONERS BETWEEN 2000 AND 2004							
	WEST				EAST			
Spelltype_	**Men**		**Women**		**Men**		**Women**	
	Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
_1: School/University	2.36	3.55	1.40	2.12	2.19	3.08	1.19	1.96
_2: Apprenticeship/Training	2.78	1.63	2.08	1.69	2.29	1.31	2.23	1.28
_3: Military/Civilian Service	0.45	1.90	0.24	2.77	0.90	1.30	0.00	0.00
_4: Full-time employed	39.58	5.31	17.61	13.77	37.59	6.05	30.88	9.59
_5: Part-time employed	0.46	1.41	8.66	10.69	0.38	1.08	4.24	7.36
_6: Unemployed	1.45	2.48	1.02	2.26	2.37	2.27	3.98	3.59
_7: House-Husband/Wife	0.36	1.24	20.26	15.46	0.08	0.38	3.18	5.63
_7n: House-Husband/Wife	0.01	0.21	15.35	14.84	0.00	0.00	2.55	5.15
_8: Retired	2.28	1.87	1.85	1.55	2.26	1.31	2.57	1.47
_9: Other	0.20	0.55	0.64	3.10	0.14	0.47	0.40	0.90
atyp_sum	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
atyp_missing	1.43	1.84	2.15	4.93	2.88	5.04	2.97	2.50
Mean Age	65.25	2.21	65.48	2.38	64.51	2.05	63.42	1.42
Sum of Weights	864379.348		1270852.36		330044.409		361513.55	
Number of cases (n=)	222		289		124		134	

(I) Comparison East vs. West: Germans only

	FIRST-TIME PENSIONERS BETWEEN 2003 AND 2004							
	WEST				EAST			
Spelltype_	Men		Women		Men		Women	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
_1: School/University	1.94	2.92	1.52	2.29	2.58	2.81	0.98	1.56
_2: Apprenticeship/Training	3.23	1.97	2.38	1.72	2.37	1.13	2.21	1.30
_3: Military/Civilian Service	0.53	1.00	0.11	0.57	0.98	1.46	0.00	0.00
_4: Full-time employed	38.68	5.60	18.29	13.88	34.39	7.25	28.74	9.77
_5: Part-time employed	0.80	2.18	8.50	10.80	0.39	1.39	3.45	5.07
_6: Unemployed	1.59	2.14	1.30	2.45	3.11	2.54	5.50	4.13
_7: House-Husband/Wife	0.38	1.15	19.78	14.81	0.00	0.00	5.07	7.34
_7n: House-Husband/Wife	0.02	0.20	14.86	14.31	0.00	0.00	4.29	6.59
_8: Retired	1.58	2.57	1.11	1.06	1.18	0.93	1.25	0.76
_9: Other	0.22	0.53	0.40	1.23	0.15	0.35	0.21	0.50
atyp_sum	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
atyp_missing	2.41	1.80	2.53	2.47	5.85	7.28	4.37	3.22
Mean Age	63.71	1.88	64.14	1.94	63.06	1.75	62.18	1.09
Sum of Weights	290681.749		539963.221		122278.889		141430.491	
Number of cases (n=)	75		110		47		55	

(J) Comparison East vs. West: Persons with a Migration History Only

	ALL PENSIONERS IN THE YEAR 2005							
	WEST				EAST			
Spelltype_	**Men**		**Women**		**Men**		**Women**	
	Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
_1: School/University	2.20	4.51	0.96	1.61	2.12	4.45	1.17	2.60
_2: Apprenticeship/Training	1.94	2.59	0.90	1.64	3.13	2.55	1.01	1.41
_3: Military/Civilian Service	1.57	3.47	0.18	1.94	2.19	1.88	0.00	0.00
_4: Full-time employed	36.62	9.32	24.95	15.38	34.48	8.40	31.37	9.67
_5: Part-time employed	0.45	1.58	4.96	8.62	0.08	0.19	3.30	7.38
_6: Unemployed	2.22	4.04	1.07	2.16	1.32	2.02	1.14	2.05
_7: House-Husband/Wife	0.12	0.67	15.03	16.00	0.00	0.00	5.76	8.92
_7n: House-Husband/Wife	0.02	0.30	11.36	13.87	0.00	0.00	2.96	5.64
_8: Retired	4.44	3.58	4.96	4.06	5.85	5.86	7.18	2.64
_9: Other	0.55	1.51	0.82	1.96	0.19	0.26	2.42	3.50
atyp_sum	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
atyp_missing	0.99	3.40	0.84	2.50	1.65	2.33	0.46	1.47
Mean Age	71.19	7.28	71.89	6.74	68.70	6.61	71.81	5.55
Sum of Weights	651013.20		516909.71		24241.33		96932.61	
Number of cases (n=)	187		155		9		25	

(K) Comparison East vs. West: Persons with a Migration History Only

	FIRST-TIME PENSIONERS BETWEEN 2000 AND 2004							
	WEST				EAST			
Spelltype_	Men		Women		Men		Women	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
_1: School/University	5.40	7.23	1.33	1.92	0.78	2.06	5.33	6.27
_2: Apprenticeship/Training	2.55	4.04	0.67	1.31	5.50	2.65	2.88	1.12
_3: Military/Civilian Service	0.65	1.31	0.00	0.00	1.55	1.18	0	0
_4: Full-time employed	32.38	12.18	24.32	13.97	35.23	3.24	7.68	4.93
_5: Part-time employed	0.42	0.85	5.22	6.07	0	0	24.45	2.15
_6: Unemployed	5.54	5.84	1.96	2.90	4.17	0.88	4.26	4.44
_7: House-Husband/Wife	0.18	0.91	18.24	16.37	0	0	0.96	1.22
_7n: House-Husband/Wife	0.00	0.00	12.74	14.08	0	0	0	0
_8: Retired	2.39	1.49	2.08	1.52	0.83	0.88	4.11	1.82
_9: Other	0.09	0.44	0.68	1.84	0.39	0.29	1.10	0.92
atyp_sum	51.00	0.00	51.00	0.00	51	0	51	0
atyp_missing	1.59	3.45	2.00	2.63	2.55	1.18	1.21	1.93
Mean Age	65.30	2.09	65.37	2.17	63.45	1.18	64.79	1.93
Sum of Weights	175196.012		116467.55		7068.87		6334.71	
Number of cases (n=)	44		39		2		3	

(L) Comparison East vs. West: Persons with a Migration History Only

	FIRST-TIME PENSIONERS BETWEEN 2003 AND 2004							
	WEST				EAST			
Spelltype_	Men		Women		Men		Women	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
_1: School/University	1.60	2.69	0.86	1.50	0.78	2.06	1	,
_2: Apprenticeship/Training	1.81	2.04	0.39	1.05	5.50	2.65	2	,
_3: Military/Civilian Service	1.49	1.68	0.00	0.00	1.55	1.18	0	,
_4: Full-time employed	37.81	8.47	22.06	13.39	35.23	3.24	14.83	,
_5: Part-time employed	0.41	1.08	7.10	5.99	0	0	21.33	,
_6: Unemployed	3.65	3.53	1.39	2.77	4.17	0.88	4.33	,
_7: House-Husband/Wife	0.50	1.69	17.28	11.00	0	0	0	,
_7n: House-Husband/Wife	0.00	0.00	14.78	10.49	0	0	0	,
_8: Retired	1.11	0.65	1.02	0.79	0.83	0.88	1.50	,
_9: Other	0.07	0.25	1.09	2.53	0.39	0.29	2	,
atyp_sum	51.00	0.00	51.00	0.00	51	0	51	,
atyp_missing	3.05	6.28	2.30	1.25	2.55	1.18	4	,
Mean Age	64.23	1.87	63.88	1.28	63.45	1.18	62	,
Sum of Weights	47556.0202		42365.4796		7068.87		1450.82	
Number of cases (n=)	16		14		2		1	

(M) Comparison Civil Servant in Last Job vs. Others[x]

	ALL PENSIONERS IN THE YEAR 2005				FIRST TIME PENSIONERS B/W 2000-2004				FIRST TIME PENSIONERS B/W 2003-2004			
Spelltype_	**Civil Servants**		**Others**		**Civil Servants**		**Others**		**Civil Servants**		**Others**	
	Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
_1: School/University	3.29	3.98	1.47	2.52	6.82	4.99	1.85	3.16	5.19	4.16	1.60	2.36
_2: Apprenticeship/Training	2.51	1.88	2.06	2.20	4.61	10.05	2.27	1.84	2.31	1.77	2.47	1.76
_3: Military/Civilian Service	1.53	2.97	0.65	2.18	0.57	0.88	0.34	2.08	0.96	1.05	0.32	0.88
_4: Full-time employed	35.68	9.37	29.15	14.26	34.89	8.80	27.82	14.60	37.84	2.38	26.52	14.30
_5: Part-time employed	0.67	2.64	3.29	7.35	0.71	2.08	4.70	8.71	0.77	1.61	5.00	8.67
_6: Unemployed	0.15	0.62	0.96	2.21	0.06	0.29	1.92	3.10	0.06	0.25	2.20	3.05
_7: House-Husband/Wife	2.26	8.29	9.61	14.66	1.71	7.26	10.15	14.80	0.16	0.44	10.79	14.15
_7n: House-Husband/Wife	1.83	7.46	7.13	12.43	0.00	0.00	7.44	12.74	0.00	0.00	8.08	12.40
_8: Retired	5.05	5.70	4.59	4.16	1.60	1.38	2.18	1.77	1.47	1.66	1.36	1.97
_9: Other	0.13	0.46	0.74	3.18	0.13	0.57	0.40	2.05	0.00	0.00	0.32	1.02
atyp_sum	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
atyp_missing	0.17	0.71	0.94	3.51	1.60	2.20	2.09	3.89	2.40	2.71	3.12	3.57
Mean Age	73.46	7.39	72.33	7.23	66.56	1.97	65.05	2.30	65.57	1.56	63.63	1.89
Sum of weights	235778.969		14836312.1		264487.34		3193841.36		120773.061		1200670.21	
Number of cases (n=)	59		3883		43		867		17		324	

(N) Comparison Self-Employed in the Last Job vs. Others[xi]

	ALL PENSIONERS IN THE YEAR 2005				FIRST TIME PENSIONERS B/W 2000-2004				FIRST TIME PENSIONERS B/W 2003-2004			
Spelltype_	**Self-Employed**		**Others**		**Self-Employed**		**Others**		**Self-Employed**		**Others**	
	Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
_1: School/University	3.13	3.67	1.46	2.51	3,35	3,89	2,26	3,46	3,46	4,58	2,28	3,02
_2: Apprenticeship/Training	2.54	1.91	2.06	2.20	1,95	1,47	2,37	2,91	1,83	1,57	2,35	1,72
_3: Military/Civilian Service	0.44	1.26	0.67	2.21	0,89	4,22	0,38	1,74	0,31	0,67	0,40	0,90
_4: Full-time employed	35.69	11.64	29.09	14.24	29,84	13,84	26,28	13,99	27,99	15,64	24,62	13,95
_5: Part-time employed	3.99	7.95	3.23	7.29	6,16	10,63	3,88	7,64	7,81	11,57	3,83	7,24
_6: Unemployed	0.52	1.49	0.96	2.20	0,81	2,14	1,66	2,85	1,25	2,49	1,77	2,79
_7: House-Husband/Wife	5.29	11.03	9.60	14.67	5,44	10,60	7,93	13,19	6,40	11,28	7,75	12,42
_7n: House-Husband/Wife	1.60	5.61	7.18	12.48	0,93	2,68	5,72	11,19	0,80	1,71	5,83	10,81
_8: Retired	1.46	2.09	4.67	4.20	1,72	1,63	2,21	1,87	1,56	2,11	1,38	2,03
_9: Other	1.00	3.74	0.73	3.14	0,76	3,58	0,64	2,41	0,16	0,40	0,59	1,47
atyp_sum	51.00	0.00	51.00	0.00	51	0,00	51	0,00	51	0,00	51	0,00
atyp_missing	0.64	1.19	0.94	3.52	4,57	7,29	5,60	8,45	5,84	8,94	7,95	9,82
Mean Age	69.42	5.94	72.42	7.25	65.93	2.27	65.12	2.31	64.70	1.80	63.75	1.94
Sum of weights	365091.839		14706999.20		292883,82		5063520,53		115466,85		2130737,99	
Number of cases (n=)	93		3847		79		1272		32		513	

Set B: SOEP Data

Note: The calculation of the average time spent in different activities in this first set of Tables is only based on individuals who have spent time a certain activity. Hence, the denominator comprises all individuals that have information > 0 for the respective category.

(A): Germans and Persons with Migration History together

	ALL PENSIONERS IN THE YEAR 2005				FIRST TIME PENSIONERS B/W 2000-2004				FIRST TIME PENSIONERS B/W 2003-2004			
Spelltype_	**Men**		**Women**		**Men**		**Women**		**Men**		**Women**	
	Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
_1: School/University	3.33	3.16	2.53	2.40	4.31	3.64	2.41	2.29	3.63	2.97	2.17	2.36
_2: Apprenticeship/Training	3.14	2.59	2.74	1.31	3.01	1.30	2.78	1.24	3.20	1.67	2.93	1.27
_3: Military/Civilian Service	3.70	3.64	2.63	4.16	2.17	2.90	11.58	21.49	2.06	1.15	3.00	.
_4: Full-time employed	38.01	7.08	24.11	13.92	39.03	5.59	21.17	13.82	37.41	6.42	20.99	13.55
_5: Part-time employed	2.45	4.52	11.29	10.08	1.73	2.18	11.63	10.58	2.51	3.19	11.12	10.54
_6: Unemployed	2.68	2.50	3.37	3.07	3.17	2.56	4.13	3.23	3.54	2.08	4.43	3.58
_7: House-Husband/Wife	2.79	3.14	21.23	15.34	2.33	2.21	20.42	14.86	1.89	2.00	20.01	14.05
_7n: House-Husband/Wife	3.66	3.13	18.69	13.86	2.69	1.35	18.15	13.98	2.00	.	16.74	13.48
_8: Retired	4.92	3.95	5.18	4.30	2.27	1.73	2.01	1.56	1.46	2.22	1.14	1.00
_9: Other	2.48	5.63	4.12	6.74	1.08	0.83	3.02	5.70	0.94	0.66	1.83	1.95
atyp_sum	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
atyp_missing	0.96	3.68	0.92	3.42	1.83	3.14	2.33	4.52	3.43	4.50	2.91	2.74
Mean Age	71.41	6.71	73.17	7.54	65.04	2.19	65.02	2.37	63.51	1.86	63.73	1.96
Sum of Weights	5761265.56		8021728.66		1194423.76		1632365.91		412960.638		681393.712	
Number of cases (n=)	1631		1933		346		423		122		165	

(B) Only Germans

	ALL PENSIONERS IN THE YEAR 2005				FIRST TIME PENSIONERS B/W 2000-2004				FIRST TIME PENSIONERS B/W 2003-2004			
Spelltype_	**Men**		**Women**		**Men**		**Women**		**Men**		**Women**	
	Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
_1: School/University	3.48	3.50	2.53	2.38	4.92	4.63	2.43	2.31	3.55	2.97	2.17	2.33
_2: Apprenticeship/Training	3.15	2.59	2.74	1.32	3.10	1.88	2.78	1.23	3.22	1.70	2.93	1.27
_3: Military/Civilian Service	3.69	3.75	2.73	4.41	2.24	2.74	11.58	21.49	2.18	1.16	3.00	.
_4: Full-time employed	37.86	7.36	24.33	13.94	38.16	7.11	21.34	13.84	37.42	6.59	21.04	13.50
_5: Part-time employed	2.40	4.35	11.20	10.08	1.67	2.05	11.39	10.39	2.54	3.10	10.91	10.28
_6: Unemployed	2.98	2.96	3.35	3.03	3.98	3.61	4.11	3.22	3.77	2.24	4.35	3.57
_7: House-Husband/Wife	2.71	3.07	21.09	15.34	2.32	2.22	20.56	14.87	2.08	2.18	20.06	13.77
_7n: House-Husband/Wife	3.52	3.09	18.74	13.78	2.69	1.35	18.32	13.89	2.00	.	16.81	13.21
_8: Retired	4.92	3.91	5.23	4.27	2.28	1.70	2.03	1.56	1.42	2.10	1.13	0.99
_9: Other	2.41	5.26	4.08	6.49	1.03	0.85	2.94	5.45	0.91	0.64	1.99	2.15
atyp_sum	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
atyp_missing	0.97	3.65	0.91	3.36	1.80	3.17	2.30	4.41	3.37	4.66	2.88	2.67
Mean Age	71.37	6.77	73.07	7.48	65.07	2.17	65.05	2.35	63.59	1.86	63.74	1.93
Sum of Weights	6436520.09		8635570.99		1376688.64		1755168.17		467585.529		725210.011	
Number of cases (n=)	1827		2113		392		465		140		180	

(C) Only Persons with a Migration History

	ALL PENSIONERS IN THE YEAR 2005				FIRST TIME PENSIONERS B/W 2000-2004				FIRST TIME PENSIONERS B/W 2003-2004			
Spelltype_	**Men**		**Women**		**Men**		**Women**		**Men**		**Women**	
	Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
_1: School/University	4.88	5.67	2.48	2.09	8.34	7.55	2.65	2.64	2.88	3.04	2.24	1.58
_2: Apprenticeship/Training	3.26	2.63	2.86	1.57	3.87	4.36	2.87	0.94	3.43	2.17	2.61	0.89
_3: Military/Civilian Service	3.57	4.40	7.92	10.88	2.73	1.05			2.66	1.15		
_4: Full-time employed	36.54	9.28	27.18	13.99	32.49	11.95	23.54	14.08	37.47	7.97	21.82	13.20
_5: Part-time employed	2.06	2.87	9.95	9.93	1.35	1.07	8.51	7.33	3.02	0.58	8.38	6.27
_6: Unemployed	5.08	4.71	3.15	2.60	8.21	5.15	3.82	3.12	5.18	2.68	3.08	3.34
_7: House-Husband/Wife	1.83	1.95	18.99	15.21	2.27	2.60	22.51	15.20	3.95	3.90	20.78	8.25
_7n: House-Husband/Wife	1.98	2.58	19.67	12.45			21.06	12.22			17.77	8.72
_8: Retired	4.92	3.56	5.89	3.72	2.33	1.49	2.18	1.59	1.07	0.65	1.04	0.78
_9: Other	2.02	2.30	3.67	3.03	0.68	0.98	2.29	2.69	0.65	0.28	3.56	3.38
atyp_sum	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
atyp_missing	1.01	3.37	0.78	2.37	1.62	3.39	1.96	2.59	2.99	5.85	2.35	1.27
Mean Age	71.10	7.26	71.88	6.55	65.22	2.09	65.34	2.14	64.13	1.82	63.82	1.30
Sum of Weights	675254.532		613842.328		182264.882		122802.26		54624.8904		43816.2995	
Number of cases (n=)	196		180		46		43		18		15	

(D) Comparison East vs. West: Germans and Persons with a Migration History together

	ALL PENSIONERS IN THE YEAR 2005							
	WEST				EAST			
Spelltype_	**Men**		**Women**		**Men**		**Women**	
	Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
_1: School/University	3.43	3.59	2.53	2.41	3.65	3.13	2.52	2.25
_2: Apprenticeship/Training	3.22	2.75	2.78	1.27	2.93	1.96	2.64	1.46
_3: Military/Civilian Service	3.75	3.78	2.93	4.66	3.51	3.66	1.38	1.51
_4: Full-time employed	37.90	7.55	22.28	14.11	37.70	6.65	30.60	11.34
_5: Part-time employed	2.57	4.72	11.95	10.15	1.88	2.97	8.67	9.41
_6: Unemployed	3.06	3.22	3.26	3.14	2.70	1.82	3.54	2.79
_7: House-Husband/Wife	2.84	3.29	23.72	15.07	2.03	1.37	8.90	9.66
_7n: House-Husband/Wife	3.76	3.32	20.68	13.81	2.30	1.05	8.93	8.45
_8: Retired	4.89	3.99	4.78	3.92	5.05	3.61	6.55	4.94
_9: Other	2.40	5.22	4.28	7.04	2.45	5.50	3.43	4.25
atyp_sum	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
atyp_missing	0.92	3.72	0.80	3.36	1.11	3.40	1.27	3.33
Mean Age	71.65	6.84	73.41	7.43	70.43	6.43	71.97	7.57
Sum of weights	4981314.42		6611747.11		1455205.67		2023823.88	
Number of cases (n=)	1268		1423		559		690	

(E) Comparison West vs. East: Germans and Persons with a Migration History together

	FIRST-TIME PENSIONERS BETWEEN 2000 AND 2004							
	WEST				EAST			
Spelltype_	**Men**		**Women**		**Men**		**Women**	
	Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
_1: School/University	5.13	4.99	2.39	2.29	4.20	3.10	2.63	2.43
_2: Apprenticeship/Training	3.22	2.06	2.84	1.30	2.73	1.13	2.62	0.96
_3: Military/Civilian Service	2.27	3.40	11.58	21.49	2.19	1.10		
_4: Full-time employed	38.37	7.44	18.83	13.70	37.54	6.00	30.48	10.00
_5: Part-time employed	1.77	2.16	12.64	10.50	1.38	1.70	6.79	8.62
_6: Unemployed	4.36	4.13	3.54	3.00	3.20	2.06	4.92	3.37
_7: House-Husband/Wife	2.37	2.32	22.98	14.46	1.81	0.57	5.82	6.51
_7n: House-Husband/Wife	2.69	1.35	19.99	13.83			6.33	6.49
_8: Retired	2.30	1.81	1.87	1.55	2.23	1.32	2.60	1.48
_9: Other	1.07	0.87	3.40	6.25	0.92	0.80	1.63	1.14
atyp_sum	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
atyp_missing	1.45	2.19	2.13	4.78	2.87	4.98	2.94	2.50
Mean Age	65.25	2.18	65.47	2.36	64.48	2.04	63.45	1.43
Sum of Weights	1039575.36		1387319.91		337113.279		367848.26	
Number of cases (n=)	266		328		126		137	

(F) Comparison East vs. West: Germans and Persons with a Migration History together

	FIRST-TIME PENSIONERS BETWEEN 2003 AND 2004							
	WEST				EAST			
Spelltype_	**Men**		**Women**		**Men**		**Women**	
	Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
_1: School/University	3.43	3.14	2.24	2.44	3.81	2.60	1.83	1.71
_2: Apprenticeship/Training	3.45	1.80	3.02	1.34	2.67	1.30	2.62	0.94
_3: Military/Civilian Service	2.00	1.16	3.00	.	2.55	1.10		
_4: Full-time employed	38.56	6.04	19.13	13.66	34.44	7.06	28.60	9.82
_5: Part-time employed	2.77	3.24	12.48	10.66	1.77	2.59	4.98	5.73
_6: Unemployed	3.67	2.31	3.31	2.98	3.95	2.13	6.24	3.81
_7: House-Husband/Wife	2.08	2.18	22.10	13.54			8.10	7.86
_7n: House-Husband/Wife	2.00	.	18.57	13.31			7.14	7.23
_8: Retired	1.51	2.40	1.11	1.04	1.16	0.92	1.25	0.76
_9: Other	1.00	0.71	2.19	2.34	0.72	0.40	1.14	0.59
atyp_sum	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
atyp_missing	2.50	2.85	2.51	2.40	5.67	7.12	4.37	3.20
Mean Age	63.78	1.88	64.12	1.90	63.08	1.71	62.18	1.09
Sum of Weights	338237.769		582328.701		129347.759		142881.311	
Number of cases (n=)	91		124		49		56	

(G) Comparison East vs. West: Germans only

	ALL PENSIONERS IN THE YEAR 2005							
	WEST				EAST			
Spelltype_	**Men**		**Women**		**Men**		**Women**	
	Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
_1: School/University	3.24	3.18	2.55	2.45	3.63	3.09	2.47	2.20
_2: Apprenticeship/Training	3.22	2.76	2.77	1.25	2.92	1.96	2.64	1.47
_3: Military/Civilian Service	3.78	3.62	2.81	4.40	3.53	3.70	1.38	1.51
_4: Full-time employed	38.10	7.23	21.93	14.01	37.75	6.61	30.56	11.42
_5: Part-time employed	2.66	4.99	12.16	10.14	1.89	2.98	8.54	9.39
_6: Unemployed	2.68	2.69	3.28	3.20	2.68	1.82	3.54	2.82
_7: House-Husband/Wife	2.94	3.38	23.85	15.07	2.03	1.37	9.06	9.68
_7n: House-Husband/Wife	3.96	3.37	20.67	13.89	2.30	1.05	8.91	8.52
_8: Retired	4.89	4.06	4.71	3.92	5.04	3.56	6.50	5.03
_9: Other	2.47	5.65	4.39	7.32	2.55	5.61	3.24	4.30
atyp_sum	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
atyp_missing	0.91	3.76	0.79	3.42	1.10	3.41	1.31	3.40
Mean Age	71.72	6.78	73.54	7.47	70.45	6.43	71.98	7.66
Sum of Weights	4330301.22		6094837.39		1430964.34		1926891.27	
Number of cases (n=)	1081		1268		550		665	

(H) Comparison East vs. West: Germans only

	FIRST-TIME PENSIONERS BETWEEN 2000 AND 2004							
	WEST				EAST			
Spelltype_	Men		Women		Men		Women	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
_1: School/University	4.35	3.82	2.39	2.31	4.21	3.12	2.50	2.20
_2: Apprenticeship/Training	3.13	1.38	2.84	1.31	2.66	1.01	2.61	0.96
_3: Military/Civilian Service	2.15	3.73	11.58	21.49	2.20	1.12		
_4: Full-time employed	39.58	5.31	18.30	13.58	37.59	6.05	30.88	9.59
_5: Part-time employed	1.88	2.36	13.17	10.70	1.38	1.70	6.33	8.23
_6: Unemployed	3.17	2.83	3.50	3.00	3.17	2.08	4.93	3.36
_7: House-Husband/Wife	2.38	2.32	22.96	14.44	1.81	0.57	5.88	6.54
_7n: House-Husband/Wife	2.69	1.35	19.91	13.95			6.33	6.49
_8: Retired	2.28	1.87	1.85	1.55	2.26	1.31	2.57	1.47
_9: Other	1.11	0.83	3.55	6.61	0.97	0.83	1.64	1.17
atyp_sum	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
atyp_missing	1.43	1.84	2.15	4.93	2.88	5.04	2.97	2.50
Mean Age	65.25	2.21	65.48	2.38	64.51	2.05	63.42	1.42
Sum of Weights	864379.348		1270852.36		330044.409		361513.55	
Number of cases (n=)	222		289		124		134	

(I) Comparison East vs. West: Germans only

	FIRST-TIME PENSIONERS BETWEEN 2003 AND 2004							
	WEST				EAST			
Spelltype_	Men		Women		Men		Women	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
_1: School/University	3.53	3.16	2.24	2.47	3.82	2.63	1.84	1.72
_2: Apprenticeship/Training	3.51	1.80	3.02	1.34	2.50	1.02	2.62	0.94
_3: Military/Civilian Service	1.77	1.06	3.00	.	2.61	1.15		
_4: Full-time employed	38.68	5.60	18.90	13.69	34.39	7.25	28.74	9.77
_5: Part-time employed	2.75	3.37	12.97	10.97	1.77	2.59	4.75	5.42
_6: Unemployed	3.27	1.98	3.34	2.96	3.94	2.20	6.26	3.83
_7: House-Husband/Wife	1.89	2.00	22.20	13.86			8.10	7.86
_7n: House-Husband/Wife	2.00	.	18.64	13.65			7.14	7.23
_8: Retired	1.58	2.57	1.11	1.06	1.18	0.93	1.25	0.76
_9: Other	1.01	0.73	2.01	2.14	0.77	0.43	1.10	0.57
atyp_sum	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
atyp_missing	2.41	1.80	2.53	2.47	5.85	7.28	4.37	3.22
Mean Age	63.71	1.88	64.14	1.94	63.06	1.75	62.18	1.09
Sum of Weights	290681.749		539963.221		122278.889		141430.491	
Number of cases (n=)	75		110		47		55	

(J) Comparison East vs. West: Persons with a Migration History Only

	ALL PENSIONERS IN THE YEAR 2005							
	WEST				EAST			
Spelltype_	**Men**		**Women**		**Men**		**Women**	
	Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
_1: School/University	4.87	5.69	2.27	1.78	2.12	4.45	4.13	3.50
_2: Apprenticeship/Training	3.22	2.65	2.94	1.68	3.13	2.55	2.55	1.00
_3: Military/Civilian Service	3.61	4.52	7.92	10.88	2.19	1.88		
_4: Full-time employed	36.62	9.32	26.35	14.58	34.48	8.40	31.37	9.67
_5: Part-time employed	2.10	2.89	9.61	9.97	0.08	0.19	13.79	9.12
_6: Unemployed	5.12	4.77	3.08	2.69	1.32	2.02	3.59	2.11
_7: House-Husband/Wife	1.83	1.95	21.82	14.93	0.00	0.00	6.78	9.38
_7n: House-Husband/Wife	1.98	2.58	20.70	12.50	0.00	0.00	9.75	6.14
_8: Retired	4.88	3.45	5.59	3.88	5.85	5.86	7.36	2.40
_9: Other	2.10	2.33	3.12	2.73	0.19	0.26	5.35	3.45
atyp_sum	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
atyp_missing	0.99	3.40	0.84	2.50	1.65	2.33	0.46	1.47
Mean Age	71.19	7.28	71.89	6.74	68.70	6.61	71.81	5.55
Sum of Weights	651013.20		516909.71		24241.33		96932.61	
Number of cases (n=)	187		155		9		25	

(K) Comparison East vs. West: Persons with a Migration History Only

	FIRST-TIME PENSIONERS BETWEEN 2000 AND 2004							
	WEST				EAST			
Spelltype_	Men		Women		Men		Women	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
_1: School/University	8.41	7.59	2.32	2.05	3.50	.	7.53	6.40
_2: Apprenticeship/Training	3.77	4.45	2.87	0.94	5.50	2.65	2.88	1.12
_3: Military/Civilian Service	2.83	1.08			2.00	.		
_4: Full-time employed	32.38	12.18	24.40	13.92	35.23	3.24	7.68	4.93
_5: Part-time employed	1.35	1.07	7.30	6.02			24.45	2.15
_6: Unemployed	8.46	5.20	3.77	3.06	4.17	0.88	4.26	4.44
_7: House-Husband/Wife	2.27	2.60	23.19	14.97			2.00	.
_7n: House-Husband/Wife			21.06	12.22				
_8: Retired	2.39	1.49	2.08	1.52	0.83	0.88	4.11	1.82
_9: Other	0.72	1.11	2.39	2.86	0.50	.	1.55	0.44
atyp_sum	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
atyp_missing	1.59	3.45	2.00	2.63	2.55	1.18	1.21	1.93
Mean Age	65.30	2.09	65.37	2.17	63.45	1.18	64.79	1.93
Sum of Weights	175196.012		116467.55		7068.87		6334.71	
Number of cases (n=)	44		39		2		3	

(L) Comparison East vs. West: Persons with a Migration History Only

	FIRST-TIME PENSIONERS BETWEEN 2003 AND 2004							
	WEST				EAST			
Spelltype_	Men		Women		Men		Women	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
_1: School/University	2.84	3.16	2.35	1.62	3.50	.	1.00	.
_2: Apprenticeship/Training	2.93	1.89	2.76	0.97	5.50	2.65	2.00	.
_3: Military/Civilian Service	2.80	1.23			2.00	.		
_4: Full-time employed	37.81	8.47	22.06	13.39	35.23	3.24	14.83	.
_5: Part-time employed	3.02	0.58	7.89	5.82			21.33	.
_6: Unemployed	5.40	2.91	2.98	3.46	4.17	0.88	4.33	.
_7: House-Husband/Wife	3.95	3.90	20.78	8.25				
_7n: House-Husband/Wife			17.77	8.72				
_8: Retired	1.11	0.65	1.02	0.79	0.83	0.88	1.50	.
_9: Other	0.86	0.31	3.75	3.57	0.50	.	2.00	.
atyp_sum	51.00	0.00	51.00	0.00	51.00	0.00	51.00	.
atyp_missing	3.05	6.28	2.30	1.25	2.55	1.18	4.00	.
Mean Age	64.23	1.91	63.88	1.28	63.45	1.18	62.00	.
Sum of Weights	47556.0202		42365.4796		7068.87		1450.82	
Number of cases (n=)	16		14		2		1	

(M) Comparison Civil Servant in Last Job vs. Others

	ALL PENSIONERS IN THE YEAR 2005				FIRST TIME PENSIONERS B/W 2000-2004				FIRST TIME PENSIONERS B/W 2003-2004			
Spelltype_	**Civil Servants**		**Others**		**Civil Servants**		**Others**		**Civil Servants**		**Others**	
	Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
_1: School/University	4.40	4.05	2.89	2.90	7.55	4.70	3.40	3.61	6.13	3.82	2.61	2.54
_2: Apprenticeship/Training	3.30	1.42	2.94	2.08	6.50	11.46	2.92	1.56	2.92	1.46	3.03	1.44
_3: Military/Civilian Service	5.06	3.39	3.44	3.92	1.39	0.88	2.86	5.42	1.44	1.01	2.18	1.08
_4: Full-time employed	35.68	9.37	30.22	13.36	34.89	8.80	28.29	14.26	37.84	2.38	26.91	14.04
_5: Part-time employed	4.85	5.70	9.32	9.85	1.77	3.03	9.52	10.39	2.08	2.12	9.49	10.02
_6: Unemployed	1.48	1.45	3.17	3.00	1.05	0.72	4.04	3.42	1.00	.	4.14	3.08
_7: House-Husband/Wife	17.10	17.14	19.69	15.55	12.56	17.08	19.36	15.47	1.18	0.54	18.53	14.16
_7n: House-Husband/Wife	22.71	15.98	18.53	13.80			18.57	14.09			16.63	13.21
_8: Retired	5.84	5.74	5.08	4.09	1.60	1.38	2.18	1.77	1.47	1.66	1.36	1.97
_9: Other	1.47	0.76	3.48	6.14	1.87	1.37	2.17	4.35			1.57	1.79
atyp_sum	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
atyp_missing	0.17	0.71	0.94	3.51	1.60	2.20	2.09	3.89	2.40	2.71	3.12	3.57
Mean Age	73.46	7.39	72.33	7.23	66.56	1.97	65.05	2.30	65.57	1.56	63.63	1.89
Sum of weights	235778.969		14836312.1		264487.34		3193841.36		120773.061		1200670.21	
Number of cases (n=)	59		3883		43		867		17		324	

(N) Comparison Self-Employed in the Last Job vs. Others

	ALL PENSIONERS IN THE YEAR 2005				FIRST TIME PENSIONERS B/W 2000-2004				FIRST TIME PENSIONERS B/W 2003-2004			
Spelltype_	**Self-Employed**		**Others**		**Self-Employed**		**Others**		**Self-Employed**		**Others**	
	Mean	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**	**Mean**	**Std**
_1: School/University	4.60	3.60	2.87	2.90	5.22	3.89	3.40	3.68	6.09	4.98	2.49	2.35
_2: Apprenticeship/Training	3.18	1.59	2.94	2.08	2.59	1.02	2.95	1.61	2.39	1.23	3.09	1.48
_3: Military/Civilian Service	3.14	1.76	3.49	3.94	1.57	0.83	2.91	5.41	1.40	0.69	2.29	1.13
_4: Full-time employed	35.69	11.64	30.17	13.34	34.50	13.02	28.51	14.08	32.27	14.79	27.27	13.74
_5: Part-time employed	7.49	9.65	9.36	9.84	8.98	9.84	9.13	10.06	13.96	12.70	8.93	9.60
_6: Unemployed	2.03	2.41	3.18	3.01	2.37	2.75	4.12	3.43	1.91	1.59	4.24	3.08
_7: House-Husband/Wife	14.60	14.23	19.78	15.56	12.16	11.03	19.06	15.28	20.74	13.03	18.22	14.24
_7n: House-Husband/Wife	8.04	10.59	18.68	13.79	3.93	3.61	18.88	13.85	3.33	1.16	17.26	13.20
_8: Retired	1.93	2.21	5.16	4.12	1.40	1.33	2.18	1.63	1.01	1.52	1.26	1.53
_9: Other	5.84	7.42	3.42	6.08	6.16	9.80	1.97	3.80	0.90	0.73	1.60	1.82
atyp_sum	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00	51.00	0.00
atyp_missing	0.64	1.19	0.94	3.52	0.92	1.22	2.15	4.01	1.23	1.40	3.18	3.65
Mean Age	69.42	5.94	72.42	7.25	65.97	2.12	65.00	2.27	65.05	1.70	63.59	1.88
Sum of weights	365091.839		14706999.20		174859.70		2956997.11		68277.96		1124517.58	
Number of cases (n=)	93		3847		48		809		17		303	

ENDNOTES

[i] Information based on Question 103 in Person_Questionnaire (wave v). For persons reporting an own pension income from the GRV (LVA, BFA or Knappschaft) in 2005, the variable `rente_new` was coded as 1. If the person was below age 60 in 2005, it is assumed that the person received either an invalidity (disability) pension or an orphan's pension. For these cases the variable `rente_new` was recoded to 0.

[ii] Information based on PBIOSPE file. A person was considered to enter the retiree status between 2000 and 2004 if the variable `beginy`>1999 and `spelltype`==8. Based on the variable `beginy` and `gebjahr`, it was checked, whether the person retired earlier than age 60. If this was the case, the person was not considered in the population of first time pensioners between 2000 and 2004. It is assumed that these individuals receive an invalidity (disability) pension or an orphan's or survivor's pension and therefore don't belong to the population of interest, namely old-age pensioners.

[iii] Information based on PBIOSPE file. A person was considered to enter the retiree status between 2003 and 2004 if the variable `beginy`>2002 and `spelltype`==8. Based on the variable `beginy` and `gebjahr`, it was checked, whether the person retired earlier than age 60. If this was the case, the person was not considered in the population of first time pensioners between 2003 and 2004. It is assumed that these individuals receive an invalidity (disability) pension or an orphan's or survivor's pension and therefore don't belong to the population of interest, namely old-age pensioners.

[iv] Spelltype_7 are all the years a person spent as "house-husband" or "hausfrau" independent of whether other activities were reported. The years remain unweighted.

[v] Spelltype_7n only counts the years as "house-husband" or "hausfrau" if no other activities were reported in a given year. Hence, if a person reports that he/she was part-time employed and "house-husband" or "hausfrau" at age 35, then only part-time employment is counted. If however, a person reports to be only "house-husband" or "hausfrau" at age 35 then this year is counted under Spelltype_7n.

[vi] The sum of years adds up to 51 years. In order to get the sum of 51 years, all spelltypes have to be summed up. Instead of using spelltype_7, which is unweighted, one needs to use spelltype_7n plus all missing values.

[vii] Missings were defined as follows: If the sum of activities reported in a given year equals 0, then this year is coded as missing. Therefore, it is not possible to distinguish between gaps in the data, and left- or right-censoring.

[8] Individuals with a migration history were defined as foreigners in this analysis. The variable migration was operationalized as follows: If a person responded that he/she had no German citizenship in 2005 (variable: `nation05`), `migration` was coded as 1. Question 137 asks whether the person has German citizenship since birth or whether the person obtained German citizenship later (`vp137`). If a person responded that he/she obtained German citizenship later, then `migration` was coded 1. The variable `germborn` indicates whether a person was born in Germany or whether the person immigrated after 1948. If a person responded that he/she immigrated after 1948, then `migration` is coded as 1. The reliability of the `migration` variable was then checked with the variable `immiyear`. If a person reported a year of immigration then the person was expected to have a migration history, the check indicated that all persons with a migration history were captured by the new variable `migration`.

[ix] The distinction between East and West was based on the variable `vbula` from the ppfad data file.

[x] The distinction between civil servant and others was based on the variable `stib00-stib05` that describes the occupational status of a person for the years 2000 and 2005. If a person reports that he/she has been a civil servant (categories 610, 620, 630 or 640) in at least one of the six years, then the variable `civilservice` is coded as 1. The variable `civilservice` was also coded as 1, if a person reported to receive monthly pension benefits from the pension scheme for civil servants.

[xi] The distinction between self-employed and others was based on the variable `stib00-stib05` that describes the occupational status of a person for the years 2000 and 2005. If a person reports that he/she has been self-employed (categories 421, 422, 423, 431, 432 or 433) in at least one of the six years, then the variable `selfemployed` is coded as 1.

Appendix B

Average Time Spent in Different Activities

Set A: VVL Data

Note: The calculation of the average time spent in different activities in this first set of Tables is based on all individuals, independent of whether they have actually spent time in the respective activity or not. Hence, the denominator is always the total number of persons in the respective demographic group.

(A): German and Persons with Migration History

	FIRST-TIME OLD-AGE PENSIONERS IN 2004 (WEST AND EAST GERMANY)			
Spelltype_	**Men**		**Women**	
	Mean	Std	Mean	Std
_1: School/University	0.99	2.09	0.45	1.32
_2: Apprenticeship/Training	2.01	1.41	0.80	1.25
_3: Non-Professional Caretaking (Pflege)	0.02	0.29	0.37	1.40
_4: Childcare/Household	0.02	0.44	8.02	6.61
_5: Invalidity and Sickness	0.29	0.58	0.22	0.50
_6: Unemployed	1.73	2.83	1.62	2.78
_7: Military and Civilian Service	0.44	0.68	0.00	0.00
_8: Marginal Employment	0.06	0.45	0.41	1.22
_9: Self-Employed	0.29	1.94	0.07	0.96
_10: Other	1.86	5.69	0.86	3.35
- 11: Employment Subject to Social Insurance Contributions	33.03	12.32	21.01	13.86
_12: Invalidity	0.32	0.63	0.26	0.74
_13: Pension	0.30	0.53	0.34	0.44
_14: Missings	10.64	11.04	17.57	10.88
Mean Age	62.86	1.89	62.79	2.14
Sum	52	0	52	0
Number of cases (n=)	14274		17200	

(B) Germans and Persons with Migration History (East vs. West)[1]

	FIRST-TIME OLD-AGE PENSIONERS IN 2004 WEST GERMANY			
Spelltype_	**Men**		**Women**	
	Mean	Std	Mean	Std
_1: School/University	0.92	2.04	0.37	1.22
_2: Apprenticeship/Training	2.06	1.46	0.64	1.19
_3: Non-Professional Caretaking (Pflege)	0.02	0.29	0.41	1.48
_4: Childcare/Household	0.03	0.45	9.31	6.60
_5: Invalidity and Sickness	0.24	0.56	0.13	0.38
_6: Unemployed	1.44	2.77	1.17	2.38
_7: Military and Civilian Service	0.43	0.67	0	0.00
_8: Marginal Employment	0.07	0.50	0.49	1.33
_9: Self-Employed	0.32	2.12	0.06	0.99
_10: Other	2.18	6.28	0.97	3.61
- 11: Employment Subject to Social Insurance Contributions	31.74	13.36	17.91	13.34
_12: Invalidity	0.31	0.68	0.26	0.79
_13: Pension	0.31	0.51	0.33	0.43
_14: Missings	11.94	12.01	19.94	10.65
Mean Age	63.00	1.89	63.19	2.10
Sum	52	0	52	0
Number of cases (n=)	10727		13486	

	First-Time Old-Age Pensioners in 2004 East Germany			
Spelltype_	**Men**		**Women**	
	Mean	Std	Mean	Std
_1: School/University	1.18	2.21	0.74	1.62
_2: Apprenticeship/Training	1.87	1.27	1.39	1.26
_3: Non-Professional Caretaking (Pflege)	0.02	0.29	0.20	1.04
_4: Childcare/Household	0.01	0.43	3.34	4.03
_5: Invalidity and Sickness	0.45	0.62	0.58	0.70
_6: Unemployed	2.59	2.84	3.26	3.42
_7: Military and Civilian Service	0.49	0.71	0.00	0.00
_8: Marginal Employment	0.03	0.25	0.12	0.63
_9: Self-Employed	0.20	1.24	0.09	0.84
_10: Other	0.89	3.10	0.44	2.15
- 11: Employment Subject to Social Insurance Contributions	36.94	7.18	32.25	9.06
_12: Invalidity	0.36	0.43	0.26	0.51
_13: Pension	0.26	0.56	0.38	0.47
_14: Missings	6.71	5.82	8.95	6.47
Mean Age	62.44	1.80	61.34	1.59
Sum	52	0	52	0
Number of cases (n=)	3547		3714	

[1] The differentiation between East and West Germany was carried out with the variable WHORT. Due to anonymization purposes, it was not possible to differentiate between East and West Berlin. Therefore, Berlin was coded as East Germany. It might be possible that the pension insurance provided the dataset with a clear differentiation between East and West Berlin.

(C): Germans only

	FIRST-TIME OLD-AGE PENSIONERS IN 2004 (EAST AND WEST GERMANY)			
Spelltype_	**Men**		**Women**	
	Mean	Std	Mean	Std
_1: School/University	0.99	2.09	0.45	1.32
_2: Apprenticeship/Training	2.09	1.39	0.82	1.26
_3: Non-Professional Caretaking (Pflege)	0.02	0.30	0.37	1.40
_4: Childcare/Household	0.02	0.40	8.15	6.62
_5: Invalidity and Sickness	0.29	0.59	0.22	0.50
_6: Unemployed	1.68	2.78	1.59	2.74
_7: Military and Civilian Service	0.44	0.65	0.00	0.00
_8: Marginal Employment	0.06	0.46	0.41	1.23
_9: Self-Employed	0.30	1.97	0.07	0.97
_10: Other	1.91	5.78	0.87	3.40
- 11: Employment Subject to Social Insurance Contributions	33.10	12.41	20.89	13.92
_12: Invalidity	0.31	0.62	0.24	0.66
_13: Pension	0.30	0.52	0.35	0.43
_14: Missings	10.50	11.09	17.56	10.90
Mean Age	62.86	1.88	62.81	2.14
Sum	52	0	52	0
Number of cases (n=)	13696		16668	

(D) Germans only (separated East vs. West)

	FIRST-TIME OLD-AGE PENSIONERS IN 2004 WEST GERMANY			
Spelltype_	**Men**		**Women**	
	Mean	Std	Mean	Std
_1: School/University	0.92	2.05	0.36	1.20
_2: Apprenticeship/Training	2.15	1.42	0.66	1.21
_3: Non-Professional Caretaking (Pflege)	0.02	0.30	0.42	1.49
_4: Childcare/Household	0.03	0.46	9.51	6.57
_5: Invalidity and Sickness	0.24	0.57	0.12	0.37
_6: Unemployed	1.37	2.70	1.11	2.31
_7: Military and Civilian Service	0.42	0.63	0.00	0.00
_8: Marginal Employment	0.07	0.51	0.50	1.34
_9: Self-Employed	0.34	2.16	0.06	1.00
_10: Other	2.25	6.41	0.99	3.66
- 11: Employment Subject to Social Insurance Contributions	31.73	13.51	17.66	13.35
_12: Invalidity	0.30	0.67	0.24	0.71
_13: Pension	0.32	0.50	0.34	0.42
_14: Missings	11.84	12.13	20.03	10.66
Mean Age	63.01	1.89	63.22	2.09
Sum	52	0	52	0
Number of cases (n=)	10196		12996	

	First-Time Old-Age Pensioners in 2004 East Germany			
Spelltype_	**Men**		**Women**	
	Mean	Std	Mean	Std
_1: School/University	1.18	2.20	0.74	1.62
_2: Apprenticeship/Training	1.89	1.26	1.40	1.26
_3: Non-Professional Caretaking (Pflege)	0.02	0.29	0.20	1.03
_4: Childcare/Household	0.00	0.07	3.35	4.04
_5: Invalidity and Sickness	0.45	0.62	0.58	0.70
_6: Unemployed	2.57	2.81	3.27	3.42
_7: Military and Civilian Service	0.50	0.71	0.00	0.00
_8: Marginal Employment	0.03	0.25	0.12	0.63
_9: Self-Employed	0.20	1.25	0.09	0.84
_10: Other	0.90	3.12	0.44	2.16
- 11: Employment Subject to Social Insurance Contributions	37.08	7.00	32.33	9.00
_12: Invalidity	0.36	0.43	0.25	0.46
_13: Pension	0.26	0.56	0.38	0.46
_14: Missings	6.57	5.63	8.85	6.32
Mean Age	62.43	1.80	61.33	1.59
Sum	52	0	52	0
Number of cases (n=)	3500		3672	

(E): Persons with a Migration History only[2]

	FIRST-TIME OLD-AGE PENSIONERS IN 2004			
Spelltype_	**Men**		**Women**	
	Mean	Std	Mean	Std
_1: School/University	0.90	2.04	0.59	1.48
_2: Apprenticeship/Training	0.23	0.72	0.12	0.44
_3: Non-Professional Caretaking (Pflege)	0.02	0.25	0.30	1.41
_4: Childcare/Household	0.07	1.05	4.01	4.87
_5: Invalidity and Sickness	0.30	0.53	0.30	0.53
_6: Unemployed	2.90	3.75	2.62	3.61
_7: Military and Civilian Service	0.59	1.15	0.00	0.00
_8: Marginal Employment	0.04	0.36	0.21	0.83
_9: Self-Employed	0.09	1.19	0.02	0.36
_10: Other	0.71	2.42	0.39	1.38
- 11: Employment Subject to Social Insurance Contributions	31.45	10.07	24.66	11.39
_12: Invalidity	0.57	0.86	0.92	1.93
_13: Pension	0.15	0.67	0.19	0.55
_14: Missings	13.98	9.25	17.68	10.24
Mean Age	62.94	1.91	62.20	1.96
Sum	52	0	52	0
Number of cases (n=)	578		532	

[2] Foreigners have been identified over the variable SA. SA differentiates between individuals with German citizenship and non-German citizenship. Based on the information provided in the dataset, it is impossible to differentiate between individuals with a migration background and individuals without a migration background. Hence, it is likely that there are significant differences between the calculated means based on pension data and GSOEP data, because in the GSOEP data it is possible to identify persons with a migration background.

(F) Persons with a Migration History Only (East vs. West)

	FIRST-TIME OLD-AGE PENSIONERS IN 2004 WEST GERMANY			
Spelltype_	**Men**		**Women**	
	Mean	Std	Mean	Std
_1: School/University	0.86	1.97	0.59	1.47
_2: Apprenticeship/Training	0.23	0.72	0.10	0.41
_3: Non-Professional Caretaking (Pflege)	0.02	0.23	0.29	1.38
_4: Childcare/Household	0.01	0.10	4.14	4.97
_5: Invalidity and Sickness	0.29	0.53	0.29	0.52
_6: Unemployed	2.75	3.68	2.59	3.61
_7: Military and Civilian Service	0.62	1.17	0.00	0.00
_8: Marginal Employment	0.04	0.37	0.21	0.85
_9: Self-Employed	0.09	1.24	0.02	0.37
_10: Other	0.74	2.49	0.39	1.42
- 11: Employment Subject to Social Insurance Contributions	31.90	9.82	24.59	11.38
_12: Invalidity	0.58	0.88	0.89	1.92
_13: Pension	0.15	0.68	0.17	0.50
_14: Missings	13.73	9.18	17.72	10.10
Mean Age	62.96	1.90	62.20	1.96
Sum	52	0	52	0
Number of cases (n=)	531		490	

	First-Time Old-Age Pensioners in 2004 East Germany			
Spelltype_	**Men**		**Women**	
	Mean	Std	Mean	Std
_1: School/University	1.37	2.68	0.54	1.62
_2: Apprenticeship/Training	0.22	0.69	0.32	0.64
_3: Non-Professional Caretaking (Pflege)	0.06	0.44	0.38	1.73
_4: Childcare/Household	0.76	3.62	2.53	3.10
_5: Invalidity and Sickness	0.38	0.54	0.41	0.61
_6: Unemployed	4.58	4.22	3.00	3.62
_7: Military and Civilian Service	0.25	0.80	0.00	0.00
_8: Marginal Employment	0.01	0.06	0.14	0.53
_9: Self-Employed	0.00	0.00	0.00	0.00
_10: Other	0.37	1.31	0.33	0.78
- 11: Employment Subject to Social Insurance Contributions	26.46	11.54	25.53	11.51
_12: Invalidity	0.52	0.55	1.33	1.93
_13: Pension	0.18	0.54	0.34	0.97
_14: Missings	16.83	9.69	17.15	11.89
Mean Age	62.79	2.03	62.14	1.97
Sum	52	0	52	0
Number of cases (n=)	47		42	

Average Time Spent in Different Activities

Set B: VVL Data

Note: The calculation of the average time spent in different activities in this first set of Tables is only based on those individuals who have actually spent time in a certain activity. Hence, the denominator is the number of persons that spent time in the respective activity.

(A): Germans and Persons with Migration History

	FIRST-TIME OLD-AGE PENSIONERS IN 2004 (WEST AND EAST GERMANY)			
Spelltype_	**Men**		**Women**	
	Mean	Std	Mean	Std
_1: School/University	3.26	2.64	2.30	2.16
_2: Apprenticeship/Training	2.82	0.73	2.49	0.79
_3: Non-Professional Caretaking (Pflege)	2.46	2.37	3.20	2.83
_4: Childcare/Household	3.51	4.15	9.04	6.32
_5: Invalidity and Sickness	0.62	0.72	0.61	0.67
_6: Unemployed	3.38	3.18	3.60	3.17
_7: Military and Civilian Service	1.28	0.51	0.00	0.00
_8: Marginal Employment	1.64	1.72	2.82	1.88
_9: Self-Employed	5.99	6.59	5.42	6.84
_10: Other	9.36	9.64	4.99	6.70
- 11: Employment Subject to Social Insurance Contributions	33.12	12.22	21.97	13.41
_12: Invalidity	0.61	0.75	0.64	1.05
_13: Pension	0.64	0.62	0.59	0.43
_14: Missings	10.64	11.04	17.57	10.88
Sum	52.00	0.00	52.00	0.00
Mean Age	62.86	1.89	62.79	2.14
Number of cases (n=)	14274		17200	

(B) Germans and Persons with Migration History (East vs. West)

	FIRST-TIME OLD-AGE PENSIONERS IN 2004 WEST GERMANY			
Spelltype_	**Men**		**Women**	
	Mean	Std	Mean	Std
_1: School/University	3.12	2.70	2.21	2.18
_2: Apprenticeship/Training	2.92	0.70	2.62	0.79
_3: Non-Professional Caretaking (Pflege)	2.39	2.46	3.26	2.83
_4: Childcare/Household	3.22	3.65	10.55	6.02
_5: Invalidity and Sickness	0.62	0.77	0.51	0.61
_6: Unemployed	3.13	3.38	3.03	3.02
_7: Military and Civilian Service	1.23	0.56		
_8: Marginal Employment	1.82	1.79	2.94	1.86
_9: Self-Employed	7.66	7.12	8.30	8.34
_10: Other	10.07	10.15	5.10	6.88
- 11: Employment Subject to Social Insurance Contributions	31.85	13.25	18.97	12.98
_12: Invalidity	0.61	0.85	0.63	1.13
_13: Pension	0.63	0.58	0.58	0.42
_14: Missings	11.94	12.01	19.94	10.65
Sum	52.00	0.00	52.00	0.00
Mean Age	63.00	1.89	63.19	2.10
Number of cases (n=)	10727		13486	

	First-Time Old-Age Pensioners in 2004 East Germany			
Spelltype_	**Men**		**Women**	
	Mean	Std	Mean	Std
_1: School/University	3.65	2.46	2.48	2.11
_2: Apprenticeship/Training	2.51	0.74	2.30	0.74
_3: Non-Professional Caretaking (Pflege)	2.66	2.11	2.76	2.77
_4: Childcare/Household	9.98	9.10	3.70	4.08
_5: Invalidity and Sickness	0.61	0.65	0.72	0.71
_6: Unemployed	3.89	2.66	4.78	3.14
_7: Military and Civilian Service	1.44	0.31		
_8: Marginal Employment	0.89	1.14	1.76	1.67
_9: Self-Employed	2.87	3.87	2.97	3.83
_10: Other	6.14	5.86	4.26	5.32
- 11: Employment Subject to Social Insurance Contributions	36.94	7.18	32.28	9.02
_12: Invalidity	0.61	0.40	0.66	0.63
_13: Pension	0.65	0.74	0.62	0.45
_14: Missings	6.71	5.82	8.95	6.47
Mean Age	52.00	0.00	52.00	0.00
Sum	62.44	1.80	61.34	1.59
Number of cases (n=)	3547		3714	

(C): Germans only

	FIRST-TIME OLD-AGE PENSIONERS IN 2004 (EAST AND WEST GERMANY)			
Spelltype_	**Men**		**Women**	
	Mean	Std	Mean	Std
_1: School/University	3.26	2.64	2.29	2.16
_2: Apprenticeship/Training	2.83	0.71	2.51	0.77
_3: Non-Professional Caretaking (Pflege)	2.45	2.42	3.17	2.81
_4: Childcare/Household	3.27	3.66	9.19	6.31
_5: Invalidity and Sickness	0.62	0.73	0.61	0.67
_6: Unemployed	3.33	3.13	3.58	3.13
_7: Military and Civilian Service	1.25	0.45		
_8: Marginal Employment	1.66	1.72	2.83	1.88
_9: Self-Employed	5.95	6.57	5.40	6.85
_10: Other	9.78	9.72	5.21	6.81
- 11: Employment Subject to Social Insurance Contributions	33.18	12.31	21.87	13.46
_12: Invalidity	0.60	0.74	0.60	0.93
_13: Pension	0.63	0.60	0.59	0.42
_14: Missings	10.50	11.09	17.56	10.90
Sum	52.00	0.00	52.00	0.00
Mean Age	62.86	1.88	62.81	2.14
Number of cases (n=)	13696		16668	

(D) Germans only (separated East vs. West)

	FIRST-TIME OLD-AGE PENSIONERS IN 2004 WEST GERMANY			
Spelltype_	**Men**		**Women**	
	Mean	Std	Mean	Std
_1: School/University	3.11	2.70	2.19	2.18
_2: Apprenticeship/Training	2.94	0.67	2.65	0.76
_3: Non-Professional Caretaking (Pflege)	2.39	2.51	3.24	2.82
_4: Childcare/Household	3.29	3.69	10.77	5.94
_5: Invalidity and Sickness	0.62	0.77	0.50	0.61
_6: Unemployed	3.06	3.33	2.96	2.95
_7: Military and Civilian Service	1.18	0.48		
_8: Marginal Employment	1.85	1.79	2.95	1.86
_9: Self-Employed	7.61	7.11	8.30	8.39
_10: Other	10.61	10.24	5.34	7.00
- 11: Employment Subject to Social Insurance Contributions	31.84	13.41	18.73	12.99
_12: Invalidity	0.60	0.85	0.59	1.01
_13: Pension	0.63	0.55	0.58	0.41
_14: Missings	11.84	12.13	20.03	10.66
Sum	52.00	0.00	52.00	0.00
Mean Age	63.01	1.89	63.22	2.09
Number of cases (n=)	10196		12996	

	First-Time Old-Age Pensioners in 2004 East Germany			
Spelltype_	**Men**		**Women**	
	Mean	Std	Mean	Std
_1: School/University	3.63	2.45	2.48	2.10
_2: Apprenticeship/Training	2.52	0.74	2.31	0.74
_3: Non-Professional Caretaking (Pflege)	2.65	2.15	2.73	2.74
_4: Childcare/Household	2.21	2.53	3.71	4.09
_5: Invalidity and Sickness	0.61	0.65	0.72	0.71
_6: Unemployed	3.86	2.62	4.79	3.14
_7: Military and Civilian Service	1.44	0.30		
_8: Marginal Employment	0.90	1.15	1.76	1.68
_9: Self-Employed	2.87	3.87	2.97	3.83
_10: Other	6.21	5.87	4.34	5.36
- 11: Employment Subject to Social Insurance Contributions	37.08	7.00	32.36	8.96
_12: Invalidity	0.61	0.40	0.64	0.54
_13: Pension	0.65	0.74	0.62	0.44
_14: Missings	6.57	5.63	8.85	6.32
Sum	52.00	0.00	52.00	0.00
Mean Age	62.43	1.80	61.33	1.59
Number of cases (n=)	3500		3672	

(E): Persons with a Migration History only

	FIRST-TIME OLD-AGE PENSIONERS IN 2004			
Spelltype_	**Men**		**Women**	
	Mean	Std	Mean	Std
_1: School/University	3.40	2.68	2.58	2.12
_2: Apprenticeship/Training	1.40	1.24	0.92	0.85
_3: Non-Professional Caretaking (Pflege)	2.53	1.07	4.83	3.24
_4: Childcare/Household	7.80	9.11	4.57	4.95
_5: Invalidity and Sickness	0.63	0.62	0.67	0.61
_6: Unemployed	4.21	3.86	4.16	3.78
_7: Military and Civilian Service	2.44	0.98		
_8: Marginal Employment	1.17	1.59	2.21	1.73
_9: Self-Employed	12.40	8.30	8.25	.
_10: Other	2.51	4.02	1.24	2.27
- 11: Employment Subject to Social Insurance Contributions	31.56	9.92	25.04	11.05
_12: Invalidity	0.70	0.90	1.17	2.10
_13: Pension	0.75	1.36	0.84	0.91
_14: Missings	13.98	9.25	17.68	10.24
Sum	52.00	0.00	52.00	0.00
Mean Age	62.94	1.91	62.20	1.96
Number of cases (n=)	578		532	

(F) Persons with a Migration History Only (East vs. West)

	FIRST-TIME OLD-AGE PENSIONERS IN 2004 WEST GERMANY			
Spelltype_	**Men**		**Women**	
	Mean	Std	Mean	Std
_1: School/University	3.25	2.63	2.58	2.07
_2: Apprenticeship/Training	1.39	1.24	0.88	0.87
_3: Non-Professional Caretaking (Pflege)	2.42	1.20	4.63	3.23
_4: Childcare/Household	1.17	0.90	4.71	5.05
_5: Invalidity and Sickness	0.63	0.62	0.65	0.61
_6: Unemployed	4.05	3.83	4.17	3.80
_7: Military and Civilian Service	2.44	0.99		
_8: Marginal Employment	1.21	1.63	2.26	1.78
_9: Self-Employed	12.40	8.30	8.25	.
_10: Other	2.53	4.08	1.23	2.32
- 11: Employment Subject to Social Insurance Contributions	32.02	9.64	25.00	11.02
_12: Invalidity	0.70	0.93	1.14	2.11
_13: Pension	0.75	1.40	0.79	0.81
_14: Missings	13.73	9.18	17.72	10.10
Mean Age	52.00	0.00	52.00	0.00
Sum	62.96	1.90	62.20	1.96
Number of cases (n=)	531		490	

	First-Time Old-Age Pensioners in 2004 East Germany			
Spelltype_	**Men**		**Women**	
	Mean	Std	Mean	Std
_1: School/University	4.96	2.89	2.53	2.81
_2: Apprenticeship/Training	1.48	1.23	1.11	0.74
_3: Non-Professional Caretaking (Pflege)	3.00	.	8.00	1.06
_4: Childcare/Household	17.75	0.82	2.95	3.16
_5: Invalidity and Sickness	0.64	0.56	0.91	0.60
_6: Unemployed	5.82	3.92	4.07	3.66
_7: Military and Civilian Service	2.38	0.96		
_8: Marginal Employment	0.42		1.52	1.00
_9: Self-Employed				
_10: Other	2.19	2.61	1.38	1.08
- 11: Employment Subject to Social Insurance Contributions	26.46	11.54	25.53	11.51
_12: Invalidity	0.69	0.53	1.55	2.00
_13: Pension	0.76	0.92	1.41	1.60
_14: Missings	16.83	9.69	17.15	11.89
Mean Age	52.00	0.00	52.00	0.00
Sum	62.79	2.03	62.14	1.97
Number of cases (n=)	47		42	

Appendix C

Average Time Spent in Different Activities

Set C: Comparison of SOEP and VVL Data

Note: The calculation of the average time spent in different activities in this first set of Tables is only based on those individuals who have actually spent time in a certain activity. Hence, the denominator is the number of persons that spent time in the respective activity.

The tables are structured according to the same logic as the tables presented so far.

- Tables (A): Germans and Persons with a Migration History together
- Tables (B): Only Germans
- Tables (C): Only Persons with a Migration History

Every set of tables displays the means for men and women, men and women in West Germany, and men and women in East Germany.

The average is only based on those individuals that have spent time in a certain activity. If a person did not spent any time in "childcare & household", then the case is not considered in the denominator. A person that spent 5 years in "childcare & household" is considered in the denominator. An average value of 12 years in "childcare & household" therefore needs to be interpreted as follows: for those people that have spent time in "childcare & household", the average duration spent with "childcare & household" amounts to 12 years.

In order to get a better feeling for how many individuals are represented by the mean values displayed in the tables, I have added a column with the number of observations for the VVL 2004 population, the SOEP First Time Pensioners Between 2000 and 2004 and the SOEP First Time Pensioners between 2003 and 2004. If $n < 30$ then the cell was colored in red. If $n > 30$ & $n < 50$, then the cell was colored in orange.

The two grey shaded columns show the differences in the average durations spent in different activities based on the VVL 2004 and the two different SOEP populations ("First time pensioners b/w 2000 & 2004" and "First time pensioners b/w 2003 & 2004"). The differences were calculated by simple subtraction. Hence, the values can be negative if the average duration spent in a certain activity based on SOEP data is higher than the average duration spent in a certain activity based on VVL data.

Average time spent in activity x_{VVL} – Average time spent in activity x_{SOEP}

(A) GERMANS AND PERSONS WITH MIGRATION HISTORY TOGETHER

Men	Mean VVL 2004	N	Mean SOEP 2000-2004	N	Δ	Mean VVL 2004	Mean SOEP 2003-2004	N	Δ
School/University	3.26	4314	4.92	192	-1.66	3.26	3.55	78	-0.29
Apprenticeship/ Training	2.82	10191	3.10	340	-0.28	2.82	3.22	122	-0.41
Hausfrau/House-Husband	3.51	95	2.69	2	0.81	3.51	2.00	1	1.51
Unemployed	3.38	7287	3.98	210	-0.60	3.38	3.77	73	-0.40
Military/ Civilian Service	1.28	4948	2.24	129	-0.96	1.28	2.18	61	-0.89
Others years other years care years invalidity and sickness	12.43	2835 107 6768	0.17	73	12.26	12.43	0.91	28	11.52
Employment years fulltime years part time years marginal employment years self employment	40.75	14237 531 696	39.83	392 192	0.91	40.75	39.96	140 36	0.79
Retired years invalidity (disability)pension years retired	1.24	7609 6642	2.28	392	-1.04	1.24	1.42	140	-0.18
Missing_Years	10.64	14274	1.03	392	9.61	10.64	3.37	140	7.26
Mean_Age	62.86	14274	65.07	392	-2.20	62.86	63.59	140	-0.72

Women	Mean VVL 2004	N	Mean SOEP 2000-2004	N	Δ	Mean VVL 2004	Mean SOEP 2003-2004	N	Δ
School/University	2.30	3373	2.43	252	-0.13	2.30	2.17	99	0.12
Apprenticeship/ Training	2.49	5550	2.78	335	-0.30	2.49	2.93	139	-0.44
Hausfrau/House-Husband	9.04	15255	18.32	309	-9.28	9.04	16.81	133	-7.76
Unemployed	3.60	7725	4.11	208	-0.50	3.60	4.35	97	-0.74
Military/ Civilian Service	0.00	0	11.58	2	-11.58	0.00	3.00	1	-3.00
Others years other years care years invalidity and sickness	8.79	3956 1985 6342	2.94	87	5.85	8.79	1.99	39	6.80
Employment years fulltime years part time years marginal employment years self employment	30.21	16443 2493 209	32.74	448 306	-2.53	30.21	31.95	176 120	-1.74
Retired years invalidity (disability)pension years retired	1.23	7136 9941	2.03	465	-0.80	1.23	1.13	180	0.09
Missing_Years	17.57	17200	2.30	465	15.26	17.57	2.88	180	14.69
Mean_Age	62.79	17200	65.05	465	-2.26	62.79	63.74	180	-0.95

Men West	Mean VVL 2004	N	Mean SOEP 2000-2004	N	Δ	Mean VVL 2004	Mean SOEP 2003-2004	N	Δ
School/University	3.12	3164	5.13	128	-2.01	3.12	3.43	48	-0.31
Apprenticeship/ Training	2.92	7559	3.22	225	-0.30	2.92	3.45	76	-0.53
Hausfrau/House-Husband	3.22	91	2.69	2	0.53	3.22	2.00	1	1.22
Unemployed	3.13	4920	4.36	116	-1.23	3.13	3.67	38	-0.53
Military/ Civilian Service	1.23	3730	2.27	68	-1.04	1.23	2.00	91	-0.77
Others years other years care years invalidity and sickness	13.08	2322 80 4161	1.07	47	12.01	13.08	1.00	17	12.08
Employment years fulltime years part time years marginal employment years self employment	41.33	10690 427 453	40.14	266 73	1.19	41.33	41.32	91 22	0.01
Retired years invalidity (disability)pension years retired	1.24	5483 5231	2.30	266	-1.06	1.24	1.51	91	-0.28
Missing_Years	11.94	10727	1.45	266	10.48	11.94	2.50	91	9.44
Mean_Age	63.00	10727	65.25	266	-2.25	63.00	63.78	91	-0.78
Women West	Mean VVL 2004	N	Mean SOEP 2000-2004	N	Δ	Mean VVL 2004	Mean SOEP 2003-2004	N	Δ
School/University	2.21	2266	2.39	175	-0.18	2.21	2.24	67	-0.04
Apprenticeship/ Training	2.62	3311	2.84	217	-0.22	2.62	3.02	88	-0.40
Hausfrau/House-Husband	10.55	11905	19.99	253	-9.44	10.55	18.57	102	-8.03
Unemployed	3.03	5191	3.54	104	-0.51	3.03	3.31	50	-0.28
Military/ Civilian Service	0	0	11.58	2	-11.58	0	3.00	1	-3.00
Others years other years care years invalidity and sickness	8.87	2570 1711 3362	3.40	59	5.47	8.87	2.19	26	6.68
Employment years fulltime years part time years marginal employment years self employment	30.21	12732 2233 96	31.47	311 209	-1.26	30.21	31.61	120 79	-1.40
Retired years invalidity (disability)pension years retired	1.21	5672 7706	1.87	328	-0.66	1.21	1.11	124	0.11
Missing_Years	19.94	13486	2.13	328	17.81	19.94	2.51	124	17.43
Mean_Age	63.19	13486	65.47	328	-2.28	63.19	64.12	124	-0.93

Men East	Mean VVL 2004	N	Mean SOEP 2000-2004	N	Δ	Mean VVL 2004	Mean SOEP 2003-2004	N	Δ
School/University	3.65	1150	4.20	64	-0.55	3.65	3.81	30	-0.16
Apprenticeship/ Training	2.51	2632	2.73	115	-0.21	2.51	2.67	46	-0.15
Hausfrau/House-Husband	9.98	4	0.00	0	9.98	9.98	0.00	0	9.98
Unemployed	3.89	2367	3.20	94	0.69	3.89	3.95	35	-0.07
Military/ Civilian Service	1.44	1218	2.19	61	-0.75	1.44	2.55	25	-1.11
Others years other years care years invalidity and sickness	9.42	518 27 2607	0.92	26	8.50	9.42	0.72	11	8.70
Employment years fulltime years part time years marginal employment years self employment	40.71	3547 104 243	38.92	126 39	1.79	40.71	36.21	49 14	4.50
Retired years invalidity (disability)pension years retired	1.26	2126 1411	2.23	126	-0.97	1.26	1.16	49	0.10
Missing_Years	6.71	3547	2.87	126	3.84	6.71	5.67	49	1.04
Mean_Age	62.44	3547	64.48	126	-2.05	62.44	63.08	49	-0.64
Women East	Mean VVL 2004	N	Mean SOEP 2000-2004	N	Δ	Mean VVL 2004	Mean SOEP 2003-2004	N	Δ
School/University	2.48	1107	2.63	77	-0.15	2.48	1.83	32	0.65
Apprenticeship/ Training	2.30	2239	2.62	118	-0.31	2.30	2.62	51	-0.32
Hausfrau/House-Husband	3.70	3350	6.33	56	-2.63	3.70	7.14	31	-3.44
Unemployed	4.78	2534	4.92	104	-0.14	4.78	6.24	47	-1.45
Military/ Civilian Service	0.00	0	0.00	0	0.00	0.00	0.00	0	0.00
Others years other years care years invalidity and sickness	7.75	386 274 2980	1.63	28	6.12	7.75	1.14	13	6.61
Employment years fulltime years part time years marginal employment years self employment	37.00	3711 260 113	37.27	137 97	-0.27	37.00	33.58	56 41	3.42
Retired years invalidity (disability)pension years retired	1.28	1464 2235	2.60	137	-1.31	1.28	1.25	56	0.04
Missing_Years	8.95	3714	2.94	137	6.01	8.95	4.37	56	4.57
Mean_Age	61.34	3714	63.45	137	-2.11	61.34	62.18	56	-0.84

(B): Only Germans

Men	Mean VVL 2004	N	Mean SOEP 2000-2004	N	Δ	Mean VVL 2004	Mean SOEP 2003-2004	N	Δ
School/University	3.26	4161	4.31	174	-1.05	3.26	3.63	70	-0.37
Apprenticeship/ Training	2.83	10097	3.01	313	-0.18	2.83	3.20	113	-0.37
Hausfrau/House-Husband	3.27	90	2.69	2	0.58	3.27	2.00	1	1.27
Unemployed	3.33	6889	3.17	178	0.16	3.33	3.54	59	-0.21
Military/ Civilian Service	1.25	4809	2.17	113	-0.92	1.25	2.06	53	-0.82
Others years other years care years invalidity and sickness	12.85	2671 102 6490	1.08	65	11.77	12.85	0.94	24	11.90
Employment years fulltime years part time years marginal employment years self employment	40.79	13661 511 692	40.76	346 101	0.04	40.79	39.92	122 33	0.87
Retired years invalidity (disability)pension years retired	1.24	7135 6528	2.27	346	-1.04	1.24	1.46	122	-0.23
	10.50	13696	1.83	346	8.67	10.50	3.43	122	7.07
	62.86	13696	65.04	346	-2.18	62.86	63.51	122	-0.66

Women	Mean VVL 2004	N	Mean SOEP 2000-2004	N	Δ	Mean VVL 2004	Mean SOEP 2003-2004	N	Δ
School/University	2.29	3251	2.41	231	-0.13	2.29	2.17	92	0.12
Apprenticeship/ Training	2.51	5480	2.78	321	-0.27	2.51	2.93	135	-0.42
Hausfrau/House-Husband	9.19	14788	18.15	280	-8.97	9.19	16.74	121	-7.56
Unemployed	3.58	7390	4.13	182	-0.56	3.58	4.43	88	-0.85
Military/ Civilian Service	0.00	0	11.58	2	-11.58	0.00	3.00	10	-3.00
Others years other years care years invalidity and sickness	8.98	2791 1952 6106	3.02	72	5.96	8.98	1.83	32	7.16
Employment years fulltime years part time years marginal employment years self employment	30.10	15919 2443 208	32.81	407 274	-2.70	30.10	32.12	161	-2.01
Retired years invalidity (disability)pension years retired	1.19	6717 9823	2.01	423	-0.82	1.19	1.14	108	0.05
Missing_Years	17.56	16668	2.33	423	15.24	17.56	2.91	165	14.65
Mean_Age	62.81	16668	65.02	423	-2.22	62.81	63.73	165	-0.93

Men West	Mean VVL 2004	N	Mean SOEP 2000-2004	N	Δ	Mean VVL 2004	Mean SOEP 2003-2004	N	Δ
School/University	3.11	3024	4.35	111	-1.23	3.11	3.53	41	-0.42
Apprenticeship/ Training	2.94	7472	3.13	200	-0.19	2.94	3.51	69	-0.57
Hausfrau/House-Husband	3.29	88	2.69	2	0.60	3.29	2.00	1	1.29
Unemployed	1.18	4559	3.17	86	-1.99	1.18	3.27	26	-2.09
Military/ Civilian Service	0.42	3596	0.45	53	-0.03	0.42	1.77	29	-1.36
Others years other years care years invalidity and sickness	13.62	2166 76 3911	1.11	40	12.50	13.62	1.01	14	12.61
Employment years fulltime years part time years marginal employment years self employment	41.31	10161 408 449	41.46	222 62	-0.15	41.31	41.43	75 19	-0.12
Retired years invalidity (disability)pension years retired	1.23	5045 5128	2.28	222	-1.05	1.23	1.58	75	-0.35
Missing_Years	11.84	10196	1.43	222	10.41	11.84	2.41	75	9.44
Mean_Age	63.01	10196	65.25	222	-2.24	63.01	63.71	75	-0.70

Women West	Mean VVL 2004	N	Mean SOEP 2000-2004	N	Δ	Mean VVL 2004	Mean SOEP 2003-2004	N	Δ
School/University	2.19	2153	2.39	156	-0.21	2.19	2.24	61	-0.05
Apprenticeship/ Training	2.65	3253	2.84	206	-0.19	2.65	3.02	85	-0.37
Hausfrau/House-Husband	10.77	11474	19.91	224	-9.14	10.77	18.64	90	-7.87
Unemployed	2.96	4887	3.50	81	-0.55	2.96	3.34	42	-0.38
Military/ Civilian Service	0.00	0	11.58	2	-11.58	0.00	3.00	1	-3.00
Others years other years care years invalidity and sickness	9.08	2415 1680 3145	3.55	46	5.54	9.08	2.01	20	7.07
Employment years fulltime years part time years marginal employment years self employment	29.99	12250 2187 95	31.47	273 180	-1.48	29.99	31.87	106 68	-1.88
Retired years invalidity (disability)pension years retired	1.17	5289 7598	1.85	289	-0.68	1.17	1.11	110	0.06
Missing_Years	20.03	12996	2.15	289	17.88	20.03	2.53	110	17.50
Mean_Age	63.22	12996	65.48	289	-2.25	63.22	64.14	110	-0.91

Men East	Mean VVL 2004	N	Mean SOEP 2000-2004	N	Δ	Mean VVL 2004	Mean SOEP 2003-2004	N	Δ
School/University	3.63	1137	4.21	63	-0.57	3.63	3.82	29	-0.19
Apprenticeship/ Training	2.52	2625	2.66	113	-0.14	2.52	2.50	44	0.02
Hausfrau/House-Husband	2.21	2	0.00	0	2.21	2.21	0.00	0	2.21
Unemployed	3.86	2330	3.17	92	0.68	3.86	3.94	33	-0.08
Military/ Civilian Service	1.44	1213	0.90	60	0.54	1.44	2.61	24	-1.18
Others years other years care years invalidity and sickness	9.46	505 26 2579	0.97	25	8.50	9.46	0.77	10	8.70
Employment years fulltime years part time years marginal employment years self employment	40.85	3500 103 243	38.97	124 39	1.89	40.85	36.16	47 14	4.69
Retired years invalidity (disability)pension years retired	1.26	2090 1400	2.26	124	-1.00	1.26	1.18	47	0.08
Missing_Years	6.57	3500	2.88	124	3.69	6.57	5.85	47	0.72
Mean_Age	62.43	3500	64.51	124	-2.07	62.43	63.06	47	-0.63

Women East	Mean VVL 2004	N	Mean SOEP 2000-2004	N	Δ	Mean VVL 2004	Mean SOEP 2003-2004	N	Δ
School/University	2.48	1098	2.50	75	-0.02	2.48	1.84	31	0.64
Apprenticeship/ Training	2.31	2227	2.61	115	-0.30	2.31	2.62	50	-0.32
Hausfrau/House-Husband	3.71	3314	6.33	56	-2.62	3.71	7.14	31	-3.43
Unemployed	4.79	2503	4.93	101	-0.14	4.79	6.26	46	-1.47
Military/ Civilian Service	0.00	0	0.0	0	-0.90	0.00	0	0	0.00
Others years other years care years invalidity and sickness	7.79	376 272 2961	1.64	26	6.15	7.79	1.10	12	6.69
Employment years fulltime years part time years marginal employment years self employment	37.08	3669 256 113	37.21	134 94	-0.12	37.08	33.49	55 40	3.59
Retired years invalidity (disability)pension years retired	1.26	1428 2225	2.57	134	-1.31	1.26	1.25	55	0.01
Missing_Years	8.85	3672	2.88	134	5.97	8.85	4.37	55	4.48
Mean_Age	61.33	3672	63.42	134	-2.09	61.33	62.18	55	-0.86

(C): Only Persons with a Migration History

Men	Mean VVL 2004	N	Mean SOEP 2000-2004	N	Δ	Mean VVL 2004	Mean SOEP 2003-2004	N	Δ
School/University	3.40	153	8.34	18	-4.95	3.40	2.88	8	0.52
Apprenticeship/ Training	1.40	94	3.87	27	-2.47	1.40	3.43	9	-2.03
Hausfrau/House-Husband	7.80	5	0.00	0	7.80	7.80	0.00	0	7.80
Unemployed	4.21	398	8.21	32	-4.00	4.21	5.18	14	-0.97
Military/ Civilian Service	2.44	139	2.73	16	-0.29	2.44	2.66	8	-0.22
Others years other years care years invalidity and sickness	5.67	164 5 278	0.68	8	4.99	5.67	0.65	4	5.02
Employment years fulltime years part time years marginal employment years self employment	45.13	576 20 4	33.84	46 11	11.29	45.13	40.49	18 3	4.64
Retired years invalidity (disability)pension years retired	1.45	474 114	2.33	46	-0.88	1.45	1.07	18	0.37
Missing_Years	13.98	578	1.62	46	12.36	13.98	2.99	18	11.00
Mean_Age	62.94	578	65.22	46	-2.28	62.94	64.13	18	-1.18

Women	Mean VVL 2004	N	Mean SOEP 2000-2004	N	Δ	Mean VVL 2004	Mean SOEP 2003-2004	N	Δ
School/University	2.58	122	2.65	21	-0.07	2.58	2.24	7	0.34
Apprenticeship/ Training	0.92	70	2.87	14	-1.95	0.92	2.61	4	-1.70
Hausfrau/House-Husband	4.57	467	21.06	29	-16.49	4.57	17.77	12	-13.20
Unemployed	4.16	335	3.82	26	0.34	4.16	3.08	9	1.09
Military/ Civilian Service	0.00	0	0.00	0	0.00	0.00	0.00	0	0.00
Others years other years care years invalidity and sickness	6.75	165 33 236	2.29	15	4.46	6.75	3.56	7	3.18
Employment years fulltime years part time years marginal employment years self employment	35.50	524 50 1	32.05	41 32	3.45	35.50	30.20	15 12	5.29
Retired years invalidity (disability)pension years retired	2.01	419 118	2.18	42	-0.17	2.01	1.04	15	0.97
Missing_Years	17.68	532	1.96	42	15.71	17.68	2.35	15	15.32
Mean_Age	62.20	532	65.34	42	-3.15	62.20	63.82	15	-1.62

Men West	Mean VVL 2004	N	Mean SOEP 2000-2004	N	Δ	Mean VVL 2004	Mean SOEP 2003-2004	N	Δ
School/University	3.25	140	8.41	17	-5.16	3.25	2.84	7	0.41
Apprenticeship/ Training	1.39	87	3.77	25	-2.38	1.39	2.93	7	-1.54
Hausfrau/House-Husband	1.17	3	0.00	0	1.17	1.17	0.00	0	1.17
Unemployed	4.05	361	8.46	30	-4.42	4.05	5.40	12	-1.36
Military/ Civilian Service	2.44	134	2.83	15	-0.39	2.44	2.80	7	-0.36
Others years other years care years invalidity and sickness	5.57	156 4 250	0.72	7	4.85	5.57	0.86	3	4.71
Employment years fulltime years part time years marginal employment years self employment	45.62	529 19 4	33.73	49 11	11.89	45.62	40.83	16 3	4.79
Retired years invalidity (disability)pension years retired	1.45	438 103	2.39	44	-0.94	1.45	1.11	16	0.34
Missing_Years	13.73	531	1.59	44	12.15	13.73	3.05	16	10.68
Mean_Age	62.96	531	65.30	44	-2.34	62.96	64.23	16	-1.27

Women West	Mean VVL 2004	N	Mean SOEP 2000-2004	N	Δ	Mean VVL 2004	Mean SOEP 2003-2004	N	Δ
School/University	2.58	113	2.32	19	0.26	2.58	2.35	6	0.23
Apprenticeship/ Training	0.88	58	2.87	11	-1.99	0.88	2.76	3	-1.88
Hausfrau/House-Husband	4.71	431	21.06	29	-16.35	4.71	17.77	12	-13.07
Unemployed	4.17	304	3.77	23	0.40	4.17	2.98	8	1.19
Military/ Civilian Service	0.00	0	0.00	0	0.00	0.00	0.00	0	0.00
Others years other years care years invalidity and sickness	6.52	155 31 217	2.39	13	4.13	6.52	3.75	6	2.77
Employment years fulltime years part time years marginal employment years self employment	35.51	482 46 1	31.70	28 29	3.81	35.51	29.95	14 11	5.56
Retired years invalidity (disability)pension years retired	1.93	383 108	2.08	39	-0.15	1.93	1.02	14	0.90
Missing_Years	17.72	490	2.00	39	15.72	17.72	2.30	14	15.42
Mean_Age	62.20	490	65.37	39	-3.17	62.20	63.88	14	-1.68

Men East	Mean VVL 2004	N	Mean SOEP 2000-2004	N	Δ	Mean VVL 2004	Mean SOEP 2003-2004	N	Δ
School/University	4.96	13	3.50	1	1.46	4.96	3.50	1	1.46
Apprenticeship/ Training	1.48	7	5.50	2	-4.02	1.48	5.50	2	-4.02
Hausfrau/House-Husband	17.75	2	0	0	17.75	17.75	0	0	17.75
Unemployed	5.82	37	4.17	2	1.65	5.82	4.17	2	1.65
Military/ Civilian Service	2.38	5	2.00	1	0.38	2.38	2.00	1	0.38
Others years other years care years invalidity and sickness	5.83	8 1 28	0.50	7	5.33	5.83	0.50	1	5.33
Employment years fulltime years part time years marginal employment years self employment	26.88	47 1	35.23	2	-8.35	26.88	35.23	2 0	-8.35
Retired years invalidity (disability)pension years retired	1.44	36 11	0.83	0	0.61	1.44	0.83	2	0.61
Missing_Years	16.83	47	2.55	2	14.28	16.83	2.55	2	14.28
Mean_Age	62.79	47	63.45	2	-0.66	62.79	63.45	2	-0.66
Women East	Mean VVL 2004	N	Mean SOEP 2000-2004	N	Δ	Mean VVL 2004	Mean SOEP 2003-2004	N	Δ
School/University	2.53	9	7.53	2	-5.01	2.53	1	1	1.53
Apprenticeship/ Training	1.11	12	2.88	3	-1.76	1.11	2	1	-0.89
Hausfrau/House-Husband	2.95	36	0	0	2.95	2.95	0	0	2.95
Unemployed	4.07	31	4.26	3	-0.19	4.07	4.33	1	-0.27
Military/ Civilian Service	0.00	0	0	0	0.00	0.00	0	0	0.00
Others years other years care years invalidity and sickness	10.28	10 2 19	1.55	2	8.73	10.28	2	1	8.28
Employment years fulltime years part time years marginal employment years self employment	27.05	42 4 0	32.13	3 3	-5.08	27.05	36.17	1 1	-9.12
Retired years invalidity (disability)pension years retired	2.96	36 10	4.11	3	-1.14	2.96	1.50	1	1.46
Missing_Years	17.15	42	1.21	3	15.94	17.15	4	1	13.15
Mean_Age	62.14	42	64.79	3	-2.65	62.14	62	1	0.14

APPENDIX D

REGRESSION RESULTS BEFORE MODIFICATIONS

SOEP & SUF VVL 2004

REGRESSION RESULTS BEFORE MODIFICATIONS: SOEP, NO CONTROLS

DEPENDENT VARIABLE: MONTHLY PUBLIC PENSION BENEFIT (TOPCODE)	(1) TOTAL	(2) TOTAL WEST	(3) TOTAL EAST	(4) TOTAL MEN	(5) TOTAL WOMEN	(6) MEN WEST	(7) MEN EAST	(8) WOMEN WEST	(9) WOMEN EAST
Years in School (topcode)	38.437 (6.26)**	32.845 (4.15)**	37.041 (4.17)**	45.763 (5.02)**	18.641 (2.24)*	48.698 (3.95)**	33.509 (2.93)**	9.446 (0.92)	47.813 (3.05)**
Years in Training (topcode)	28.756 (7.75)**	25.037 (2.61)**	34.272 (2.90)**	19.932 (1.68)	30.243 (3.21)**	26.341 (1.67)	9.723 (0.57)	25.332 (2.23)*	62.655 (3.37)**
Years in Employment	4.350 (2.89)	1.106 (0.31)	10.392 (2.14)*	5.154 (1.05)	-4.118 (1.22)	5.047 (0.76)	12.103 (1.99)*	-4.421 (1.16)	13.773 (1.31)
Years in Unemployment	-25.665 (4.82)**	-34.147 (4.82)**	-13.203 (1.77)	-41.671 (4.35)**	-24.019 (4.13)**	-42.005 (3.39)**	-28.898 (2.17)*	-22.218 (2.77)**	0.640 (0.05)
Years in Homeproduction	-13.412 (4.75)**	-14.661 (4.30)**	-8.769 (1.69)		-18.402 (5.85)**			-18.325 (5.10)**	-4.697 (0.47)
Retired (dummy)	-2.791 (0.04)	-100.616 (1.11)	133.231 (1.47)	-219.778 (1.65)	32.109 (0.44)	-329.964 (1.89)	38.557 (0.23)	-27.867 (0.30)	101.332 (0.94)
Other (dummy)	-186.666 (2.49)*	-196.845 (2.33)*	-31.622 (0.14)	-424.818 (2.22)*	-175.032 (2.44)*	-380.221 (1.73)	dropped	-167.303 (2.11)*	62.353 (0.27)
Missing (dummy)	-56.724 (1.70)	-47.902 (1.00)	-87.217 (2.32)*	-100.974 (1.87)	-60.537 (1.57)	-113.654 (1.45)	-40.067 (0.67)	4.337 (0.08)	-109.666 (2.25)*
Receives Civil Servant Pension (dummy)	-671.797 (6.35)**	-782.488 (6.13)**	-230.859 (1.34)	-721.529 (6.26)**	dropped	-833.168 (5.82)**	-212.360 (1.17)	dropped	dropped
Receives Private Pension (dummy)	-110.076 (1.71)	-135.850 (1.80)	-52.813 (0.44)	-210.942 (2.42)*	79.264 (0.85)	-224.659 (2.17)*	-180.922 (0.99)	49.441 (0.44)	38.052 (0.24)
Worked as Civil Servant or Self-employed (dummy)	-258.515 (6.63)**	-300.494 (6.23)**	-118.463 (2.05)*	-335.448 (6.43)**	-159.982 (2.75)**	-386.817 (5.82)**	-184.381 (2.49)*	-181.444 (2.52)*	13.858 (0.15)
Educational Attainment: low Reference category: medium	-41.808 (1.27)	-37.169 (0.92)	-53.743 (1.06)	-108.931 (1.71)	-8.766 (0.26)	-106.680 (1.30)	-68.358 (0.80)	-20.394 (0.50)	-30.535 (0.49)
Educational Attainment: high Reference category: medium	104.383 (2.19)*	147.519 (2.39)*	86.075 (1.37)	92.803 (1.52)	34.846 (0.45)	152.602 (1.87)	24.105 (0.32)	-25.940 (0.26)	188.165 (1.55)
Educational Attainment: missing Reference category: medium	36.228 (0.51)	-133.669 (1.19)	196.652 (2.69)**	-124.008 (1.15)	199.664 (2.33)*	-265.026 (1.50)	34.184 (0.34)	-13.828 (0.10)	397.929 (3.63)**
Educational Attainment: unknown Reference category: medium	-106.643 (2.18)*	-89.247 (1.64)	-133.362 (0.58)	-220.776 (2.70)**	3.815 (0.07)	-216.211 (2.35)*	dropped	-5.116 (0.08)	-61.265 (0.27)
Sex – Female=1 (Dummy)	-349.582 (13.69)**	-458.074 (13.09)**	-187.980 (6.23)**						
West – West=1 (dummy)	100.678 (3.96)**			222.938 (5.94)**	-30.647 (0.96)				
Constant	895.672 (6.43)**	1203.751 (7.00)**	523.258 (2.37)*	857.542 (3.67)**	939.924 (5.89)**	1067.965 (3.29)**	578.985 (2.08)*	941.145 (5.27)**	56.850 (0.12)
Observations	949	662	287	443	506	304	139	358	148
R-Square	0.59	0.62	0.55	0.40	0.44	0.40	0.33	0.40	0.55

REGRESSION RESULTS BEFORE MODIFICATIONS: SUF VVL 2004, NO CONTROLS

DEPENDENT VARIABLE MONTHLY PUBLIC PENSION BENEFIT[1]	(1) TOTAL	(2) TOTAL WEST	(3) TOTAL EAST	(4) TOTAL MEN	(5) TOTAL WOMEN	(6) MEN WEST	(7) MEN EAST	(8) WOMEN WEST	(9) WOMEN EAST
Years in School (topcode)	39.602 (48.71)***	31.365 (31.37)***	55.205 (42.44)***	45.465 (42.74)***	38.244 (32.93)***	39.619 (30.66)***	54.717 (30.10)***	59.172 (31.33)***	29.904 (20.73)***
Years in Training (topcode)	24.709 (26.13)***	23.518 (21.17)***	20.141 (11.78)***	41.319 (29.90)***	5.209 (4.43)***	43.507 (27.34)***	30.397 (11.27)***	11.110 (5.26)***	4.720 (3.41)***
Years in Employment	26.615 (186.00)***	27.049 (169.43)***	22.734 (61.76)***	31.711 (159.25)***	20.750 (112.52)***	32.207 (145.93)***	24.977 (43.23)***	20.958 (45.59)***	20.856 (100.84)***
Years in Unemployment	0.389 (0.85)	-0.323 (0.57)	-1.415 (1.91)*	1.939 (2.79)***	1.533 (2.79)***	1.777 (2.17)**	-3.920 (3.01)***	0.647 (0.76)	3.180 (4.51)***
Years in Homeproduction	8.999 (29.28)***	10.573 (31.14)***	4.263 (4.92)***		1.933 (6.19)***			1.850 (2.06)**	2.071 (6.14)***
Retired (Dummy)	229.809 (11.99)***	231.493 (9.92)***	227.662 (7.65)***	268.592 (9.88)***	178.167 (7.40)***	246.880 (7.56)***	322.420 (7.12)***	132.841 (3.50)***	198.437 (6.70)***
Other (Dummy)	96.500 (25.46)***	101.958 (23.82)***	47.791 (6.10)***	154.445 (26.39)***	64.185 (14.57)***	171.341 (25.65)***	50.984 (4.28)***	44.513 (4.38)***	67.527 (13.78)***
Missing (Dummy)	-125.583 (23.94)***	-153.178 (23.78)***	-77.427 (9.59)***	-61.586 (10.71)***	-180.386 (10.51)***	-76.170 (10.91)***	-48.799 (5.13)***	-104.194 (4.55)***	-232.392 (9.99)***
Sex – Female = 1 (Dummy)	-197.208 (59.37)***	-231.164 (55.77)***	-132.136 (26.41)***						
West – West =1 (Dummy)	166.819 (53.67)***			233.646 (52.75)***	66.463 (16.87)***				
Educational Attainment: low (Reference Category: medium)	-114.930 (21.73)***	-128.184 (21.19)***	-52.728 (5.11)***	-151.069 (16.50)***	-69.389 (12.12)***	-163.455 (15.93)***	-79.718 (3.97)***	-39.853 (3.52)***	-73.893 (11.19)***
Educational Attainment: high (Reference Category: medium)	188.765 (28.14)***	200.363 (21.84)***	178.025 (20.61)***	146.457 (17.40)***	191.018 (17.90)***	150.438 (13.86)***	163.576 (13.27)***	180.243 (14.78)***	142.427 (8.04)***
Educational Attainment: missing (Reference Category: medium)	-60.144 (17.52)***	-52.591 (12.81)***	-64.516 (11.40)***	-105.538 (20.69)***	-37.767 (9.08)***	-106.693 (17.54)***	-81.710 (9.29)***	-56.231 (7.86)***	-31.230 (6.29)***
Educational Attainment: unknown (Reference Category: medium)	-81.834 (18.99)***	-85.979 (16.63)***	-57.922 (8.29)***	-90.819 (13.71)***	-70.801 (14.06)***	-97.525 (12.25)***	-57.996 (5.22)***	-64.143 (7.46)***	-72.439 (12.04)***
Constant	73.134 (7.69)***	269.830 (26.65)***	138.025 (6.97)***	-225.425 (17.72)***	193.267 (9.91)***	5.316 (0.40)	28.677 (0.95)	98.061 (3.07)***	306.550 (12.27)***
Observations	30965	23749	7216	14036	16929	10504	3532	3684	13245
R-Square	0.78	0.80	0.70	0.74	0.70	0.76	0.61	0.67	0.67

[1] Absolute value of t-statistics in parentheses/ * significant at 10% level **significant at 5% level *** significant at 1% level

APPENDIX E

REGRESSION RESULTS AFTER MODIFICATIONS

REGRESSION RESULTS AFTER MODIFICATIONS: SOEP, NO CONTROLS

DEPENDENT VARIABLE: MONTHLY PUBLIC PENSION BENEFIT	TOTAL	TOTAL WEST	TOTAL EAST	TOTAL MEN	TOTAL WOMEN	MEN WEST	MEN EAST	WOMEN WEST	WOMEN EAST
Years in School (Topcode)	35.346 (6.07)**	38.135 (5.03)**	24.194 (3.18)**	37.354 (4.00)**	20.407 (2.64)**	44.559 (3.48)**	13.578 (1.27)	19.370 (1.98)*	33.343 (2.88)**
Years in Training (Topcode)	31.312 (4.17)**	28.653 (3.06)**	32.149 (2.94)**	16.527 (1.38)	42.474 (4.61)**	24.836 (1.56)	2.891 (0.18)	36.861 (3.22)**	59.432 (4.24)**
Years in Employment	13.353 (7.99)**	12.625 (6.17)**	14.826 (5.43)**	3.468 (0.71)	12.060 (7.59)**	4.200 (0.63)	11.840 (2.04)*	12.220 (6.48)**	13.967 (4.90)**
Years in Unemployment	-12.894 (2.95)**	-18.604 (3.10)**	-5.629 (1.05)	-46.251 (4.90)**	-3.895 (0.81)	-47.075 (3.83)**	-27.002 (2.12)*	-0.315 (0.04)	2.910 (0.50)
Years in Homeproduction	-10.200 (2.70)**	-7.720 (1.71)	8.880 (0.53)		-6.590 (1.89)			-9.618 (2.45)*	
Retired (Dummy) if years in retirement > 4	58.941 (0.85)	8.632 (0.10)	170.575 (1.99)*	-232.418 (1.75)	142.457 (1.98)*	-340.300 (1.93)	56.282 (0.35)	114.260 (1.23)	174.965 (1.79)
Other (Dummy) if years in other activities > 3	-94.753 (1.28)	-84.926 (1.01)	-8.716 (0.04)	-431.427 (2.25)*	-29.606 (0.43)	-376.950 (1.69)	0.000 (0.00)**	-23.229 (0.30)	72.744 (0.33)
Missing (Dummy) if years missing > 3	-20.311 (0.65)	-12.407 (0.28)	-74.798 (2.13)*	-94.195 (1.74)	-11.565 (0.34)	-105.246 (1.33)	-42.620 (0.76)	41.066 (0.89)	-81.432 (1.90)
Worked as Civil Servant or Self-Employed (Dummy)	-302.669 (7.66)**	-351.942 (7.07)**	-141.289 (2.55)*	-341.986 (6.56)**	-151.734 (2.57)*	-392.859 (5.88)**	-212.287 (3.04)**	-179.715 (2.43)*	-0.089 (0.00)
Educational Attainment: Low Reference Category: Intermediate	-54.301 (1.64)	-55.801 (1.36)	-46.355 (0.94)	-98.479 (1.55)	-32.320 (0.94)	-96.917 (1.18)	-48.785 (0.59)	-50.987 (1.23)	-17.064 (0.29)
Educational Attainment: High Reference Category: Intermediate	162.520 (4.27)**	147.190 (2.71)**	212.110 (5.05)**	139.762 (2.70)**	219.483 (4.17)**	155.443 (2.12)*	168.495 (3.02)**	92.197 (1.16)	295.065 (4.78)**
Educational Attainment: Missing Reference Category: Intermediate	-94.716 (0.69)	-123.466 (0.75)	138.267 (0.60)	-231.194 (1.31)	0.536 (0.00)	-231.010 (1.01)	60.772 (0.26)	-4.455 (0.02)	0.000 (0.00)**
Educational Attainment: Unknown Reference Category: Intermediate	-130.087 (1.74)	-124.274 (1.49)		-189.058 (1.54)	-42.205 (0.50)	-173.114 (1.25)	0.000 (0.00)**	-62.835 (0.68)	0.000 (0.00)**
Gender (Female = 1)	-331.502 (12.35)**	-431.668 (11.27)**	-197.502 (6.67)**						
West (West = 1)	119.825 (4.48)**			229.683 (6.11)**	12.727 (0.37)				
Receives Civil Servant Pension (Dummy)				-705.688 (6.14)**		-819.666 (5.68)**			
Receives Private Pension (Dummy)				-203.679 (2.33)*		-205.209 (1.97)*			
Constant	458.298 (5.41)**	659.207 (6.36)**	312.191 (2.49)*	921.928 (3.94)**	171.991 (2.14)*	1094.230 (3.34)**	586.616 (2.22)*	197.825 (2.15)*	20.804 (0.16)
Observations	949	662	287	443	506	304	139	358	148
R-squared	0.58	0.59	0.57	0.40	0.42	0.39	0.36	0.36	0.56

REGRESSION RESULTS AFTER MODIFICATIONS: SOEP (WITHOUT PRIVATE AND CIVIL SERVANT PENSIONS), NO ADDITIONAL CONTROLS

DEPENDENT VARIABLE: VP10301 MONTHLY PUBLIC PENSION BENEFIT	TOTAL	TOTAL WEST	TOTAL EAST	TOTAL MEN	TOTAL WOMEN	MEN WEST	MEN EAST	WOMEN WEST	WOMEN EAST
Years in School (Topcode)	44.250 (7.48)**	42.099 (5.57)**	39.580 (4.86)**	41.944 (4.39)**	34.030 (4.23)**	43.860 (3.36)**	30.186 (2.72)**	25.002 (2.48)*	48.568 (3.79)**
Years in Training (Topcode)	34.034 (4.49)**	30.739 (3.27)**	38.997 (3.45)**	15.670 (1.26)	45.912 (4.97)**	15.522 (0.94)	12.378 (0.74)	39.987 (3.56)**	62.686 (4.14)**
Years in Employment	13.496 (8.02)**	12.691 (6.19)**	16.250 (5.80)**	7.825 (1.54)	11.972 (7.48)**	6.300 (0.90)	13.886 (2.33)*	12.116 (6.42)**	16.589 (5.70)**
Years in Unemployment	-13.684 (3.10)**	-17.480 (2.86)**	-7.263 (1.32)	-37.122 (3.71)**	-6.481 (1.34)	-39.569 (3.02)**	-26.005 (1.97)	-1.127 (0.16)	0.786 (0.13)
Years in Homeproduction	-10.580 (2.79)**	-8.015 (1.78)	4.110 (0.24)		-7.521 (2.16)*			-10.242 (2.62)**	
Retired (Dummy)	70.113 (1.00)	15.140 (0.17)	150.334 (1.67)	-173.955 (1.25)	136.472 (1.86)	-275.144 (1.49)	45.846 (0.27)	115.205 (1.23)	107.443 (1.01)
Other (Dummy)	-109.370 (1.48)	-98.603 (1.18)	-6.405 (0.03)	-429.380 (2.14)*	-39.679 (0.57)	-412.433 (1.76)	DROPPE D	-25.666 (0.34)	62.339 (0.28)
Missing (Dummy)	-27.863 (0.89)	-15.039 (0.34)	-79.148 (2.18)*	-87.926 (1.55)	-19.078 (0.56)	-97.922 (1.18)	-44.531 (0.76)	41.842 (0.90)	-84.678 (1.93)
Worked as Civil Servant or Self-Employed (Dummy)	-296.824 (7.46)**	-344.948 (6.92)**	-136.585 (2.38)*	-384.271 (7.09)**	-154.646 (2.59)**	-438.296 (6.26)**	-206.449 (2.85)**	-179.291 (2.41)*	9.533 (0.10)
Educational Attainment: Low Reference Category: Intermediate	-50.431 (1.49)	-52.679 (1.26)	-55.983 (1.10)	-113.741 (1.70)	-20.059 (0.57)	-117.760 (1.35)	-72.646 (0.85)	-40.652 (0.96)	-28.647 (0.46)
Educational Attainment: High Reference Category: Intermediate	87.280 (1.79)	128.102 (2.00)*	82.280 (1.31)	67.772 (1.06)	67.578 (0.84)	118.590 (1.37)	29.533 (0.40)	17.606 (0.17)	206.773 (1.73)
Educational Attainment: Missing Reference Category: Intermediate	64.348 (0.88)	-99.095 (0.84)	211.290 (2.91)**	-94.741 (0.84)	232.144 (2.63)**	-227.085 (1.21)	53.241 (0.53)	16.141 (0.12)	401.257 (3.90)**
Educational Attainment: Unknown Reference Category: Intermediate	-93.645 (1.86)	-81.985 (1.45)	-125.925 (0.54)	-221.699 (2.59)**	23.266 (0.41)	-220.096 (2.25)*		10.525 (0.17)	-67.248 (0.30)
Sex – Female = 1 (Dummy)	-332.450 (12.28)**	-432.380 (11.29)**	-193.994 (6.27)**						
West – West = 1	106.753 (3.97)**			212.988 (5.43)**	-11.133 (0.32)				
Constant	465.900 (5.46)**	657.393 (6.32)**	263.034 (2.04)*	754.023 (3.10)**	188.411 (2.34)*	1032.331 (3.04)**	499.318 (1.84)	190.462 (2.07)*	-56.540 (0.42)
Observations	949	662	287	443	506	304	139	358	148
R-squared	0.57	0.59	0.54	0.34	0.41	0.32	0.32	0.36	0.55

REGRESSION RESULTS AFTER MODIFICATIONS: SUF VVL 2004, NO ADDITIONAL CONTROLS

DEPENDENT VARIABLE: MONTHLY PUBLIC PENSION BENEFIT (LOG)	TOTAL	TOTAL WEST	TOTAL EAST	TOTAL MEN	TOTAL WOMEN	WEST MEN	EAST MEN	WEST WOMEN	EAST WOMEN
Years in School (Topcode)	36.738 (42.39)**	27.999 (26.01)**	53.514 (39.92)**	43.477 (37.20)**	35.416 (29.53)**	37.347 (25.92)**	53.417 (28.32)**	26.772 (17.94)**	56.823 (29.40)**
Years in Training (Topcode)	23.907 (23.70)**	23.457 (19.59)**	18.805 (10.66)**	41.755 (27.46)**	4.547 (3.75)**	44.205 (24.85)**	28.721 (10.29)**	4.359 (3.05)**	10.483 (4.83)**
Years in Employment	25.214 (165.44)**	25.247 (147.92)**	22.627 (58.50)**	30.046 (138.19)**	20.102 (105.08)**	30.014 (123.99)**	24.910 (40.84)**	20.047 (93.68)**	21.084 (43.69)**
Years in Unemployment	6.579 (13.54)**	6.934 (11.42)**	2.856 (3.78)**	11.770 (15.67)**	4.473 (7.94)**	12.964 (14.37)**	2.821 (2.13)*	6.493 (8.98)**	3.304 (3.79)**
Years in Homeproduction	9.703 (29.14)**	11.070 (29.83)**	3.748 (4.18)**		2.514 (7.65)**			2.705 (7.60)**	1.432 (1.55)
Retired (Dummy)	147.405 (7.20)**	134.762 (5.37)**	190.753 (6.15)**	209.226 (7.02)**	81.895 (3.28)**	170.889 (4.70)**	307.068 (6.56)**	92.148 (3.02)**	70.571 (1.77)
Other (Dummy)	-37.382 (4.75)**	-33.788 (3.86)**	-73.801 (4.01)**	-62.771 (1.57)	-50.971 (7.58)**	-82.847 (1.68)	-10.038 (0.16)	-49.106 (6.67)**	-83.670 (4.76)**
Missing (Dummy)	-120.168 (21.41)**	-147.032 (21.11)**	-72.427 (8.67)**	-58.608 (9.24)**	-110.073 (6.18)**	-73.517 (9.40)**	-47.086 (4.76)**	-124.925 (5.19)**	-79.062 (3.34)**
Educational Attainment: Low Reference Category: Intermediate	-128.253 (22.73)**	-141.582 (21.73)**	-68.926 (6.31)**	-159.917 (15.91)**	-77.006 (13.02)**	-172.511 (15.09)**	-100.136 (4.76)**	-80.590 (11.82)**	-51.233 (4.32)**
Educational Attainment: High Reference Category: Intermediate	201.585 (28.08)**	222.587 (22.49)**	183.626 (20.51)**	152.399 (16.47)**	195.338 (17.53)**	162.829 (13.44)**	171.279 (13.39)**	149.346 (8.09)**	181.670 (14.34)**
Educational Attainment: Missing Reference Category: Intermediate	-129.566 (36.99)**	-134.930 (32.03)**	-100.023 (17.75)**	-208.322 (39.40)**	-72.479 (17.40)**	-228.653 (36.03)**	-128.653 (14.64)**	-70.015 (14.09)**	-79.887 (11.17)**
Educational Attainment: Unknown Reference Category: Intermediate	-108.725 (23.75)**	-120.558 (21.75)**	-67.889 (9.44)**	-120.235 (16.56)**	-87.901 (17.00)**	-132.011 (14.90)**	-67.158 (5.85)**	-93.385 (15.12)**	-73.947 (8.42)**
Gender (Female = 1)	-198.275 (55.85)**	-229.388 (51.26)**	-129.814 (25.11)**						
West (West = 1)	144.817 (43.91)**			200.766 (41.57)**	53.087 (13.11)**				
Constant	118.616 (11.59)**	307.647 (27.82)**	126.633 (6.06)**	-157.161 (11.23)**	141.170 (6.88)**	67.117 (4.51)**	15.000 (0.47)	209.057 (7.97)**	62.570 (1.86)
Number of Observations	30829	23656	7173	13983	16846	10463	3520	13193	3653
R^2	0.78	0.80	0.70	0.74	0.70	0.76	0.61	0.67	0.67

APPENDIX F

COMPARISON OF OBSERVATIONS & PREDICTIONS SOEP

Density Plots with two lines for all nine models:

First Graph – Observed vs. Predicted without Randomly Assigned Residuals

- solid line → observed
- dashed line → prediction without randomly assigned residuals

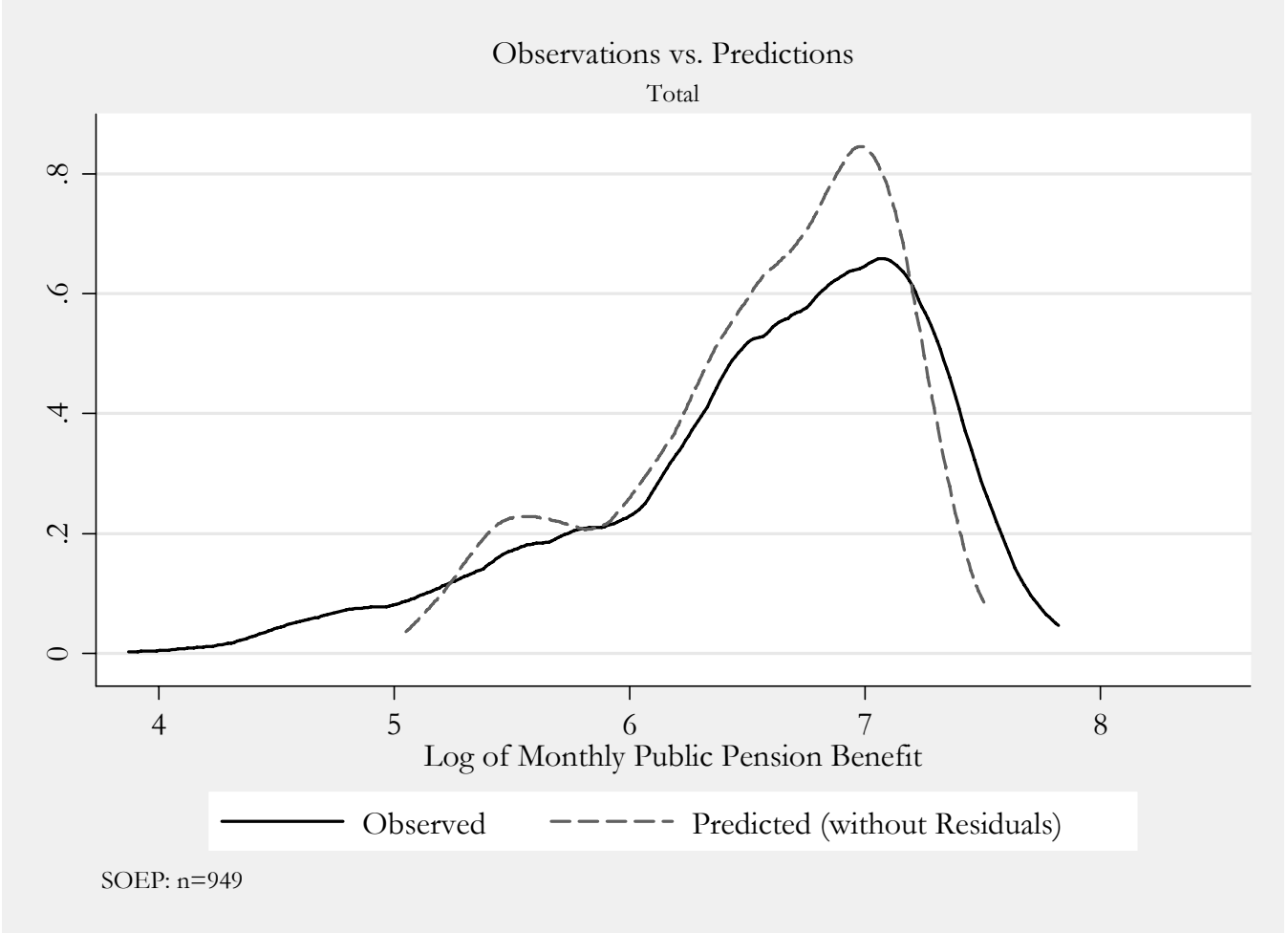
Second Graph – Observed vs. Predicted with Randomly Assigned Residuals

- solid line → observed
- dashed line → prediction with randomly assigned residuals

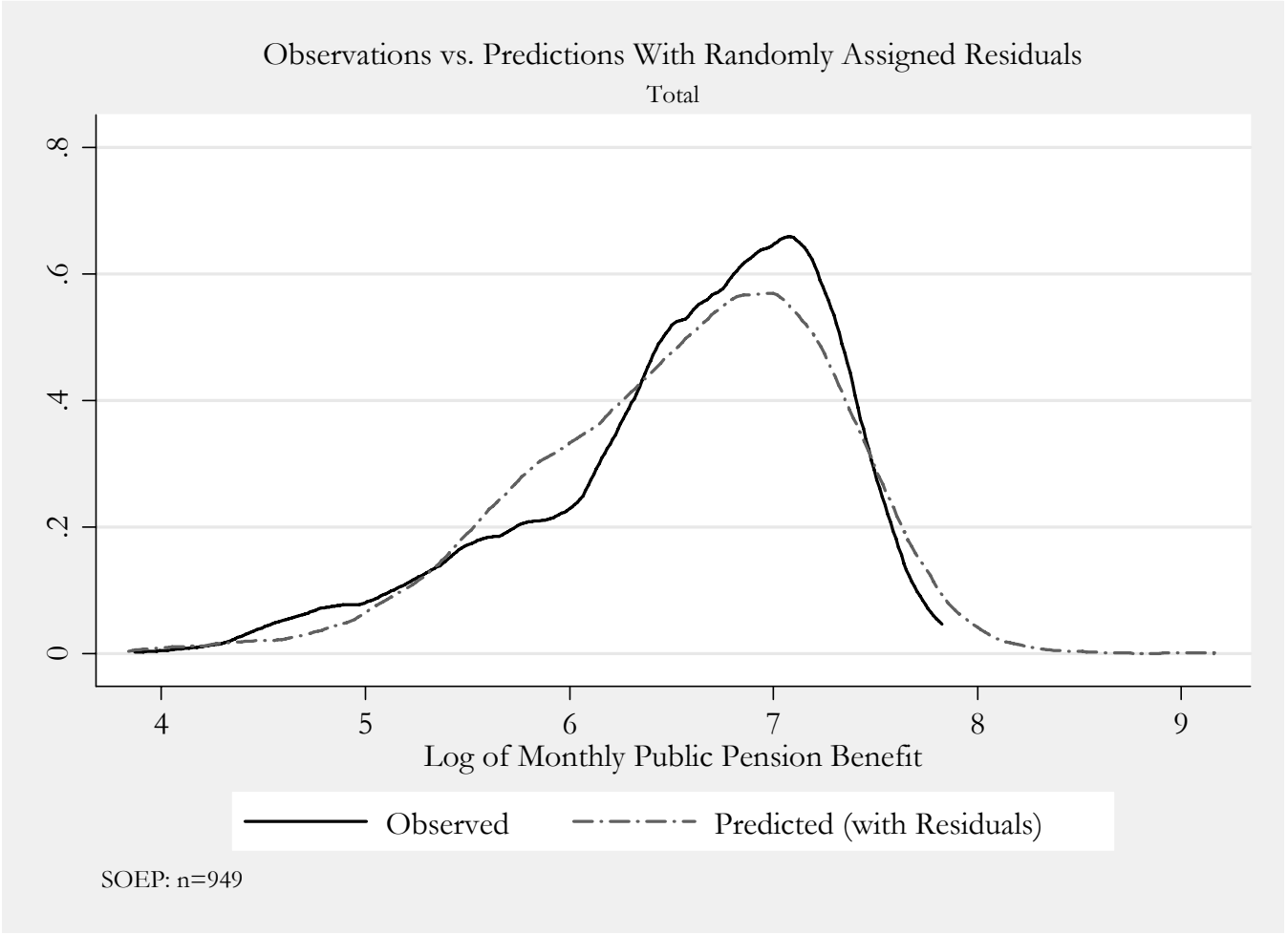
Third Graph – Confidence Bands

- solids line → upper and lower confidence band of observations
- dashed line → prediction with randomly assigned residuals

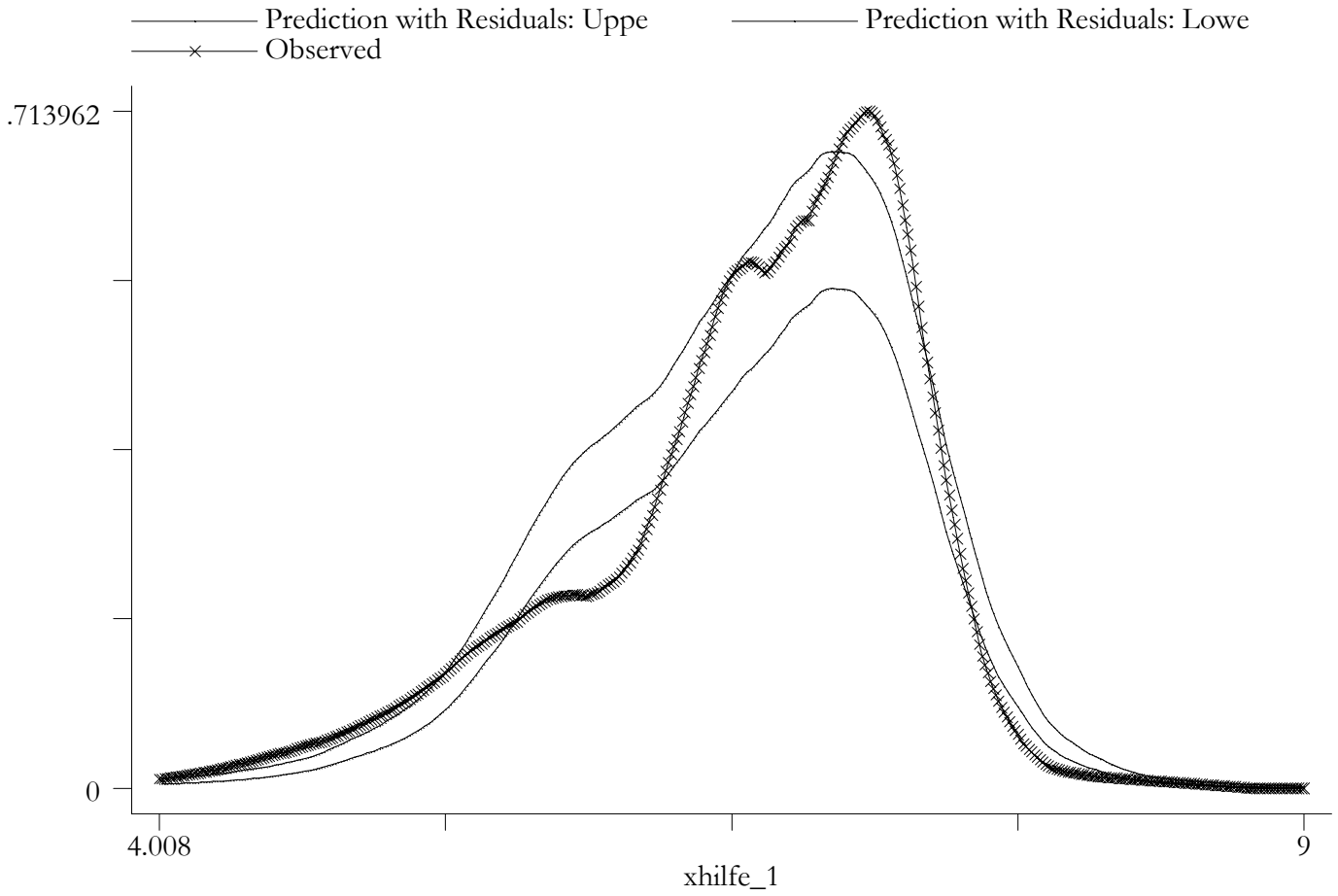
TOTAL – FIRST TIME PENSIONERS SOEP 2000 – 2004
Observed vs. Predicted without Residuals



TOTAL – FIRST TIME PENSIONERS SOEP 2000 – 2004
Observed vs. Predicted with Residuals

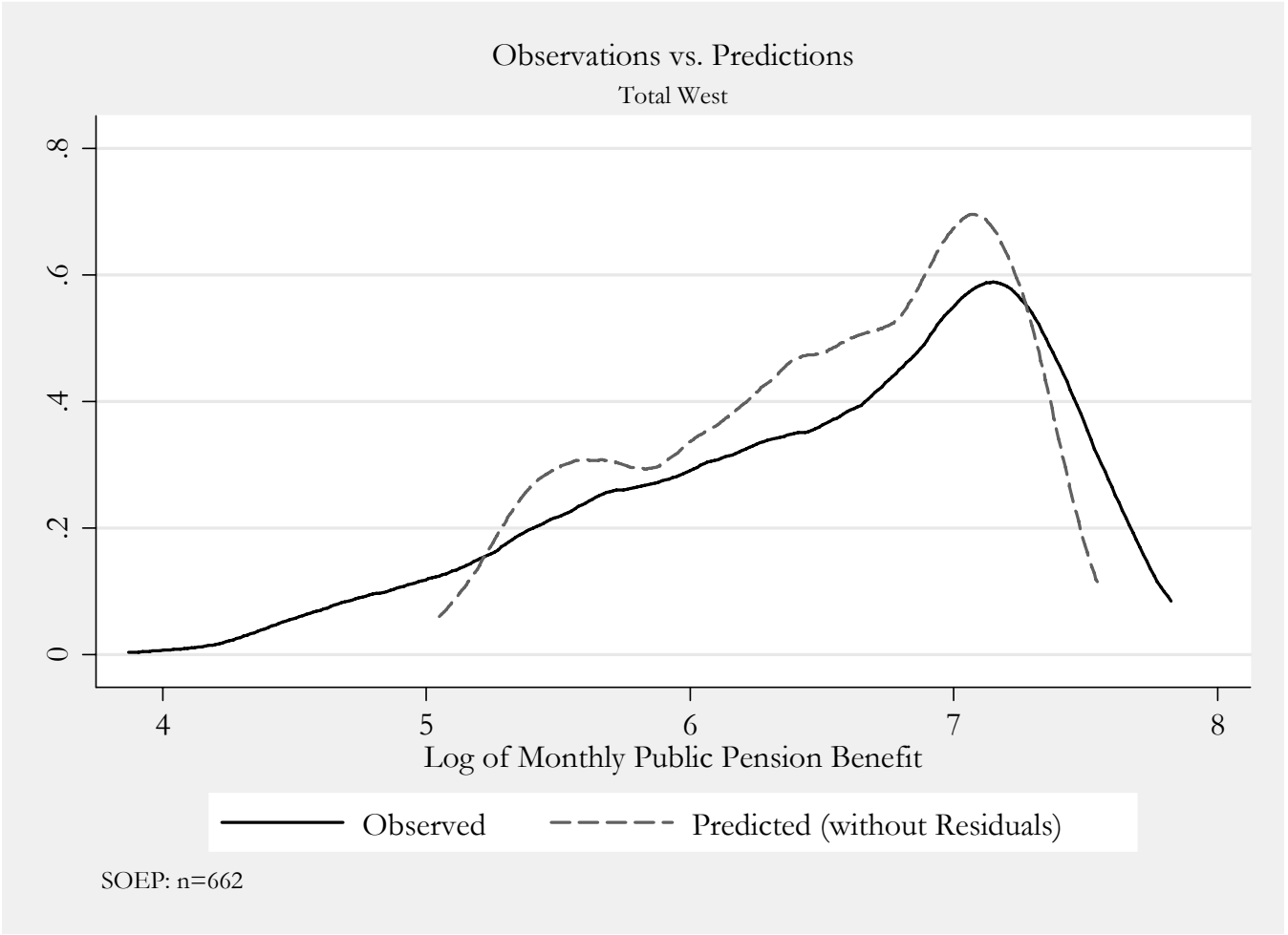


TOTAL – FIRST TIME PENSIONERS SOEP 2000 – 2004
Confidence Bands

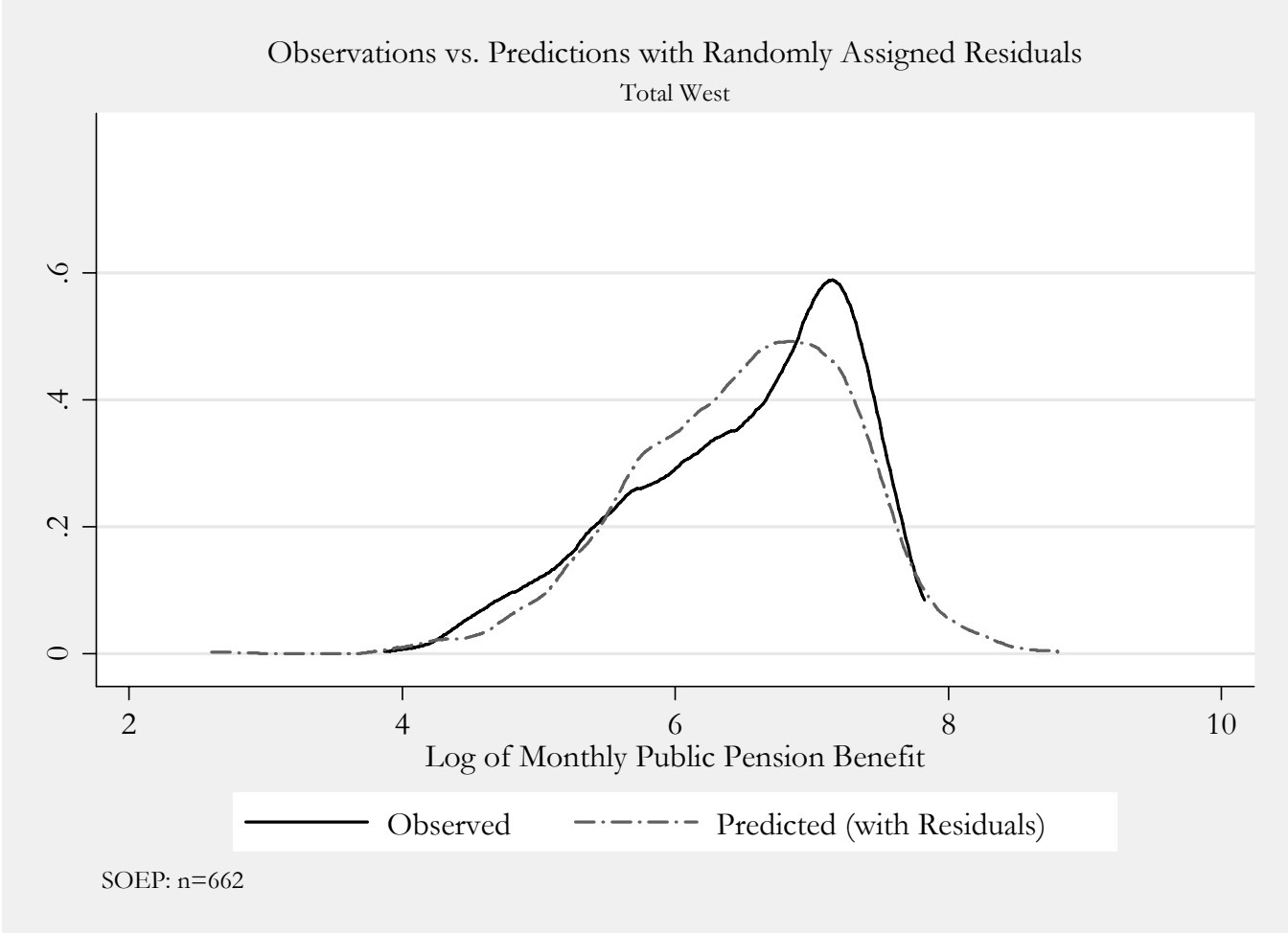


TOTAL WEST – FIRST TIME PENSIONERS SOEP 2000 – 2004

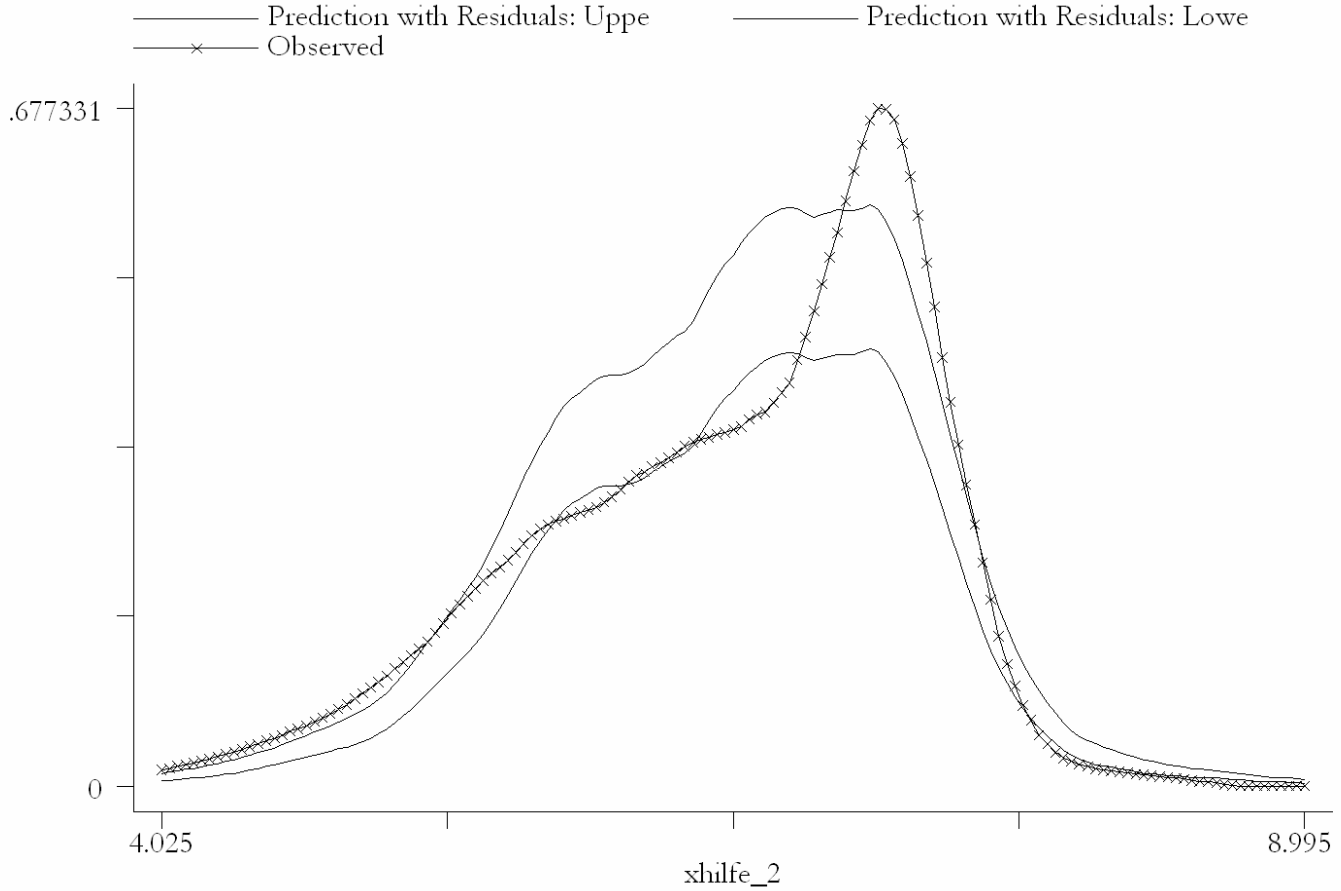
Observed vs. Predicted without Residuals



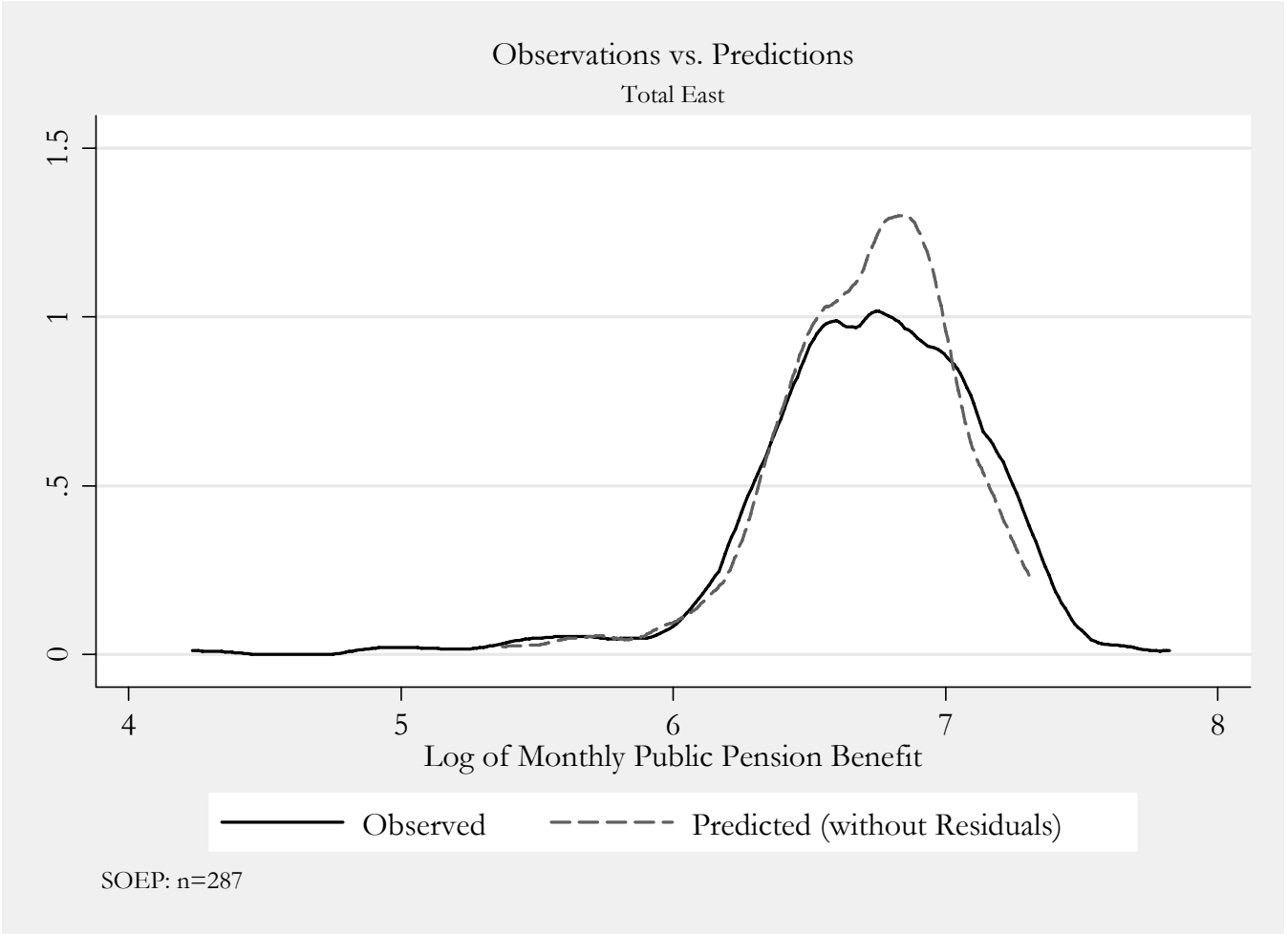
TOTAL WEST – FIRST TIME PENSIONERS SOEP 2000 – 2004
Observed vs. Predicted with Residuals



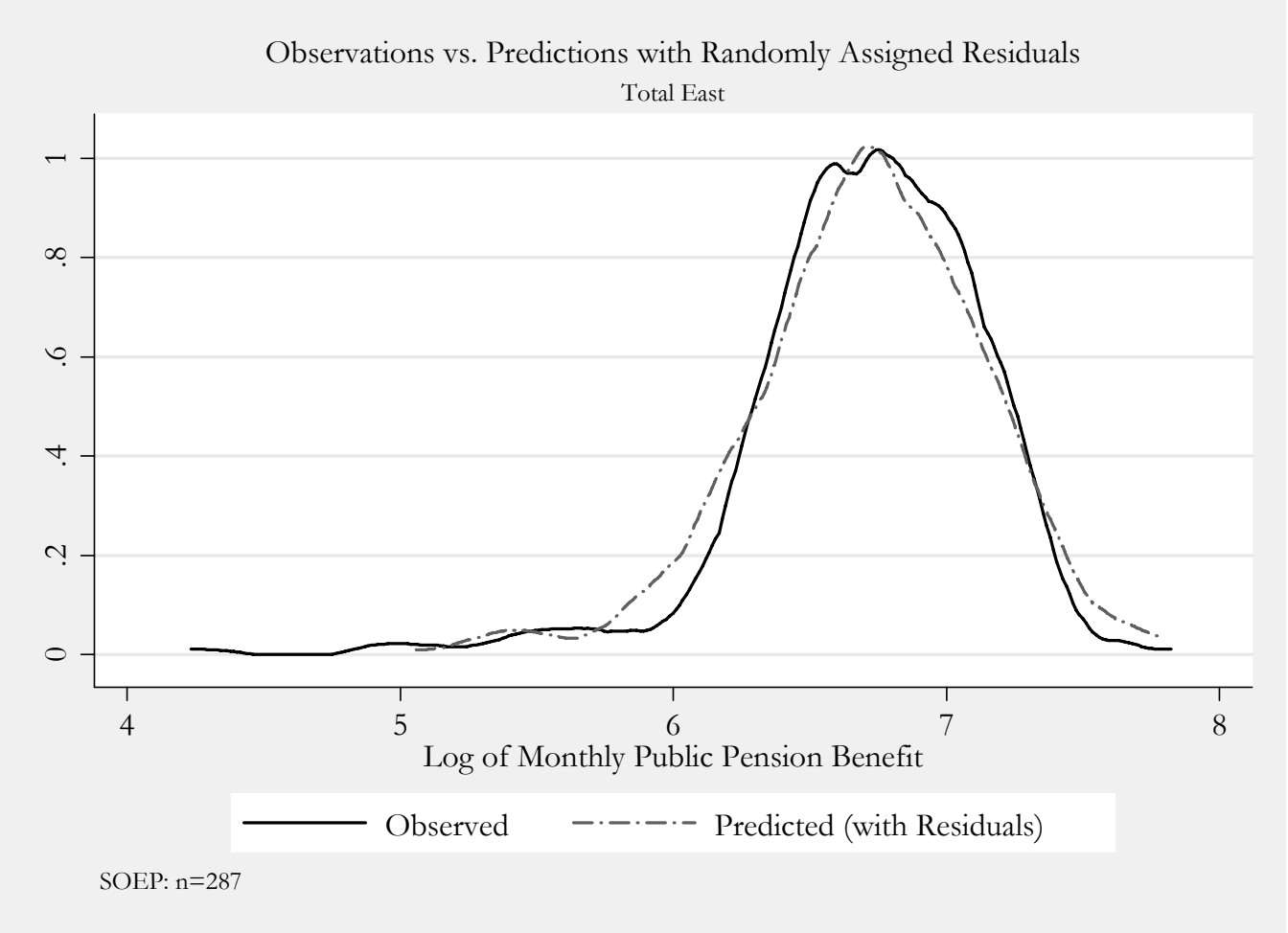
TOTAL WEST – FIRST TIME PENSIONERS SOEP 2000 – 2004
Confidence Bands



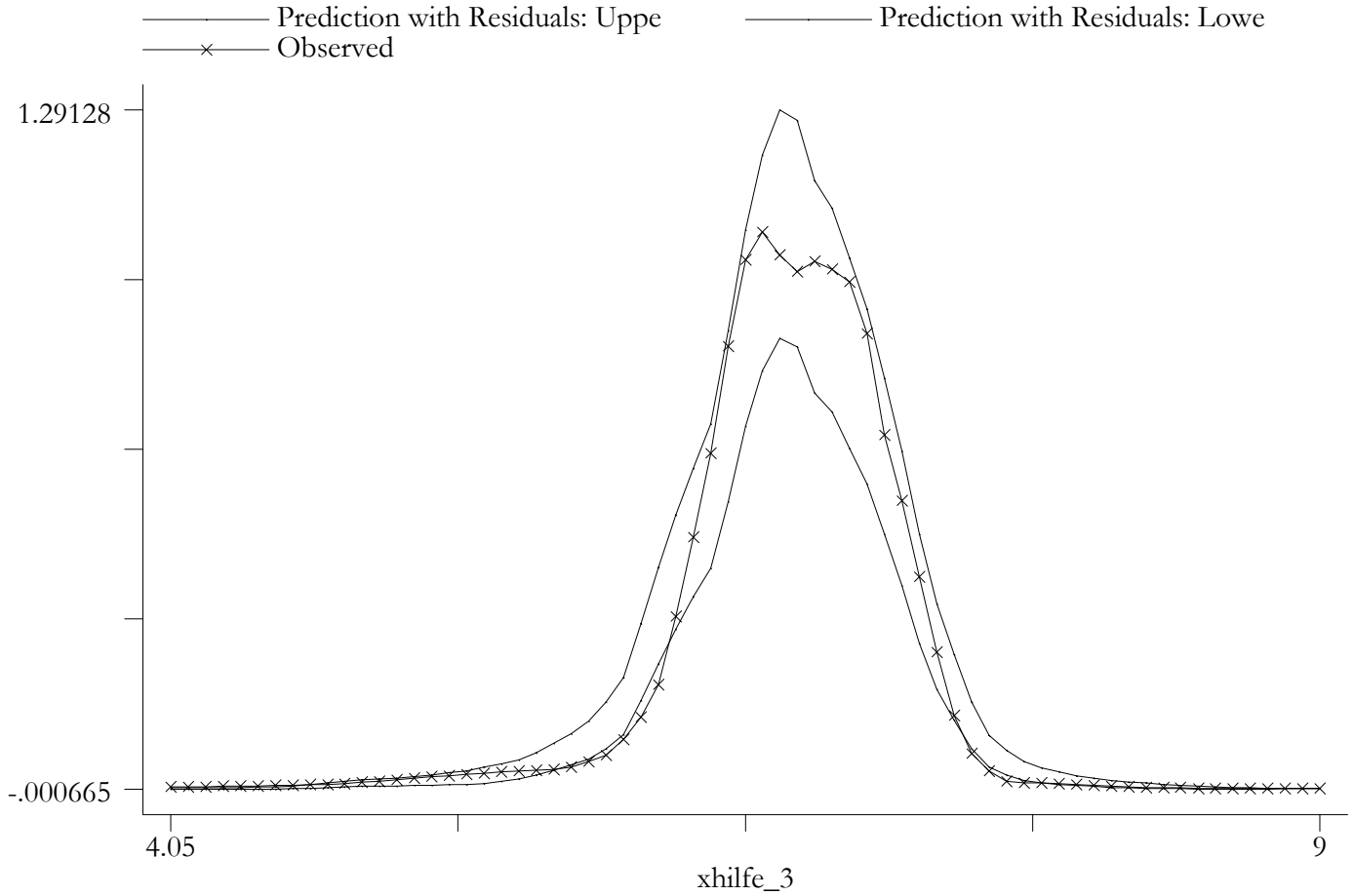
TOTAL EAST – FIRST TIME PENSIONERS SOEP 2000 – 2004
Observed vs. Predicted without Residuals



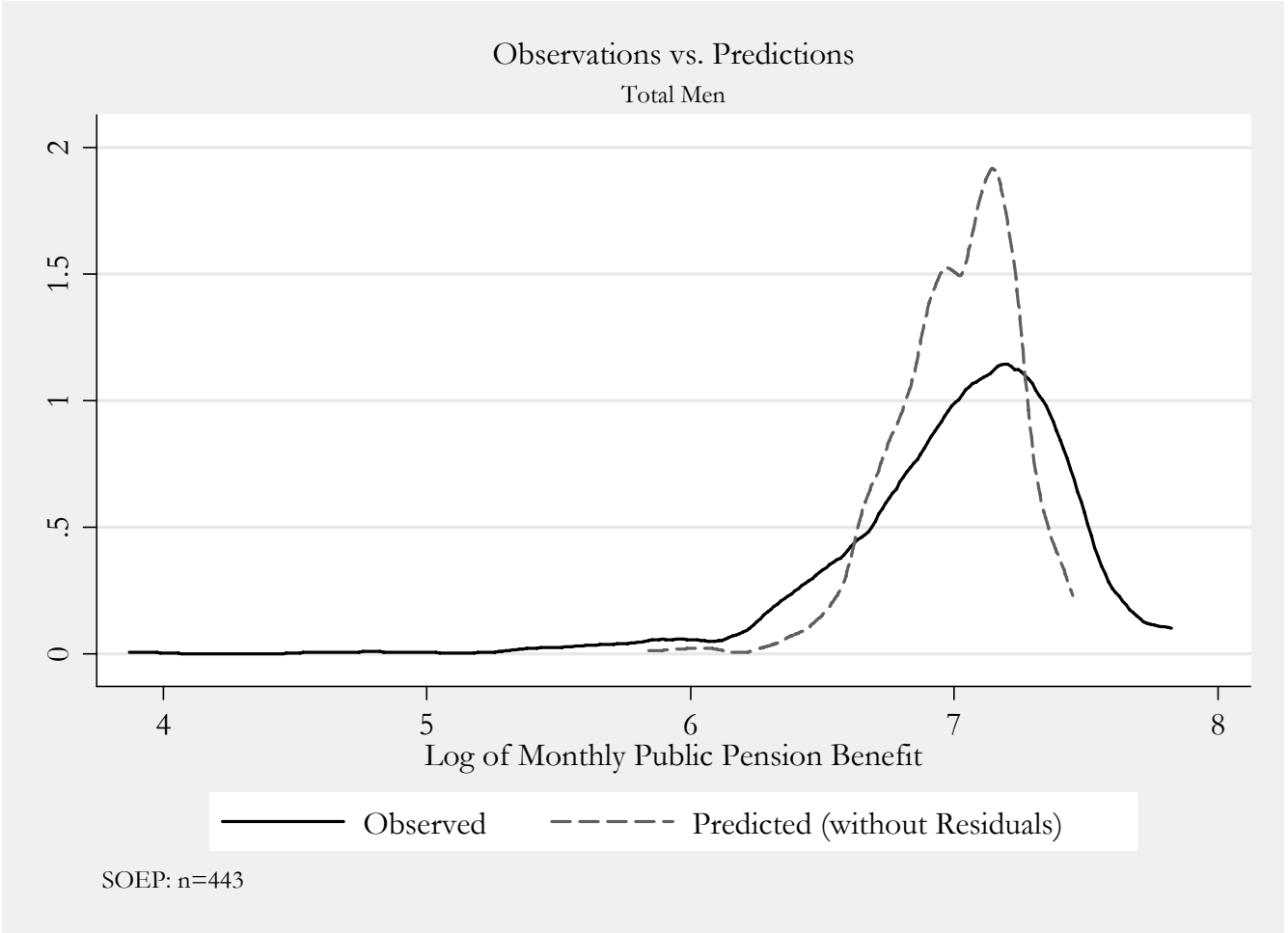
TOTAL EAST – FIRST TIME PENSIONERS SOEP 2000 – 2004
Observed vs. Predicted with Residuals



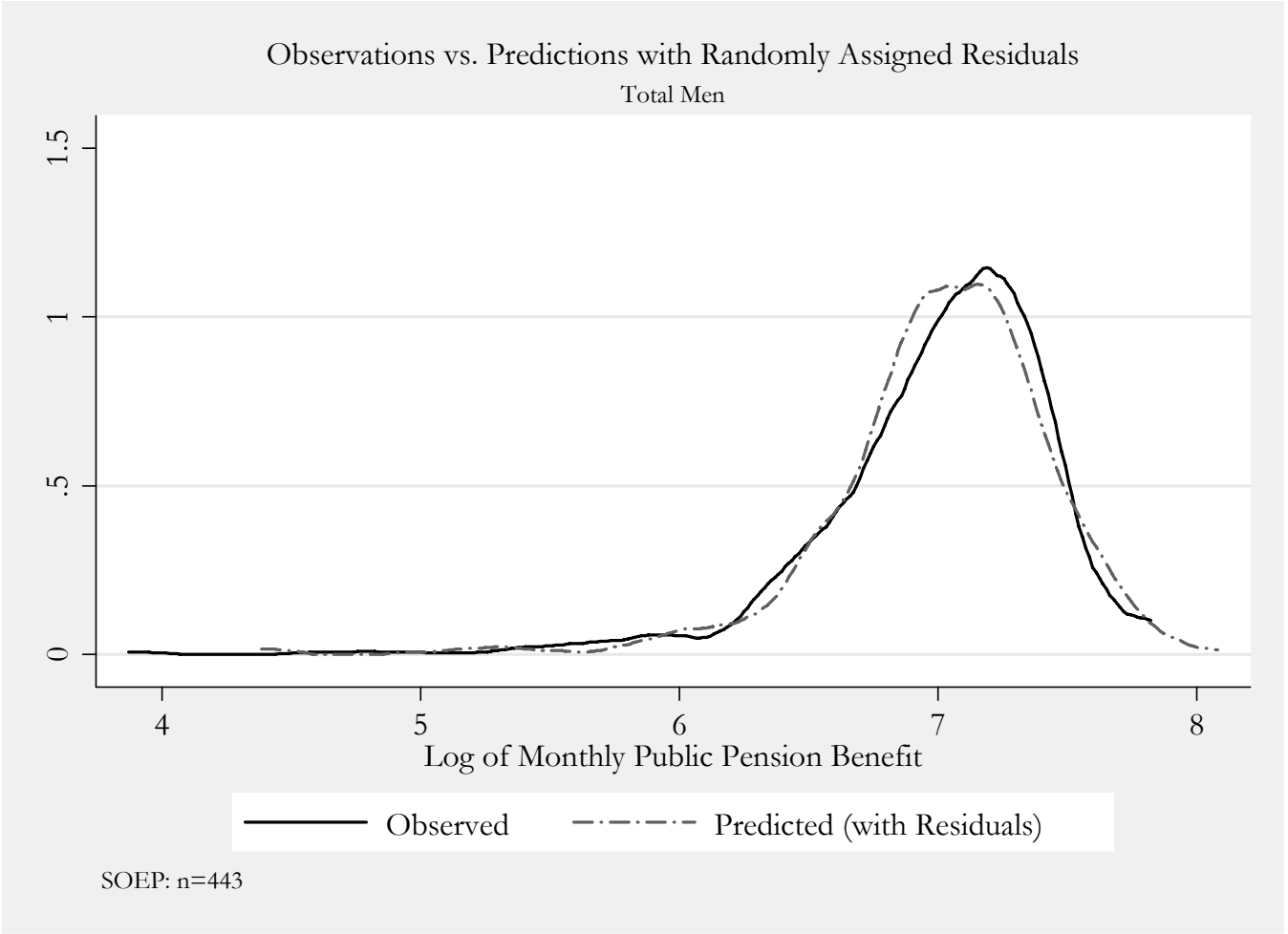
TOTAL EAST – FIRST TIME PENSIONERS SOEP 2000 – 2004
Confidence Bands



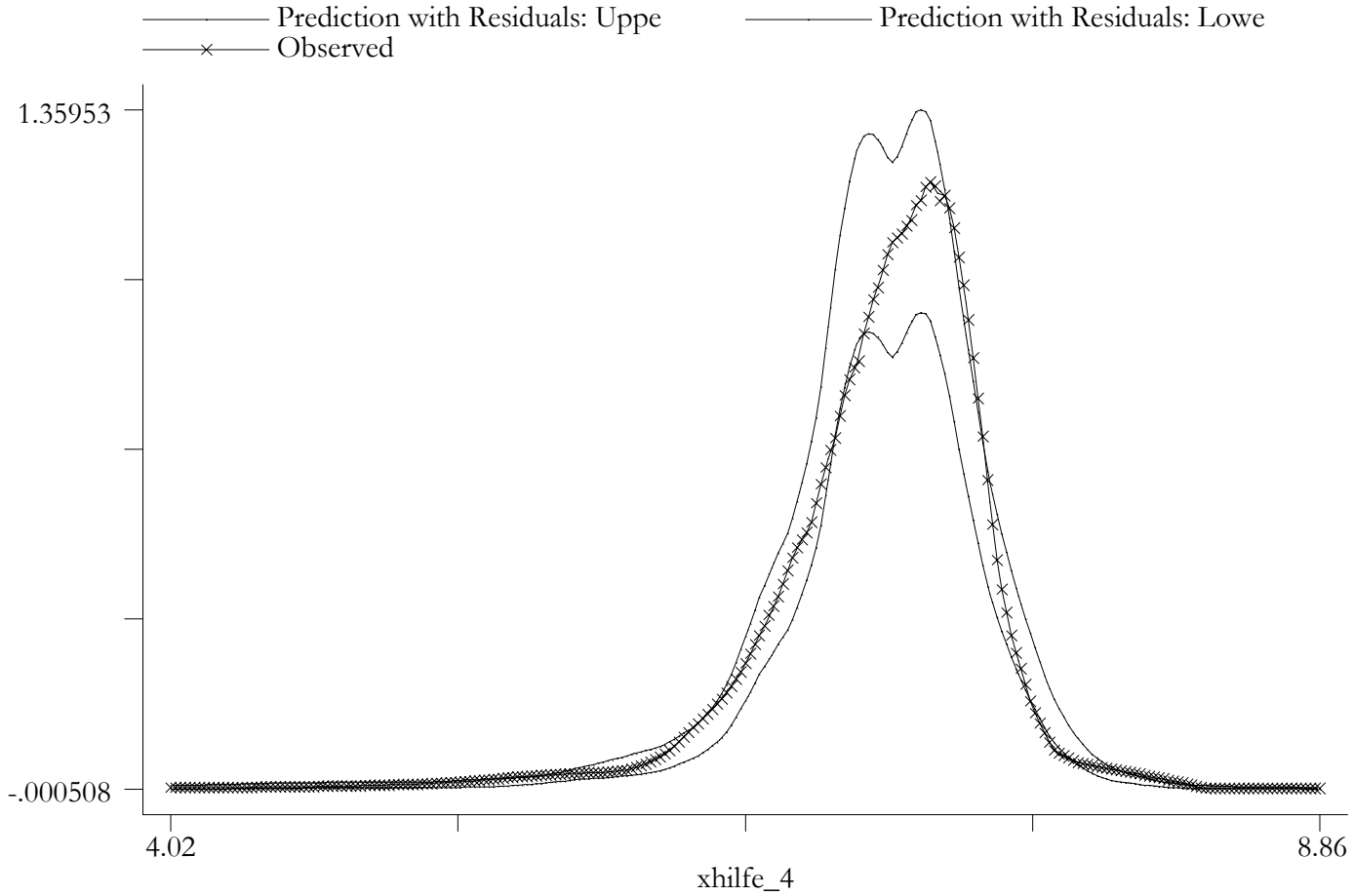
TOTAL MEN – FIRST TIME PENSIONERS SOEP 2000 – 2004
Observed vs. Predicted without Residuals



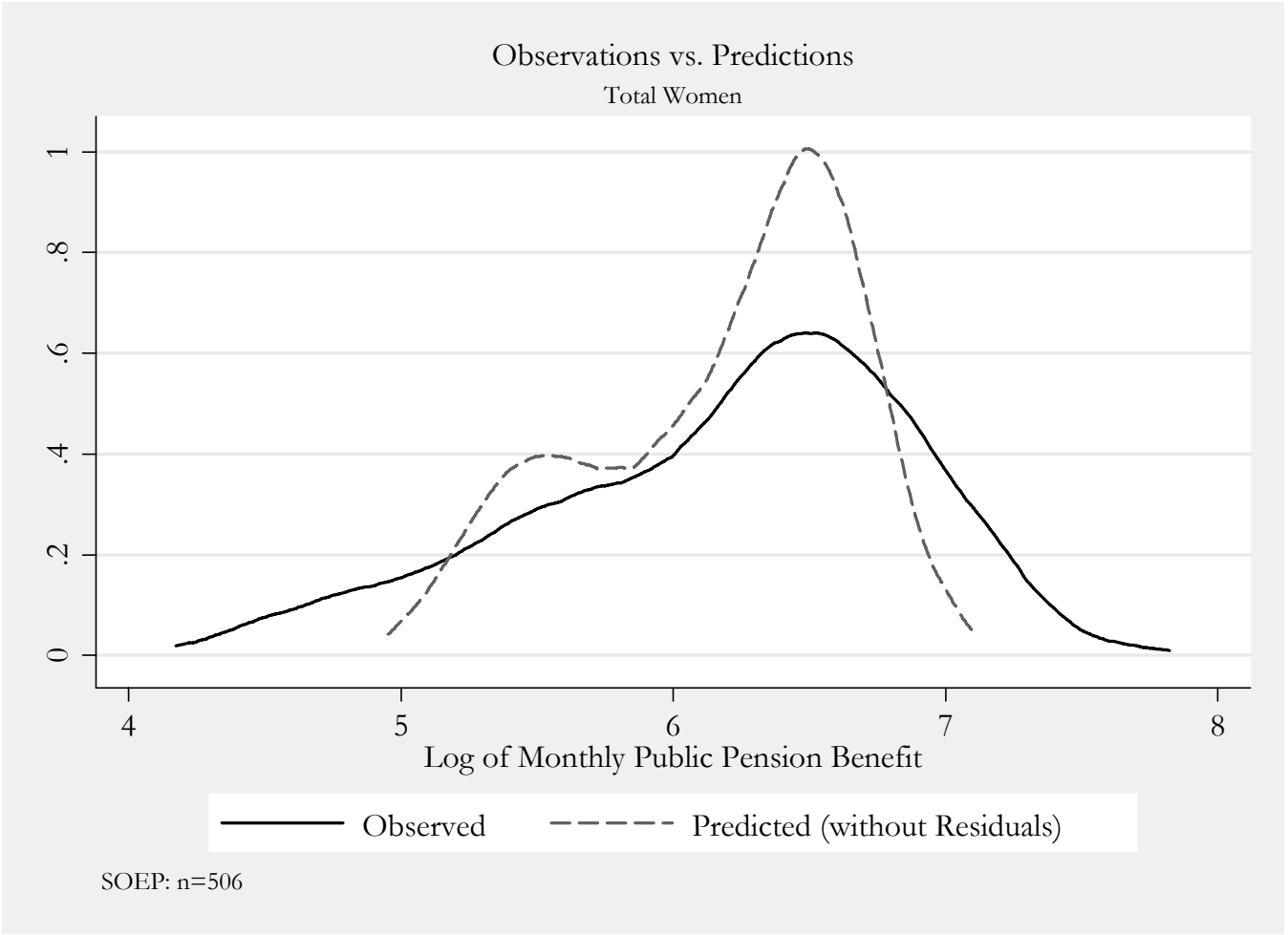
TOTAL MEN – FIRST TIME PENSIONERS SOEP 2000 – 2004
Observed vs. Predicted with Residuals



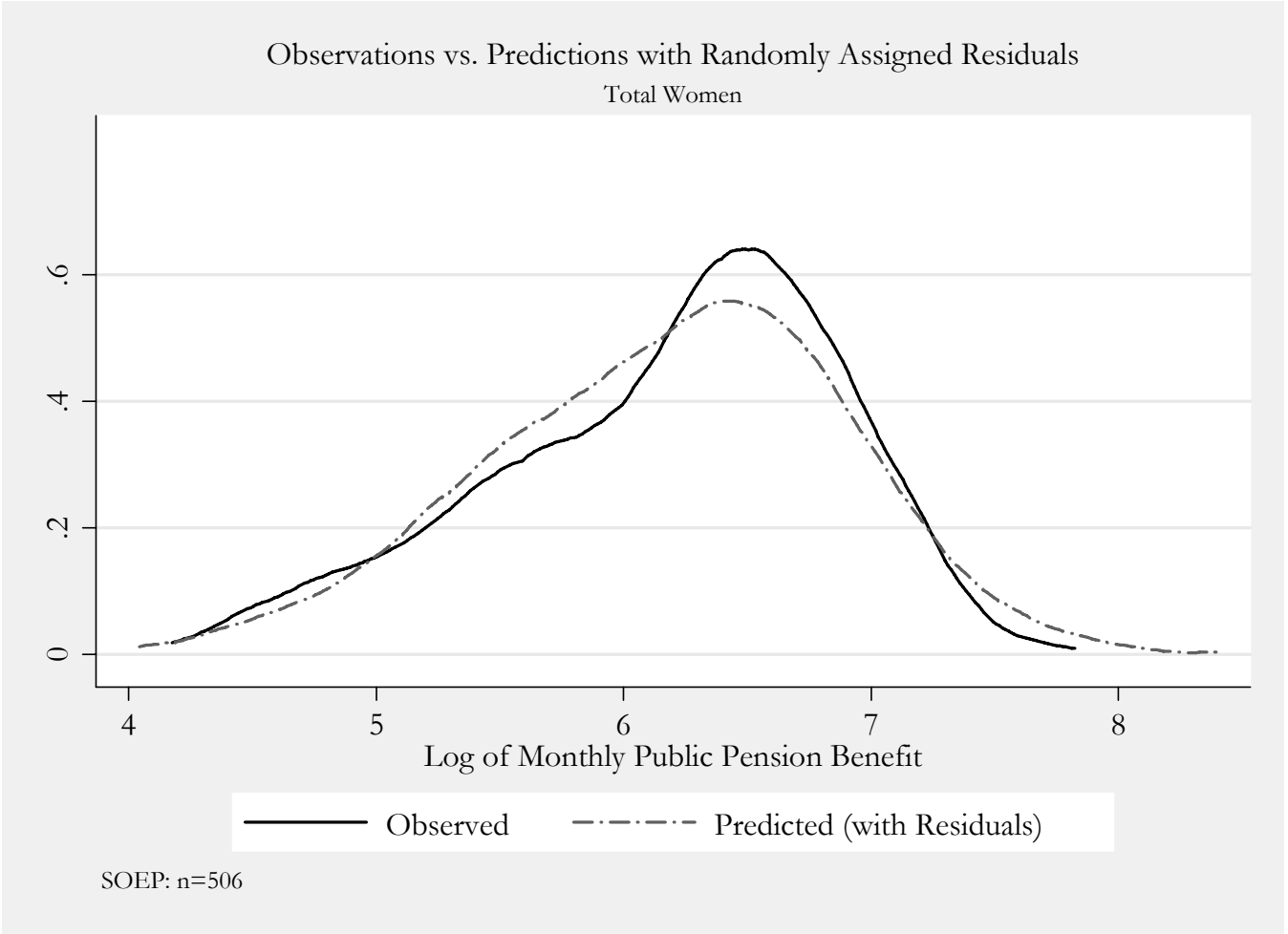
TOTAL MEN – FIRST TIME PENSIONERS SOEP 2000 – 2004
Confidence Bands



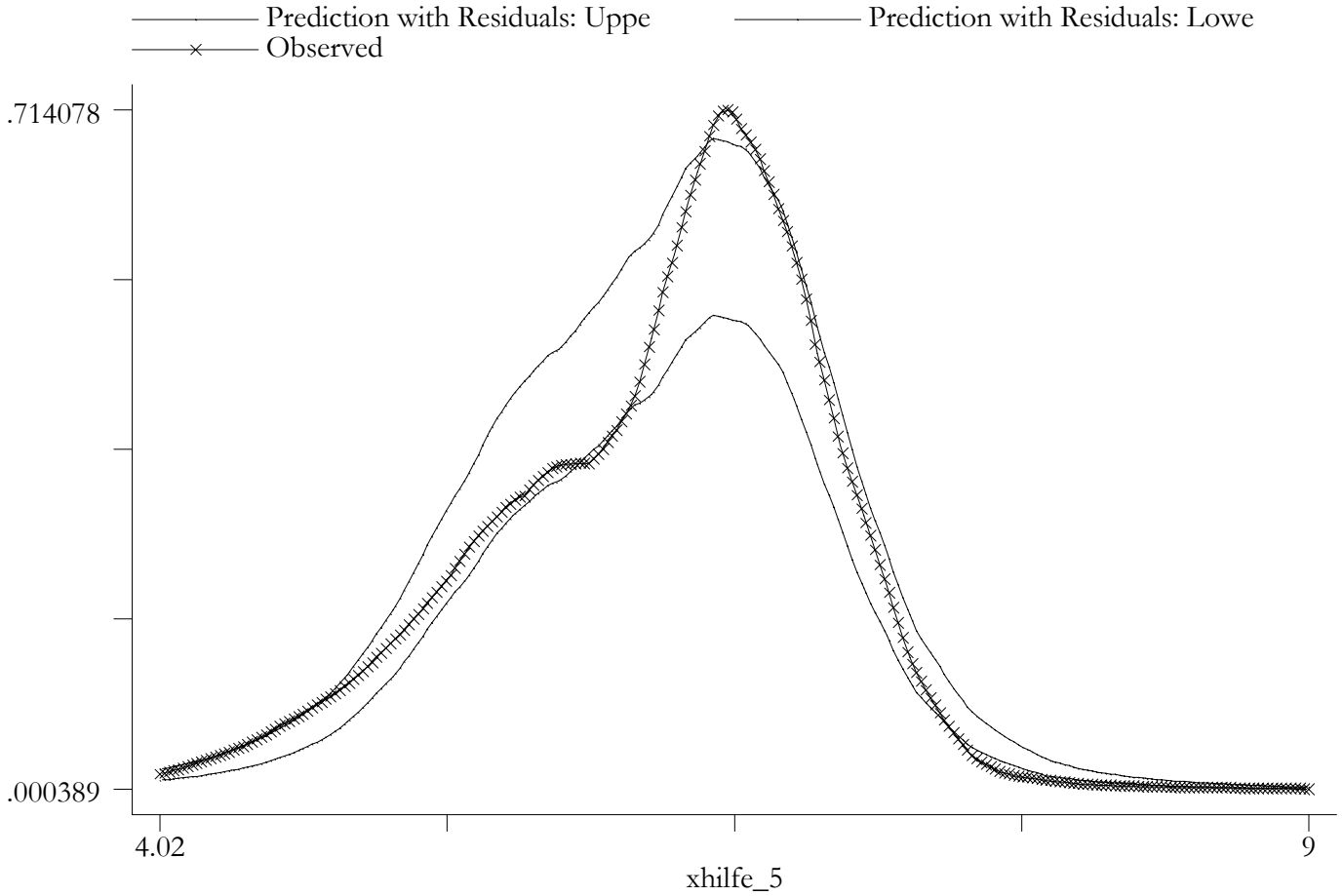
TOTAL WOMEN – FIRST TIME PENSIONERS SOEP 2000 – 2004
Observed vs. Predicted without Residuals



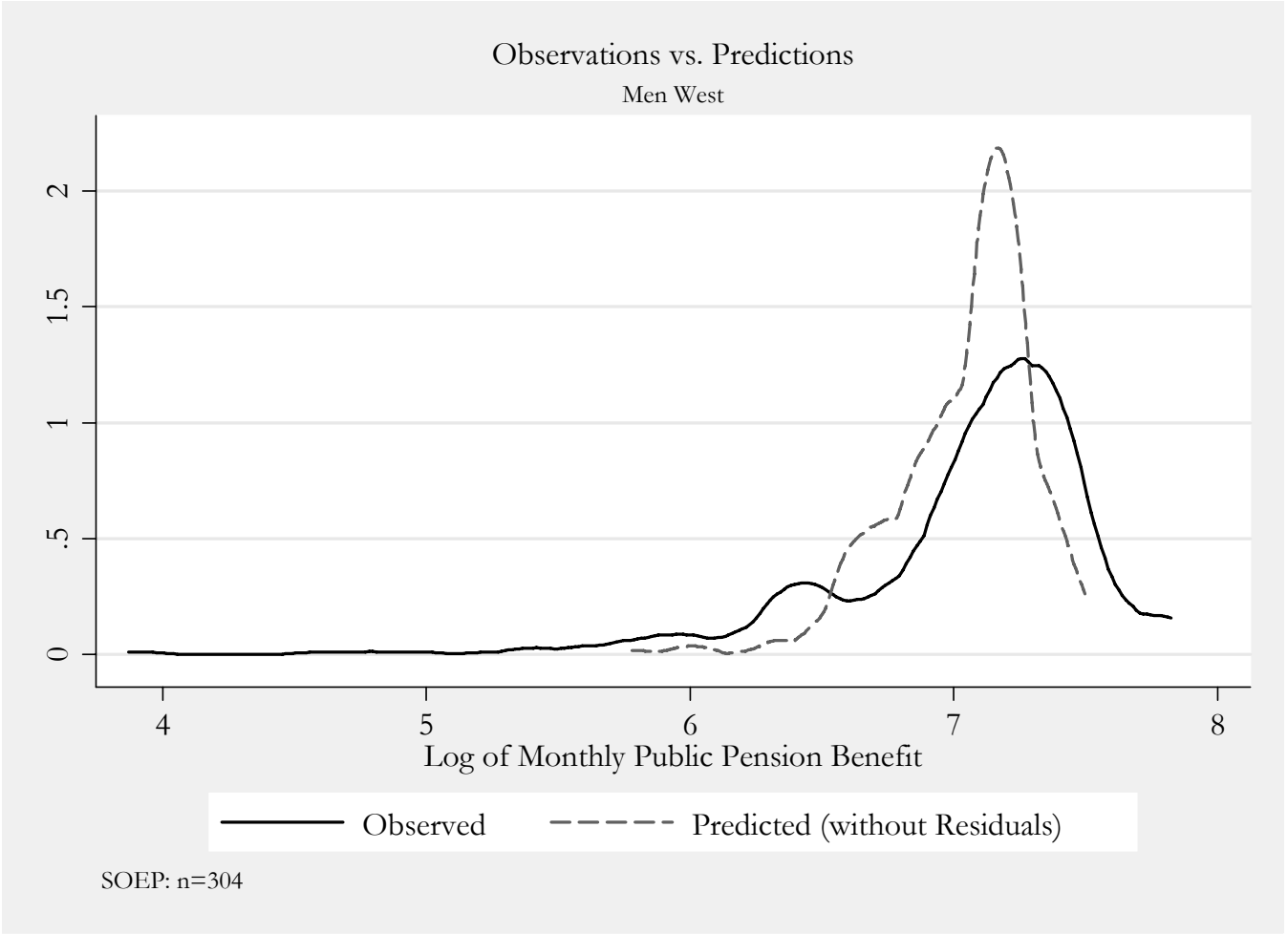
TOTAL WOMEN – FIRST TIME PENSIONERS SOEP 2000 – 2004
Observed vs. Predicted with Residuals



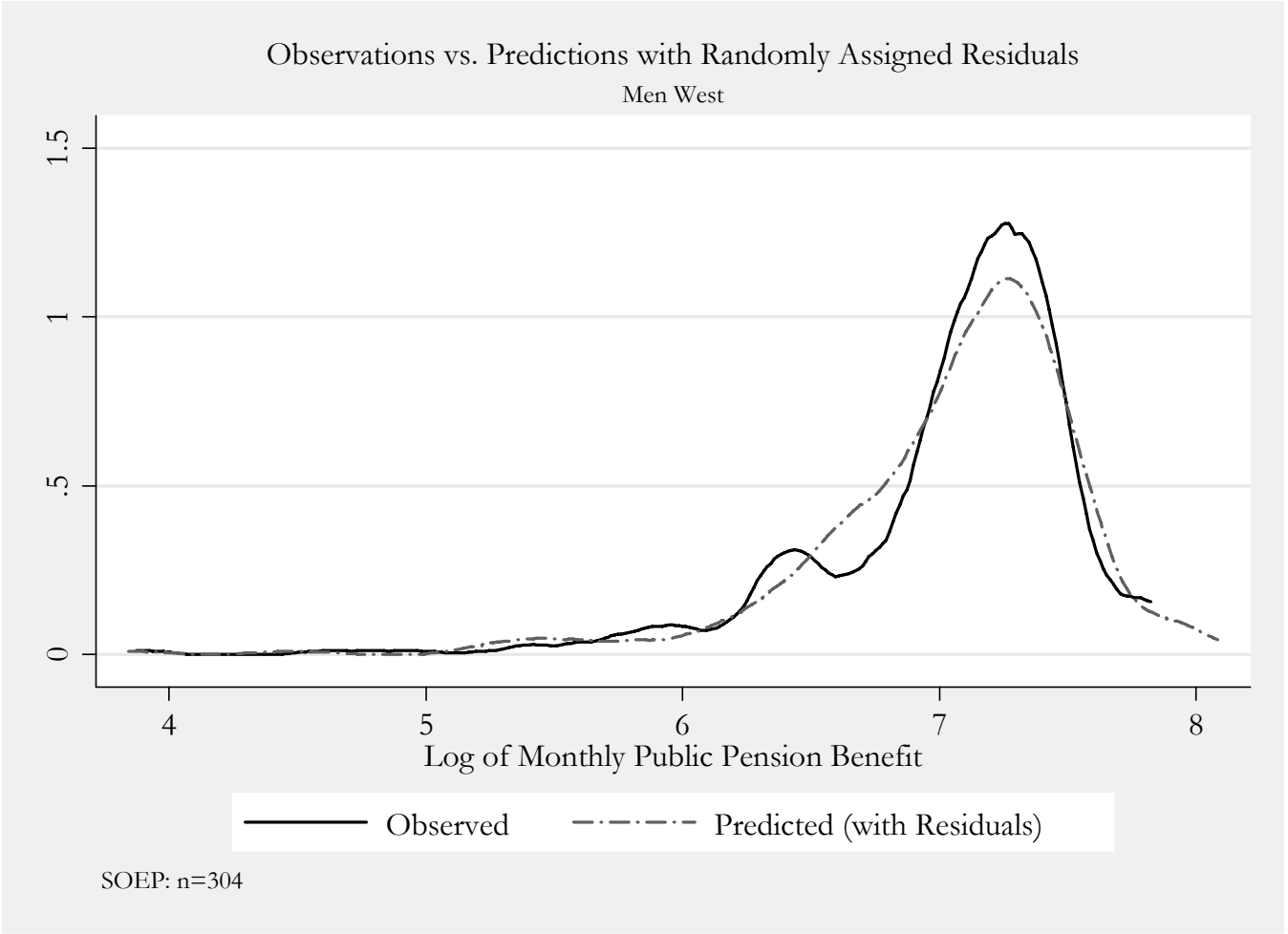
TOTAL WOMEN – FIRST TIME PENSIONERS SOEP 2000 – 2004
Confidence Bands



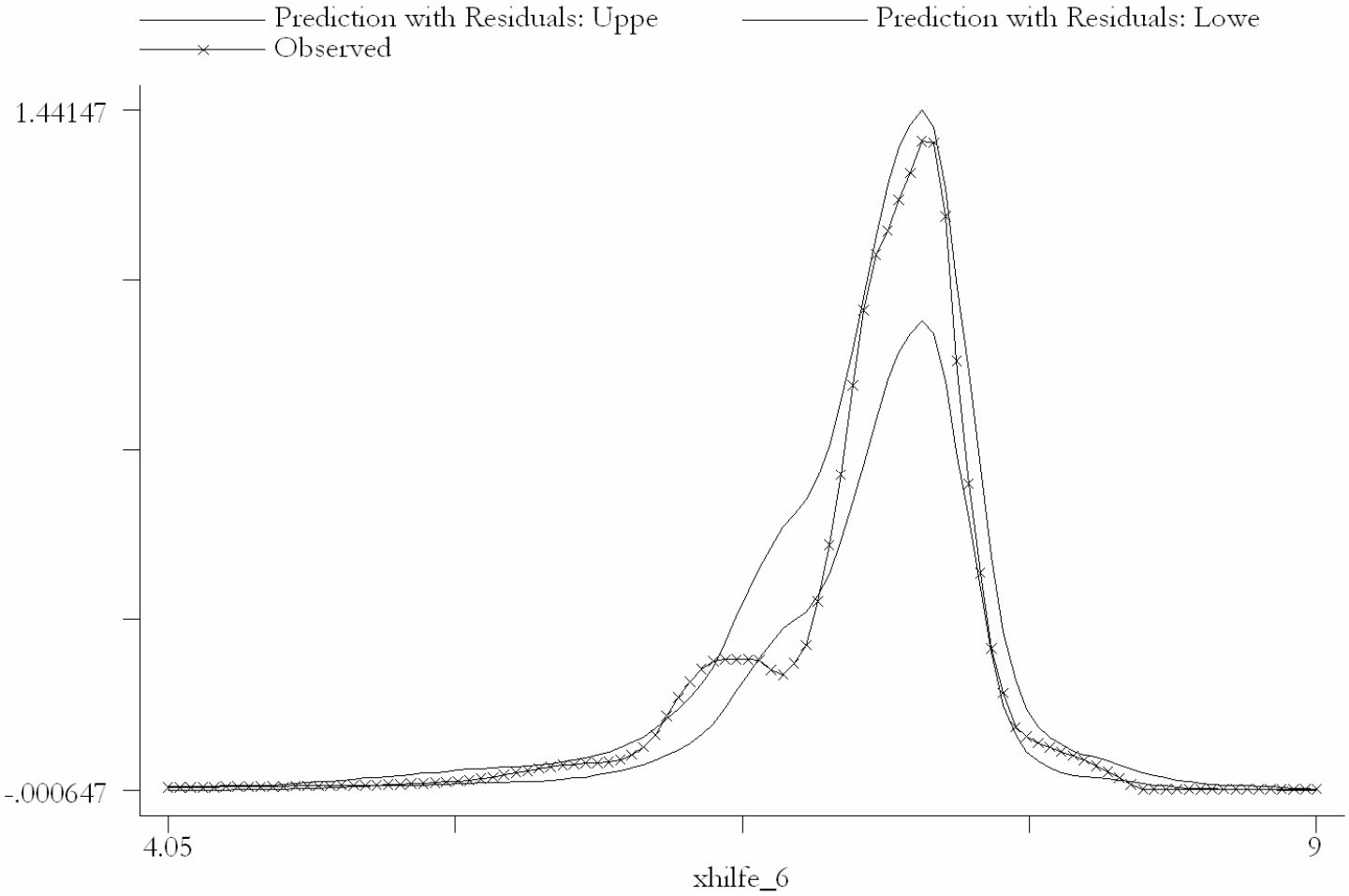
MEN WEST – FIRST TIME PENSIONERS SOEP 2000 – 2004
Observed vs. Predicted without Residuals



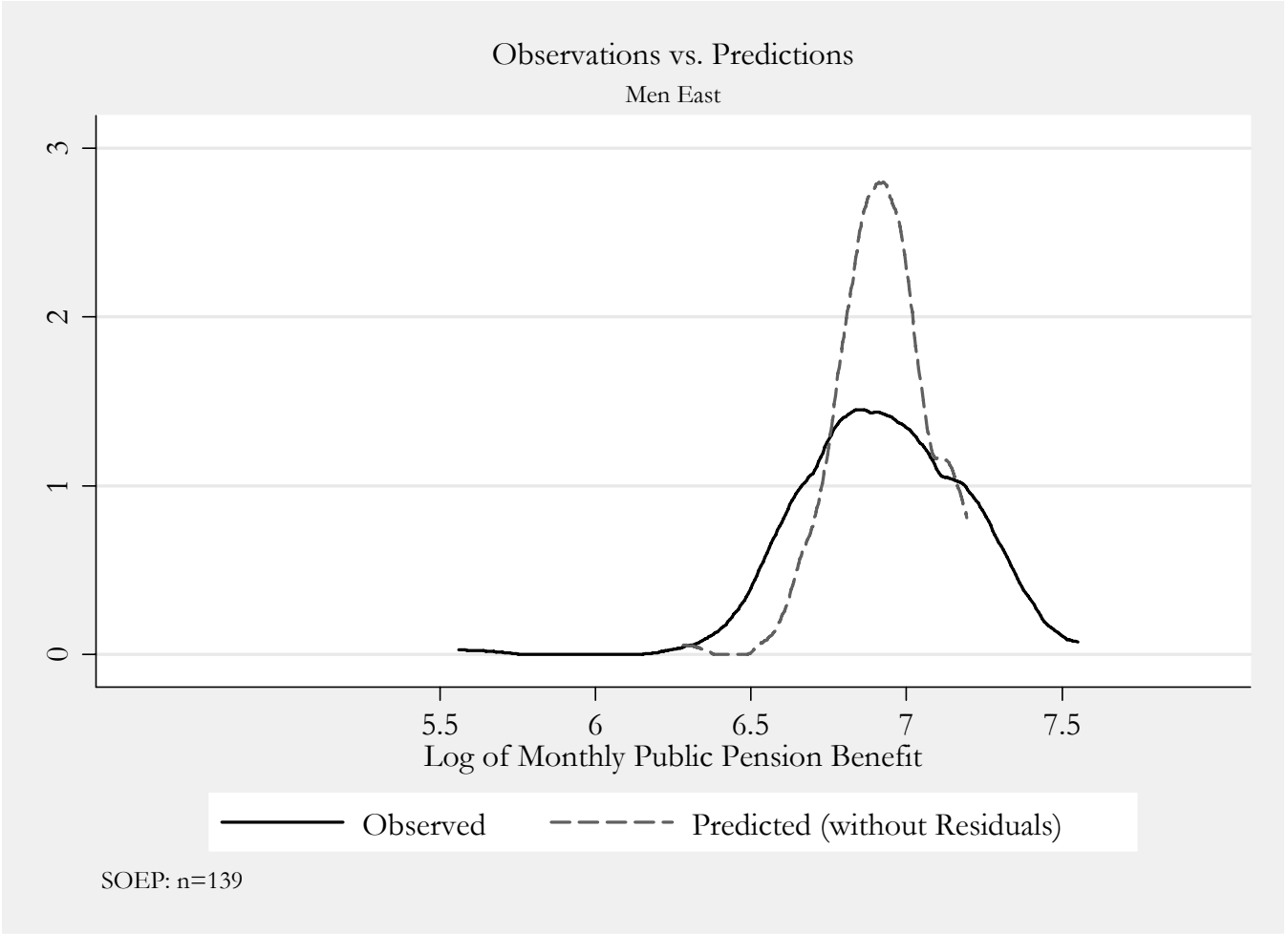
MEN WEST – FIRST TIME PENSIONERS SOEP 2000 – 2004
Observed vs. Predicted with Residuals



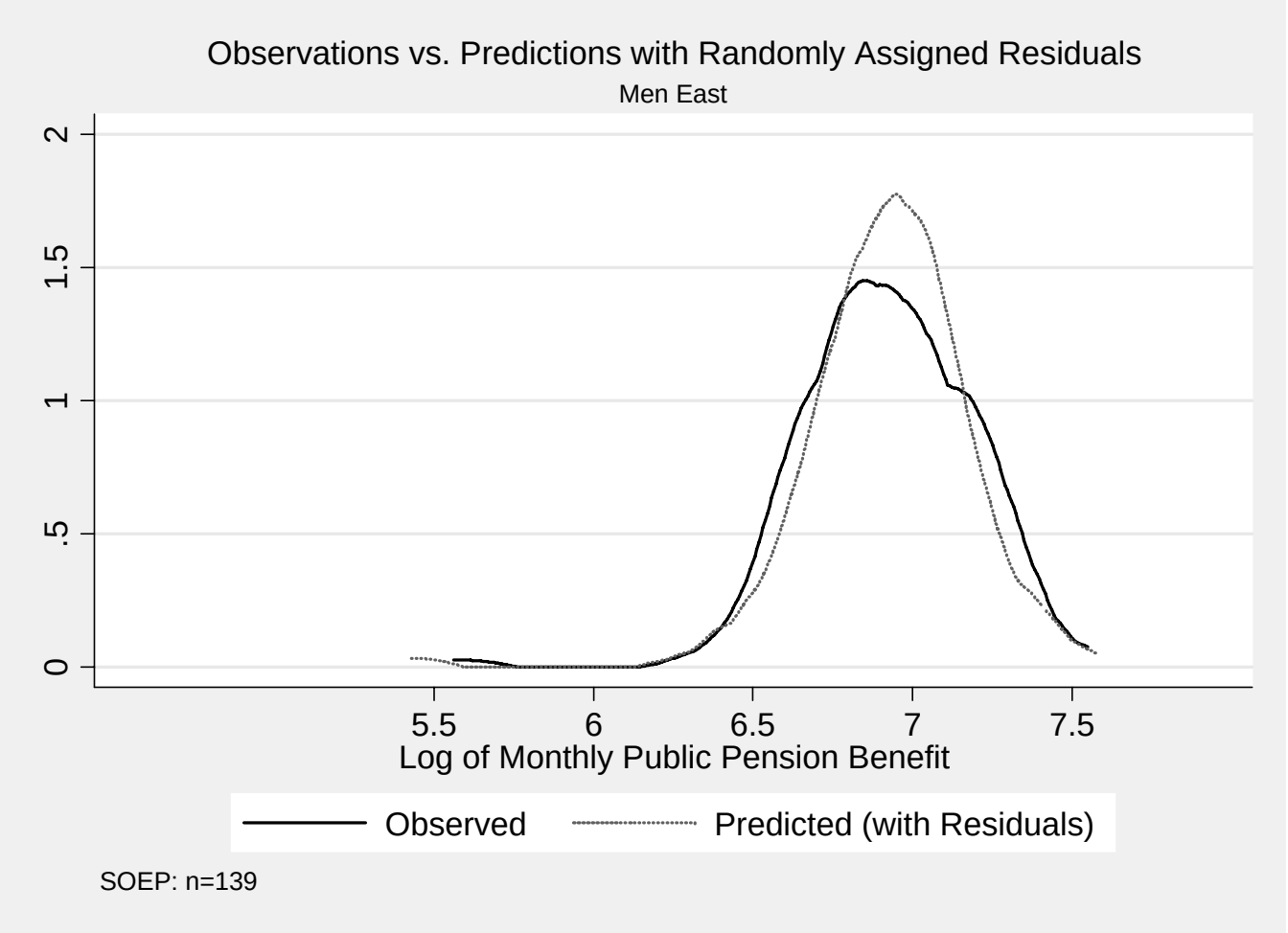
MEN WEST – FIRST TIME PENSIONERS SOEP 2000 – 2004
Confidence Bands



MEN EAST – FIRST TIME PENSIONERS SOEP 2000 – 2004
Observed vs. Predicted without Residuals

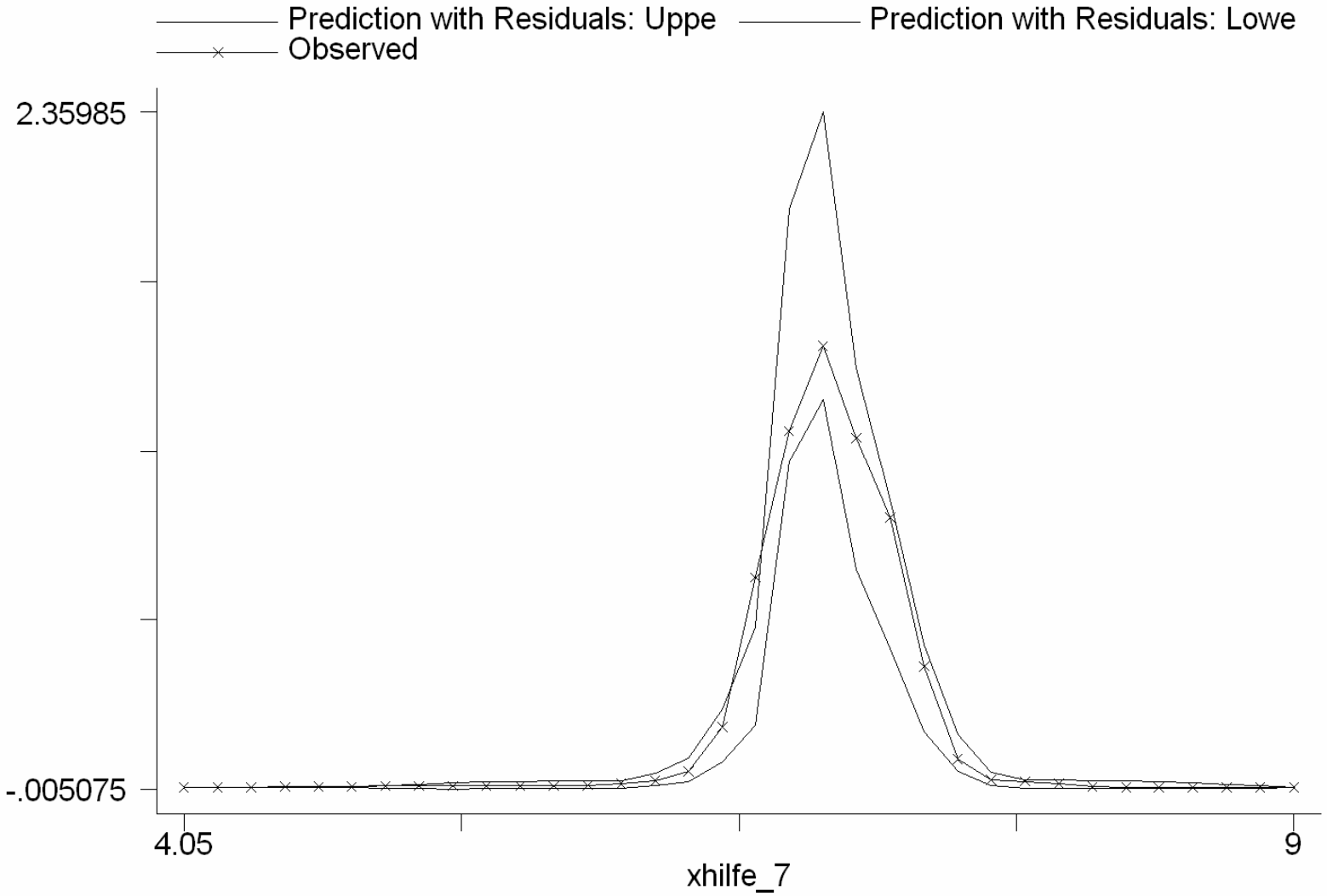


MEN EAST – FIRST TIME PENSIONERS SOEP 2000 – 2004
Observed vs. Predicted with Residuals

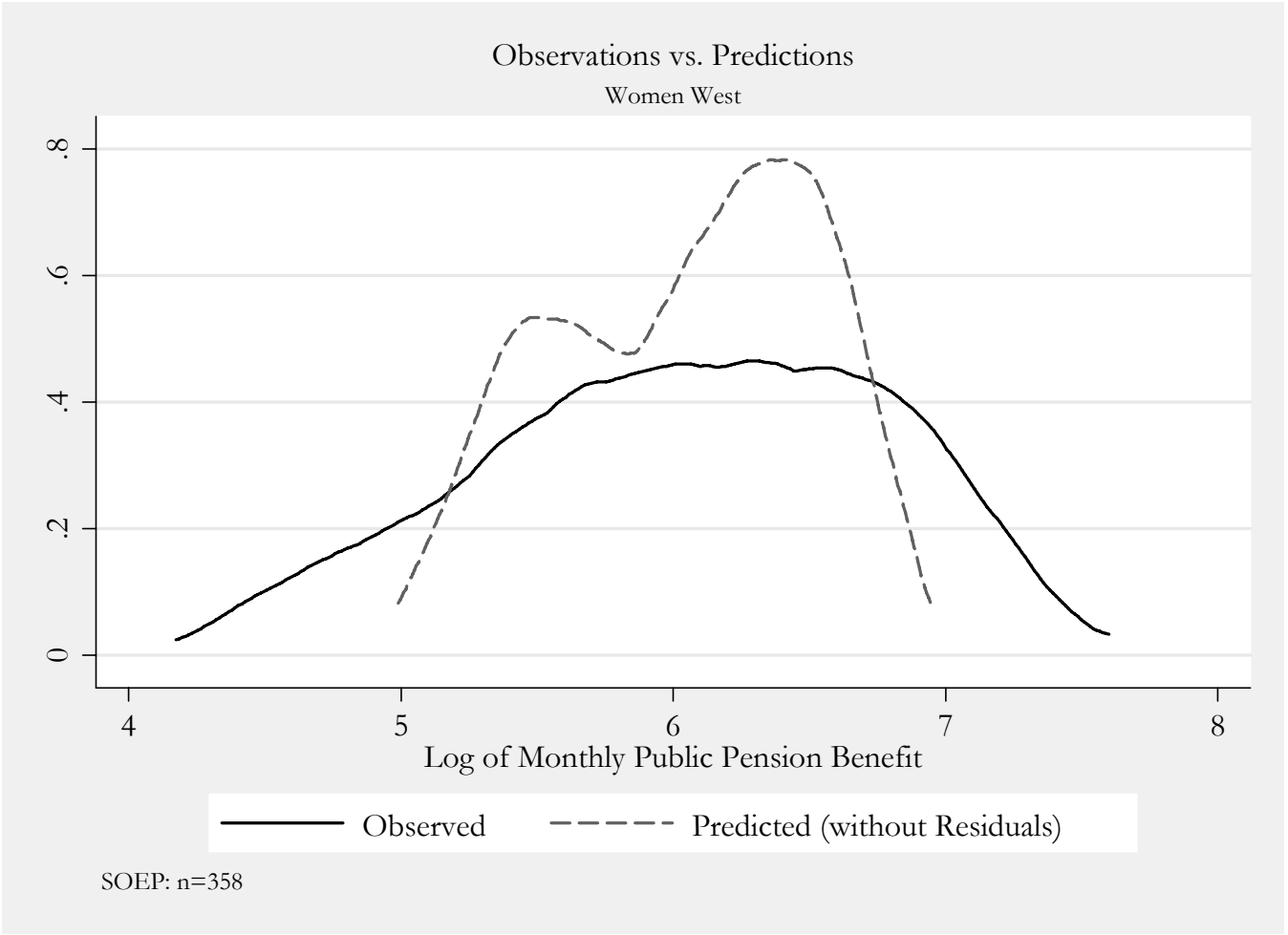


MEN EAST – FIRST TIME PENSIONERS SOEP 2000 – 2004

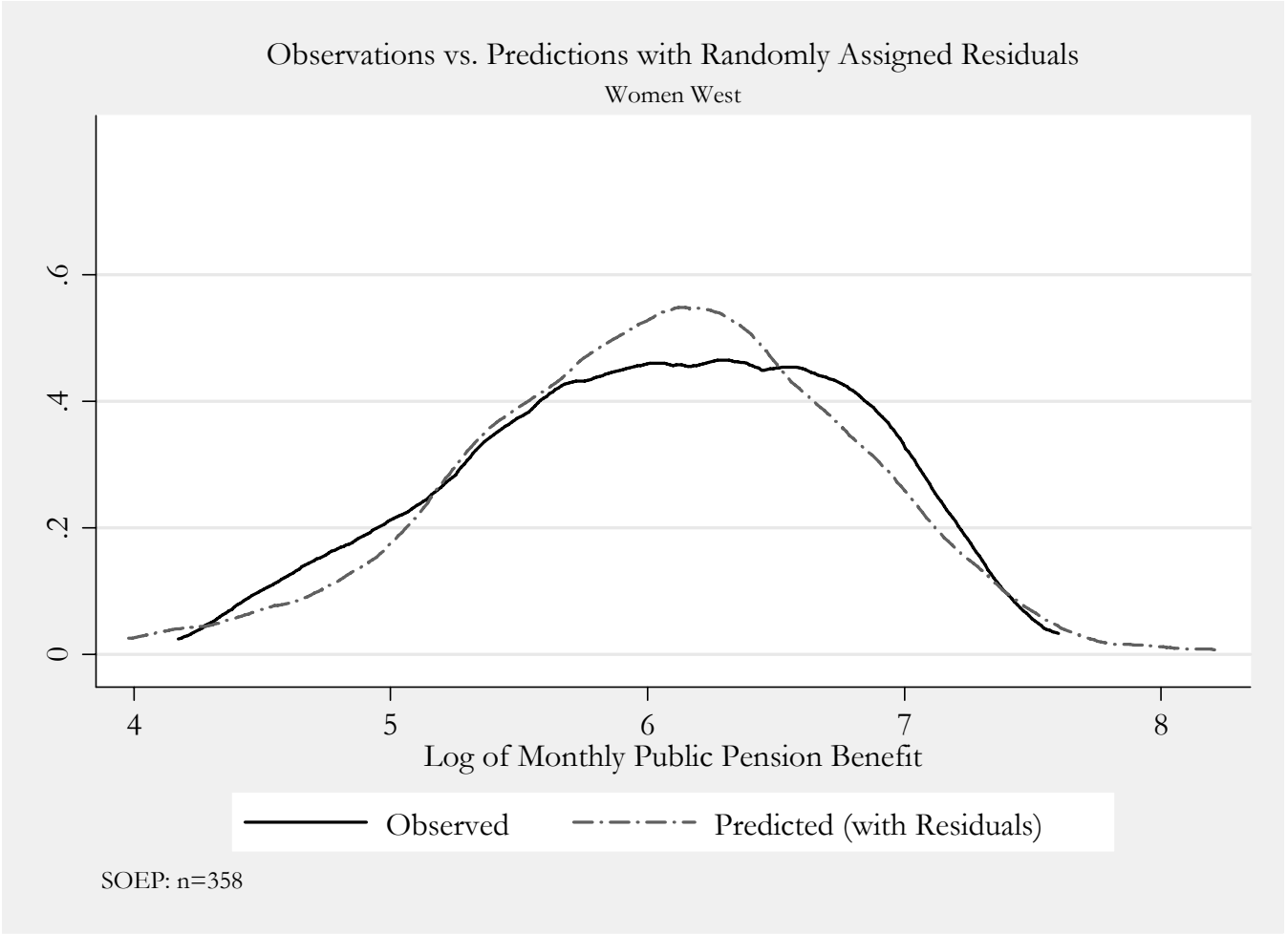
Confidence Bands



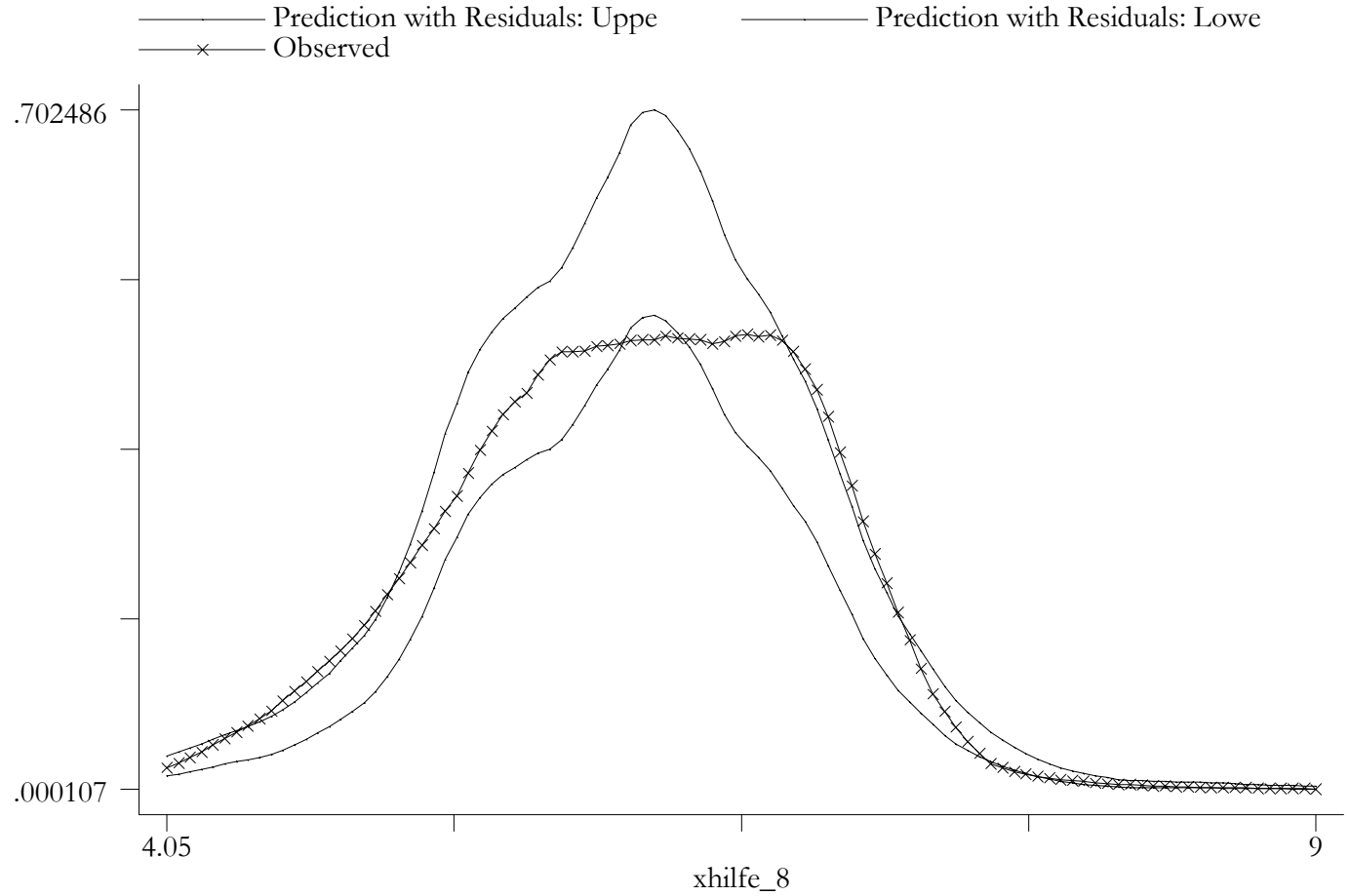
WOMEN WEST – FIRST TIME PENSIONERS SOEP 2000 – 2004
Observed vs. Predicted without Residuals



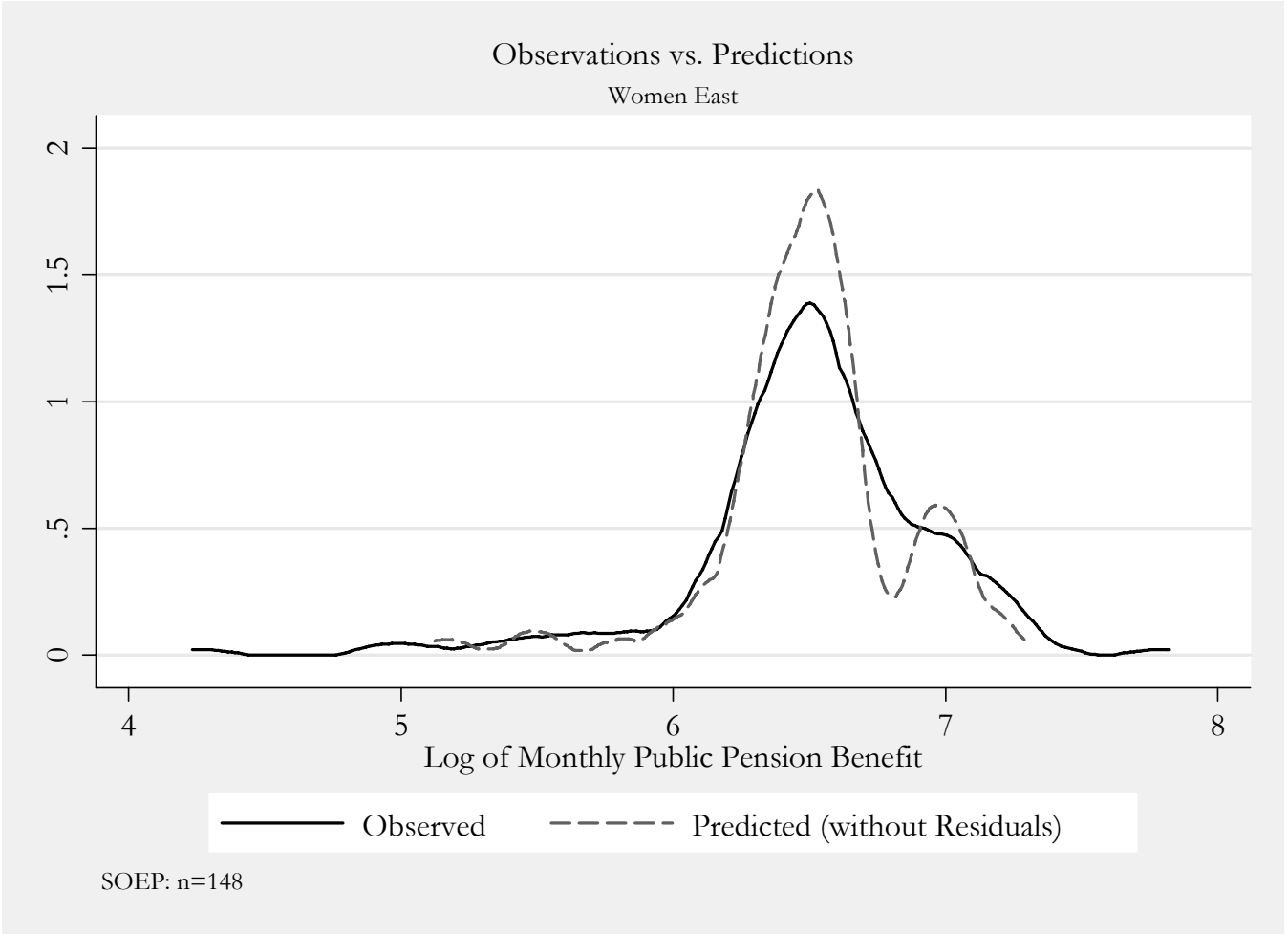
WOMEN WEST – FIRST TIME PENSIONERS SOEP 2000 – 2004
Observed vs. Predicted with Residuals



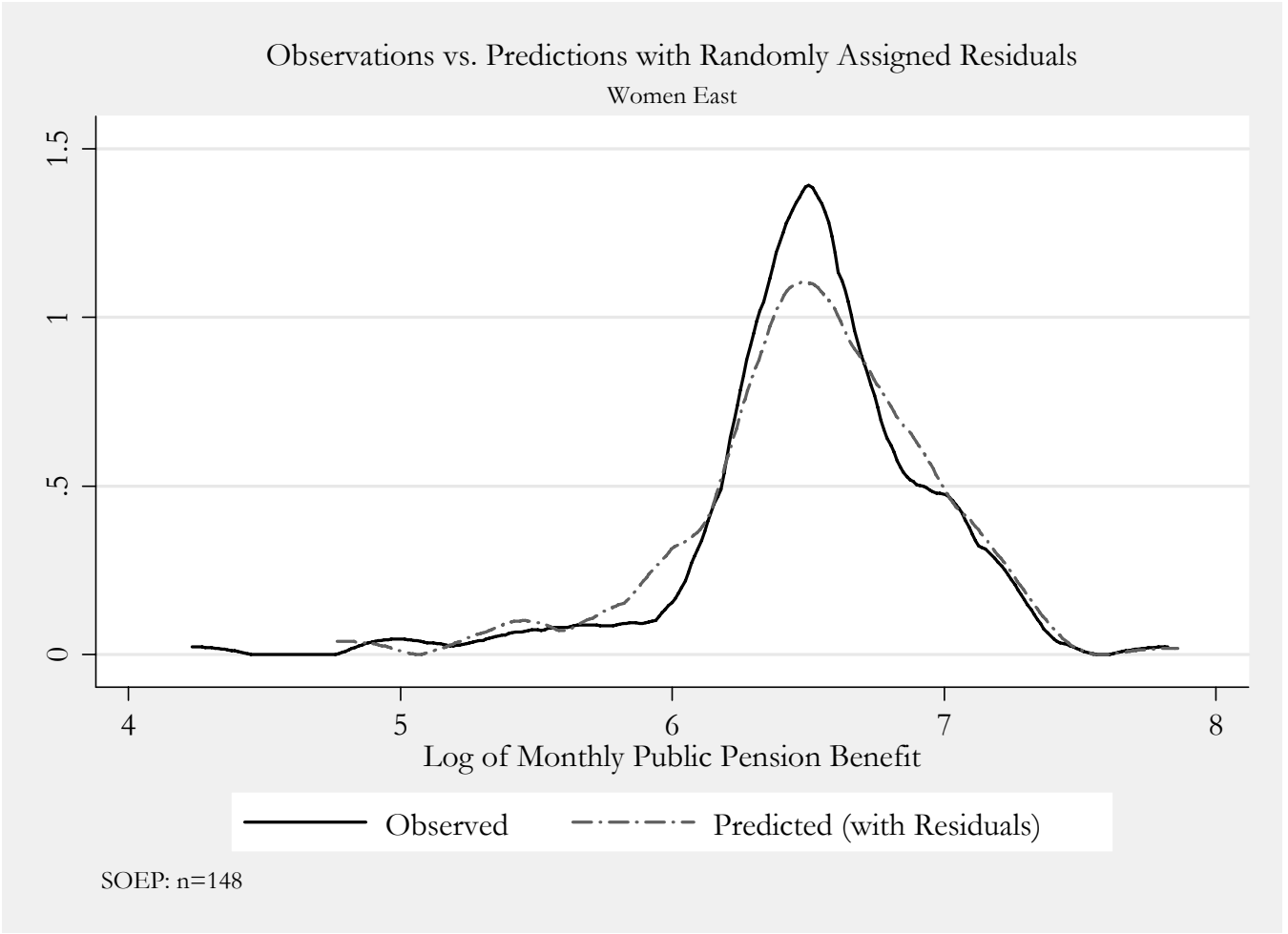
WOMEN WEST – FIRST TIME PENSIONERS SOEP 2000 – 2004
Confidence Bands



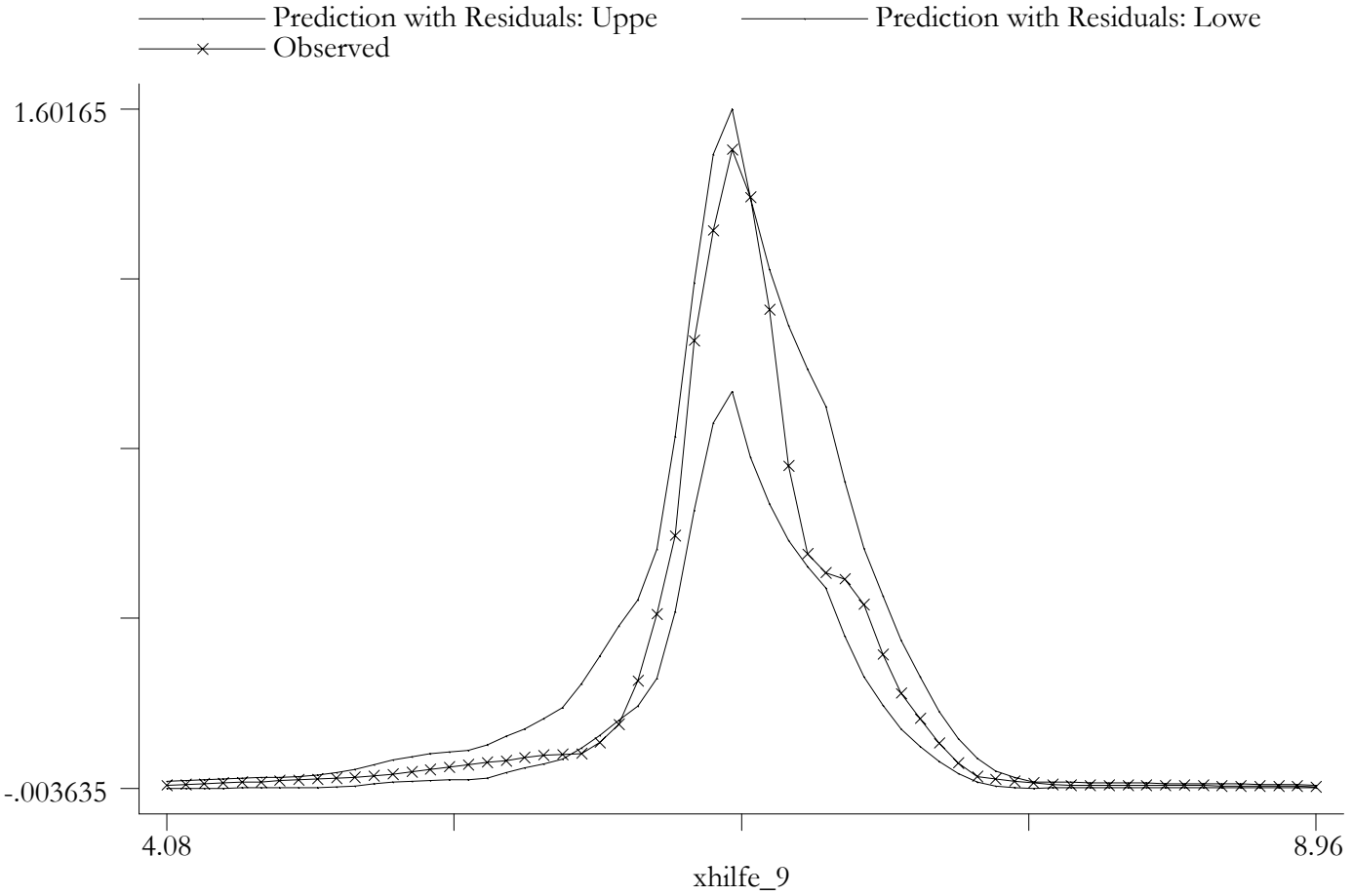
WOMEN EAST – FIRST TIME PENSIONERS SOEP 2000 – 2004
Observed vs. Predicted without Residuals



WOMEN EAST – FIRST TIME PENSIONERS SOEP 2000 – 2004
Observed vs. Predicted with Residuals



WOMEN EAST – FIRST TIME PENSIONERS SOEP 2000 – 2004
Confidence Bands



APPENDIX G

Observations vs. Out-of-Sample Predictions

Density Plots with two lines for four models:

First Graph – Observations vs. Out of Sample Predictions without Randomly Assigned Residuals

- solid line → observed
- dashed line → out of sample prediction without randomly assigned residuals

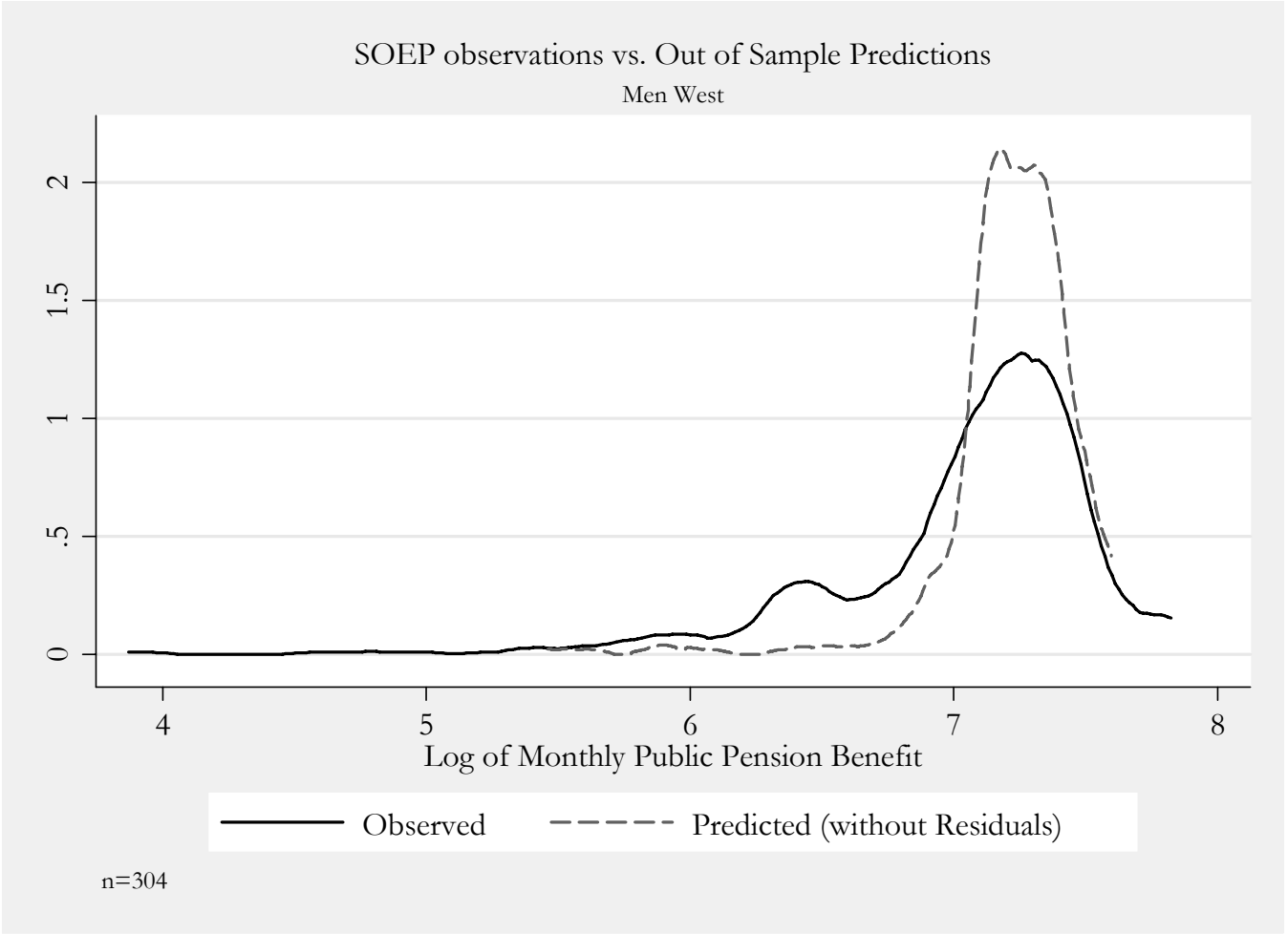
Second Graph – Observed vs. Out of Sample Predictions with Randomly Assigned Residuals

- solid line → observed
- dashed line → out of sample prediction with randomly assigned residuals

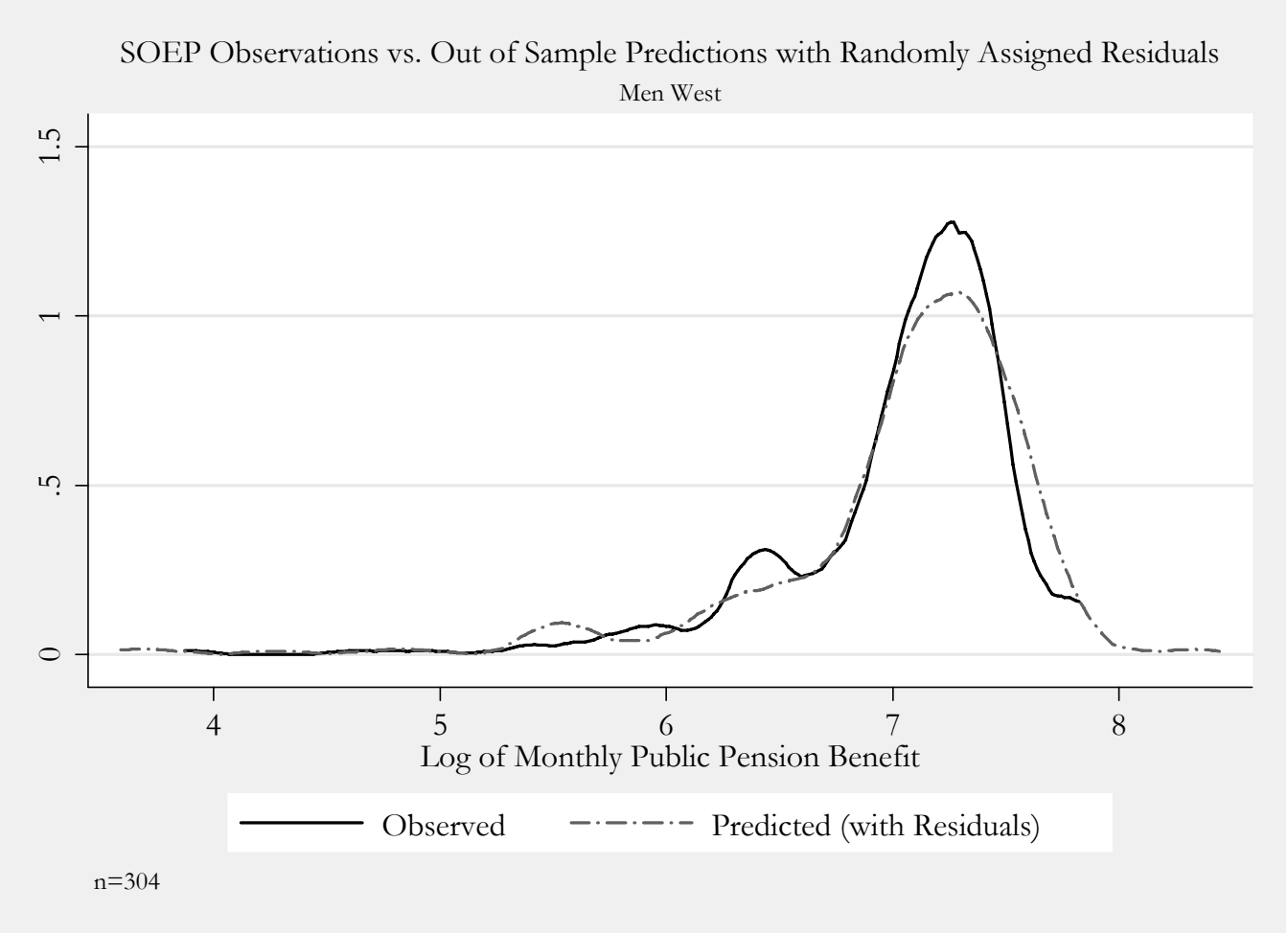
Third Graph – Confidence Bands

- solids line → upper and lower confidence band of observations
- dashed line → out of sample prediction with randomly assigned residuals

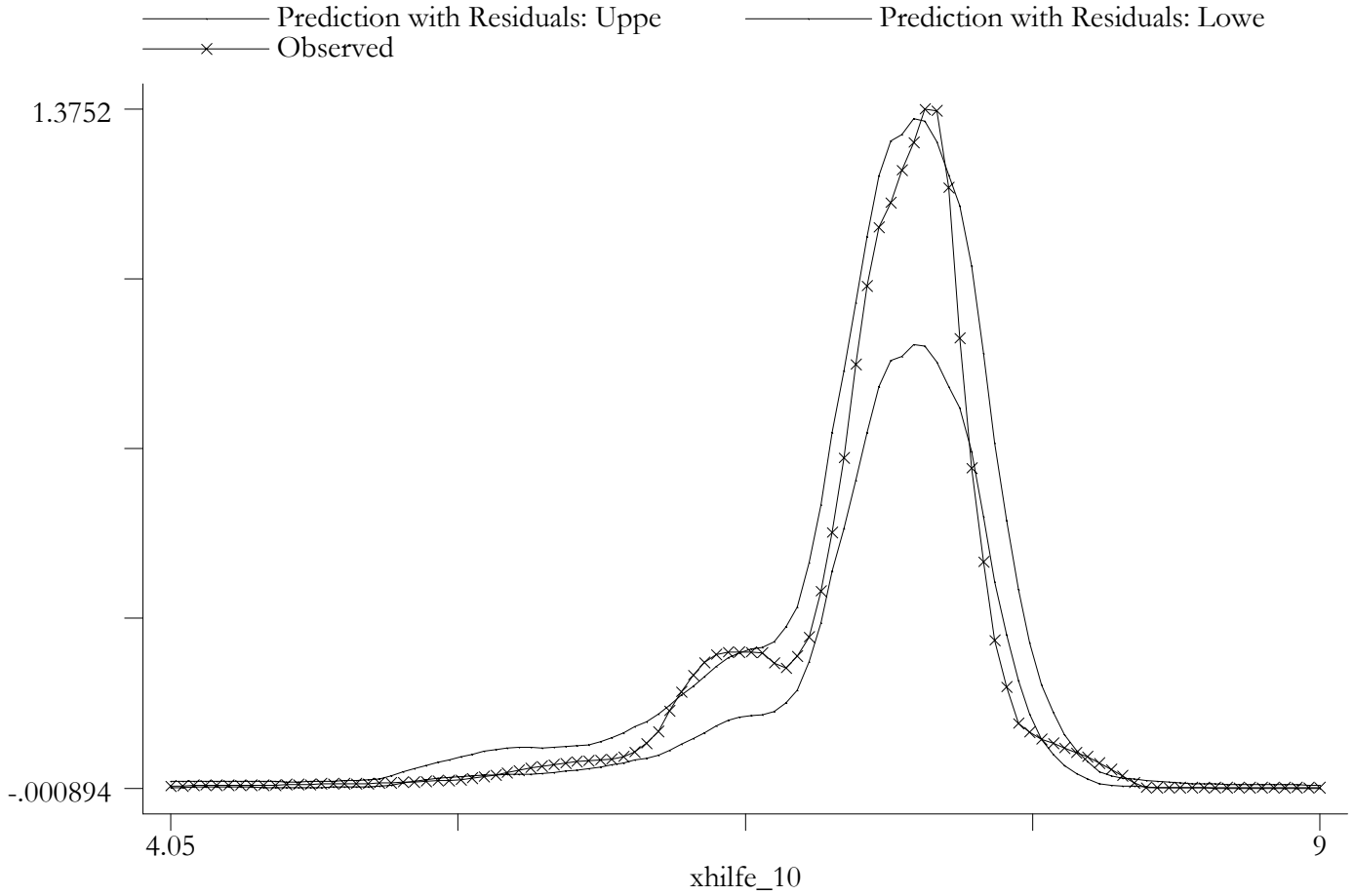
MEN WEST – FIRST TIME PENSIONERS SOEP 2000 – 2004
Observations vs. Out of Sample Predictions without Residuals



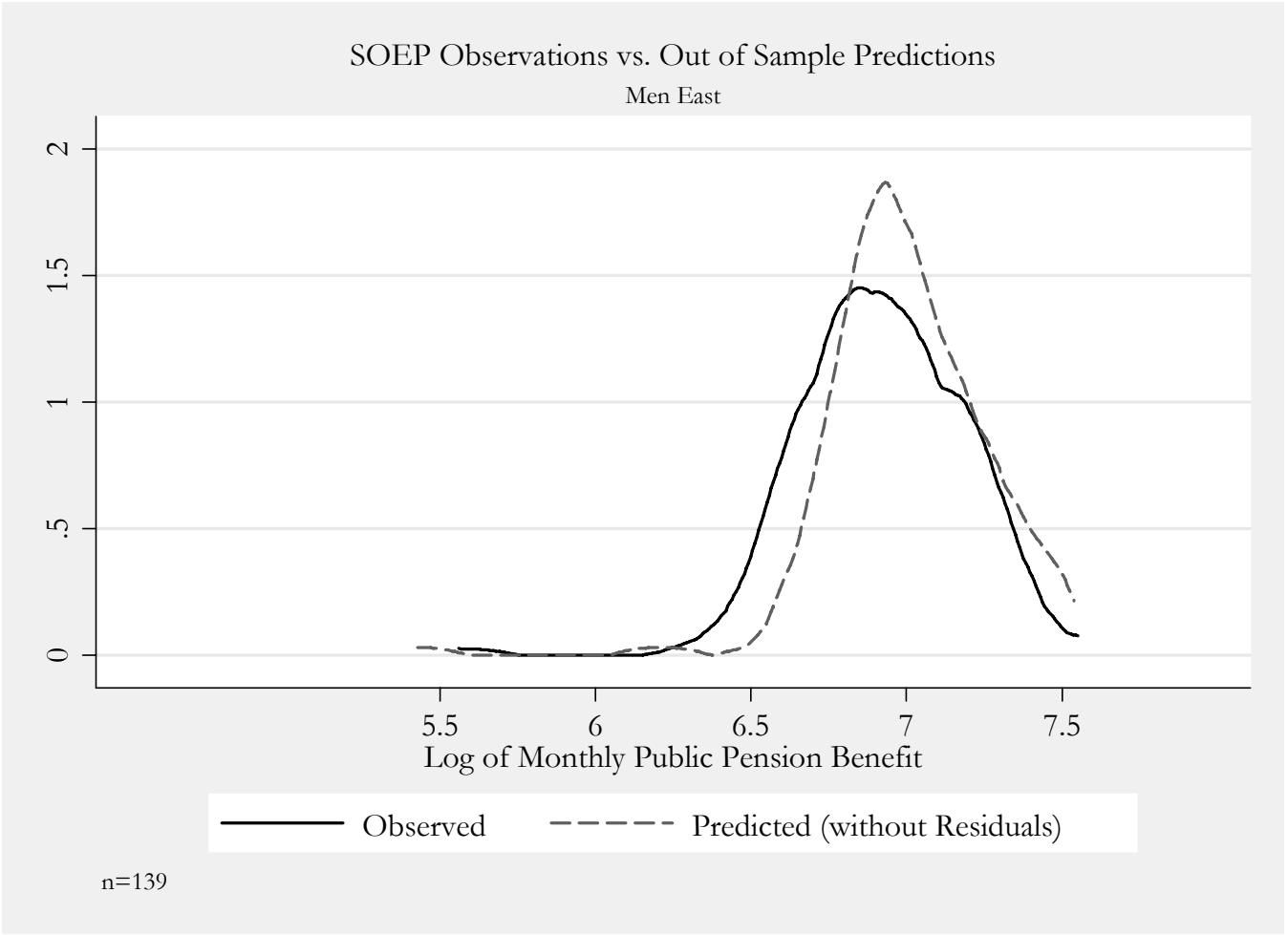
MEN WEST – FIRST TIME PENSIONERS SOEP 2000 – 2004
Observations vs. Out of Sample Predictions with Residuals



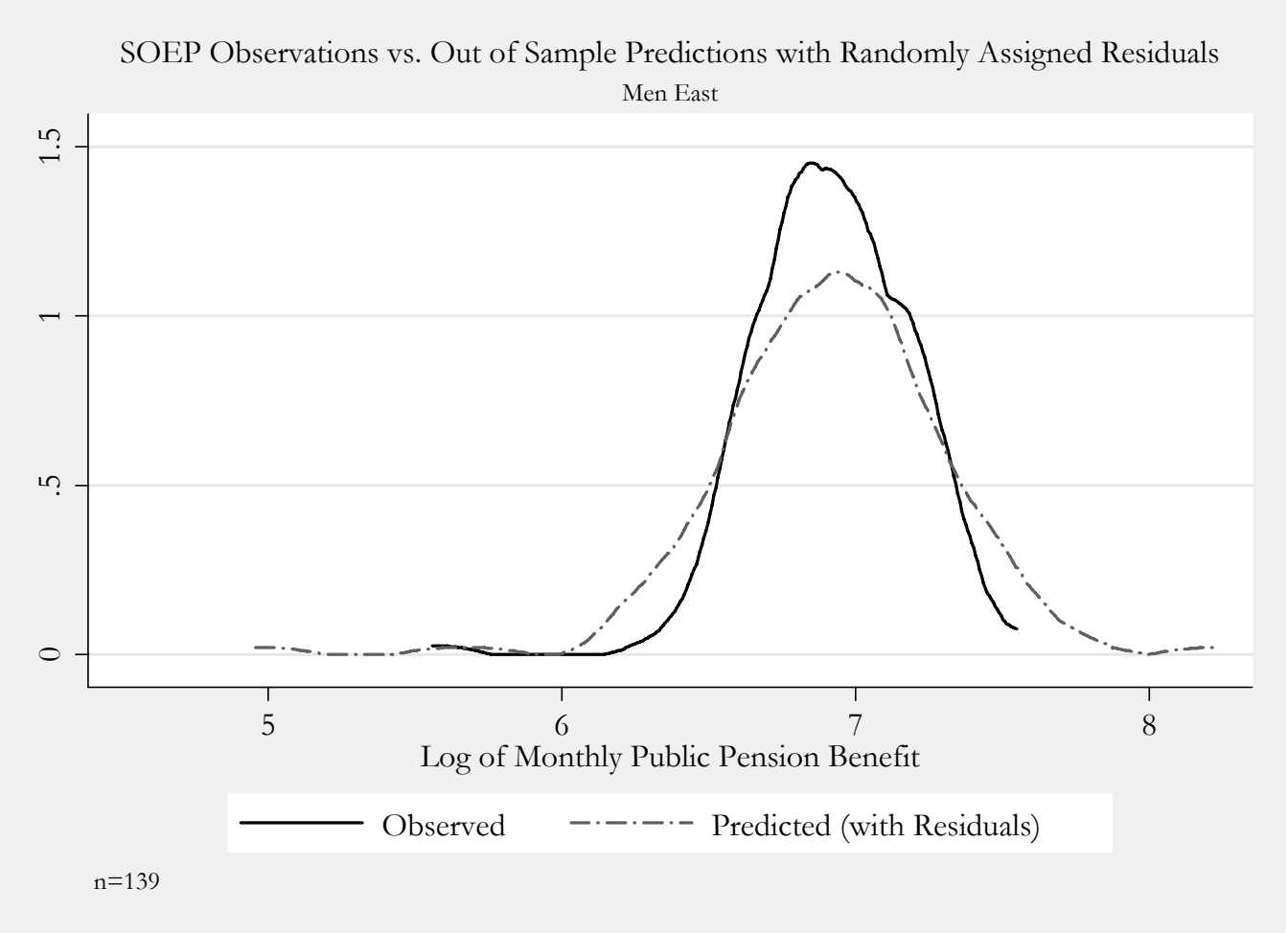
MEN WEST – FIRST TIME PENSIONERS SOEP 2000 – 2004
Confidence Bands



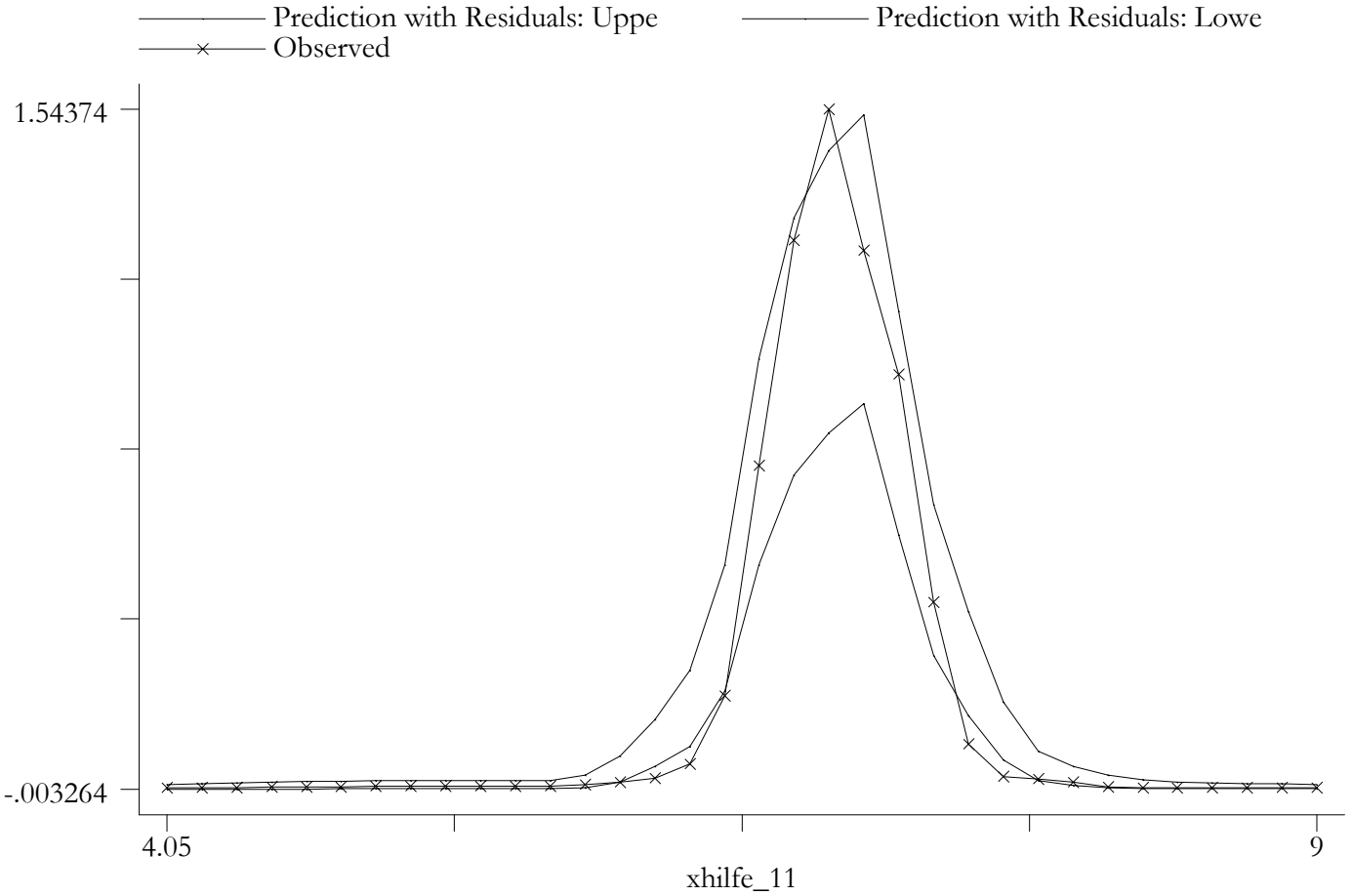
MEN EAST – FIRST TIME PENSIONERS SOEP 2000 – 2004
Observations vs. Out of Sample Predictions without Residuals



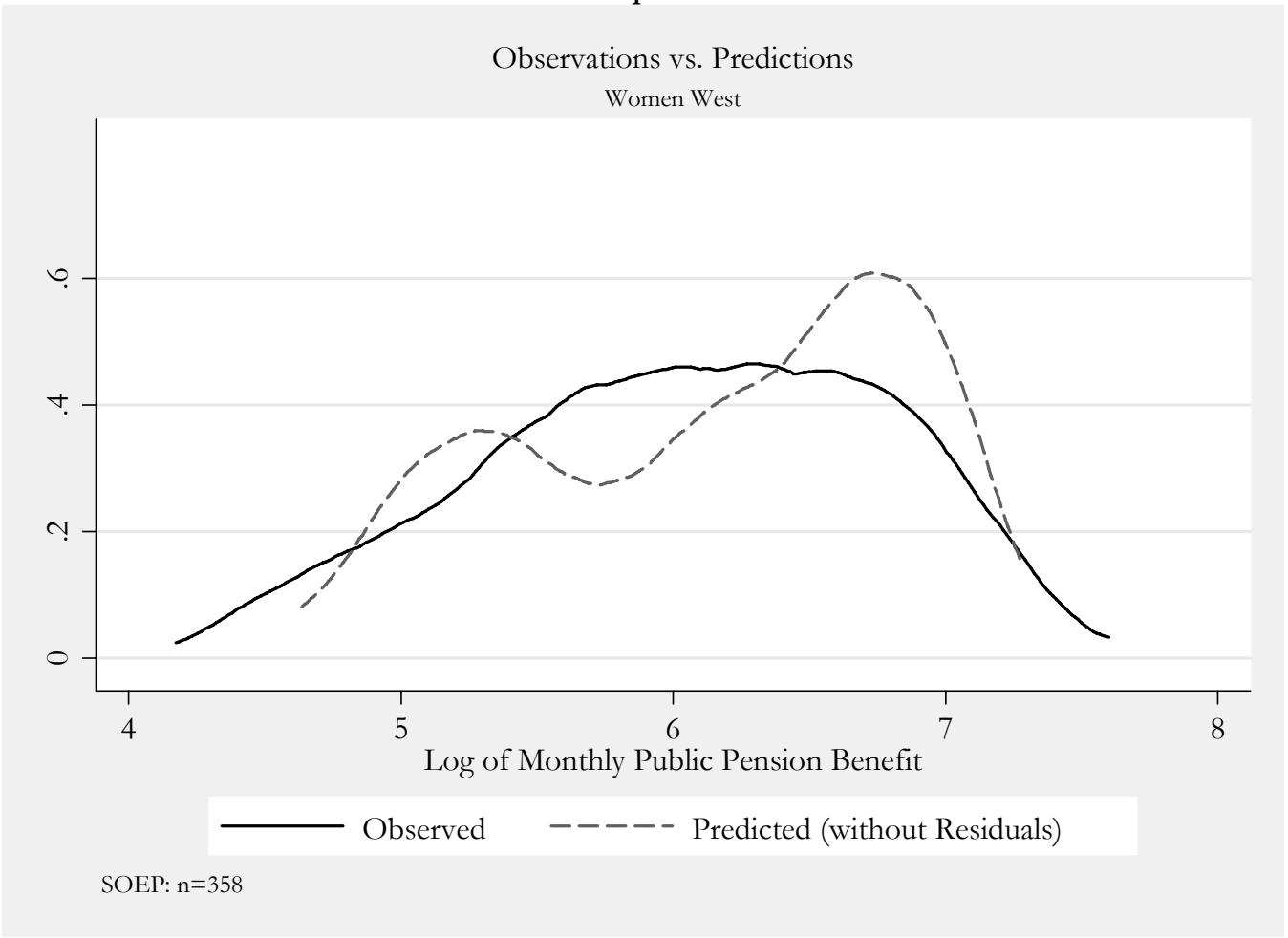
MEN EAST – FIRST TIME PENSIONERS SOEP 2000 – 2004
Observations vs. Out of Sample Predictions Randomly Assigned Residuals



MEN EAST – FIRST TIME PENSIONERS SOEP 2000 – 2004
Confidence Bands

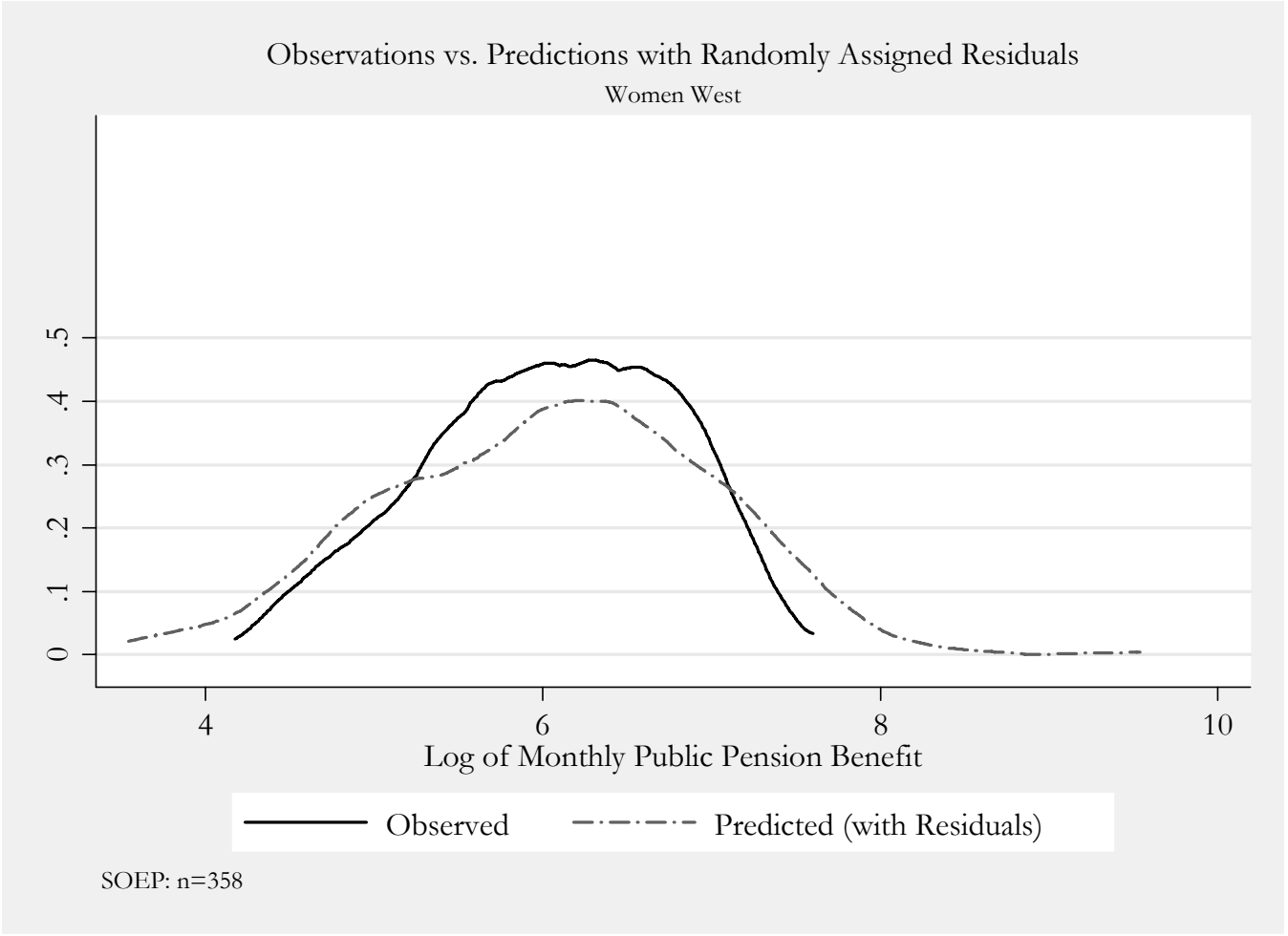


WOMEN WEST – FIRST TIME PENSIONERS SOEP 2000 – 2004
Observations vs. Out of Sample Predictions without Residuals

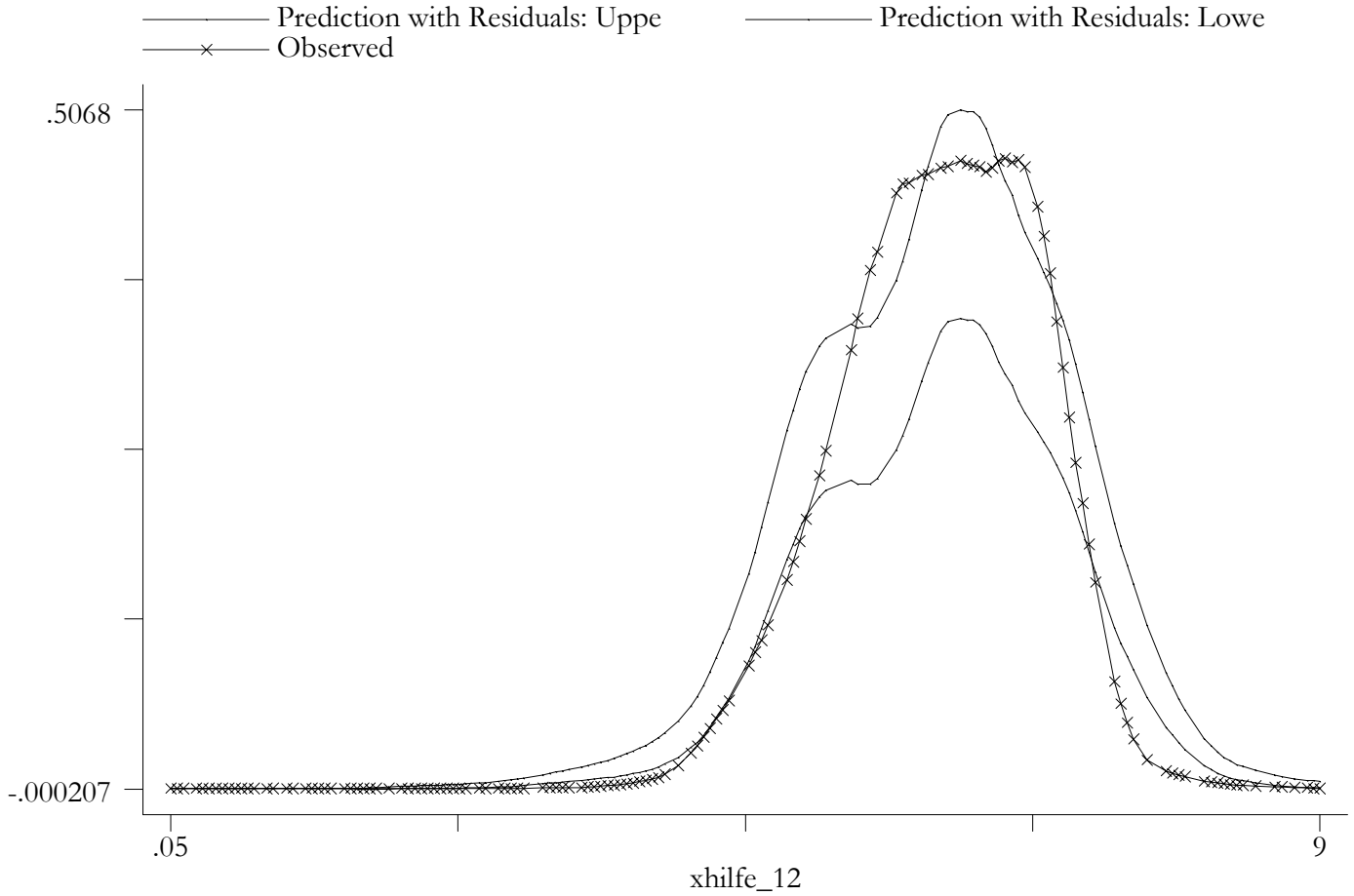


WOMEN WEST – FIRST TIME PENSIONERS SOEP 2000 – 2004

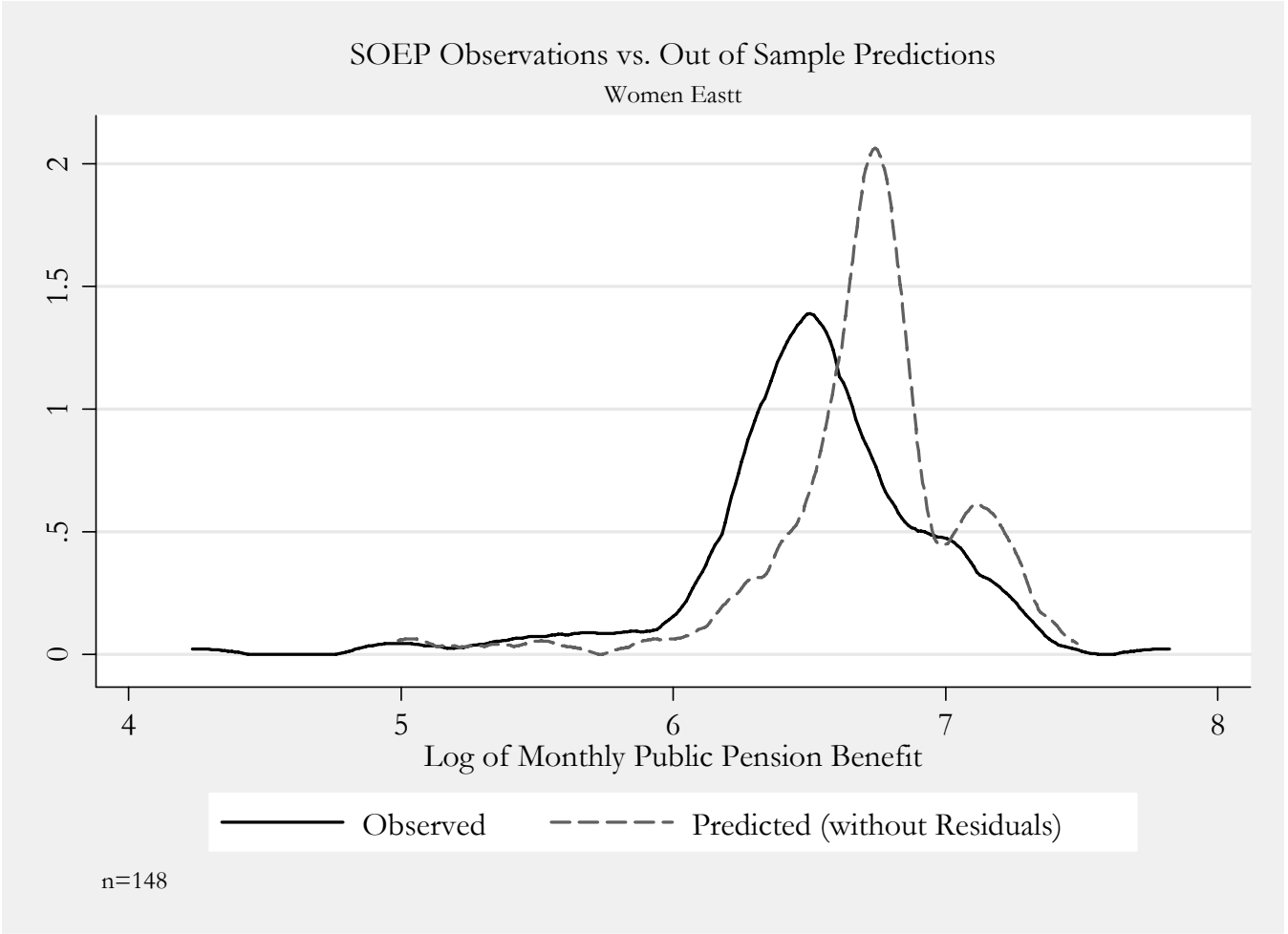
Observations vs. Out of Sample Predictions with Randomly Assigned Residuals



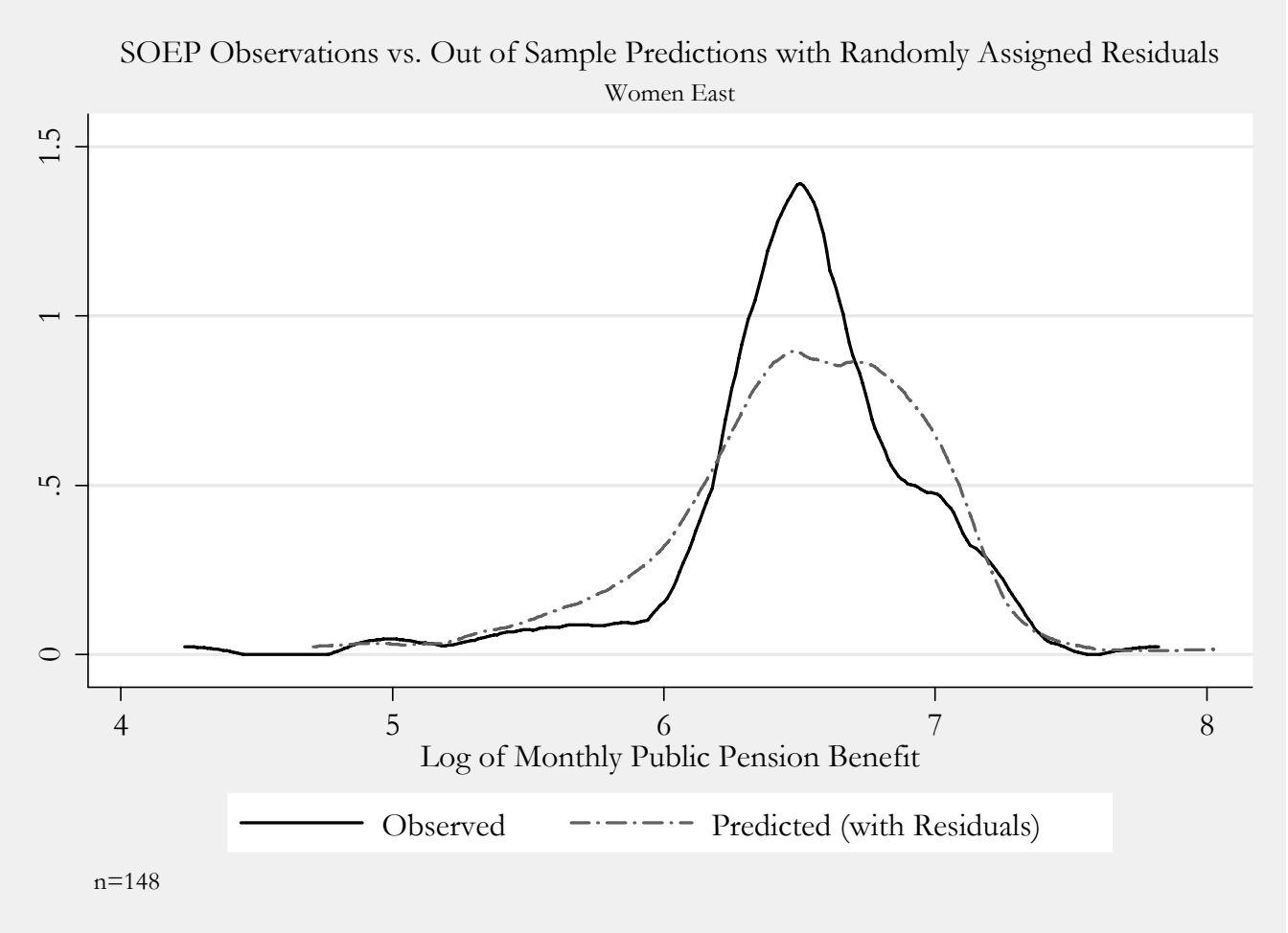
WOMEN WEST – FIRST TIME PENSIONERS SOEP 2000 – 2004
Confidence Bands



WOMEN EAST – FIRST TIME PENSIONERS SOEP 2000 – 2004
Observations vs. Out of Sample Predictions without Residuals



WOMEN EAST – FIRST TIME PENSIONERS SOEP 2000 – 2004
Observations vs. Out of Sample Predictions with Randomly Assigned Residuals



WOMEN EAST – FIRST TIME PENSIONERS SOEP 2000 – 2004
Confidence Bands