

Abstract

The paper introduces a general approach to distributing households by size, which is based on recurrent procedure of estimating conditional proportions of households of a given size among households with the same or larger size. Based on formal study and on empirical evidence, models and methods are proposed for deriving the distribution of households from their average size. The methods are tested and compared using census data and projections of Austrian and EU households. The approach proposed is also developed to distribute households both by size and by socio-demographic characteristics of the household head. Methods proposed are consistent and robust and may find a wide range of applications in household modelling and projecting, in enriching household projections by headship rates and other extrapolative methods, and also in studying the economic, ecological and other implications of household dynamics.

Keywords

Household modelling, household projection, headship rates, distribution by size, distribution by age of household head

Authors

Dalkhat M. Ediev is a research scholar at the Vienna Institute of Demography of the Austrian Academy of Sciences.

Acknowledgements

This work contains results which were developed in part within the EU-TACIS project “System of current estimation, analysis and forecasting of number and structure of households in the periods between censuses”. The author is thankful to Richard Gisser, Sergei Scherbov, and Alexander Hanika for useful comments and discussion of the method and also to Magdalena Muszynska for valuable help.

On Projecting the Distribution of Private Households by Size

Dalkhat M. Ediev

1. Introduction

In recent decades many advances have been made in household modelling and in developing household projecting methods (e.g., Akkerman 1977, 1980; Bell et al. 1995; Bongaarts et al. 1987; Hooimeijer and Heida 1995; Jiang and O'Neill 2004; Keilman et al. 1988; Lutz 1997; Nelissen 1991; Van Imhoff and Keilman 1992; Van Imhoff et al. 1995; Zeng et al. 1998; Zeng et al. 1997). Despite this, however, statistical agencies widely use the simplest method by headship rates and its extensions. This method—based on age/sex-specific proportions of household heads¹ in the population of corresponding age and sex—provides a very simple and yet robust method of inferring about the consequences for household dynamics of changing the age/sex composition of a population. In its traditional form, the method ends up with projecting the overall number of households and with distributing households by age and sex of the head. For applications, however, it is also important to have more detailed distributions of households, e.g., by size (e.g., O'Neill and Chen 2001; Prskawetz et al. 2004).

Although the headship rates method by itself does not provide such distributions, several extensions were developed to enrich projections by these distributions. Household propensity rates, household membership rates, household position rates, and household association rates methods were developed to project both the number and distribution of households and also the distribution of the population by status within the household (Mason 1987; Linke 1988; Boleslawski 1997; McDonald and Kippen 1998; O'Leary 1998). Being relatively simple, these methods result in inconsistent projections, however, as observed proportions of households of different types are usually inconsistent with projected population structures. Concerning the distribution of households by size, these inconsistencies may result, for example, in households with k and more persons to have average sizes of less than k . Despite the application of reconciliation procedures to fix such problems, the inconsistencies remain in official projections and even in some official *estimates* for past years. To avoid these problems while keeping the simplicity of the method, it was suggested to distribute households by size *after* estimating their number from the headship rates method (Gisser 1986a, 1986b). In this extension of the headship rates method, the total number of households and the number of single households are derived from the population age/sex composition using single and multi-person household headship rates, respectively. Then, multi-person households are distributed according the number of members from empirical relations between the average size of such households and the proportions of multi-person households of different sizes. Finally, regional adjustments to projected distributions are applied in order to take into account regional diversity observed from the last census. Although this approach avoids estimating numbers of households of different sizes directly from headship rates, occasionally it also results in

¹ One may use *household marker or representative* concept instead of *household head*. From the formal point of view, however, both concepts are identical. This paper uses terminology of household headship, although, the results are equally applicable to study of any household representatives instead of heads.

inconsistent distributions. First, due to peculiarities of the estimating techniques, the total number of households may occasionally deviate from the sum of households of different sizes. This inconsistency is usually fixed by appropriate scaling. Another inconsistency is related to an unlucky distribution of population among households of different sizes. If, for example, the number of small-sized households is biased downward, that will deflate per household population for households with larger sizes. As a result, there might not be enough people for a given number of largest households, and their average size might become inconsistently low. Another disadvantage of the method is its inability to provide a distribution of households both by household size and by age of the household head. These distributions, though potentially inconsistent, are available from other extensions (mentioned above) of the headship rates method.

This paper contributes to the field by proposing models and methods of deriving an internally consistent distribution of households by size from the average size of households. It follows the approach proposed in the works by Gisser cited above, who pointed out the importance of average household size as an indicator of their distribution by size. Based on analytical study, the models and the very methodology of distributing households by size are being revised in several ways. First, a framework for household distribution by size is proposed, which—without any reconciliation procedures—prevents from getting the sum of households of individual sizes to differ from the total number of households to be distributed. The idea is to derive households of different sizes step by step in a recurrent procedure, with households of k members being derived from the number of households of the same or of a larger size as well as from the average size of such households. The proposed framework allows the analytical study of the sources of possible inconsistencies in household distributions. Based on the results of this study and also on empirical observations, simple models are proposed which proved to be robust and efficient in projecting household distribution by size. The proposed models proved their efficiency even in projecting one-person households from the average size of households, i.e., without using age/sex-specific one-person household headship rates. The method may be applied independently of the headship rates method. In particular, it is possible to derive the number or average size of households by a simple extrapolation technique and then use the methodology presented herein to distribute households by size. When the traditional headship rates method is used, however, the method and analytical results provide distributing households of different sizes by age and sex of the household head, which significantly enriches the projection. For distributing households of different sizes by age and sex (or any other characteristic) of the head, the total number of households with such heads and also a provisional estimates for average sizes of households are made available by the method.

The paper is organised as follows. Section 2 describes the general approach to distributing households by size and some formal findings. Next, empirical evidence is presented in a way that is suggested by theoretical findings. Based on empirical patterns and also on theoretical findings, models and methods for distributing households by size are presented in the fourth section. Section 5 presents methods for distributing households both by size and by socio-demographic characteristics of the household head. Next, alternative methods for distributing households by size are discussed, which may also be developed within the general approach proposed. A final section addresses questions on the robustness of the method and on error accumulation, and the paper is concluded by a brief discussion of applications and of further developments of the method.

2. General Approach. Analytical Study

The main idea of the framework proposed here is to avoid working explicitly with proportions of households of different sizes. Instead, it is proposed to work with *conditional shares*, i.e., proportions of households of, say, size k among all the households of the same or larger size. Having obtained these conditional shares, one may derive the absolute shares of the entire set of households. The framework proposed has the advantage of giving the possibility of *independently* deriving and correcting the conditional shares. Unlike absolute shares of households, which must sum up to one, conditional shares are interlinked to a lesser extent and may be changed independently, without breaking the consistency of the overall distribution. Assuming the average size of households to bear predictive information about the distribution of households by size, it is also possible to consider the conditional share to be a function of the *conditional* average size of households of the given or larger size. This makes models for deriving numbers of households of different size even more independent. Additionally the *conditional size-share* approach enables developing analytical study of causes for inconsistency of distributions of households and establishing theoretical constraints to be taken into account in modelling and projecting households.

The main assumption underlying models presented here is that the distribution of households by size is a function of the average size of households:

$$v_k = \frac{H_k}{H} = v_k(n), \quad (1)$$

here v_k is the (unconditional) share of private households of size k , i.e., the ratio of the number of such households (H_k) to the total number of private households (H); $n = \frac{N}{H}$ is the average size of private households, i.e., the ratio of the population in private households (N) to the number of private households. Putting model (1) into the framework presented above, one may note that conditional shares are determined by unconditional shares (1) and, vice versa, conditional shares determine the unconditional ones:

$$v_{k/k+} = \frac{H_k}{H_{k+}} = \frac{v_k}{v_k + v_{k+1} + v_{k+2} + \dots} = \frac{v_k}{1 - (v_1 + v_2 + \dots + v_{k-1})}; \quad (2)$$

$$v_1 = v_{1/1+};$$

$$v_k = v_{k/k+} \cdot (1 - (v_1 + v_2 + \dots + v_{k-1})) = v_{k/k+} \cdot v_{k-1} \left(\frac{1}{v_{(k-1)/(k-1)+}} - 1 \right), \quad k = 2, 3, \dots \quad (3)$$

here $v_{k/k+}$ - is the *conditional share* of private households of size k among all households with size equal to or higher than k (H_{k+}). Introducing *conditional average sizes*

$$n_{k+} = \frac{N_{k+}}{H_{k+}}, \quad (4)$$

one may also note that the distribution of households and any of conditional average sizes determine all other average sizes (here N_{k+} is the size of the population residing in

households of k and more members). Therefore, model (1) may equivalently be substituted by the following set of relations for the conditional shares $v_{k/k+}$ as functions of corresponding conditional average sizes:

$$v_{k/k+} = f_k(n_{k+}), \quad (5)$$

where $f_k(\cdot)$ are some functions to be derived from observations.

Based on general considerations, one may impose some restrictions on functions (5). First of all, none of conditional shares (5) may exceed unity or be negative. An important restriction follows from the simple proposition that the average size of households sized $k+$ may not be lower than k :

$$n_{k+} \geq k. \quad (6)$$

In order to study implications of the constraint, one may exploit the following general relation²:

$$n_{(k+1)+} = \frac{N_{(k+1)+}}{H_{(k+1)+}} = \frac{N_{k+} - N_k}{H_{k+} - H_k} = \frac{n_{k+} H_{k+} - k \cdot v_{k/k+} H_{k+}}{H_{k+} - v_{k/k+} H_{k+}} = \frac{n_{k+} - k \cdot v_{k/k+}}{1 - v_{k/k+}}. \quad (7)$$

Substituting (6) into (7) and taking into account (7), we have:

$$n_{(k+1)+} = \frac{n_{k+} - k \cdot v_{k/k+}}{1 - v_{k/k+}} \geq k + 1, \quad (8)$$

which finally implies:

$$v_{k/k+} \geq 1 - (n_{k+} - k). \quad (9)$$

Hence, we have derived the general lower-bound constraint to be imposed on conditional shares (5). What is important for the framework proposed here is that the constraint involves conditional shares and conditional average sizes of the same order only, i.e., it may be applied to households of different sizes *independently*. It is also valuable that the constraint depends only on the average size of households of k and more members, i.e., it may be explicitly taken into account when developing model (5), which involves the same variable. Inequality (9) also suggests that *conditional average size excesses* ($n_{k+} - k$) might be useful in establishing functional relations (5). Indeed, this is the case, as it follows from the results presented further in the work. Therefore, for the sake of simplicity, we use special short notations for these conditional average excesses: $\eta_k = n_{k+} - k$. Hence, the lower-bound restriction may be written in the notations introduced as

$$v_{k/k+} \geq 1 - \eta_k. \quad (10)$$

One may also simplify relation (7) in new notations:

$$\eta_{k+1} = \frac{\eta_k + k - k \cdot v_{k/k+}}{1 - v_{k/k+}} - k - 1 = \frac{\eta_k}{1 - v_{k/k+}} - 1. \quad (11)$$

Other constraints presented further down in this section are not of the same generality as (9), (10) but are based on considerations derived from empirical observations which are assumed to be reasonable in general, but might not work in some occasions.

It is safe to assume that household size has some upper limit. In reality most households have small sizes, the share of large households quickly decreases as their size grows, and perhaps only a few households may exceed the size of, say, 20. Taking into account this strict limitation of the household size and the observed tendency of household

² The relation may also be derived from the following identity: $n_{k+} = k \cdot v_{k/k+} + n_{(k+1)+} \cdot (1 - v_{k/k+})$.

distribution to be closer to minimal sizes, we may assume that the following relation will, in general, be true:

$$\eta_k \geq \eta_{k+1}, \quad (12)$$

which means that while approaching the upper limit of household size, the conditional average excess should tend to decrease. Simple analogy might be found in life tables: limitation of the life span implies that in general the expected years to be lived decrease as the age increases. Eqs. (11) and (12) imply:

$$v_{k/k+} \leq \frac{1}{1+\eta_k} = \frac{1}{1+(n_k-k)}. \quad (13)$$

Note that the existence of the upper limit is not the only condition, which may result in (12), (13). The inequality will be observed whenever $k+1$ is closer yet still lower than the average or the most common household size above k . When the conditional average excess η_k is already less than unity, however, inequality (12) may not be applied, as distributions of households with $k+$ members and with $(k+1)+$ members may be much less related. As a general heuristic rule one might consider that constraint (12)-(13) is relevant to situations when the conditional average excess is higher than unity only. Empirical evidence presented in the next section suggests that for average excesses η_k lower than unity, the constraint (12)-(13) does not work. In fact, the line corresponding to (13) may well be used as an approximate for regression line of empirical observations in case of $\eta_k < 1$. However, one may develop another upper bound for this situation from the empirical evidence and the tendency of the share of large households to decrease rapidly as the size grows. This suggests that $\eta_{k+1} < 1$ whenever $\eta_k < 1$. Taking Eq. (11) into account, this yields:

$$v_{k/k+} < 1 - \frac{\eta_k}{2} = 1 - \frac{(n_k - k)}{2}. \quad (14)$$

Constraints (13) and (14) are consistent at $\eta_k = 1$. Therefore, one may use the following combined upper-bound constraint for conditional shares:

$$v_{k/k+} \leq \begin{cases} \frac{1}{1+\eta_k}, \eta_k \geq 1 \\ 1 - \frac{\eta_k}{2}, \eta_k < 1 \end{cases} = \begin{cases} \frac{1}{1+(n_k-k)}, n_k \geq k \\ 1 - \frac{(n_k-k)}{2}, n_k < k \end{cases} \quad (15)$$

Turning to another heuristic constraint for the conditional shares (5), let us consider the hypothetical situation when households with sizes k and above are distributed according to the Poisson distribution as regards their size's excess over k . In that case the share of the smallest of these households, i.e., of households of size k , would be equal to

$$v_{k/k+}^P = e^{-\eta_k} = e^{-(n_k-k)}. \quad (16)$$

Empirical evidence presented further in the work suggests that conditional shares were usually *above* Poisson levels (16). Again, as well as constraint (15), this observation is not as fundamental as (7)-(8). However, it holds for most of empirical data. A possible explanation of this observation may be the following. Poisson distribution might be observed in case when all ' $k+$ ' households are formed independently in similar random fashion as a result of independent scarce random events ('appearance of the excessive household members') with similar probabilities and during equal time durations. As a

result of such hypothetical process household population would be uniform and Poisson-distributed by size. As a matter of fact, such uniformity of the household population is broken due to several factors: the real population consists of different groups with differing life strategies concerning household formation (number of children, tendency to live with the parents, and so on); the real population is a mixture of households at different stages of their formation with consequently different average sizes; regional differences across the country may also contribute to heterogeneity of real households. As a result, the real distribution of households by size is more dispersed and—given the same average size—shares of both largest and smallest households are higher in reality compared to the Poisson distribution. From this explanation one may expect that Poisson shares (16) work well as a lower-bound estimate for large and heterogeneous populations. In fact, this expectation is confirmed by empirical evidence.

It is noteworthy that all the constraints presented may never come to a conflict. Indeed, the following inequalities hold, assuring the consistency of the constraints introduced:

$$1 - \eta_k \leq e^{-\eta_k} \leq \frac{1}{1 + \eta_k} . \quad (17)$$

3. Empirics

We will start the study of empirical evidence from the case of Austria. As the theoretical findings presented above suggest, household distribution may better be presented in terms of conditional shares and conditional average sizes. Figures 1 to 6 present these indicators for Austria, its regions and census districts for the years 1961/2001 (enlarged dots correspond to the country level; indicators are calculated from Gisser 1986a, 1986b; Statistics Austria 1986a, 1986b, 1996, 2005, and 2006). The figures are supplied with levels obtained from minimal possible shares (9), (10), Poisson shares (16), the upper estimate (15), and the ‘constant conditional excess in average size’ estimates (13).

One may note that the constraints discussed above work quite well. Also the tendency may be noted for average conditional excesses in household size to decrease from census to census. As a result, conditional shares, especially for largest households, get closer to the lower bound (9), (10). Hence, it becomes more and more important for projecting techniques to take care of this lower bound restriction in order to provide consistent results. Figures for smaller populations (regions and census tracts) are naturally more volatile and may easily break two other constraints presented above (bearing in mind that these constraints are not as fundamental as (9), (10) and may reflect the general tendency only). The dynamics of conditional shares seen from the graphs presented are clear and seem to follow simple exponent-like curves, which may be easy to exploit in projections. We will return to the implications on projection methodology in the next section, after completing the empirical review.

Figures 7-9 present same indicators for conditional household shares and sizes in EU countries and regions for the 1991 and 2001 census rounds (calculated from the data retrieved from Eurostat (2006); incidentally, these include also data from some non-EU countries, e.g., Turkey). Enlarged dots correspond to aggregate European figures. Indeed, the EU is more heterogeneous compared to Austria, and its regional patterns are more diverse. However, heuristic constraints imposed by the Poisson distribution and by

monotonically decreasing conditional excesses work in general quite well. Inequality (13) works well for conditional excesses in household size above unity, while for lower levels of excesses it may be used as a reference line only. Similarly to the Austrian case, conditional shares for households of size three seem to be significantly lower compared to the trend of other households. This might be explained by the ‘instability’ of such households compared to larger ones. Indeed, if the bulk of three-person households are those of two parents and a child and if most couples stop enlarging their households only after having two children, then the share of three-person households will be suppressed. At the same time, four-person households will be relatively stable (and, therefore, more abundant) in this case as it might take a long time before these households start decreasing (due to children leaving home, parents dying, etc.). These two factors may result in a sharp distribution of households of size three and above with higher shares for 4-person households and lower shares for all other households. This hypothesis is supported by the observation that the average excess in size for households with three and more members exhibits downward shifts only when it is close to one, i.e., when the average size of such households is close to four (comparing patterns from Austrian censuses 1961 and 1971 to those from 1981-2001 censuses).

The same interpretations may be applied to the case of Russia. Distributions of Russian regional rural and urban households by size are presented in Figures 10 and 11 (data from 1994 micro-census and 2002 census). Despite the socio-economic changes of the last few decades, the distribution of Russian households fits well the general patterns both in 1994 and 2002.

The next three Figures (12 to 14) present indicators of distributions of households for Canada in 1996 and 2001 (Statistics Canada 2006) and historical estimates for Japan in 1975-2000 (National Institute of Population and Social Security Research 2006). Again, one may observe the same general regularities as in the cases addressed above.

Pakistan and India also follow these general trends, see Figures 15 and 16 (Federal Bureau of Statistics of Pakistan 2006, Registrar General and Census Commissioner of India 2006). The example of these countries is important as they represent populations with extended average household sizes. Note that three-person households do not show any special patterns in these examples, which indirectly supports the interpretations given above for countries with small average sizes of households.

The last case—historical estimates for the US population—is presented in Fig. 17 (U.S. Census Bureau 2006). Amazingly, US patterns for large households seem to fall out of the general trend predicted by theoretical findings and observed in a variety of countries with different demographic profiles. Seemingly, this would reflect problems with data quality rather than some peculiarities in US household formation. This view is suggested by the fact that some figures for the US violate even the lower-bound constraint (9)-(10), which cannot be attributed to anything but data inconsistency. Indeed, the average size of households with sizes above 6 is 5.1 and equals 4.1 for households with size 7 and above in 1968. These inconsistent results remain even after checking for rounding errors in the initial data source.

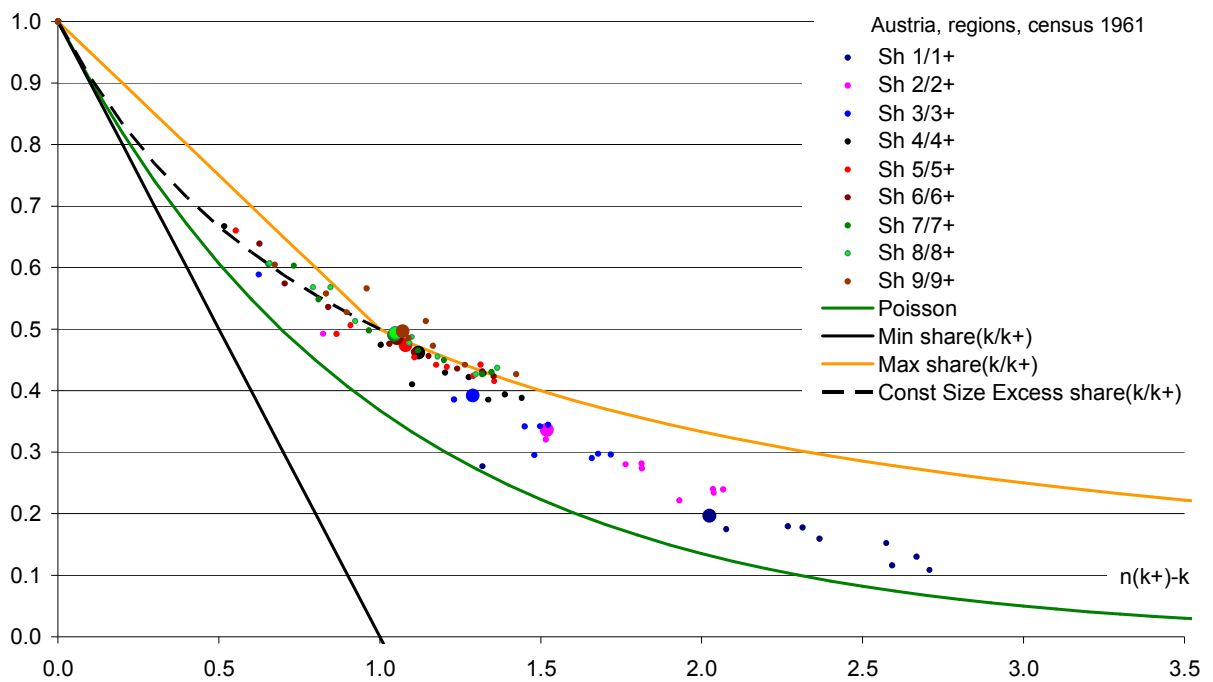


Fig. 1. Census population of Austria and regions, 1961. Conditional shares of households of different sizes ($k=1, 2, 3, \dots, 9$) among households of the same or larger sizes against the average size of such households minus k .

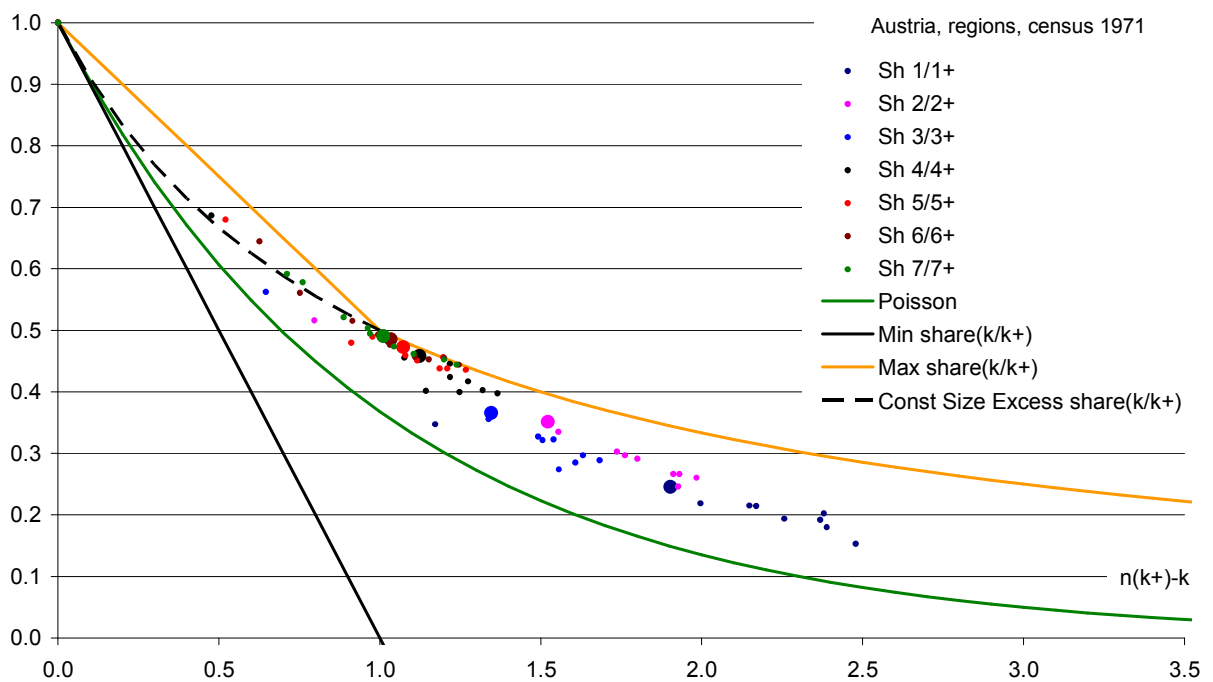


Fig. 2. Census population of Austria and regions, 1971. Conditional shares of households of different sizes ($k=1, 2, 3, \dots, 7$) among households of the same or larger sizes against the average size of such households minus k .

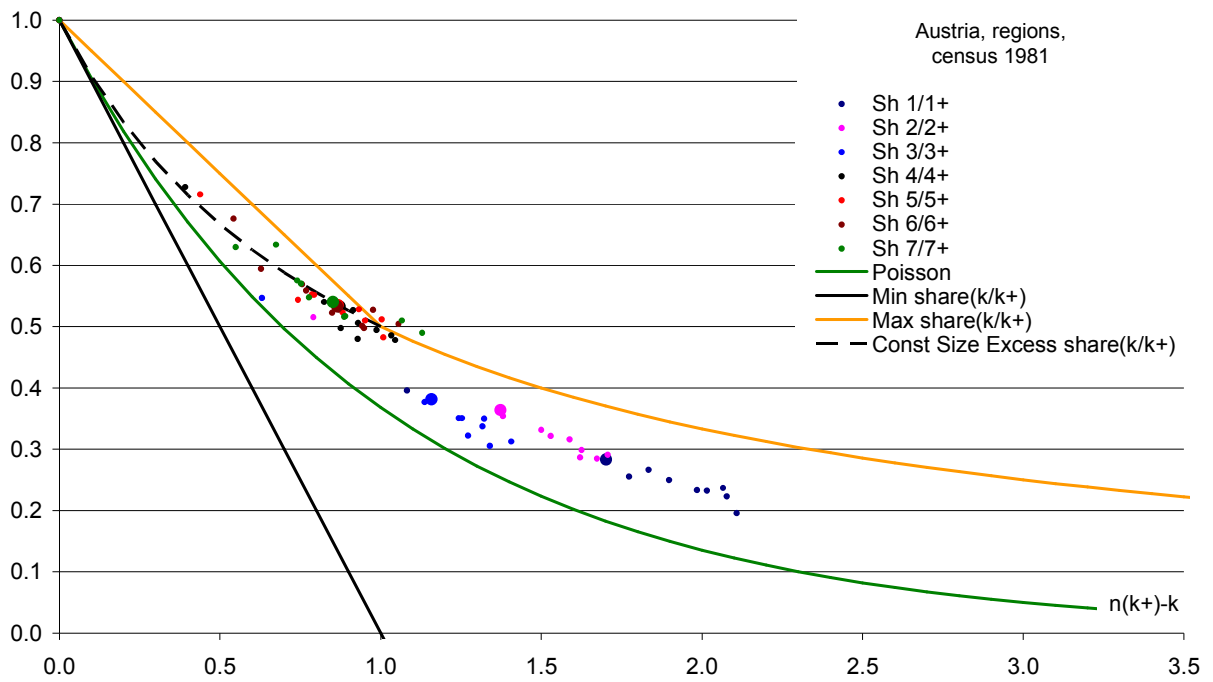


Fig. 3. Census population of Austria and regions, 1981. Conditional shares of households of different sizes ($k=1, 2, 3, \dots, 7$) among households of the same or larger sizes against the average size of such households minus k .

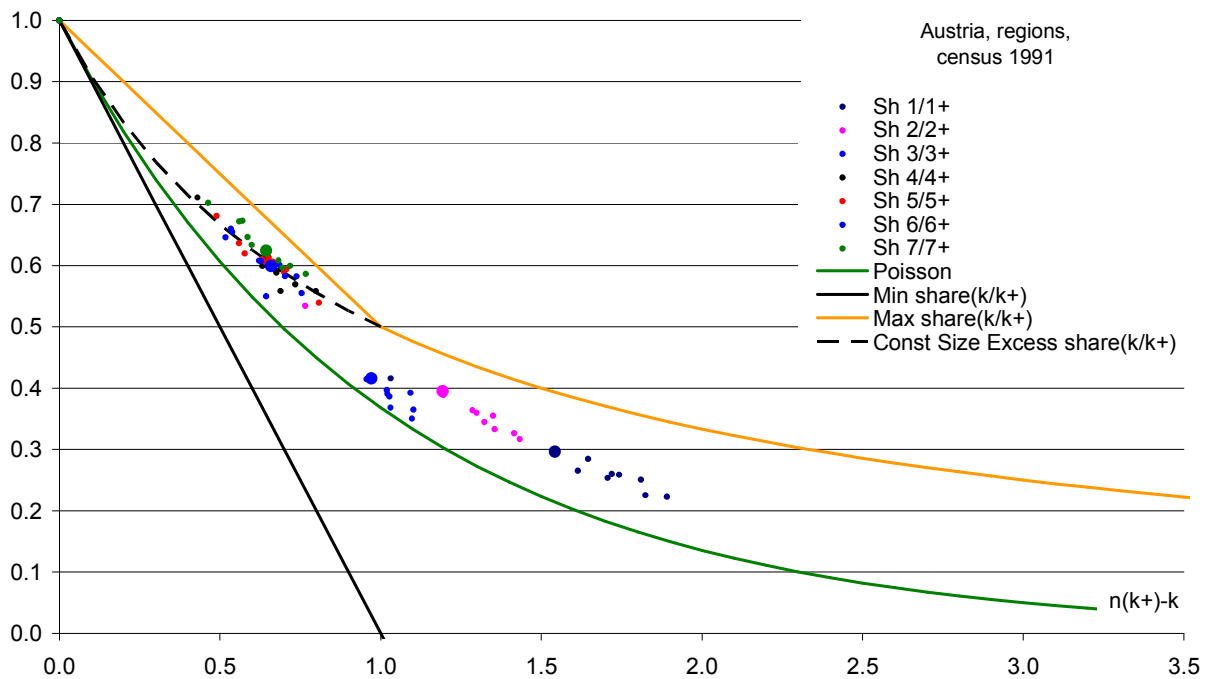


Fig. 4. Census population of Austria and regions, 1991. Conditional shares of households of different sizes ($k=1, 2, 3, \dots, 7$) among households of the same or larger sizes against the average size of such households minus k .

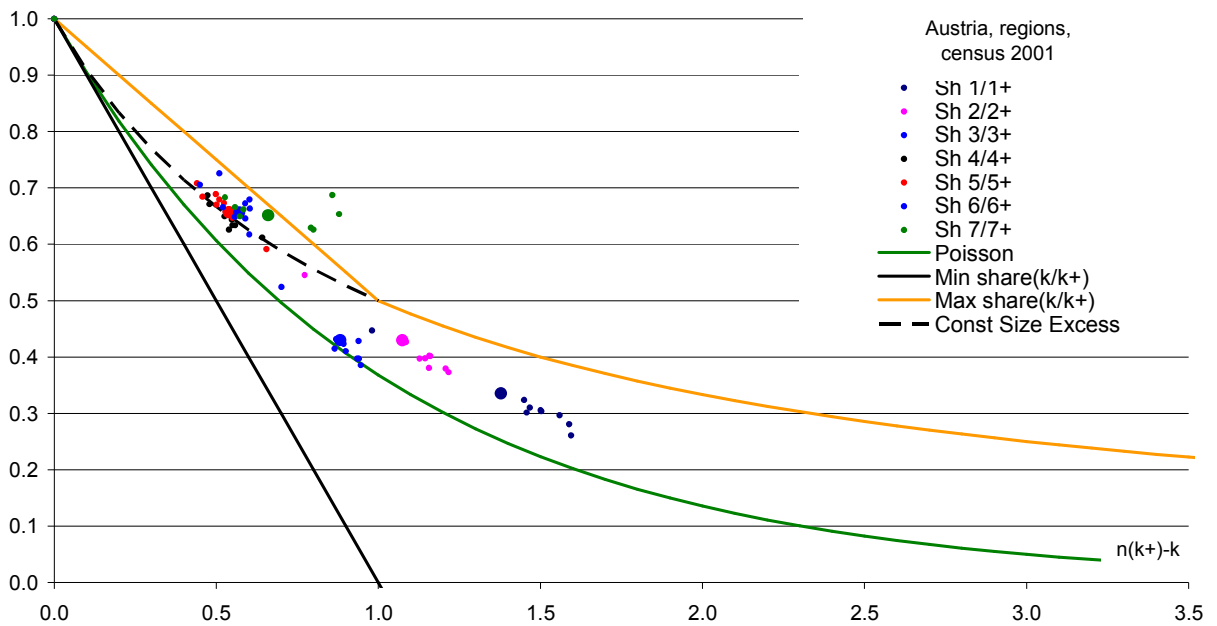


Fig. 5. Census population of Austria and regions, 2001. Conditional shares of households of different sizes ($k=1, 2, 3, \dots, 7$) among households of the same or larger sizes against the average size of such households minus k .

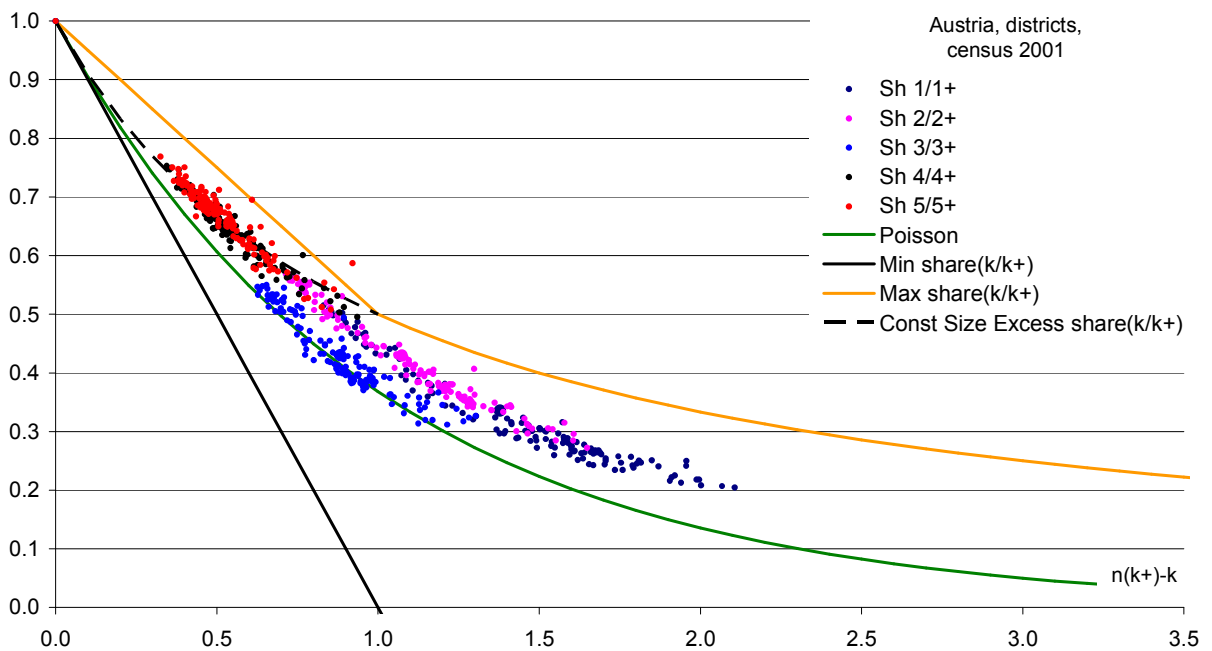


Fig. 6. Census population of Austria and census districts, 2001. Conditional shares of households of different sizes ($k=1, 2, 3, 4, \text{ or } 5$) among households of the same or larger sizes against the average size of such households minus k .

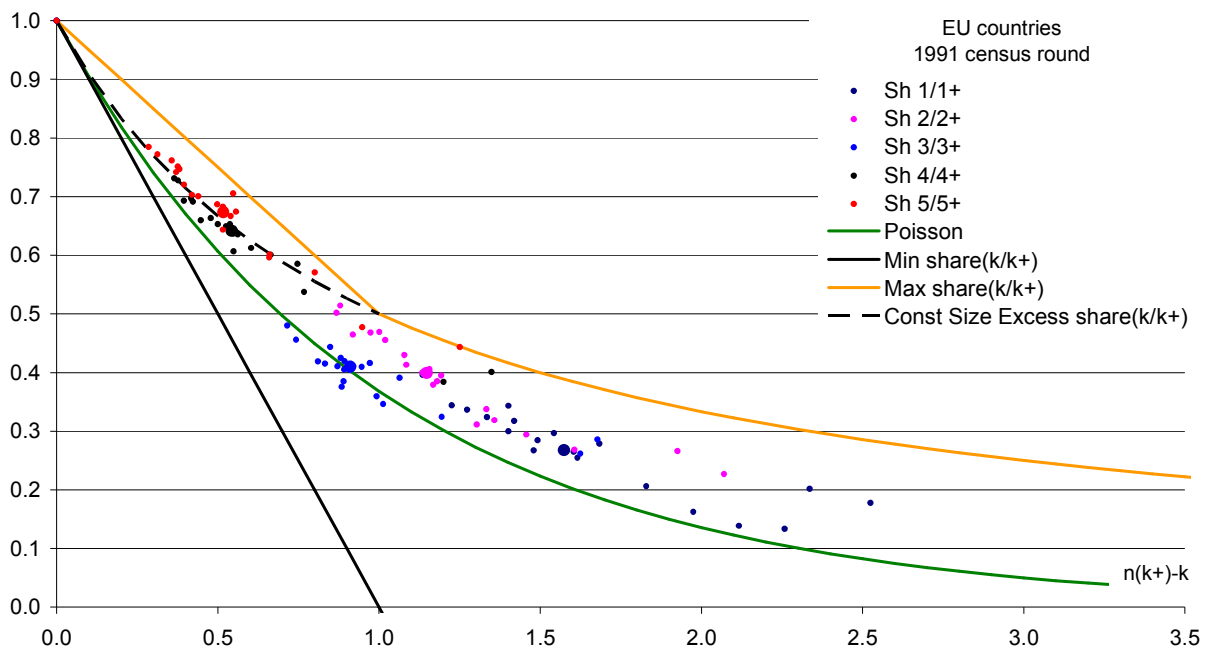


Fig. 7. Census populations of EU countries, 1991 census round. Conditional shares of households of different sizes ($k=1, 2, 3, 4$, or 5) among households of the same or larger sizes against the average size of such households minus k .

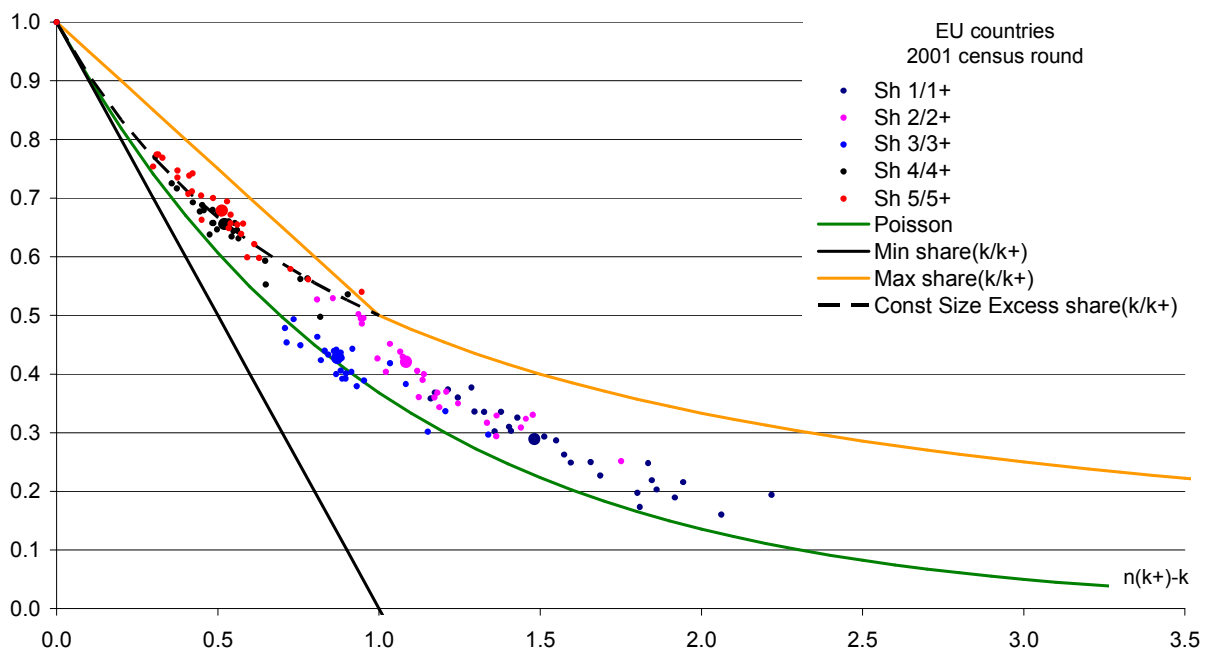


Fig. 8. Census populations of EU countries, 2001 census round. Conditional shares of households of different sizes ($k=1, 2, 3, 4$, or 5) among households of the same or larger sizes against the average size of such households minus k .

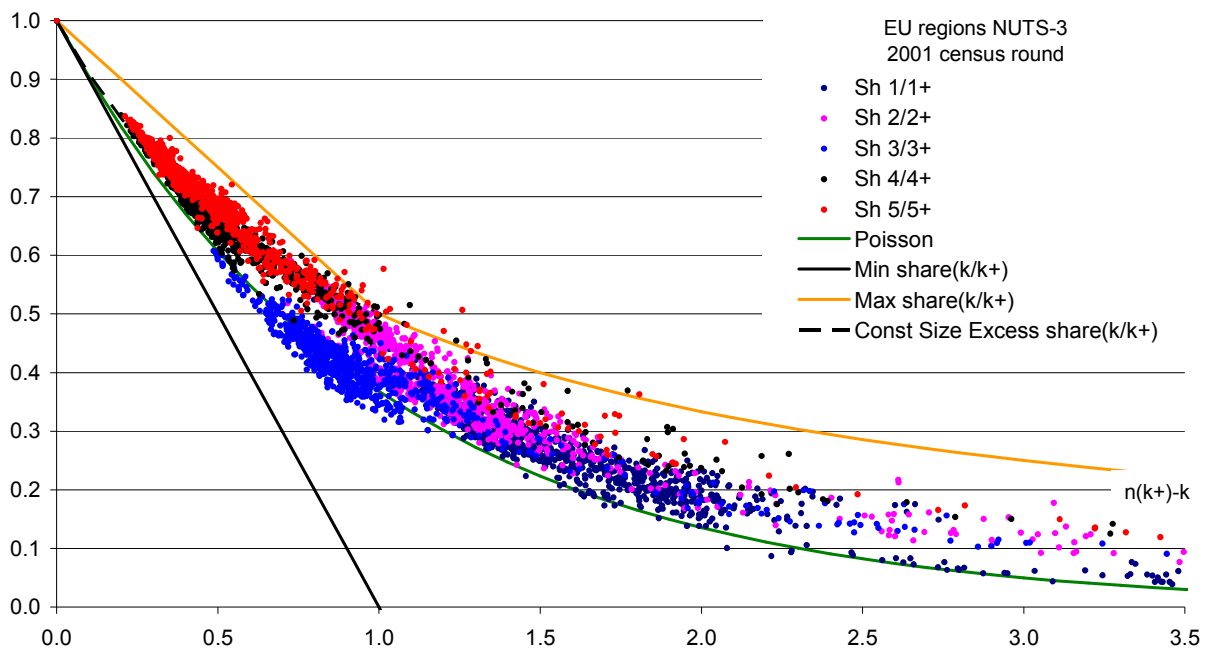


Fig. 9. Census populations of EU countries and regions, NUTS-3 level, 2001 census round. Conditional shares of households of different sizes ($k=1, 2, 3, 4,$ or 5) among households of the same or larger sizes against the average size of such households minus k .

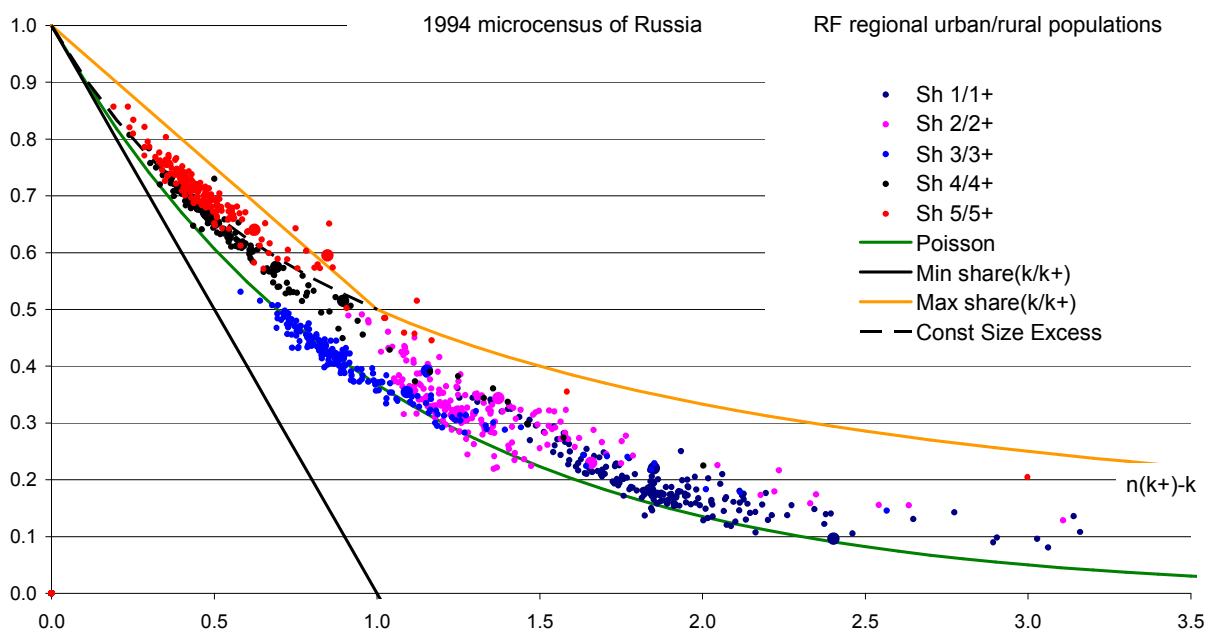


Fig. 10. Urban and rural micro-census populations of Russian regions, 1994. Conditional shares of households of different sizes ($k=1, 2, 3, 4,$ or 5) among households of the same or larger sizes against the average size of such households minus k .

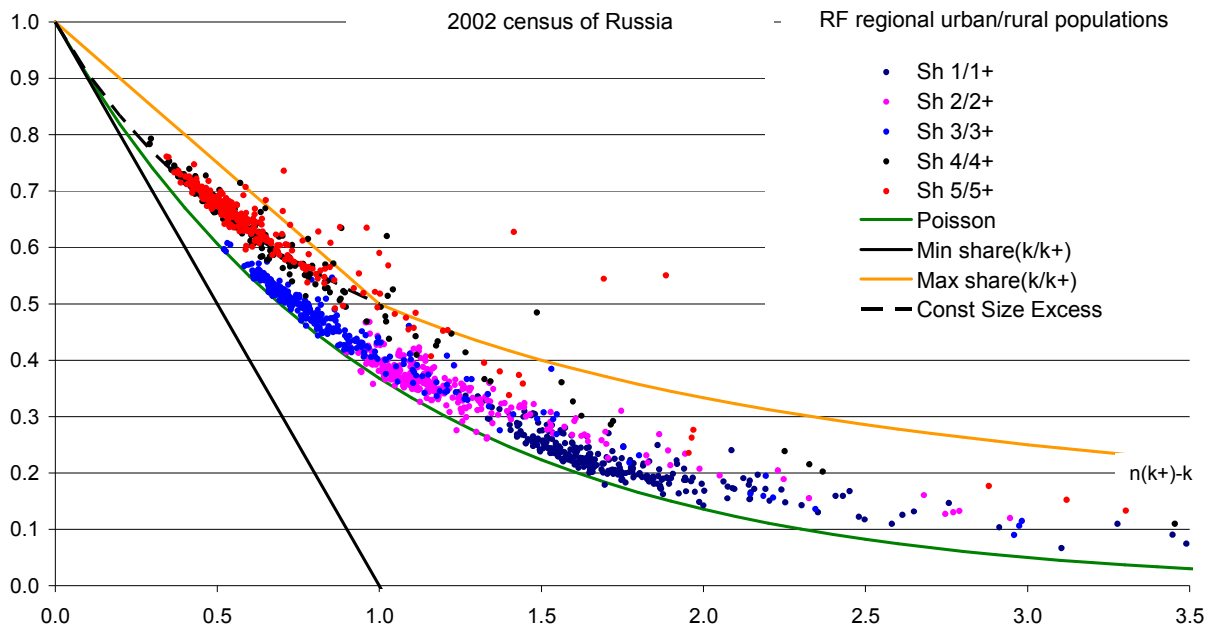


Fig. 11. Urban and rural census populations of Russian regions, 2002. Conditional shares of households of different sizes ($k=1, 2, 3, 4,$ or 5) among households of the same or larger sizes against the average size of such households minus k .

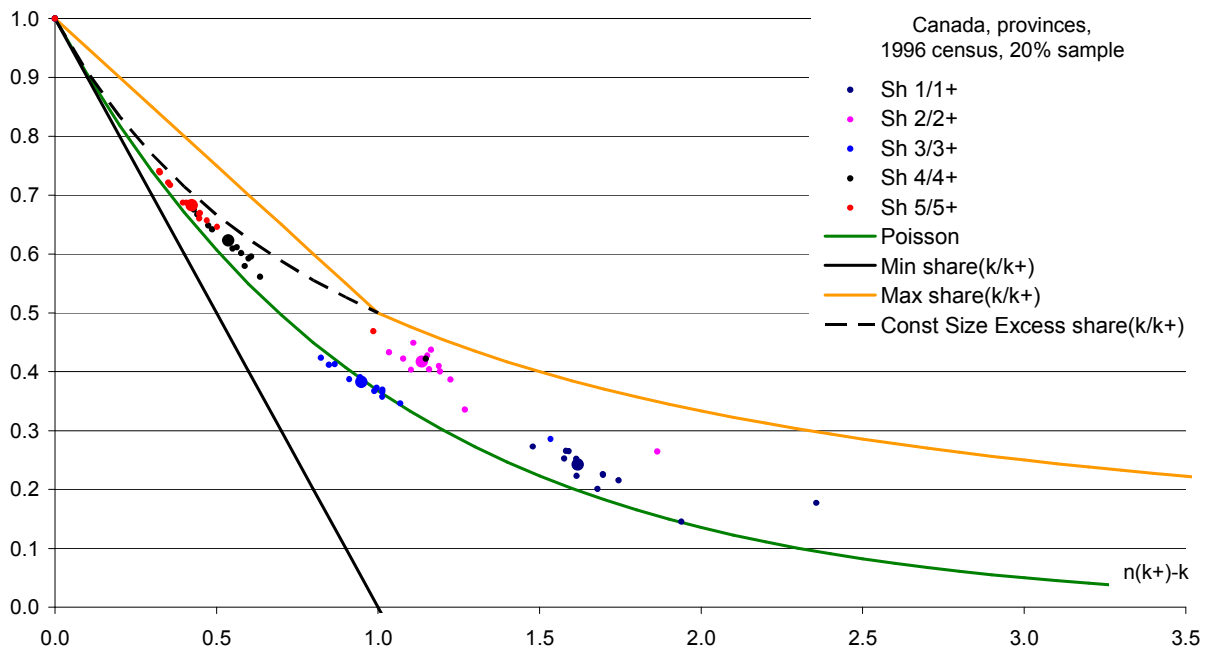


Fig. 12. Census populations of Canada and of its regions, 1996 census, 20% sample data. Conditional shares of households of different sizes ($k=1, 2, 3, 4,$ or 5) among households of the same or larger sizes against the average size of such households minus k .

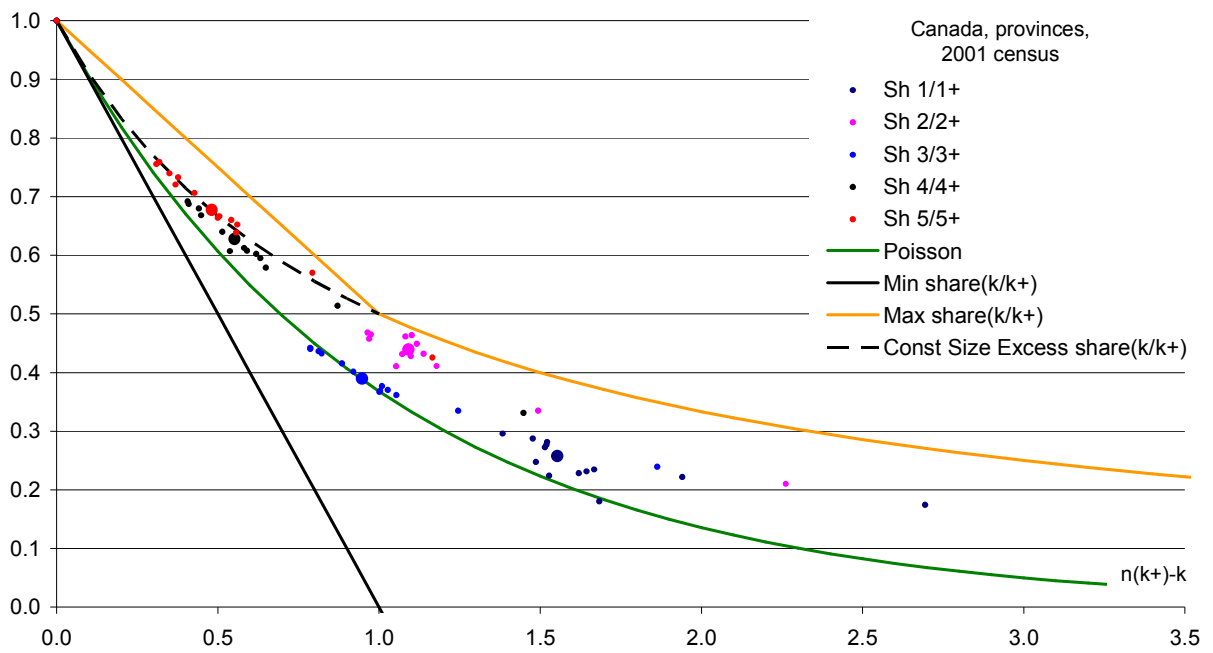


Fig. 13. Census populations of Canada and of its regions, 2001 census. Conditional shares of households of different sizes ($k=1, 2, 3, 4,$ or 5) among households of the same or larger sizes against the average size of such households minus k .

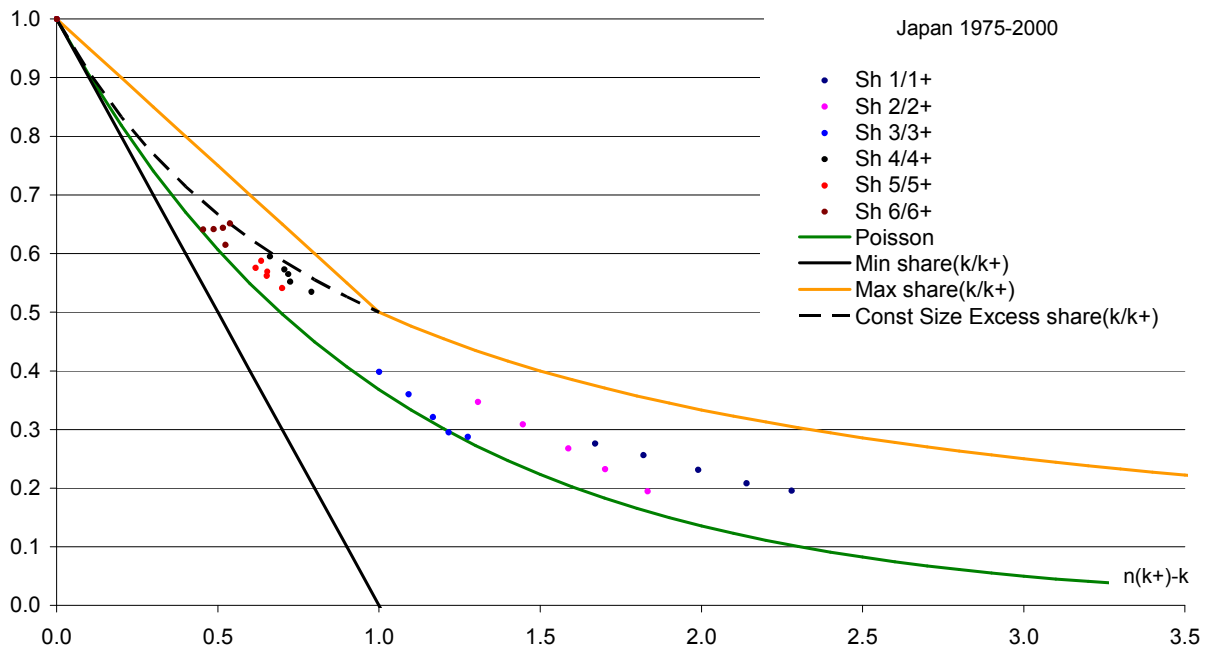


Fig. 14. 1975-2000 Census populations of Japan. Conditional shares of households of different sizes ($k=1, 2, 3, 4,$ or 5) among households of the same or larger sizes against the average size of such households minus k .

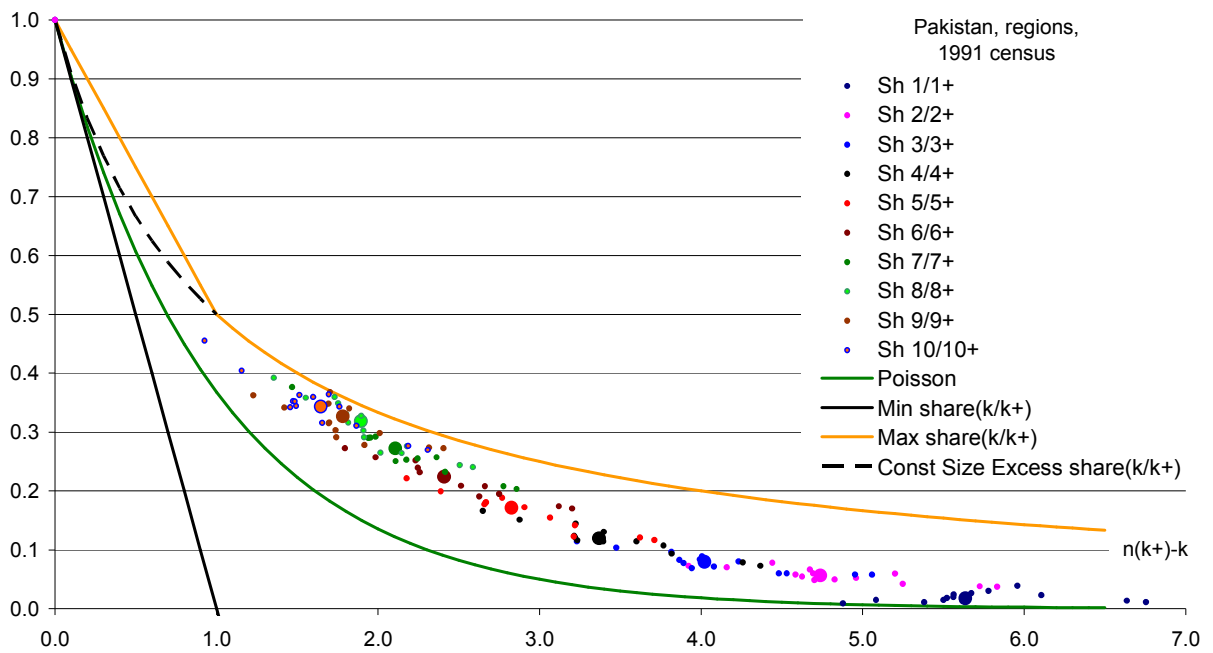


Fig. 15. Census populations of Pakistan and of its regions, 2001 census. Conditional shares of households of different sizes ($k=1, 2, 3, \dots, 10$) among households of the same or larger sizes against the average size of such households minus k .

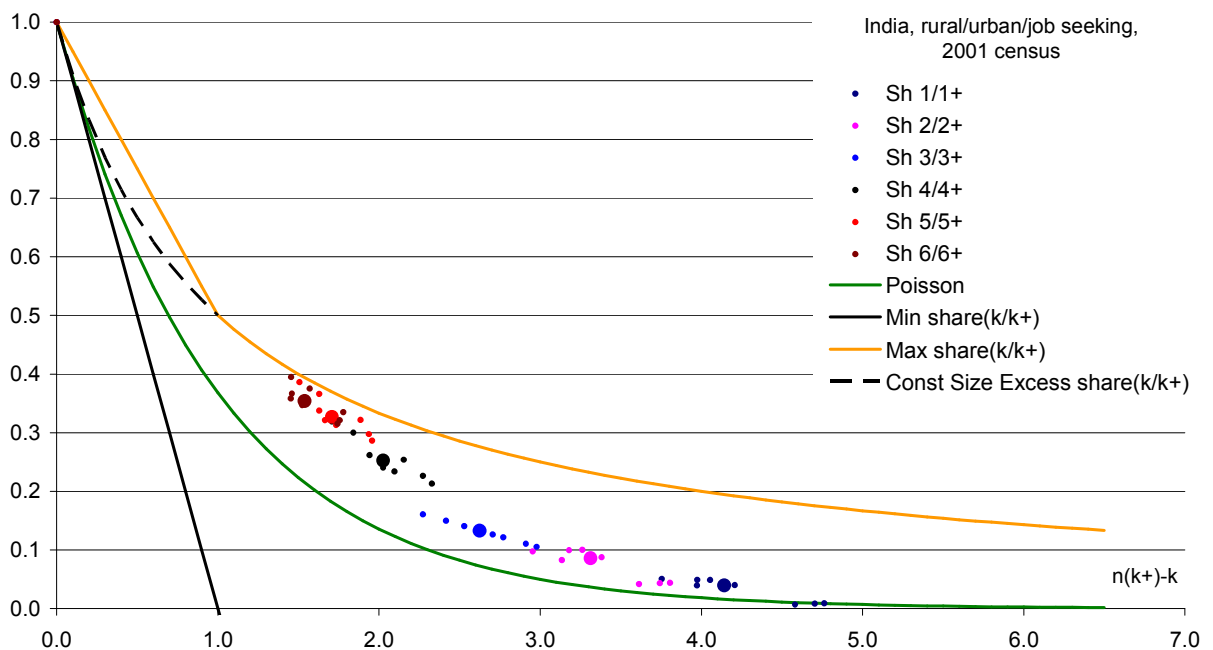


Fig. 16. Census populations of India and subpopulations grouped according to the residence status and job-seeking household members available, 2001 census. Conditional shares of households of different sizes ($k=1, 2, 3, \dots, 6$) among households of the same or larger sizes against the average size of such households minus k .

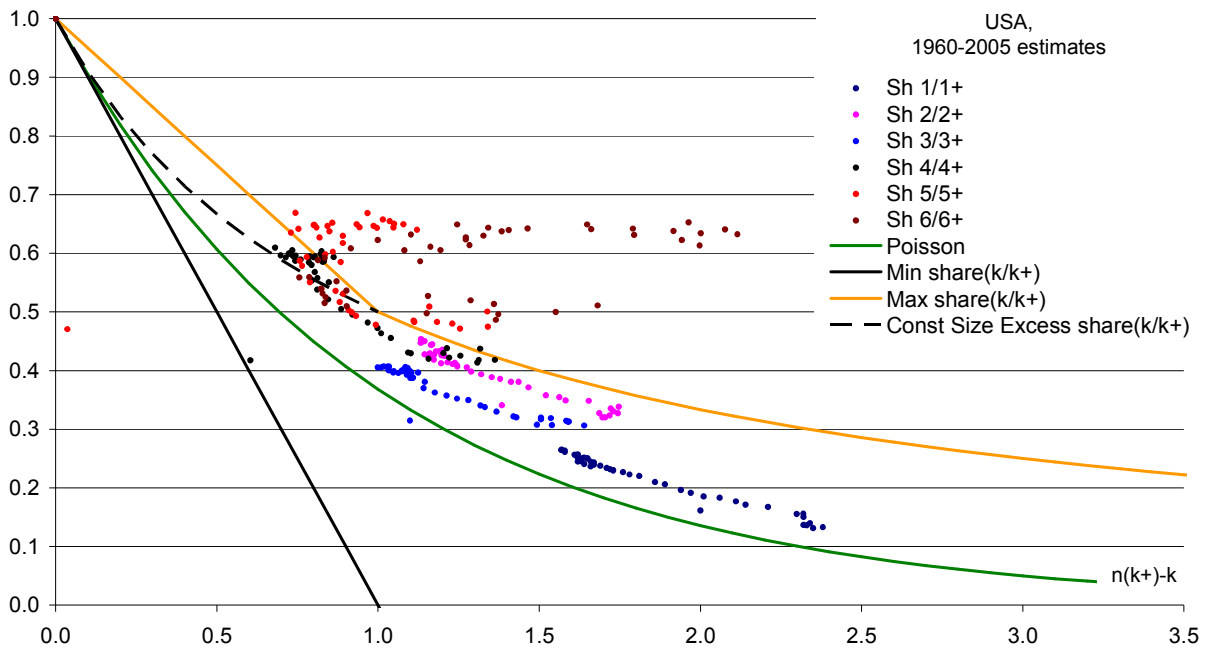


Fig. 17. Estimates for USA in 1960-2005. Conditional shares of households of different sizes ($k=1, 2, 3, \dots, 6$) among households of the same or larger sizes against the average size of such households minus k .

4. Projection Methodology for Distributing Households by Size

Based on theoretical study and empirical observations, several approaches to projecting the distribution of households by size are proposed here. The first approach (' α -method') relies on exponential-type relations between the conditional average sizes and shares of households observed in empirical data. Two other approaches (the ' η -method' and the ' θ -method') are based on relation (11), which allows deriving the conditional shares from conditional average household sizes. The last approach (' ν -method') relies on directly projecting the conditional shares based on knowledge about the average household size. For each of these methods three more variants may be proposed: (a) the simplest approach, with model parameters fixed at levels observed from the last census (or survey/simulation) available; (b) estimating model parameters from regressions derived from data available by regions and/or for different periods; (c) applying regional adjustments to model parameters estimated from regressions. As a result, we have a variety of alternative methods, which may fit different needs for projecting the households depending on data availability, heterogeneity of the population addressed, phase of social development that the population has achieved, etc. All methods work within the same recurrent framework of deriving households of size k only after having derived households of size $k-1$. Here we consider in more detail the first method, while other methods and their comparison to the α -method are discussed a little further down. The recurrent procedure, which equally applies to all methods proposed, is described in detail in the following section concerning the α -method.

4.1. The α -method

Apart from certain country peculiarities seen on the graphs above, one may note that exponential functions may be a good fit for the conditional shares of households (5) as functions of the conditional average sizes. This expectation is supported by empirical observations as illustrated, for example, in Fig. 18. The Figure presents conditional shares for Austrian districts in 2001 presented above in Fig. 6, with absolute values of logarithmic shares put on the graph instead of the shares themselves. One may observe that an approximation of logarithmic shares with straight lines crossing the origin of the coordinates seems to work well³:

$$\ln(v_{k/k+}) = -\alpha_k \cdot \eta_k . \quad (18)$$

One may also observe from Fig. 18 that considering the slopes α_k as depending on the conditional average size may provide some improvements in the approximation.

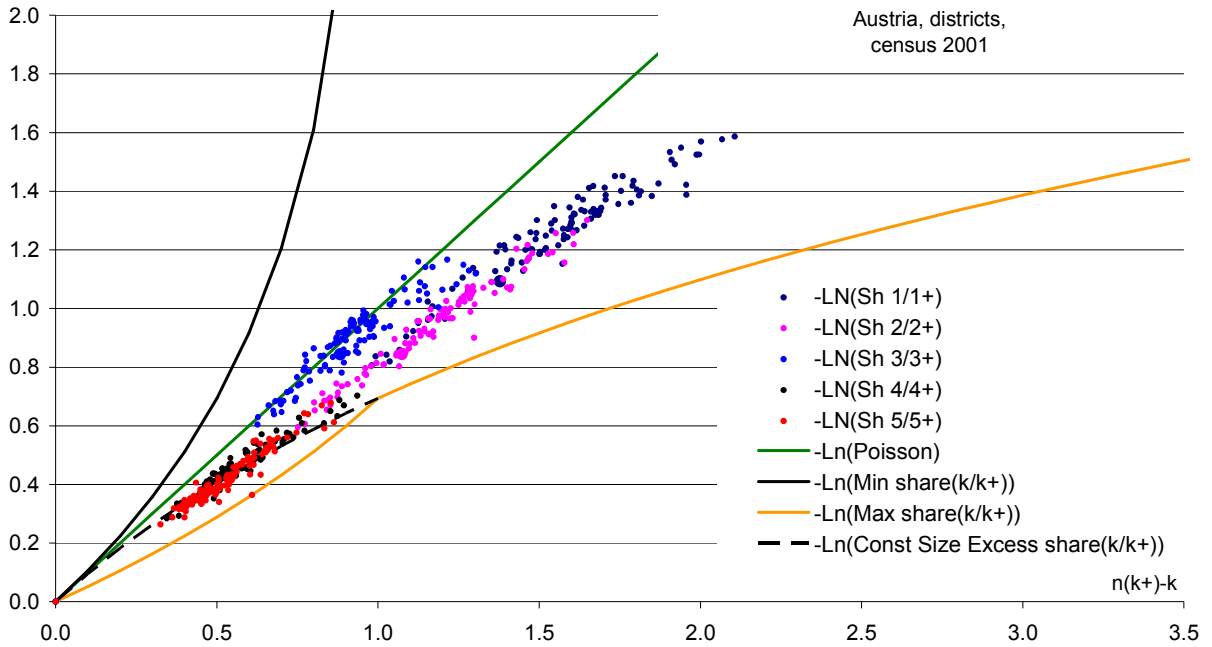


Fig. 18. Census population of Austria and census districts, 2001. Absolute logarithmic conditional shares of households of different sizes ($k=1, 2, 3, 4,$ or 5) among households of the same or larger sizes against the average size of such households minus k .

Having obtained a projection of the slopes α_k for each population group of interest, we may apply them for projecting the distribution of households by size in those population groups:

$$v_{k/k+} = e^{-\alpha_k \cdot \eta_k} . \quad (19)$$

Note that for slopes above unity (i.e., higher than for the Poisson shares) and for small

³ Interestingly, the same slope α seems to fit almost all household sizes. We do not follow this possibility here, although it might be useful at some occasions, especially for extending households distributions to larger sizes.

conditional excesses in size η_k , Eq. (19) may lead to shares below the minimal possible level (9), (10). In that case conditional shares should be set to the minimal level $v_{k/k+} = 1 - \eta_k$. Hence, for practical calculations (19) might be replaced by

$$v_{k/k+} = \max \begin{cases} e^{-\alpha_k \cdot \eta_k}, \\ 1 - \eta_k. \end{cases} \quad (20)$$

Calculations should start with the smallest size of households (we start with one-person households, treating all the households within the same framework):

$$v_{1/1+} = \max \begin{cases} e^{-\alpha_1 \cdot \eta_1}, \\ 1 - \eta_1, \end{cases} \quad (21)$$

where $\eta_1 = n_{1+} - 1 = n - 1 = \frac{N}{H} - 1$ as all the households under study should have a size equal to or larger than the smallest size. After estimating the projected conditional share of smallest households we may use relation (11) to obtain conditional average excess in size for the next smallest households:

$$\eta_2 = \frac{\eta_1}{1 - v_{1/1+}} - 1 \quad (22)$$

and then proceed to the conditional shares using (20):

$$v_{2/2+} = \max \begin{cases} e^{-\alpha_2 \cdot \eta_2}, \\ 1 - \eta_2. \end{cases} \quad (23)$$

This recurrent procedure is to be applied until there are no more households to be distributed ($v_{k/k+} = 1$) or we have reached households of the largest size K of interest, when the conditional share of largest households among these same households is to be set equal to unity by definition:

$$v_{K+/K+} = 1. \quad (24)$$

Once, we have projections of the total population and of the total number of households of interest, the recurrent procedure described will provide projections for all conditional shares and average sizes. Then one may derive unconditional shares of households of different sizes, using, e.g., Eq. (3)⁴. Finally, the numbers of households by size may be obtained from the unconditional shares:

$$H_k = v_k \cdot H. \quad (25)$$

In the method presented one starts by projecting the slopes α_k . As mentioned above, three approaches may be developed to do that. Further on we investigate these possibilities. We use census data presented above and projections made by Statistics Austria for the years 1990, 2000 from the 1981 census (Gisser 1986a, 1986b), and for 2001 from the 1991 census (Hanika 1997) to study the efficiency of the projecting techniques

⁴ It might be more convenient in practice to use the following relations too: $v_1 = v_{1/1+}$;

$$v_k = v_{k/k+} \cdot v_{k-1} \left(\frac{1}{v_{(k-1)/(k-1+)}} - 1 \right), \quad k = 2, 3, 4, \dots$$

proposed. Projections by Statistics Austria are important, as they are based on the aforementioned method proposed by Gisser, which was used as a starting point for the methods presented here.

(a) *Constant Slopes*

As already noted, regression lines approximating the logarithmic conditional shares should cross the origin of the coordinates. Hence, these lines have only one parameter (the slope), which can be obtained directly from the latest census or survey available:

$$\alpha_k = \tilde{\alpha}_k = -\frac{\ln(\tilde{v}_{k/k+})}{\tilde{\eta}_k}. \quad (26)$$

Hereinafter we use the ‘~’ mark for initial data used in estimating model parameters (e.g., census, survey, or simulation results).

Figs. 19-21 present conditional shares of households of different sizes for Austria in 1991 and 2001 according to the censuses and projections using the standard technique of Statistics Austria and the technique presented above. Census data are represented by dots, official projections by hyphens, and projections obtained using (20)-(25) and (26) by circles. Note that we use the total population and number of households from the official projection of Statistics Austria. Hence, any errors of official projection as regards the overall number of households are imputed in our projection too. Tables 1, 3, and 5 present projection errors for the official projection of households. Tables 2, 4, and 6, in turn, present errors for household distributions obtained using the method proposed here. Note that for Austria as a whole two approaches are presented. First, Austria is considered a separate entity using the projecting method as described above. Secondly, Austrian figures are obtained as totals of regional numbers.

Both the official projection and the method presented here result in errors of the same magnitude. In 1990 the official projection was more precise in general. Note, however, that one-person households were projected separately in the official projection, and only multi-person households were distributed by size after projecting their overall number and average size. In the method presented here all households are distributed in the same way based on a single parameter—the average size of households. This might explain the higher accuracy of the official projection of one-person households in 1990. Our results, however, were more accurate for 2- and 5-person households. Also, in two other projections (2000 from 1981 and 2001 from 1991) the method proposed here was more accurate. The method for projecting the distribution of households presented here seems to be more robust as it is seen from projection results for 2000 (from census 1989, i.e., with a 20-year projecting horizon). This is true even for one-person households, although we apply to them only the general method based on the average size of households, without taking into account the dynamics of the headship rates for such households.

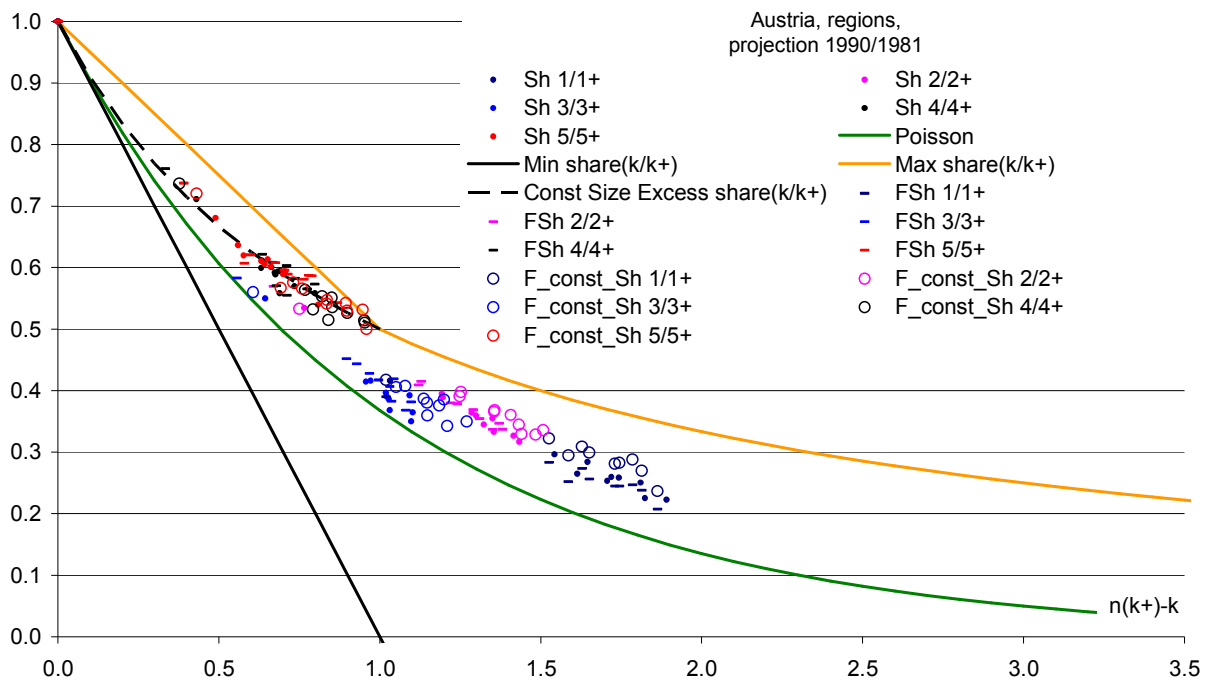


Fig. 19. Austria, projections to 1990 from 1981 versus census 1991. Shares of households of different sizes ($k=1, 2, 3, 4,$ or 5) among households of the same or larger sizes against the average size of such households minus k . Census data are represented by dots, official projections by hyphens, and projections obtained using (20)-(25) and (26) by circles.

Table 1. Projection errors of Statistics Austria for the distribution of households in 1990 compared to census figures in 1991, base census for the projection: 1981.

Region	Austria, regions, projection 1990/1981									
	Austria	Burgenland	Carinthia	Lower Austria	Upper Austria	Salzburg	Styria	Tyrol	Vorarlberg	Vienna
Household size										
<i>Absolute percentage errors in numbers of households</i>										
1	6%	10%	7%	7%	2%	5%	5%	2%	8%	11%
2	5%	1%	3%	5%	5%	6%	6%	6%	5%	5%
3	3%	3%	1%	5%	7%	6%	7%	2%	1%	3%
4	6%	8%	5%	8%	7%	5%	2%	6%	7%	9%
5	10%	2%	1%	12%	12%	14%	7%	6%	8%	24%
6+	8%	7%	2%	11%	6%	3%	8%	3%	2%	42%
Average	6.6%	5.2%	3.4%	8.1%	6.6%	6.5%	5.8%	4.1%	5.2%	15.8%
<i>Absolute percentage errors in shares of households</i>										
1	5%	8%	6%	5%	1%	4%	5%	2%	7%	6%
2	7%	4%	5%	7%	4%	7%	6%	7%	4%	11%
3	5%	6%	3%	7%	6%	6%	7%	2%	0%	2%
4	5%	5%	3%	6%	8%	5%	2%	6%	8%	4%
5	9%	4%	0%	11%	13%	13%	7%	5%	9%	20%
6+	7%	10%	4%	10%	7%	3%	9%	3%	3%	39%
Average	6.1%	6.2%	3.5%	7.6%	6.5%	6.3%	6.0%	4.0%	5.2%	13.8%

Table 2. Errors of distributing the households by size in 1990 according to (20)-(26). Projections are compared to 1991 census figures. Average sizes of households are based on the projection by Statistics Austria based on 1981 census.

Region	Austria, total	Austria, as a separate reg.	Burgenland	Carinthia	Lower Austria	Upper Austria	Salzburg	Styria	Tyrol	Vorarlberg	Vienna
Household size											
<i>Absolute percentage errors in numbers of households</i>											
1	6%	7%	2%	7%	9%	19%	8%	9%	14%	22%	5%
2	4%	5%	5%	5%	5%	5%	2%	1%	3%	2%	6%
3	7%	8%	6%	9%	8%	7%	7%	5%	10%	9%	4%
4	11%	12%	13%	13%	12%	14%	11%	10%	14%	13%	4%
5	3%	3%	5%	4%	1%	4%	7%	0%	3%	7%	11%
6+	19%	22%	30%	32%	19%	25%	24%	17%	25%	21%	26%
Average	8.3%	9.2%	10.2%	11.5%	9.2%	12.1%	9.9%	7.1%	11.7%	12.2%	9.1%
<i>Absolute percentage errors in shares of households</i>											
1	8%	9%	5%	9%	11%	18%	9%	9%	15%	21%	0%
2	3%	3%	3%	3%	4%	6%	1%	2%	2%	3%	1%
3	5%	6%	3%	8%	6%	7%	6%	6%	10%	10%	2%
4	10%	10%	11%	11%	11%	14%	11%	11%	14%	14%	1%
5	1%	1%	7%	5%	1%	5%	6%	1%	2%	7%	6%
6+	21%	24%	34%	34%	21%	24%	25%	16%	26%	20%	22%
Average	7.9%	8.8%	10.5%	11.7%	9.0%	12.3%	9.7%	7.2%	11.6%	12.4%	5.3%

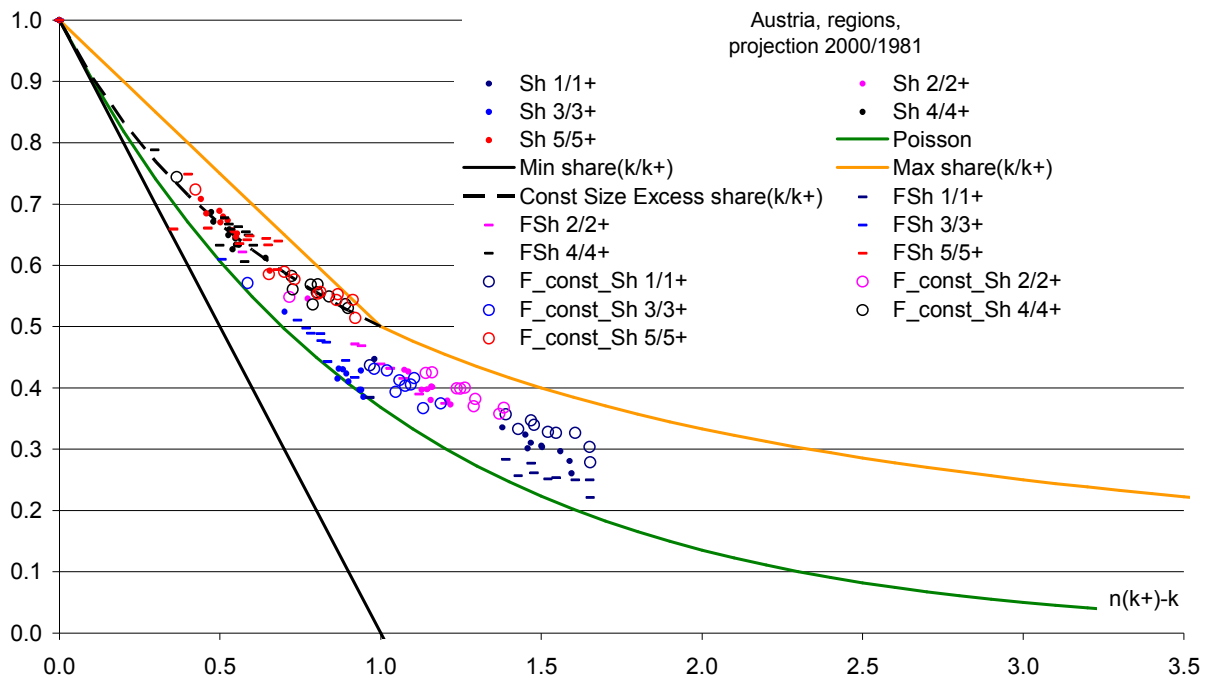


Fig. 20. Austria, projections to 2000 from 1981 versus census 2001. Shares of households of different sizes ($k=1, 2, 3, 4$, or 5) among households of the same or larger sizes against the average size of such households minus k . Census data are represented by dots, official projections by hyphens, and projections obtained using (20)-(25) and (26) by circles.

Table 3. Projection errors of the Statistics Austria for the distribution of households in 2000 compared to census figures in 2001, base census for the projection: 1981.

Region	Austria	Burgenland	Carinthia	Lower Austria	Upper Austria	Salzburg	Styria	Tyrol	Vorarlberg	Vienna
Household size										
<i>Absolute percentage errors in numbers of households</i>										
1	20%	22%	22%	21%	15%	17%	20%	19%	15%	23%
2	11%	1%	8%	10%	14%	13%	13%	9%	2%	13%
3	10%	2%	10%	10%	21%	13%	14%	10%	9%	4%
4	13%	14%	8%	19%	14%	12%	5%	10%	8%	22%
5	24%	2%	8%	29%	27%	23%	19%	13%	6%	50%
6+	18%	10%	5%	28%	15%	2%	19%	1%	17%	64%
Average	16.1%	8.5%	10.1%	19.5%	17.5%	13.3%	15.1%	10.4%	9.3%	29.4%
<i>Absolute percentage errors in shares of households</i>										
1	15%	15%	18%	15%	14%	14%	18%	16%	11%	14%
2	18%	8%	13%	18%	16%	16%	16%	14%	3%	27%
3	16%	11%	15%	18%	23%	17%	17%	15%	14%	8%
4	8%	7%	4%	13%	12%	9%	3%	6%	4%	13%
5	19%	6%	3%	24%	26%	21%	17%	10%	2%	43%
6+	14%	19%	0%	23%	13%	1%	17%	3%	22%	60%
Average	15.1%	11.1%	9.0%	18.5%	17.3%	13.1%	14.7%	10.5%	9.2%	27.4%

Table 4. Errors of distributing households by size in 2000 according to (20)-(25) and (26). Projections are compared to 2001 census data. Average sizes of households are based on the projection by Statistics Austria based on 1981 census.

Region	Austria, total	Austria, as a separate reg.	Burgenland	Carinthia	Lower Austria	Upper Austria	Salzburg	Styria	Tyrol	Vorarlberg	Vienna
Household size											
<i>Absolute percentage errors in numbers of households</i>											
1	1%	0%	2%	0%	3%	10%	4%	5%	5%	3%	13%
2	9%	10%	13%	11%	11%	7%	6%	6%	10%	13%	9%
3	7%	8%	13%	11%	9%	3%	8%	8%	13%	9%	2%
4	18%	19%	21%	17%	20%	20%	17%	15%	20%	16%	12%
5	6%	5%	8%	10%	8%	9%	7%	0%	1%	2%	29%
6+	39%	44%	66%	70%	27%	50%	67%	37%	72%	72%	43%
Average	13.2%	14.4%	20.5%	19.8%	13.0%	16.4%	18.2%	12.0%	19.9%	19.0%	17.9%
<i>Absolute percentage errors in shares of households</i>											
1	5%	6%	7%	5%	11%	12%	7%	7%	10%	8%	2%
2	4%	4%	5%	6%	5%	6%	3%	4%	6%	9%	2%
3	1%	3%	6%	6%	2%	1%	5%	6%	9%	5%	10%
4	13%	14%	14%	13%	15%	19%	15%	13%	16%	12%	1%
5	0%	0%	17%	16%	1%	7%	4%	3%	4%	6%	20%
6+	47%	53%	80%	79%	36%	52%	72%	41%	79%	80%	36%
Average	11.8%	13.4%	21.6%	20.8%	11.6%	16.1%	17.7%	12.3%	20.6%	19.9%	12.0%

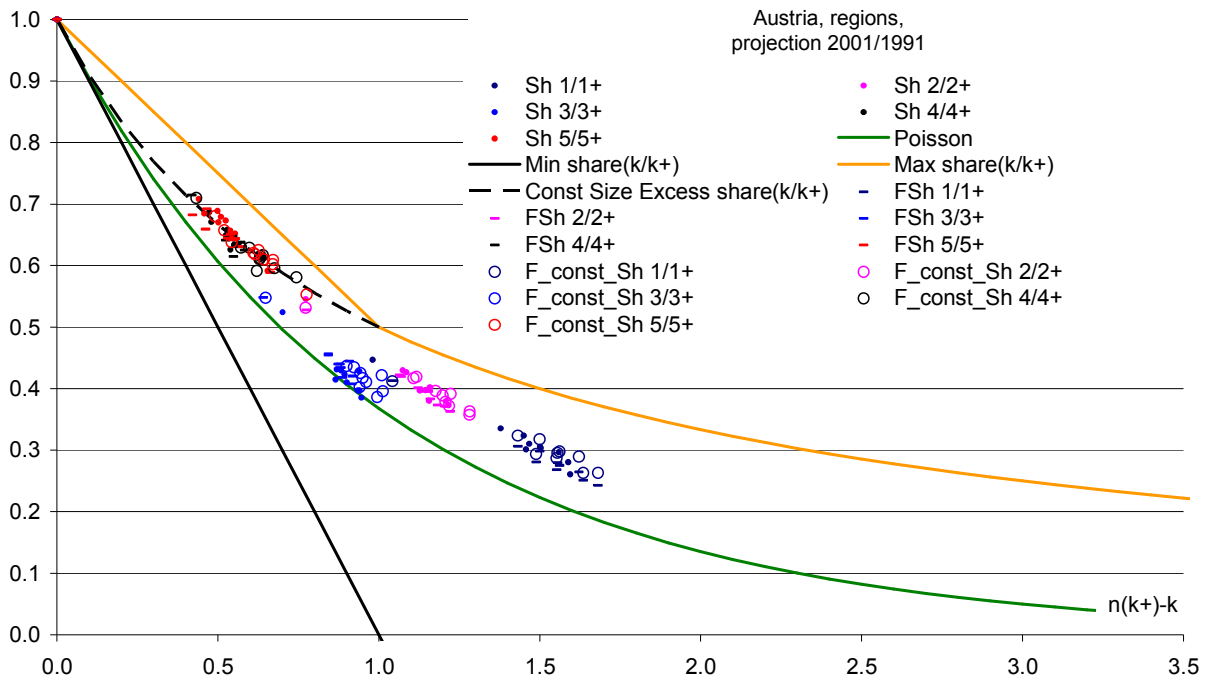


Fig. 21. Census population of Austria, 2001 and projections to year 2001 from census 1991. Conditional shares of households of different sizes ($k=1, 2, 3, 4,$ or 5) among households of the same or larger sizes against the average size of such households minus k . Census data are represented by dots, official projections by hyphens, and projections obtained using (20)-(25) and (26) by circles.

Another test example is the case of EU countries, for nineteen of which we have data on household distribution by size for two census rounds (1991 and 2001) (Eurostat 2006). As we have no forecasts to compare them with, the method is applied using the slopes from the first census and average household sizes from the next one. Hence, in spite of using the projection of the number of households and of population, we use actual figures. At the same time, the distribution of households by size is estimated solely by the method presented, based on the average size of households only. Summary results for absolute percentage errors are presented in Table 7. Errors for Austria are slightly lower than those presented in Table 6, as we used actual household size in spite of taking it from the projection.

In general, the method works well enough and—given its simplicity, robustness, and safety for inconsistency—it may be used for projecting and estimating the distribution of households by size in different research contexts. Although errors for largest households seem to be high on the percentage basis, one should note that the absolute number of such households is extremely low. Therefore, errors in absolute numbers are much lower.

Table 5. Projection errors of the Statistics Austria for the distribution of households in 2001 compared to 2001 census figures, base census for the projection: 1991.

Region	Austria	Burgenland	Carinthia	Lower Austria	Upper Austria	Salzburg	Styria	Tyrol	Vorarlberg	Vienna
Household size	Austria	Burgenland	Carinthia	Lower Austria	Upper Austria	Salzburg	Styria	Tyrol	Vorarlberg	Vienna
<i>Absolute percentage errors in numbers of households</i>										
1	10%	7%	12%	9%	12%	9%	11%	13%	16%	7%
2	1%	4%	2%	1%	4%	2%	2%	1%	3%	3%
3	11%	0%	6%	10%	16%	9%	8%	8%	10%	16%
4	0%	6%	1%	3%	3%	3%	2%	3%	1%	9%
5	3%	4%	6%	5%	7%	3%	0%	4%	0%	3%
6+	4%	16%	13%	1%	4%	15%	1%	9%	13%	8%
Average	4.8%	6.0%	6.7%	4.7%	7.7%	6.9%	4.1%	6.5%	7.3%	7.6%
<i>Absolute percentage errors in shares of households</i>										
1	9%	4%	10%	7%	12%	8%	10%	11%	13%	8%
2	3%	1%	1%	1%	5%	3%	3%	4%	1%	3%
3	12%	3%	9%	12%	16%	10%	9%	11%	14%	15%
4	1%	3%	2%	1%	2%	2%	3%	1%	2%	9%
5	2%	7%	9%	3%	7%	2%	0%	2%	3%	3%
6+	5%	20%	16%	1%	5%	16%	1%	12%	17%	8%
Average	5.2%	6.2%	7.8%	4.2%	7.7%	6.9%	4.6%	6.7%	8.4%	7.6%

Table 6. Errors of distributing households by size in 2001 according to (20)-(25) and (26). Projections are compared to 2001 census data. Average sizes of households are based on the projection by Statistics Austria based on 1991 census.

Region	Austria, total	Austria, as a separate region	Burgenland	Carinthia	Lower Austria	Upper Austria	Salzburg	Styria	Tyrol	Vorarlberg	Vienna
Household size	Austria, total	Austria, as a separate region	Burgenland	Carinthia	Lower Austria	Upper Austria	Salzburg	Styria	Tyrol	Vorarlberg	Vienna
<i>Absolute percentage errors in numbers of households</i>											
1	6%	5%	2%	7%	4%	6%	3%	3%	5%	9%	8%
2	1%	2%	6%	6%	3%	1%	3%	2%	5%	7%	4%
3	5%	4%	5%	1%	4%	8%	1%	0%	1%	3%	15%
4	2%	3%	6%	3%	4%	4%	5%	2%	6%	3%	8%
5	0%	2%	5%	8%	1%	3%	1%	4%	0%	4%	3%
6+	20%	25%	29%	30%	13%	22%	34%	22%	32%	31%	4%
Average	5.9%	6.6%	8.9%	9.2%	5.0%	7.3%	7.9%	5.7%	8.1%	9.5%	6.8%
<i>Absolute percentage errors in shares of households</i>											
1	5%	4%	1%	5%	2%	5%	2%	3%	2%	6%	8%
2	0%	1%	3%	3%	1%	1%	1%	1%	2%	3%	4%
3	6%	5%	2%	4%	6%	9%	2%	1%	2%	6%	15%
4	1%	2%	3%	0%	2%	4%	4%	1%	3%	0%	8%
5	2%	3%	8%	11%	0%	2%	3%	5%	3%	7%	3%
6+	22%	26%	34%	34%	15%	23%	36%	23%	36%	36%	4%
Average	6.0%	6.7%	8.5%	9.5%	4.6%	7.3%	8.0%	5.8%	8.0%	9.8%	6.8%

Table 7. Absolute percentage errors in numbers of households of selected EU countries distributed by size according to (20)-(26). Projections are based on actual average sizes of households in 2001-round censuses with distribution parameters taken from 1991-round censuses.

Household size Region	Denmark	Germany	Greece	Spain	France	Ireland	Italy	Luxembourg	Netherlands	Austria	Portugal	Finland	United Kingdom	Liechtenstein	Norway	Switzerland	AVERAGE
1	2%	3%	4%	6%	1%	22%	1%	5%	2%	1%	7%	0%	2%	4%	1%	3%	4.0%
2	1%	8%	3%	11%	2%	8%	0%	3%	1%	0%	3%	5%	1%	6%	3%	1%	3.4%
3	15%	10%	2%	1%	4%	13%	1%	13%	2%	3%	6%	9%	1%	3%	6%	11%	6.2%
4	3%	2%	3%	10%	1%	15%	5%	3%	6%	6%	8%	9%	4%	2%	7%	1%	5.4%
5	16%	7%	2%	5%	4%	4%	1%	13%	8%	4%	2%	10%	6%	7%	10%	10%	6.9%
6+	25%	9%	7%	14%	2%	36%	29%	6%	17%	16%	31%	20%	8%	2%	15%	17%	15.8%
Aver.	10%	7%	4%	8%	2%	16%	6%	7%	6%	5%	10%	9%	4%	4%	7%	7%	6.9%

(b) Regression of the logarithmic slope across regions with similar demographics without regional corrections

As noted above, some corrections of the logarithmic conditional shares' slopes α_k might result in improving the accuracy of projecting the distribution of households by size. One may note that patterns presented in Fig. 18, for example, exhibit some non-linear trends which were ignored in the simplistic method outlined above. It seems that it was this ignorance that resulted in the accumulation of errors in the projection of the households of highest size. Therefore we introduce the second approach to estimating the slopes α_k , which may be used when data on different populations with similar demographic patterns are available. One may develop a regression of the slopes α_k against the average size of households from the data available:

$$\alpha_k = a_k + b_k \cdot \eta_k + c_k \cdot (\eta_k)^2. \quad (27)$$

These regressions are to be used as substitution for relations (26). Alternatively, one may use the conditional average excess for households 1+ as a regression variable:

$$\alpha_k = a_k + b_k \cdot \eta_1 + c_k \cdot (\eta_1)^2. \quad (28)$$

When applied to the data available, no significant outperforming of any of these two regressions was observed. Also, one may consider both the ordinary and the weighted least squares methods for estimating parameters in (27), (28). Ordinary LS reflect all the variety observed in the data, while weighted LS may ignore important outliers depicting the overall trend. On the other hand, ordinary LS may be sensitive to random fluctuations if regions with tiny populations are concerned.

Further on we present results for ordinary LS estimates for regression (28), which is simpler to operate in practice. Regression coefficients a_k , b_k , and c_k are to be estimated from empirical observations. In the case of Austria, for example, we derived estimates from census data on the distribution of households in 121 administrative districts. Results from estimating the regression parameters in (28) for Austria as of the 1981, 1991, and 2001 censuses, are presented in Table 8. Although in some cases parameters of the

regression line are not significant, in most of cases presented in Table 8, regression (28) explains a significant part of the variation in the regional slopes α_k . At the same time, an equally significant part of the variance may not be explained by regression, which provides the rationale for further development of the models (see the next subsection).

In the case where the regression constant a_k exceeds unity, another regression might be better to fit, setting $a_k = 1$ in order to prevent projections from getting close to the lower-bound constraint (9), (10):

$$\alpha_k = 1 + b_k \cdot \eta_k + c_k \cdot (\eta_k)^2 \quad (29)$$

or, if regression (27) is implemented,

$$\alpha_k = 1 + b_k \cdot \eta_1 + c_k \cdot (\eta_1)^2. \quad (30)$$

Having obtained slopes (28) for households with minimal size, one may project the number of these households and the average excess in size for the next group of households and proceed with the recurrent procedure as outlined in (20)-(25). Efficiency of such an approach in the case of Austria is presented in Fig. 22-24 and in Tables 9-11. A projection based on the proposed method seems to be more accurate and robust compared to the traditional method, especially in the long run, as can be seen from the results for year 2000 (Fig. 23 and Table 10). It is worth noting that the method was more effective even for one-person households, which were addressed more rigorously in the projections by Statistics Austria.

Table 8. Parameters of regression (28) estimated from census data on the distribution of households by size in *administrative districts* of Austria. k is the smallest size of households, for which the parameters are presented, a , b , and c are regression coefficients, and $se(\cdot)$ is the standard error for the appropriate coefficient.

Year:	1981					1991					2001				
	k	a	b	c	$se(b)$	$se(c)$	a	b	c	$se(b)$	$se(c)$	a	B	C	$se(b)$
1	0.94	-0.06	-0.01	0.03	0.01	0.72	0.22	-0.09	0.05	0.02	0.55	0.44	-0.17	0.07	0.02
2	0.85	0.01	-0.03	0.04	0.01	0.67	0.22	-0.08	0.04	0.01	0.60	0.29	-0.10	0.06	0.02
3	0.98	0.03	-0.04	0.04	0.01	0.65	0.44	-0.15	0.07	0.02	0.49	0.70	-0.24	0.11	0.04
4	0.87	-0.03	-0.01	0.04	0.01	0.67	0.21	-0.08	0.06	0.02	0.69	0.19	-0.07	0.10	0.03
5	0.77	0.03	-0.02	0.05	0.01	0.69	0.14	-0.05	0.06	0.02	0.86	-0.13	0.05	0.12	0.04
6	0.71	0.06	-0.02	0.06	0.02	0.71	0.08	-0.02	0.09	0.03	1.07	-0.59	0.24	0.20	0.07
7	0.70	0.07	-0.02	0.10	0.03	0.59	0.15	-0.03	0.12	0.04	1.43	-1.24	0.47	0.34	0.12

The regression approach introduced in this section is of course based on the assumption that the dataset used in estimating regression parameters is more or less uniform. If we put populations into the dataset with significantly different profiles concerning family formation, housing, etc., the projection results might be poor. To illustrate this point we present errors of projections made for the EU countries mentioned above with regression lines (28) estimated from data for all these countries combined in a single dataset. The errors are presented in Table 12. As can be seen in the table, the regression approach in general performs worse compared to the simpler ‘constant slopes’ approach presented above (Table 7). As a general rule one should use the regression approach whenever data are available for a number of subpopulations with similar demographic and housing patterns, and the simpler ‘constant slopes’ approach otherwise.

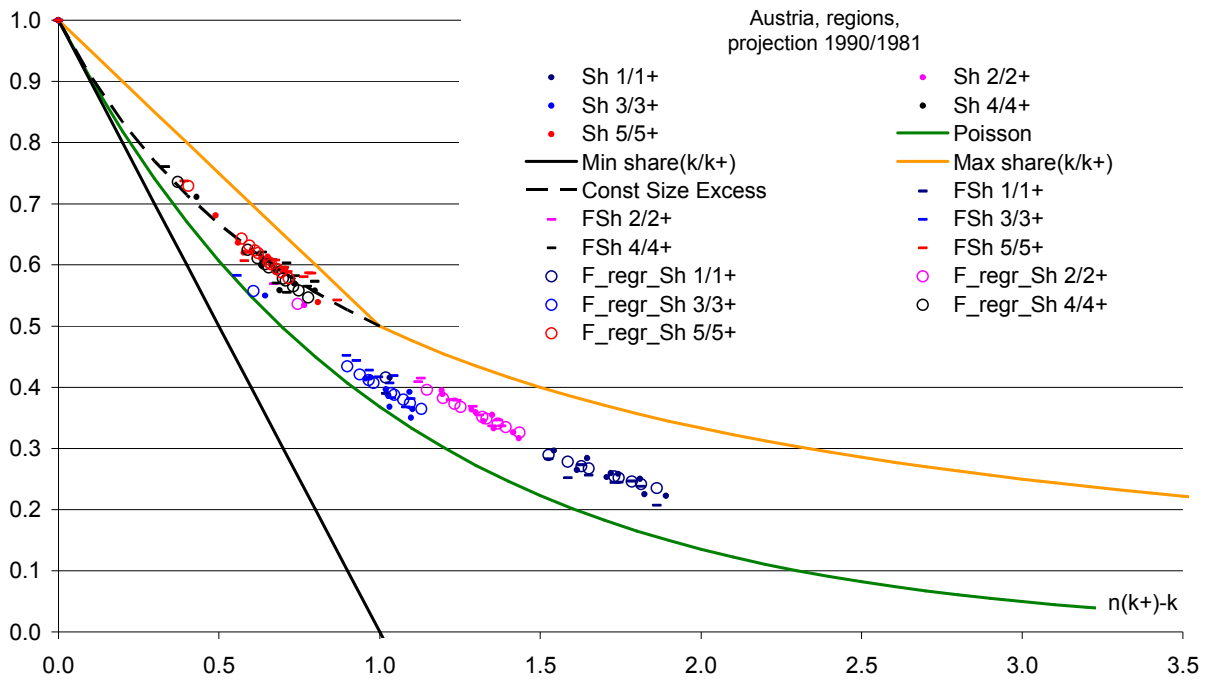


Fig. 22. 1991 census of Austria and projections to 1990 from 1981. Conditional shares of households of different sizes ($k=1, 2, 3, 4$, or 5) among households of the same or larger size against the average size of such households minus k . Census data are represented by dots, official projections by hyphens, and projections (20)-(25) and (28) by circles.

Table 9. Errors of distributing the households by size in 1990 according to (20)-(25) and (28). Projections are compared to 1991 census. Average sizes of households are based on the projection by Statistics Austria based on 1981 census.

Region	Austria, total	Austria, as a separate reg.	Burgenland	Carinthia	Lower Austria	Upper Austria	Salzburg	Styria	Tyrol	Vorarlberg	Vienna
Household size											
	<i>Absolute percentage errors in numbers of households</i>										
1	1%	4%	2%	4%	3%	6%	5%	1%	2%	9%	5%
2	2%	0%	6%	1%	5%	0%	5%	0%	5%	4%	5%
3	0%	4%	4%	3%	1%	2%	6%	1%	2%	2%	4%
4	3%	0%	11%	3%	4%	3%	4%	6%	5%	8%	4%
5	1%	3%	3%	8%	0%	4%	6%	9%	3%	15%	9%
6+	8%	19%	23%	2%	9%	9%	10%	16%	3%	4%	28%
Average	2.6%	5.1%	8.0%	3.5%	3.8%	4.2%	5.8%	5.6%	3.4%	7.1%	9.3%
	<i>Absolute percentage errors in shares of households</i>										
1	1%	2%	4%	3%	5%	5%	5%	2%	2%	9%	0%
2	0%	1%	3%	2%	4%	1%	6%	0%	5%	3%	0%
3	1%	5%	1%	2%	1%	2%	6%	1%	3%	1%	1%
4	1%	2%	8%	1%	2%	4%	3%	5%	5%	9%	1%
5	1%	2%	6%	9%	2%	5%	5%	9%	3%	16%	4%
6+	7%	18%	26%	4%	7%	10%	9%	17%	3%	4%	24%
Average	1.8%	5.0%	8.2%	3.6%	3.3%	4.4%	5.6%	5.5%	3.2%	6.8%	5.1%

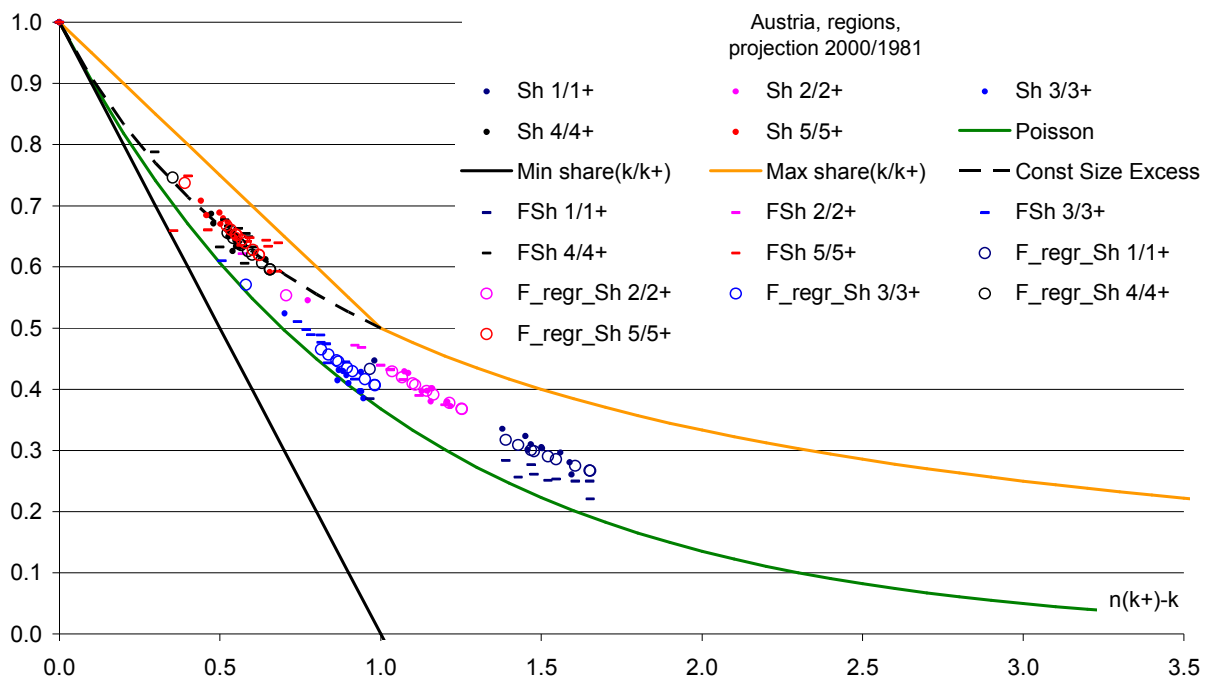


Fig. 23. 2001 census of Austria and projections to 2000 from 1981. Conditional shares of households of different sizes ($k=1, 2, 3, 4$, or 5) among households of the same or larger sizes against the average size of such households minus k . Census data are represented by dots, official projections by hyphens, and projections (20)-(25) and (28) by circles.

Table 10. Errors of distributing the households by size in 2000 according to (20)-(25) and (28). Projections are compared 2001 census data. Average sizes of households are based on the projection by Statistics Austria based on 1981 census.

Region	Austria, regions, projection 2000/1981										
	Austria, total	Austria, as a separate reg.	Burgenland	Carinthia	Lower Austria	Upper Austria	Salzburg	Styria	Tyrol	Vorarlberg	Vienna
Household size											
<i>Absolute percentage errors in numbers of households</i>											
1	9%	11%	6%	13%	4%	3%	10%	7%	11%	9%	14%
2	4%	3%	12%	3%	9%	0%	3%	1%	0%	6%	7%
3	2%	5%	9%	1%	1%	10%	7%	2%	3%	5%	2%
4	9%	7%	16%	6%	12%	9%	10%	1%	9%	9%	12%
5	8%	11%	5%	9%	11%	13%	10%	4%	5%	10%	29%
6+	6%	15%	40%	13%	14%	5%	6%	17%	16%	35%	46%
Average	6.5%	8.6%	14.6%	7.6%	8.5%	6.8%	7.6%	5.5%	7.4%	12.1%	18.5%
<i>Absolute percentage errors in shares of households</i>											
1	3%	5%	2%	8%	3%	1%	7%	5%	7%	5%	3%
2	1%	3%	4%	2%	3%	2%	7%	1%	4%	1%	4%
3	8%	11%	1%	4%	8%	12%	10%	5%	7%	10%	10%
4	3%	2%	9%	1%	6%	8%	7%	4%	5%	5%	1%
5	3%	5%	14%	15%	4%	11%	7%	7%	1%	6%	20%
6+	1%	10%	52%	19%	8%	4%	9%	15%	21%	41%	40%
Average	3.3%	6.0%	13.7%	8.3%	5.3%	6.4%	7.8%	5.9%	7.7%	11.1%	13.0%

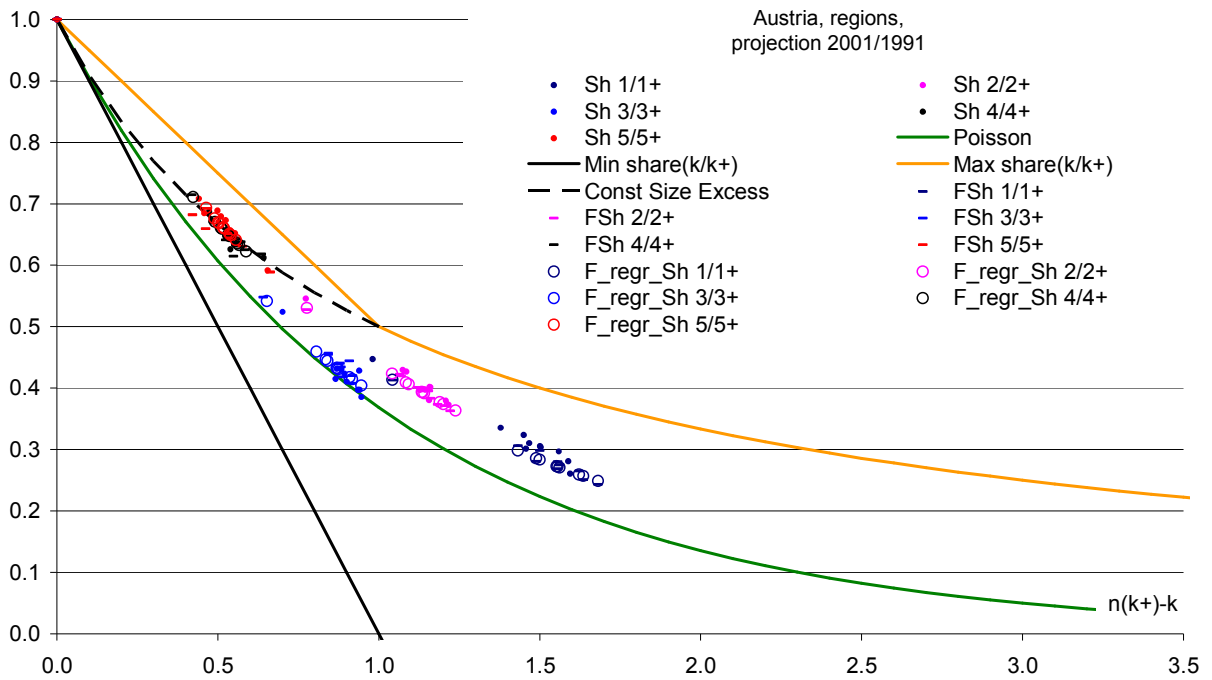


Fig. 24. 2001 census of Austria and projections to 2001 from 1991. Conditional shares of households of different sizes ($k=1, 2, 3, 4$, or 5) among households of the same or larger sizes against the average size of such households minus k . Census data are represented by dots, official projections by hyphens, and projections (20)-(25) and (28) by circles.

Table 11. Errors of distributing the households by size in 2001 according to (20)-(25) and (28). Projections are compared to census data. Average sizes of households are based on the projection by Statistics Austria based on 1991 census.

Region	Austria, regions, projection 2001/1991											
	Austria, total	Austria, as a separate reg.	Burgenland	Carinthia	Lower Austria	Upper Austria	Salzburg	Styria	Tyrol	Vorarlberg	Vienna	
Household size												
	<i>Absolute percentage errors in numbers of households</i>											
1	10%	12%	4%	15%	7%	11%	14%	12%	15%	14%	7%	
2	1%	3%	4%	1%	4%	2%	7%	1%	3%	3%	3%	
3	10%	13%	2%	3%	9%	15%	12%	6%	7%	9%	14%	
4	3%	4%	7%	2%	1%	2%	1%	11%	2%	0%	9%	
5	2%	5%	4%	8%	2%	6%	8%	7%	6%	8%	0%	
6+	5%	13%	22%	2%	10%	3%	1%	20%	4%	24%	7%	
Average	5.0%	8.2%	7.4%	5.2%	5.5%	6.3%	7.1%	9.8%	6.1%	9.9%	6.8%	
	<i>Absolute percentage errors in shares of households</i>											
1	9%	11%	1%	12%	5%	10%	12%	11%	13%	11%	8%	
2	2%	4%	1%	4%	2%	2%	8%	2%	7%	0%	3%	
3	11%	14%	1%	6%	11%	15%	13%	7%	10%	12%	13%	
4	4%	5%	4%	5%	3%	2%	0%	12%	1%	3%	9%	
5	0%	4%	8%	12%	0%	6%	7%	8%	3%	5%	1%	
6+	4%	12%	26%	5%	8%	2%	2%	20%	7%	29%	7%	
Average	5.1%	8.3%	7.0%	7.3%	5.0%	6.3%	7.2%	10.2%	6.7%	10.1%	6.8%	

Table 12. Absolute percentage errors in numbers of households of selected EU countries distributed by size according to (20)-(25) and (28). Projections are based on actual average sizes of households in 2001-round censuses with regression (28) parameters estimated from the entire dataset available for 1991-round censuses. Errors are calculated against actual figures from 2001-round censuses.

Household size Region	Denmark	Germany	Greece	Spain	France	Ireland	Italy	Luxembourg	Netherlands	Austria	Portugal	Finland	United Kingdom	Liechtenstein	Norway	Switzerland	AVERAGE
1	6%	10%	8%	1%	1%	11%	3%	6%	2%	6%	23%	0%	6%	8%	8%	0%	6.1%
2	9%	10%	4%	4%	2%	2%	8%	6%	6%	7%	5%	3%	10%	11%	13%	3%	6.4%
3	2%	13%	5%	4%	2%	15%	13%	5%	13%	1%	20%	1%	2%	3%	7%	9%	7.3%
4	2%	4%	5%	7%	11%	18%	7%	1%	0%	7%	0%	14%	9%	1%	1%	0%	5.5%
5	8%	28%	17%	8%	7%	13%	9%	9%	4%	0%	28%	3%	0%	12%	22%	3%	10.7%
6+	6%	30%	7%	13%	20%	23%	59%	12%	3%	30%	27%	26%	11%	7%	7%	8%	18.1%
Aver.	5%	16%	8%	6%	7%	14%	17%	6%	5%	9%	17%	8%	6%	7%	10%	4%	9.0%

(c) Regional corrections

As was suggested above, two methods for estimating the slopes α_k might be appropriate for two different situations. If we have regional census data with regions presumably similar in respect to demographic and household formation regimes, the more sophisticated regression approach seems to be better. On the other hand, if no such data are available or if regions differ substantially in their demographic and household profiles, then the simpler ‘constant slopes’ approach might be more robust and even more precise. The latter approach is also useful for incorporating the regional differentials observed from a census, while the regression approach incorporates only part of regional differences, which may be reflected in the regression. As can be seen from the census data on Austria presented above, however, there are regional differentials unexplained by regressions (27), (28), in what concerns the distribution of households. These differentials reflect regional peculiarities in household formation history, in population age composition, in housing availability, etc. In the projections methodology developed by Gisser (1986a, 1986b) a useful tool for incorporating such regional differences was proposed. Namely, household projections were corrected for regional deviations observed in the last census; the corrections being reduced in projection at some fixed annual rate (2% of the original correction per year). This method allows both fitting the observed differences at the year of census and reflecting the convergence of regional distributions in future. The convergence reflects both actual convergence observed from census to census and decrease of knowledge as one projects more deeply into the future. We borrow this approach by introducing regional corrections in the values of the logarithmic conditional shares’ slopes α_k :

$$\alpha_k(t) = \hat{\alpha}_k(t) + (1 - \beta)^t (\tilde{\alpha}_k - \hat{\alpha}_k(t)), \quad (31)$$

here $\alpha_k(t)$ are parameters used to project the distribution of households at year t , $\hat{\alpha}_k(t)$ are estimates of the parameters obtained using regressions (29), (30) or other methods, i.e., without consideration of convergence; $\tilde{\alpha}_k$ are parameters estimated from the most recent census; and β is annual rate of decrease of regional residuals. Having data on successive censuses, one may estimate a convergence parameter β for each size of households directly from empirical data, see Table 13.

Table 13. Least square estimates of convergence parameters β for households of different sizes from the combined dataset of the 1961, 1971, 1981, 1991, and 2001 censuses of Austria.

Household size	Annual depreciation rate	Decennial depreciation rate	Annual linear depreciation rate
1	1.9%	18%	1.8%
2	4.3%	35%	3.5%
3	2.5%	22%	2.2%
4	3.7%	32%	3.2%
5	4.0%	33%	3.3%
Average	3.3%	28%	2.8%

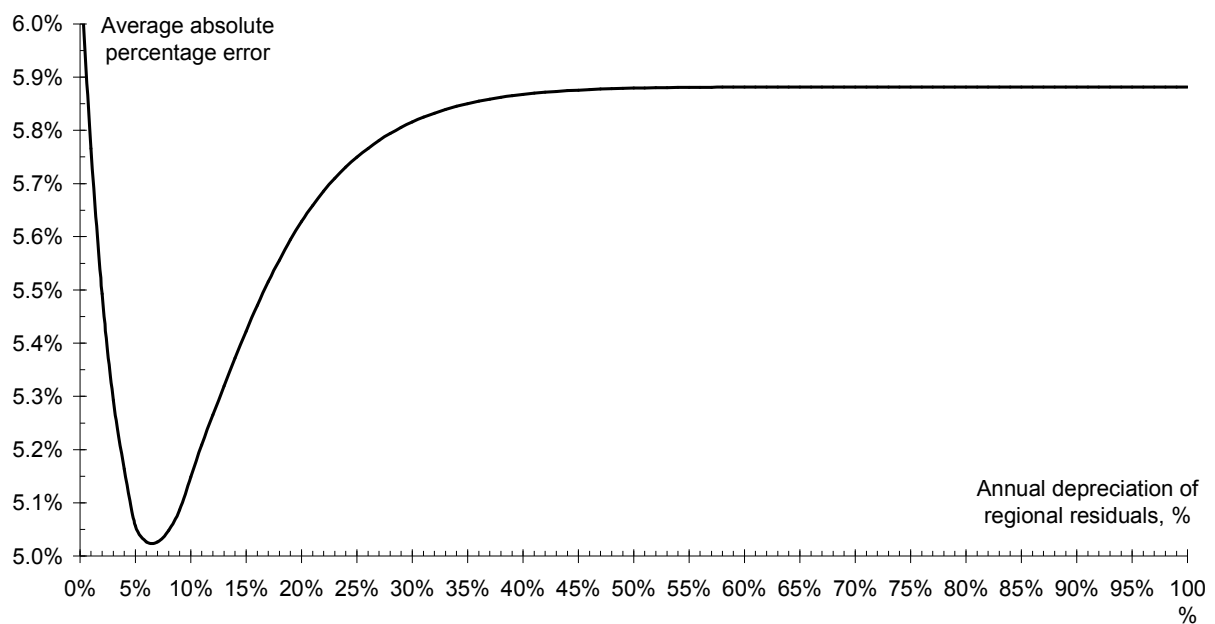


Fig. 25. Overall average decennial absolute percentage errors for three projections of Austrian distribution of households by size (1981 to 1991, 1981 to 2000, and 1991 to 2001) obtained using different convergence parameters.

One may note that the observed convergence rate is quite close to the one used in the aforementioned work (annual linear depreciation of regional residuals by 2%). Yet it

was somewhat higher (2.8% on a linear basis) and, additionally, variant calculations show that it might be better to apply even higher convergence rates in projections. Fig. 25 depicts average decennial absolute percentage errors for the three projections of Austrian households combined obtained using different convergence parameters. A convergence rate of 0% corresponds to keeping unchanged regional residuals compared to the regression estimates, while depreciation at 100% corresponds to complete ignorance of regional residuals, i.e., to applying the simple regression approach described above. In any case, the overall average percentage error varies between 5% and 6%, which is not as much as in the case of keeping constant regional slopes (9%). Yet using a depreciation rate of about 6% might work better in projections for Austria. A possible explanation of why the ‘optimal’ convergence rate is higher than the one observed from censuses might be that projections use regressions obtained from the *previous* census, as the actual regression line for the forthcoming census is not available in the projection.

In long run, projections obtained by the regression method supplemented with regional corrections are close to the simple regression method as reflected in Fig. 25 (compare also Figs. 22-24 and 26-28). Still, this method is bound to perform better at years close to the base census, as it is consistent with actual regional variations observed in the census. Hence, one may use regional corrections with depreciating regional residuals by 3-10% yearly.

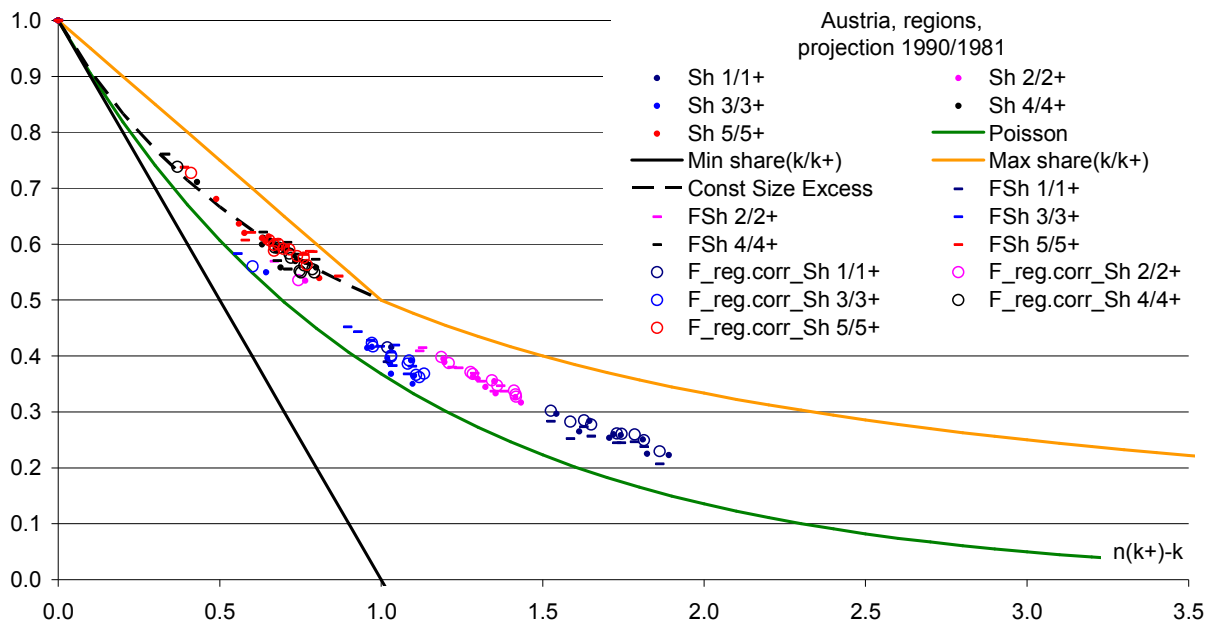


Fig. 26. Census population of Austria, 1991 and projections to year 1990 from census 1981. Conditional shares of households of different sizes ($k=1, 2, 3, 4,$ or 5) among households of the same or larger size against the average size of such households minus k . Census data are represented by dots, official projections by hyphens, and projections obtained using (20)-(25), (28), and (31) by circles.

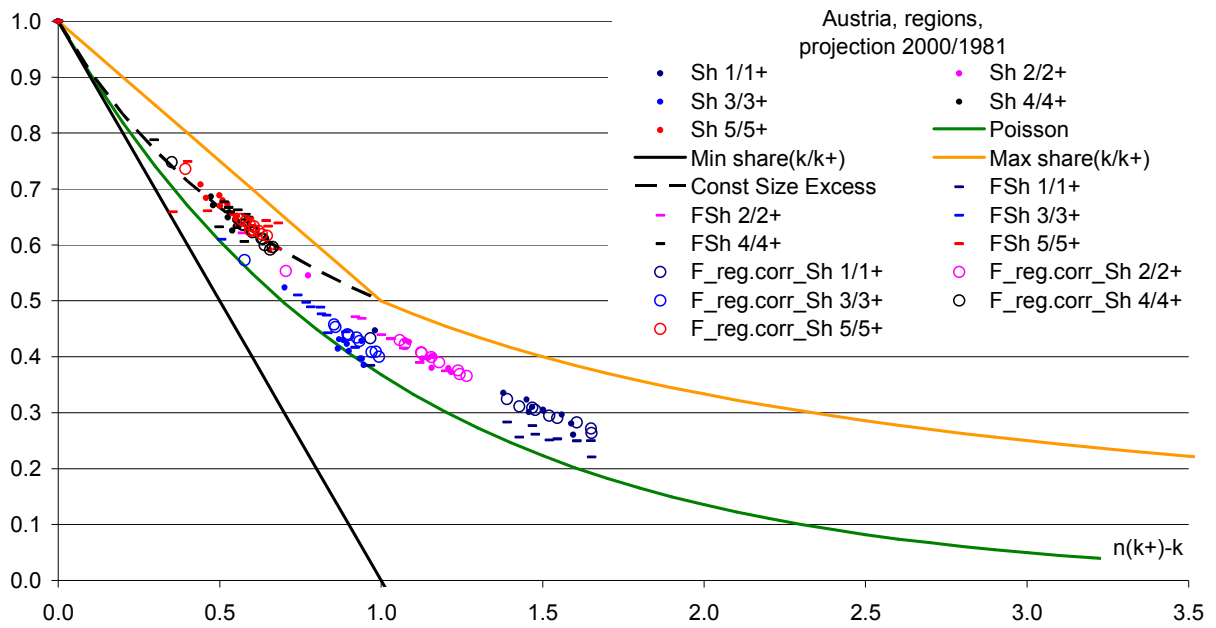


Fig. 27. 2001 census of Austria and projections to 2000 from 1981. Conditional shares of households of different sizes ($k=1, 2, 3, 4,$ or 5) among similar or larger households against the average size of such households minus k . Census data are represented by dots, official projections by hyphens, and projections (20)-(25), (28), and (31) by circles.

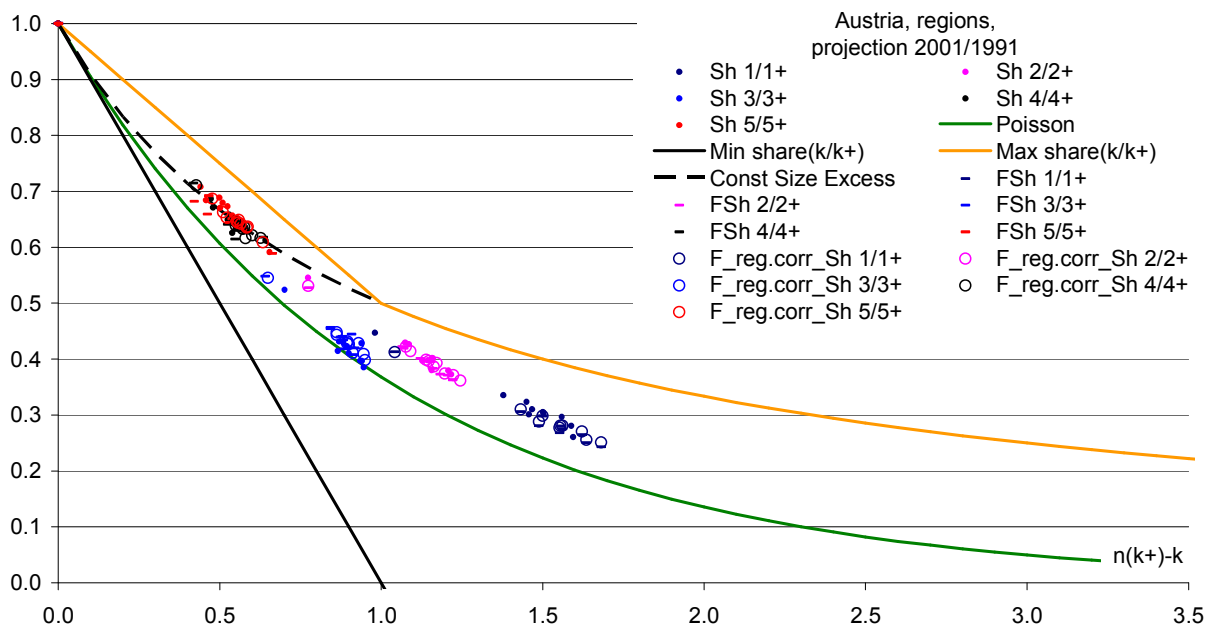


Fig. 28. 2001 census of Austria and projections to 2001 from 1991. Conditional shares of households of different sizes ($k=1, 2, 3, 4,$ or 5) among similar or larger households against the average size of such households minus k . Census data are represented by dots, official projections by hyphens, and projections (20)-(25), (28), and (31) by circles.

5. Distribution by age, sex, and other characteristics of the head of household

Some applications of household projections may require—in addition to distribution by size—also a distribution of households by age, sex, and, perhaps, by other characteristics of the household head. Some of the methods mentioned in the introduction allow having such projections. Again, however, simple headship-rates-based methods will usually lead to even more inconsistencies compared to the case of distributing households only by size. More sophisticated dynamic methods, on the other hand, pose considerable demand on model assumptions and on data to derive model parameters. Here the approach described above is utilised to develop a method of distributing households by demographic characteristics of the head, in addition to size of the household. For the sake of simplicity, only one variable, x , is introduced to reflect characteristics of the head. Hence, for example, $H_k(x)$ denotes number of households of size k with head of type x (which may stand for age or sex, or for age and sex combined, etc.). The exact meaning of x is of no importance for the method. However, readers might find it helpful to read it as age of the household head.

It is assumed that headship rates are applied to every subpopulation of interest:

$$H(x) = h(x)P(x), \quad (32)$$

here $h(x)$ is the headship rate for subpopulation of type x and $P(x)$ is the size of the subpopulation. Note that the population residing in households with head of type x (we denote it by $N(x)$) is not necessarily equal to the subpopulation of type x , $N(x) \neq P(x)$, as a non-head member of household headed by person of, say, age x may be of another age.

The total number of households may be derived from head-status-specific numbers (32):

$$H = \sum_x H(x) = \sum_x h(x)P(x), \quad (33)$$

and may be distributed by size according to one of the methods described above, based on the average size of households:

$$n = \frac{P}{H} = \frac{\sum_x P(x)}{\sum_x h(x)P(x)}. \quad (34)$$

Hereinafter, omission of the x variable means that the number refers to the total population. Distributing households by size gives numbers that are consistent with the overall number of households (33):

$$\sum_k H_k = H = \sum_x H(x). \quad (35)$$

Hence, we have consistent yet separate distributions by head type x and household size k . The purpose is to develop a consistent distribution by both variables combined, $H_k(x)$, from which both separate distributions might be obtained⁵:

⁵ One might consider a possibility to project the general distribution first, with deriving marginal distributions from it. However, applying the model proposed here to age/sex-specific distributions of households by size revealed that such bottom-up aggregation method of obtaining distribution by size works not as neatly compared to the disaggregating approach proposed herein. See the text for explanation of this paradoxical situation.

$$\begin{aligned}
H_k &= \sum_x H_k(x), \\
H(x) &= \sum_k H_k(x).
\end{aligned}
\tag{36}$$

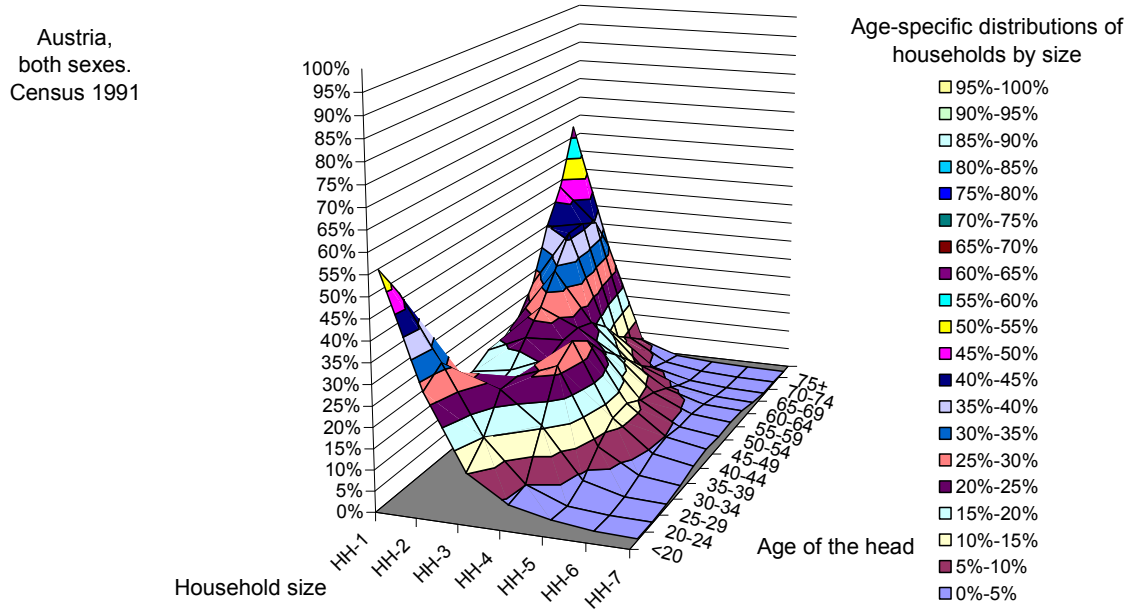


Fig. 29. Distributions by size of households with heads of a given age (Austria, both sexes combined, census 1991).

The idea of the method proposed here is to apply the same methodology to distributing x -specific households by size as it was described before for distributing the total number of households. The rationale behind this is that the age and other relevant characteristics of the head might mark the type of household, the stage of its development. Hence, one may expect households with heads of different characteristics to have different and persistent regularities in their distribution by size. See Fig. 29 for empirical support of such an expectation.

To be able to apply the general methodology of distributing by size to households subgroups, we assume the following *provisional* estimates for average sizes of households by type of the head available: $n(x) = \frac{N(x)}{H(x)} = \frac{N(x)}{h(x)P(x)}$. Note again that $N(x)$ is the population residing in households $H(x)$ which differs, generally speaking, from the population that is of the same type as the household head, i.e., from $P(x)$. In practice, household-head status-specific average sizes may be *provisionally* projected from historical data in the same manner as headship rates $h(x)$. These estimates—based on extrapolation of past trends, on experience of other populations, on separate models, etc.—are indeed provisional, as they might be inconsistent with headship rates and with the projected structure of the population. In particular, they may violate the following identity to be held in any consistent projection:

$$P = N = \sum_x P(x) = \sum_x n(x)H(x) = \sum_x n(x)h(x)P(x). \quad (37)$$

In case, when this identity is violated, one should correct either the average sizes or the headship rates implied. This work suggests the former, i.e., to correct the average sizes. It is a rational conjecture that headship rates are less sensitive to fertility assumptions, while the average sizes of households are directly linked to the fertility assumptions imputed in projection. Indeed, headship rates depend rather on marriages, divorces and mortality than on the fertility level, although fertility and marriages are also interrelated. Household size, on the other hand, is determined both by the same factors mentioned and, to a great deal, also by the fertility level which determines the number of children in households. This is of more importance for households with heads at childbearing ages, although it may affect other households as well. Given the sensitivity of the average size to fertility assumptions used in population projections and also taking into account that it is fertility which is the most uncertain component in shaping population projections, it might be better to rely on headship rates in reconciling the balance identity (37)⁶. It seems that it is the sensitivity to fertility of age/sex-specific distributions of households by size which resulted in the paradoxical situations mentioned above: the distribution of the total number of households—especially in the long run—is more robust compared to calculating the totals of households of different sizes as a sum of age/sex-specific numbers obtained separately, e.g., using the size/age/sex-specific headship rates. Indeed, the overall average size of households explicitly reflects fertility assumptions implied in the population projection. Hence, the distribution of households by size according to the overall average size will incorporate information on fertility assumptions. If, for example, a population projection implies lower fertility, the average size of households will be depressed and, therefore, the proportion of one-person households will increase in the models proposed here. By contrast, one-person households obtained *directly* from the projected population by age and sex and from extrapolated one-person household headship rates will not reflect the decline in births, as that does not affect age groups, from which household heads are estimated and—if no explicit model for a relation between fertility and headship rates is developed—will not affect headship rates either. These points are illustrated in Fig. 30 and 31, where household-head age-specific headship rates and average household sizes are presented for Austria in 1981-2004. As one may note, age-specific headship rates were more stable compared to the average sizes. One may also note, however, that at older ages the headship rates seem to be more volatile—perhaps due to an effect of mortality.

⁶ One may also consider the role of migration, which might affect both headship rates and households' average sizes at young working ages. The effect of migration on headship rates, however, will be less important if migrants have headship rates close to those of the population in general and, of course, if migration is low.

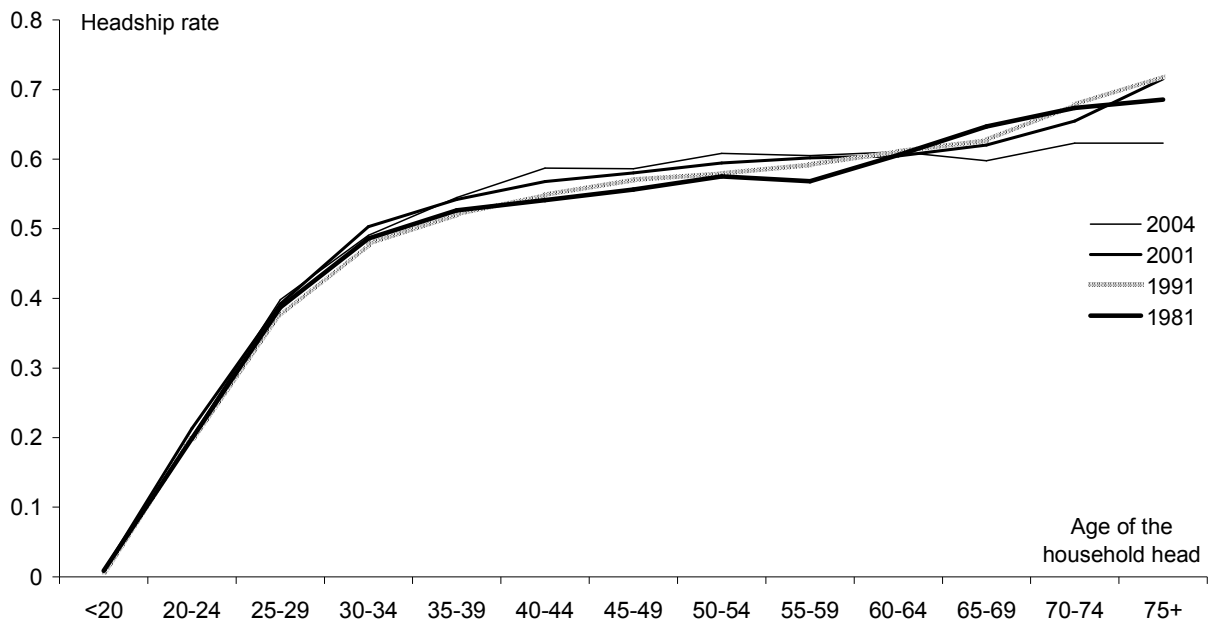


Fig. 30. Age-specific headship rates from census and micro-census populations of Austria, 1981-2004.

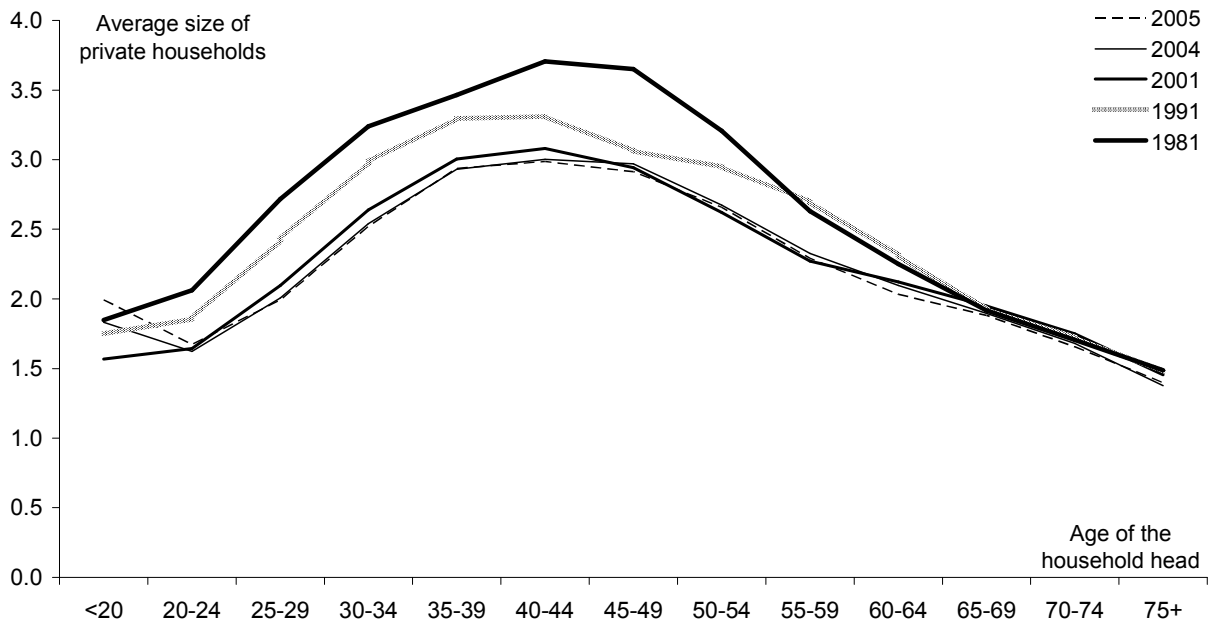


Fig. 31. Average sizes of private households by age of the head from census and micro-census populations of Austria, 1981-2005.

Taking into account the considerations presented, let us turn to the procedure for correcting the provisional average sizes in order to reconcile the balance identity (37). If the identity is violated, the average sizes are either deflated or inflated. In the simplest

form, the same correction rate may be applied to all average sizes. It should be taken into account, however, that the average size of households must not be less than one. Therefore it is better to adjust average excesses of the size of households above the minimal size (one), i.e., $\eta_1(x) = n(x) - 1$, instead of average sizes $n(x)$ themselves:

$$\eta_1(x) = \frac{P - H}{\sum_y \hat{\eta}_1(y) H(y)} \hat{\eta}_1(x), \quad (38)$$

here $\hat{\eta}_1(x) = \hat{n}(x) - 1$ are provisional estimates, which may not fit balance (37). To get (38), identity (37) may be transformed into the following equivalent form for non-head members of households:

$$P - H = N - H = \sum_x \eta_1(x) H(x). \quad (39)$$

In terms of average sizes, then, the corrections look as follows:

$$n(x) = 1 + \frac{N - H}{\sum_y \hat{n}(y) H(y) - H} (\hat{n}(x) - 1). \quad (40)$$

Alternatively, one may use more rigorous correction procedures, based on models that relate average sizes of households to fertility and, perhaps, to mortality. Apparently these are households with young and middle-aged heads, which are most likely to change in size as a result of changes in fertility. This may happen both directly, when fertility affects the number of children in families, and also indirectly, when marriage and fertility postponements at young ages result in a higher proportion of singles at older ages. Such suggestions are supported by observations for Austria, see Fig. 31, where average sizes were most volatile for households with young and middle-aged heads, while households of elderly persons were more stable in size. Therefore, one might improve corrections (38), (40) by differentiating them by age, sex and other relevant characteristics of household heads. In particular, one may use some scaling function $s(x)$, which is assumed to be proportional to possible variations of the average sizes of households and use corrections of the following form:

$$\eta_1(x) = \hat{\eta}_1(x) + \delta \cdot s(x) \hat{\eta}_1(x), \quad (41)$$

where δ is a constant to be derived from the balance identity (39)⁷:

$$\delta = \frac{N - H - \sum_x \hat{\eta}_1(x) H(x)}{\sum_x s(x) \hat{\eta}_1(x) H(x)}. \quad (42)$$

Hence, corrections for average sizes are given by

$$\eta_1(x) = \left(1 + \frac{N - H - \sum_y \hat{\eta}_1(y) H(y)}{\sum_y s(y) \hat{\eta}_1(y) H(y)} \cdot s(x) \right) \hat{\eta}_1(x) \quad (43)$$

and

⁷ In case of negative δ and when the scaling function $s(x)$ and provisional estimates $\hat{n}(x, t)$ are only rough approximations, one should be aware of getting possibly negative average size excesses in (41).

$$n(x) = \hat{n}(x) + \frac{N - \sum_y \hat{n}(y)H(y)}{\sum_y s(y)\hat{n}(y)H(y) - \sum_y s(y)H(y)} \cdot s(x)(\hat{n}(x) - 1). \quad (44)$$

The scaling function $s(x)$ may be derived in different ways. It may be obtained from empirical observations on average sizes variations, from simulations using dynamic household models, from projections developed for other populations (e.g., functions for separate regions may be derived from the projection developed for the country as a whole), or even from simplified theoretical considerations, e.g., as proportional to the accumulated fertility from the base census year. In case, when there is a time series of observed average sizes $n(x, t)$, one may consider Eq. (41) as model and fit its parameters:

$$\frac{n(x, t) - n(x, t_0)}{n(x, t_0) - 1} = \delta(t) \cdot s(x) \quad (45)$$

for some reference year t_0 . For the projecting purposes it might be convenient to use the year of the projection's base census as the reference year. In model (45) $\delta(t)$ and $s(x)$ may be interpreted as interacting time and age effects on variations of the average sizes of households. They may be estimated using the following scaling assumption:

$$\overline{\delta(t)}^{(t)} = 1, \quad (46)$$

hereinafter the upper line denotes taking average over the variable indicated in parentheses next to the line. Hence, from (46) one may get the scaling function as

$$s(x) = \left(\frac{\overline{n(x, t) - n(x, t_0)}^{(t)}}{n(x, t_0) - 1} \right)^{(x)}, \quad (47)$$

and the time effect may be derived from least squares⁸:

$$\delta(t) = \frac{1}{s^2(x)^{(x)}} \left(s(x) \frac{\overline{n(x, t) - n(x, t_0)}^{(t)}}{n(x, t_0) - 1} \right)^{(x)}. \quad (48)$$

In particular, the application of (47) to 1981, 1991, and 2001 census data on Austrian distributions of households by size and by age of the head gives results presented in Table 14 (trend values are obtained by fitting linear regression line and with setting to zero negative values obtained for the group 75+).

Table 14. Results of estimating the scaling function $s(x)$ for correcting average sizes of households with heads at age x . Estimates are based on Austrian census data from 1981, 1991, and 2001 (both sexes combined).

Age	<20	20-24	25-29	30-34	35-39	40-44	45-49	50-54	55-59	60-64	65-69	70-74	75+
s(x) estim.	0.274	0.328	0.288	0.192	0.125	0.138	0.141	0.188	0.205	0.094	-0.021	-0.031	0.042
s(x) trend	0.293	0.272	0.248	0.224	0.200	0.177	0.153	0.129	0.105	0.082	0.058	0.034	0

⁸ Model (45) and its extensions may be used to *project* the average sizes of households based on a projection for time effect (48), in the same manner as it is done in the Lee & Carter (1992a, 1992b) method for mortality projections.

After obtaining consistent estimates for average sizes of households by status of the head, one may use one of the techniques presented above to distribute such households by size. These methods are to be corrected, however, in order to produce a distribution of households which is consistent with separate distributions by size and head's characteristic x . To follow this task, the following recurrent procedure is proposed which follows the same general framework proposed above.

First, smallest-size households are provisionally obtained from x -specific average sizes. Whatever method is used for that purpose, it does not matter for the procedure described here. Hereinafter we will follow model (19), (20) without specifying how the model parameters are derived. For the latter purpose one may use any of the three approaches presented above (constant slopes method, regression approach, and regression combined with regional corrections). Hence, the x -specific conditional shares of smallest-size households (here, of size one) are *provisionally* estimated:

$$\hat{v}_1(x) = \hat{v}_{1/1+}(x) = \max \begin{cases} e^{-\alpha_1(x)\eta_1(x)}, \\ 1 - \eta_1(x). \end{cases} \quad (49)$$

Here $\eta_1(x) = n(x) - 1$, and cups above expressions reflect the provisional character of the estimates which thereby are indicated as possibly inconsistent. These shares are used to obtain the provisional distribution by the head's characteristic (x) of one-person households:

$$\hat{H}_1(x) = \hat{v}_{1/1+}(x)H(x). \quad (50)$$

After completing these calculations, the provisional total number of one-person households may be calculated, the overall balance may be checked, and the following reconciliation corrections—if needed—may be implemented:

$$H_1(x) = \begin{cases} \hat{H}_1(x) - (\hat{H}_1(x) - H_1^{MIN}(x)) \frac{\hat{H}_1 - H_1}{\hat{H}_1 - H_1^{MIN}}, & \text{if } \hat{H}_1 > H_1, \\ \hat{H}_1(x) + (H_1^{MAX}(x) - \hat{H}_1(x)) \frac{H_1 - \hat{H}_1}{H_1^{MAX} - \hat{H}_1}, & \text{if } \hat{H}_1 < H_1. \end{cases} \quad (51)$$

here:

$$H_1^{MIN}(x) = \max(0, (1 - \eta_1(x))H(x)),$$

$$H_1^{MAX}(x) = H(x),$$

$$\hat{H}_1 = \sum_x \hat{H}_1(x),$$

$$H_1^{MIN} = \sum_x H_1^{MIN}(x),$$

$$H_1^{MAX} = \sum_x H_1^{MAX}(x).$$

Note that it is a simple scaling within the feasible range of values which is proposed by (51). The scaling is developed in such a way, however, that none of essential theoretical constraints will be breached due to the correction. In principle, one may consider other, more sophisticated correction procedures, based on models presented and on empirical observations. These possibilities are not considered here, however, as the accuracy of the method proposed seems to be sufficient for the task addressed.

After completing with one-person households, the calculations proceed recurrently to other sizes of households, using the following relations:

$$H_{k+}(x) = H_{(k-1)+}(x) - H_{k-1}(x), \quad (52)$$

$$\eta_k(x) = \frac{\eta_{k-1}(x)}{1 - H_{k-1}(x)/H_{(k-1)+}(x)} - 1, \quad (53)$$

$$\hat{v}_{k/k+}(x) = \max \begin{cases} e^{-\alpha_k(x)\eta_k(x)}, \\ 1 - \eta_k(x), \end{cases} \quad (54)$$

$$\hat{H}_k(x) = \hat{v}_{k/k+}(x)H_{k+}(x), \quad (55)$$

$$H_k(x) = \begin{cases} \hat{H}_k(x) - (\hat{H}_k(x) - H_k^{MIN}(x)) \frac{\hat{H}_k - H_k^{MIN}}{\hat{H}_k - H_k^{MIN}}, & \text{if } \hat{H}_k > H_k, \\ \hat{H}_k(x) + (H_k^{MAX}(x) - \hat{H}_k(x)) \frac{H_k - \hat{H}_k}{H_k^{MAX} - \hat{H}_k}, & \text{if } \hat{H}_k < H_k, \end{cases} \quad (56)$$

where:

$$H_k^{MIN}(x) = \max(0, (1 - \eta_k(x))H_{k+}(x)),$$

$$H_k^{MAX}(x) = H_{k+}(x),$$

$$\hat{H}_k = \sum_x \hat{H}_k(x),$$

$$H_k^{MIN} = \sum_x H_k^{MIN}(x),$$

$$H_k^{MAX} = \sum_x H_k^{MAX}(x).$$

As a result of the described recurrent procedure, a distribution of households by size and by characteristic of the head will be generated, which is consistent with the distributions by the head's status projected from the headship rates and also with the distribution by size projected separately as described above.

The method was applied to distributing Austrian households from the 2001 census by size and age of the head given the census distributions separately by size and by age in order to test its accuracy. The distribution was done using simplest 'constant slopes' method for age-size-specific parameters $\alpha_k(x)$, which were derived from the 1991 census and, alternatively, from the 1981 census. Provisional average sizes of households by age of the head were derived from the same censuses. For each census both equal scaling (40) and age-specific scaling (44) were applied with scaling function taken from Table 14 (trend values). In addition, actual age-of-the-head-specific average sizes were also used for comparative purposes. All in all, we have six alternative scenarios for computations: three for each of the 1981 and 1991 censuses. Age-specific scaling was more effective, although all methods, including the one based on actual average sizes, were less effective for households with more than three members⁹. These households, however, are less prevalent in the population, and the lower effectiveness of their reconstruction has less effect on the

⁹ In part, effectiveness of distributing large households was affected by the simplicity of the method: $\alpha_k(x)$ were fixed at base census levels. While this was an adequate assumption for households with up to three members, the model parameters for larger households varied considerably in the data used.

overall effectiveness of the distribution. Weighted size-specific absolute percentage errors of distributions obtained are presented in Figs. 32 and 33 (numbers of households by age are used as weights). Age-specific average absolute percentage errors (weighted by the number of households by size) are presented in Figs. 34 and 35. One may note that the method works better for smaller households (with three or less members). More detailed patterns of errors (basis: year 1991) with age-specific scaling are presented in Fig. 36. The overall weighted average percentage errors were: 16% for base year 1981 with equal scaling; 12% for the same base census with age-specific scaling; 10% for the same base census with actual 2001 average sizes used; 11% for base census 1991 with equal scaling; 9% with age-specific scaling; and 7% with actual 2001 average sizes combined with $\alpha_k(x)$ parameters from the 1991 census.

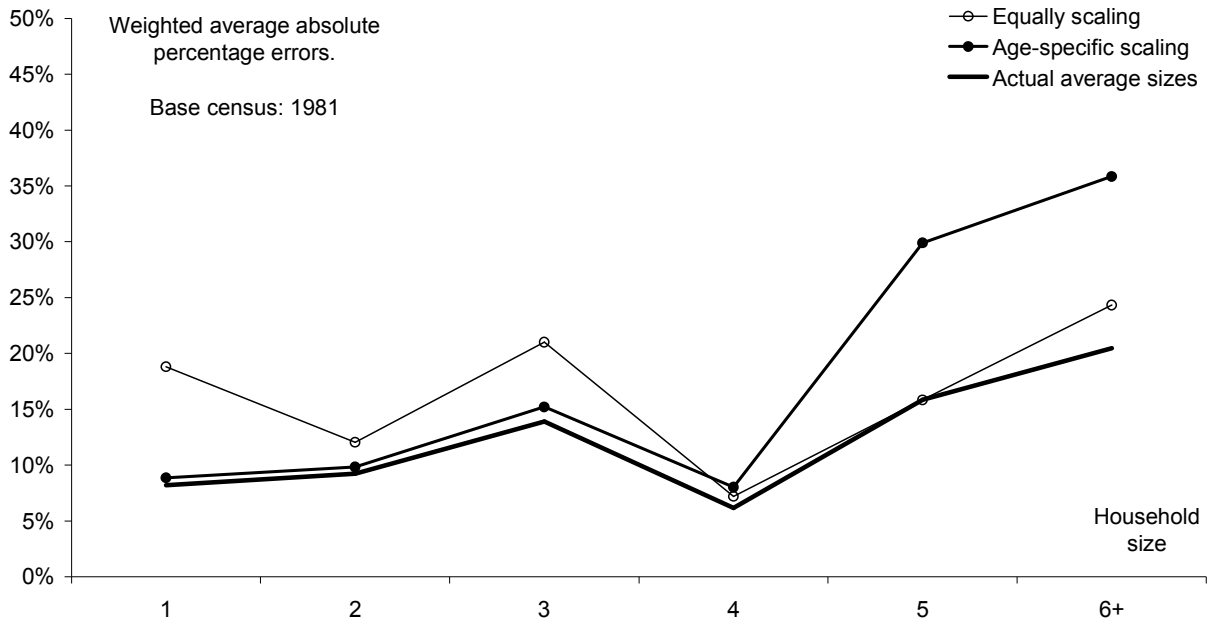


Fig. 32. Absolute percentage errors in household number by size weighted by the number of households by age of the head. Austria, 2001 census. Base census: 1981.

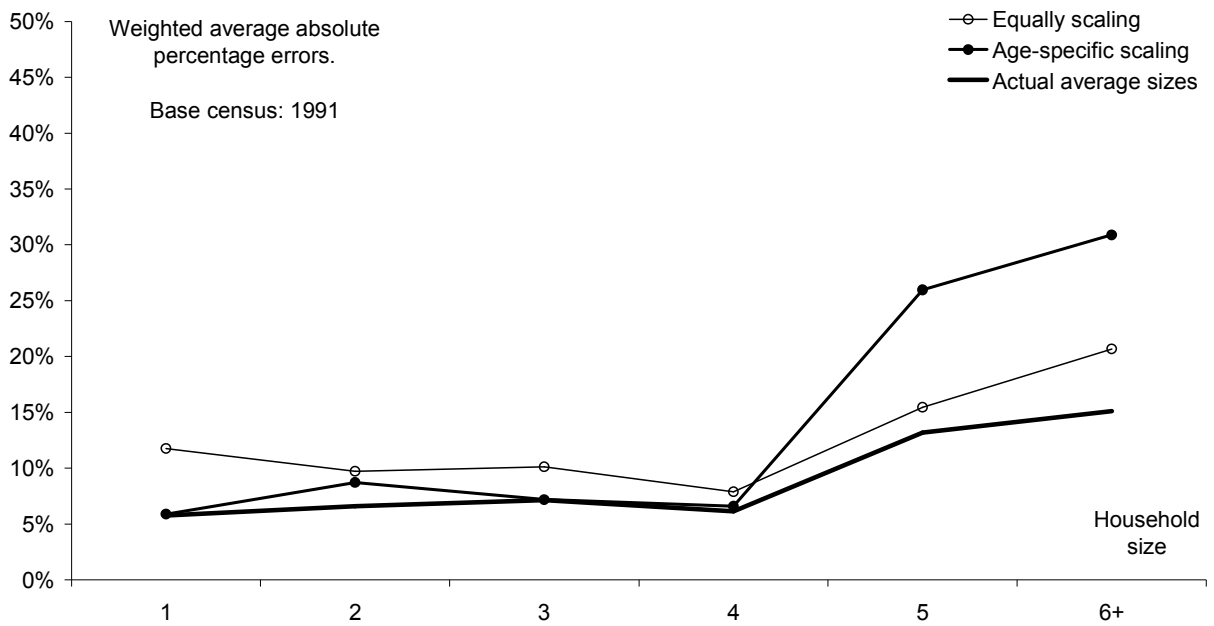


Fig. 33. Absolute percentage errors in household number by size weighted by the number of households by age of the head. Austria, 2001 census. Base census: 1991.

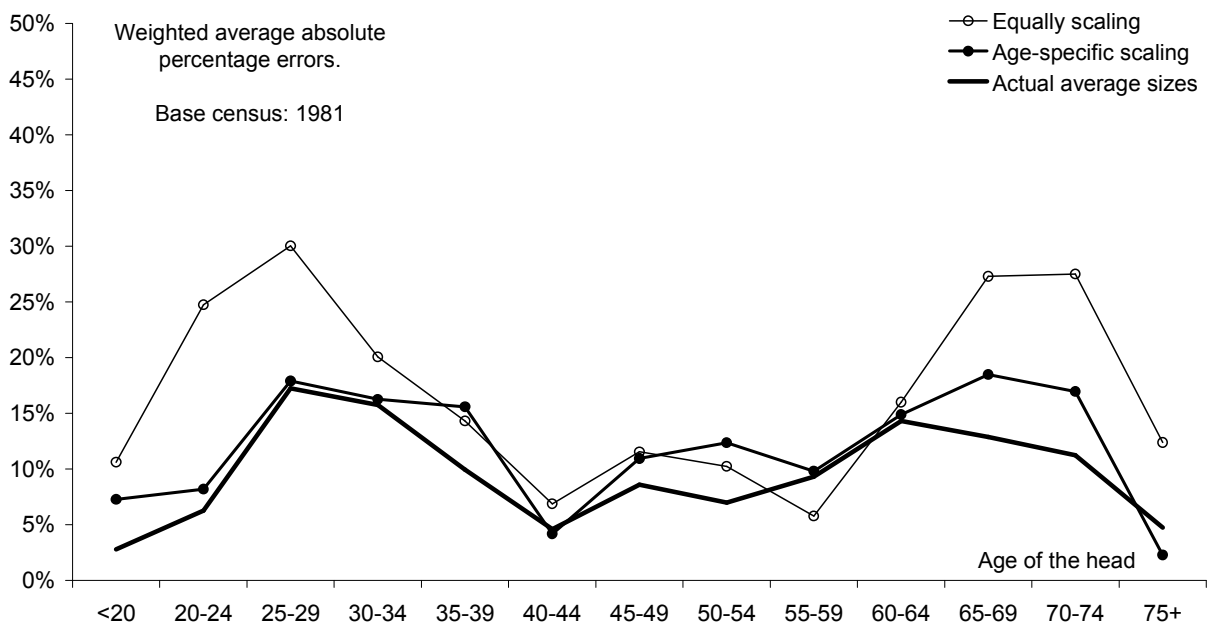


Fig. 34. Absolute percentage errors in household number by age of the head weighted by the number of households by size. Austria, 2001 census. Base census: 1981.

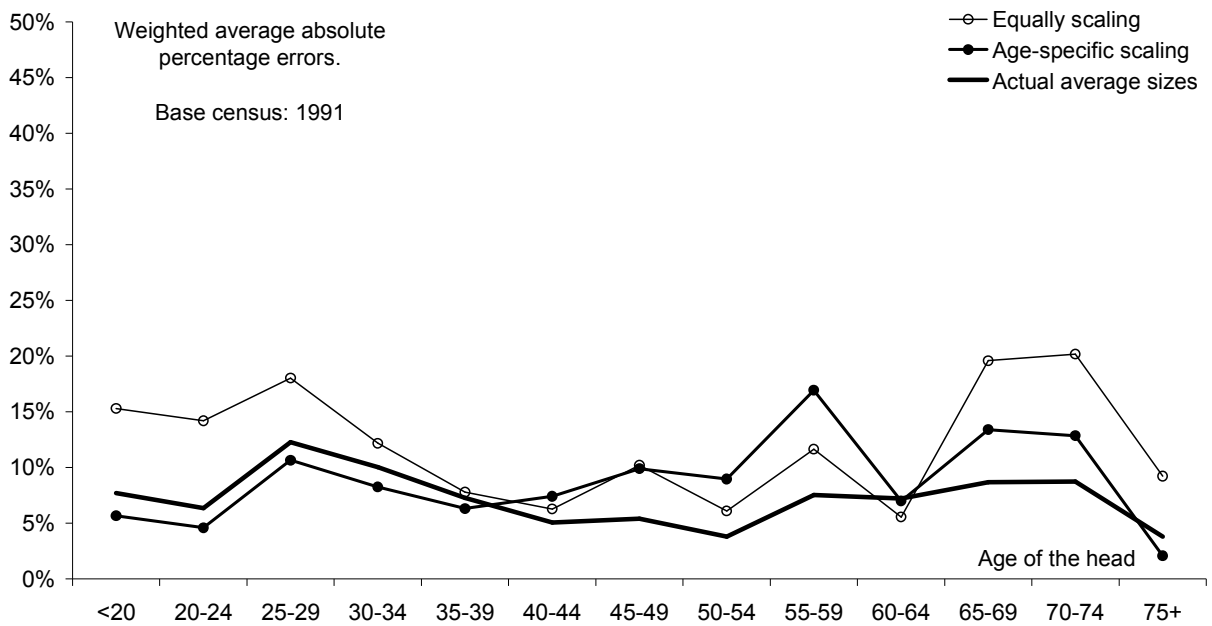


Fig. 35. Absolute percentage errors in household number by age of the head weighted by the number of households by size. Austria, 2001 census. Base census: 1991.

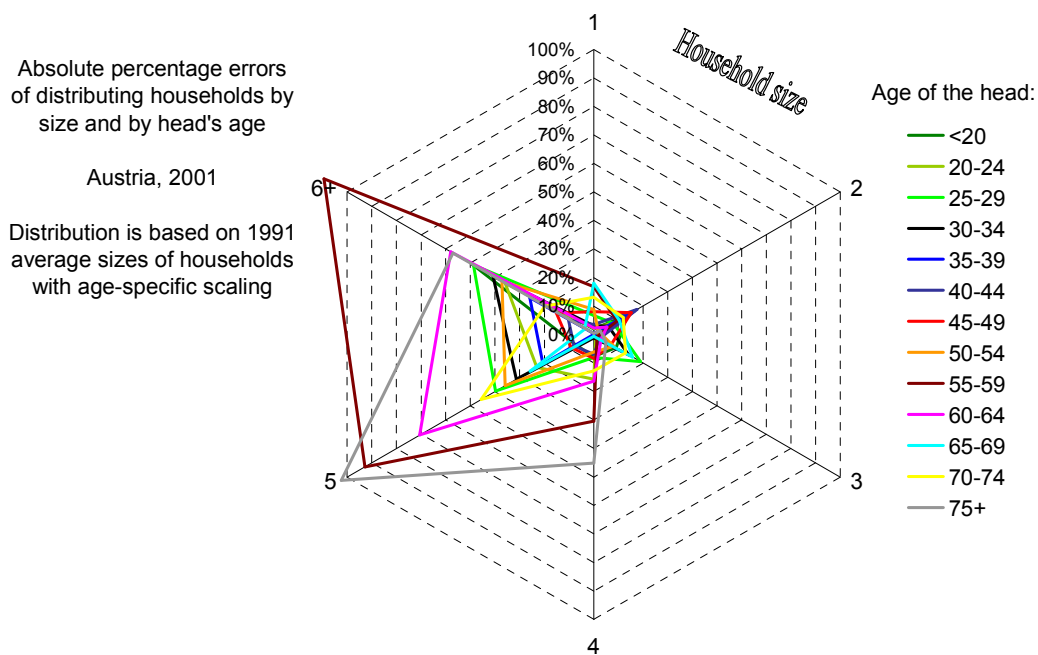


Fig. 36. Absolute percentage errors in household number by size and by age of the head. Austria, 2001 census. Base census: 1991.

6. Alternative Methods

In this section we briefly discuss some alternative methods for distributing households by size, which may be proposed within the same general framework of conditional shares and averages utilised above. Although the methods described there seem to be more efficient and robust, the methods presented further down may also be of importance for theoretical studies, and in some special applications as well.

One of alternative methods, which we may call *the Δ -method*, is based on deriving the household shares from Eq. (7), (11):

$$v_{k/k+} = 1 - \frac{n_{k+} - k}{n_{(k+)+} - k} = 1 - \frac{\eta_k}{1 + \eta_{k+1}} = \frac{\Delta_k}{\Delta_k + \eta_k}, \quad (57)$$

where $\Delta_k = n_{(k+)+} - n_{k+}$. The idea is to use Δ_k as parameters instead of α_k used in the models described previously. Note that under fixed values of Δ_k , relation (57) has the same shape as is suggested by empirical observations, i.e., it is a decreasing convex-down function, which approaches unity as η_k approaches zero. This relation, however, may occasionally violate constraint (9), (10) imposed from analytical considerations. Hence, when using model (57) one should take care of the implications of the constraint for parameters Δ_k :

$$\frac{\Delta_k}{\Delta_k + \eta_k} \geq 1 - \eta_k, \text{ i.e., } \Delta_k \geq 1 - \eta_k. \quad (58)$$

Besides, parameters $\Delta_k = n_{(k+)+} - n_{k+}$ may not be negative, as average size of households with k and more members may not exceed that of households with $k+1$ and more members:

$$\Delta_k \geq 0. \quad (59)$$

Deriving η_k from (58) and substituting it into (57), one may also obtain the following inequality:

$$v_{k/k+} \leq \Delta_k. \quad (60)$$

Model (57) might be of importance when $\Delta_k = n_{(k+)+} - n_{k+}$ are relatively constant and may therefore easily be projected into the future, either as constants or from some simple regressions. While this is close to the observed patterns for households with small average sizes, for other cases one might rather assume that ratios of successive conditional average sizes η_{k+1} and η_k might be more stable and more robustly modelled.

In that case, one may develop *the θ -method* as an alternative to (57), based on the following relation:

$$v_{k/k+} = 1 - \frac{1}{\frac{1}{\eta_k} + \theta_k}, \quad (61)$$

where $\theta_k = \frac{\eta_{k+1}}{\eta_k} \geq 0$. Using these non-negative coefficients will guarantee that constraint (58), which is equivalent to the non-negativity of η_{k+1} , will always be held. However, another constraint for parameters θ_k is to be taken into account, which is equivalent to (58):

$$\theta_k \geq \frac{\eta_k - 1}{\eta_k} = 1 - \frac{1}{\eta_k}. \quad (62)$$

This inequality becomes crucial when households of average sizes larger than unity are concerned.

The last approach, *the ν -method*, which is proposed here within the general framework is perhaps the simplest in its formulation. It relies simply on developing direct polynomial or other good-fit regressions for relations between conditional shares and conditional average sizes or, alternatively, the overall average size of households:

$$v_{k/k+} = a_0 + a_1\eta_k + a_2\eta_k^2 + \dots + a_m\eta_k^m, \quad (63)$$

or

$$v_{k/k+} = a_0 + a_1n + a_2n^2 + \dots + a_mn^m. \quad (64)$$

This approach ignores the regularities suggested from theoretical findings and from empirics. Yet, given a rich and long-term historical data and taking into account the general restriction (9), (10), this approach may also provide robust projections.

Concluding this brief description of alternatives, it is worth mentioning that these methods are less relevant to empirical observations compared to the α -method described before. The ν -method ignores the theoretical and empirical findings quite explicitly, and we will not focus on it. As for methods (57) and (61), their relevance may be checked looking into performance of the simplest variants of these methods, which may be obtained by stating that model parameters are constant. For the Δ -method, this yields:

$$\Delta_k = \frac{\eta_k \cdot v_{k/k+}}{1 - v_{k/k+}} \approx const. \quad (65)$$

The θ -method will rely on another assumption:

$$\theta_k = \frac{1}{1 - v_{k/k+}} - \frac{1}{\eta_k} \approx const. \quad (66)$$

And it is also worth to look into the α -method described before, which yields:

$$\alpha_k = -\frac{\ln v_{k/k+}}{\eta_k} \approx const. \quad (67)$$

These relations suggest ‘natural’ transformations, which should result in nearly constant expressions that vary only modestly from population to population and also from year to year. Table 15 presents the results of estimating parameters (65)-(67) from data on 892 NUTS-3 regions of EU (including some non-EU countries as well, e.g., Turkey and Norway) and also average absolute percentage errors of estimating the conditional shares of households of different sizes using the constant parameters assumptions (65)-(67). The table also contains standard deviations and variances of the parameters. Estimates presented may be used as rough approximations for simplistic models of the distribution of households by size. Apart from this, Table 15 reflects the relevance of the models and also gives an impression about errors’ magnitudes for the constant-parameters projections. Average absolute percentage errors presented in the table are obtained by applying the average values of parameters obtained from the dataset itself. The constant α assumption seems to work robustly for all types of households, while Δ has low variance for larger households only (i.e., with small conditional excesses in size), and θ is less volatile only for small-size households (with large conditional average excesses in size). Turning to average errors, however, it is easy to note that both alternative methods work with less

accuracy for small-sized households. Even the θ -method is not effective for small households, since relation (61) is more sensitive for variations in θ when big average sizes are concerned.

Table 15. Average values, standard deviations and variances of models parameters and average absolute percentage errors of estimates of conditional shares of households for EU NUTS-3 regions, 2001 census round.

	Household size k :					average
	1	2	3	4	5	
Averages of Parameters						
α_k	0.860	0.806	0.976	0.830	0.783	
Δ_k	0.524	0.729	0.616	0.931	1.027	
θ_k	0.723	0.796	0.593	0.904	1.113	
Standard Deviations of Parameters						
α_k	0.0670	0.0736	0.0978	0.0747	0.0606	
Δ_k	0.1478	0.1629	0.0832	0.1068	0.1211	
θ_k	0.0508	0.0950	0.0977	0.1313	0.1958	
Coefficients of Variation of Parameters, %						
α_k	7.8%	9.1%	10.0%	9.0%	7.7%	8.7%
Δ_k	28.2%	22.3%	13.5%	11.5%	11.8%	17.5%
θ_k	7.0%	11.9%	16.5%	14.5%	17.6%	13.5%
Average Absolute % Errors in estimating conditional shares						
α_k -method	10.0%	8.0%	8.5%	4.5%	3.1%	6.8%
Δ_k -method	29.7%	15.0%	7.4%	4.5%	4.1%	12.1%
θ_k -method	28.7%	9.6%	16.5%	3.3%	5.4%	12.7%

7. Sensitivity of the Method. Errors Accumulation

This section addresses a very important issue for any recurrent procedure: the issue of errors accumulation within the method. In the framework used in this work households are estimated recurrently, size after size. Therefore, if households, say, of size k happen to be estimated with an error ε_k , this will also affect estimates for all larger households. This mechanism may well explain why in the projections presented above the performance of the models was usually the worst for largest households. It is important, therefore, to investigate how such errors are transmitted from one size of households to another. If these errors are amplified due to the procedure, that will make the method less robust, especially for larger households. If on the other hand errors are being reduced or, at least, are not amplified, the method might work robustly.

To study errors transmission within the method, one may use the recurrent relation (11) and focus on errors in conditional average excesses η_k from which all other quantities are derived in the models presented. Considering the α -method, one may rewrite the relation in the following way:

$$\eta_{k+1} = \frac{\eta_k}{1 - \exp(-\alpha_k \eta_k)} - 1 = g_k(\eta_k), \quad (68)$$

where we denote by $g_k(\cdot)$ the non-linear function in the right-hand side of the equation. It is this function which determines how errors are transmitted from one size of households to the next. If this function has a derivative below unity, errors will be reduced and if the derivative is above unity, errors will be amplified:

$$Err(\eta_{k+1}) \approx g_k'(\eta_k) Err(\eta_k), \quad (69)$$

when errors in distributing households of size $k+1$ are caused by errors in distributing smaller households only, i.e., the model for households of $k+1$ members is correct.

Hence, one should consider a derivative of (68):

$$\begin{aligned} g_k'(\eta_k) &= \frac{\partial}{\partial \eta} \left(\frac{\eta}{1 - \exp(-\alpha_k \eta)} - 1 \right)_{\eta=\eta_k} = \left(\frac{1 - \exp(-\alpha_k \eta) - \eta \alpha_k \exp(-\alpha_k \eta)}{(1 - \exp(-\alpha_k \eta))^2} \right)_{\eta=\eta_k} = \\ &= \frac{1 - \exp(-\alpha_k \eta_k) - \eta_k \alpha_k \exp(-\alpha_k \eta_k)}{(1 - \exp(-\alpha_k \eta_k))^2} = \frac{1 - v_{k/k+} + v_{k/k+} \ln(v_{k/k+})}{(1 - v_{k/k+})^2}. \end{aligned} \quad (70)$$

It is quite convenient that the derivative is solely determined by the conditional share $v_{k/k+}$. This share, apparently, may not be negative nor may it exceed unity. Within this range of conditional shares, derivative (70) falls within unity and 0.5. As a result, errors are being reduced in the recurrent procedure by up to 50%, as illustrated in Fig. 37. For levels of conditional shares usually observed in real populations (15-70%) errors are reduced by 20-45% when proceeding from households with k members to households with $k+1$ members.

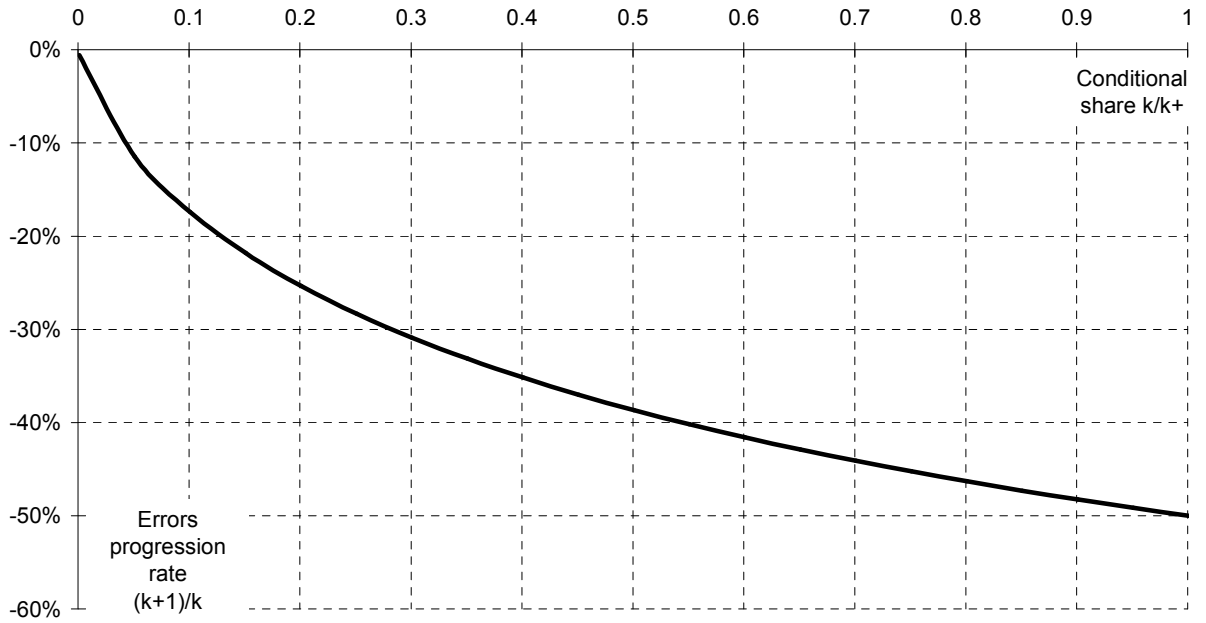


Fig. 37. Errors in conditional average size η_{k+1} caused by errors in η_k as compared to errors in η_k at different values of conditional shares $v_{k/k+}$.

Hence, the method proposed here does not amplify errors in input data and in the model itself. Rather, errors in estimating households of smaller sizes tend to be reduced as the procedure moves to larger households.

In reality it is also important that errors do not originally occur at a single stage of the method alone. Instead, errors may arise at all stages of modelling. Therefore, the total sum of errors originated from distributing households of different sizes may in fact rise within the method, even though the impact of errors originated from a specific size of households decreases when larger households are concerned. The relation for overall errors for households with $k+1$ members is the following:

$$Err(\eta_{k+1}) \approx g'_k(\eta_k)Err(\eta_k) + \varepsilon_{k+1}, \quad (71)$$

where ε_{k+1} are errors originated from biases of the model applied to households of size $k+1$, i.e., errors in η_{k+1} , which are not related to errors in modelling smaller households.

Assuming, for simplicity, the derivative in (71) to be nearly constant, $g'_k(\eta_k) \approx const = \varphi$, we may derive errors (71) as an explicit function of errors originated at different stages of the procedure:

$$\begin{aligned} Err(\eta_{k+1}) &\approx \varphi \cdot Err(\eta_k) + \varepsilon_{k+1} \approx \varphi^2 \cdot Err(\eta_{k-1}) + \varphi \cdot \varepsilon_k + \varepsilon_{k+1} \approx \dots \approx \\ &\approx \varphi^k \cdot \varepsilon_1 + \dots + \varphi^2 \cdot \varepsilon_{k-1} + \varphi \cdot \varepsilon_k + \varepsilon_{k+1}. \end{aligned} \quad (72)$$

Assuming ‘elementary’ errors ε_k to be uncorrelated with a similar standard deviation σ , one may derive the standard deviation of total errors (72):

$$\sigma_{k+1} \approx \sigma \sqrt{\varphi^{2k} + \dots + \varphi^4 + \varphi^2 + 1} = \sigma \sqrt{\frac{1 - \varphi^{2k+2}}{1 - \varphi^2}}. \quad (73)$$

This function starts with unity and grows quite rapidly to the asymptotic level $\frac{\sigma}{\sqrt{1 - \varphi^2}}$,

which may be used as an upper-bound estimate for standard errors of estimating conditional average excesses. For usual levels of conditional shares (15-70%), when φ is about 0.8-0.55, this upper-bound estimate for errors accumulation is about 1.2-1.7.

Hence, the recurrent procedure might be subject to a moderate accumulation of errors. Still, random errors may cumulate by no more than 20-70% for largest households compared to the errors for smallest households.

Conclusion

The methods presented in this study allow distributing the projected/reconstructed number of households by size and by socio-demographic characteristics of the household head. Such information on household distributions may significantly enrich the projection and its applications. In particular, the economic performance of the household is apparently related to the age, sex, and other characteristics of the head and also to size of the household. Therefore, the projection of households by size and by type of the head may be important for studies in savings, consumption, labour, housing, economic productivity, population-related ecological load, etc.

Sometimes statistical agencies and consumers of their projections are interested in distributions of households by type of household (not by type of the head). Although we do not consider this issue in the work, the results presented here might be important for developing such distributions. There is apparently a linkage between the distribution of households by type and their distribution by size and by type of the head. Indeed, one may expect to find many households formed of the mother and her child among households of two persons headed by a young woman, while the chances of finding such households among those of two persons headed by a man are rather low. Having census proportions of households of different types among households with a given size and a given status of the head and using distributions of households by size and the head's status, the task of estimating households by type may straightforwardly be addressed.

The methods presented in this paper are a good supplement to the headship rates method. Yet they may also be used in other contexts. In particular, projections might benefit from a combination of the methods described here with micro-simulation and other dynamic models with latter models used to derive parameters of the household distribution model.

References

- Akkerman, A. 1977. The Household Composition Matrix and Its Application to Migration Analysis and Population Projection. *General Systems* 22: 105-109.
- Akkerman, A. 1980. On the Relationship between Household Composition and Population Age Distribution. *Population Studies* 34 (6): 524-534.
- Bell, M., J. Cooper, M. Les. 1995. Household and family forecasting models – A review. Commonwealth Department of Housing and Regional Development. Commonwealth of Australia.
- Boleslawski, L. 1997. An operative method for household projections. *Studia Demograficzne* 3(129): 3-21.
- Bongaarts, J., T. Burch and K. W. Wachter. 1987. *Family Demography: Methods and Applications*. Oxford, Clarendon Press.
- Eurostat. 2006. Data downloaded from Eurostat webpage: <http://epp.eurostat.ec.europa.eu/>
- Federal Bureau of Statistics of Pakistan. 2006. Number of households by size and urban rural residence. Downloaded from the site <http://www.statpak.gov.pk/depts/index.html> on 30.08.2006.
- Gisser, R. 1986a. Vorausschätzung der Haushalte nach Bundesländern bis 2015. *Statistische Nachrichten*, 41. Jahrgang 1986 (Neue Folge), Heft 4: 279-288.
- Gisser, R. 1986b. Vorausschätzung der Haushalte nach Größe und Bundesländern bis 2015. *Statistics Austria*, 1986b. *Statistische Nachrichten*, 41. Jahrgang 1986 (Neue Folge), Heft 5: 378-383.
- Hanika, A. 1997. Vorausschätzung der Haushalte 1991-2030 nach Größe und Bundesländern, *Statistik Austria*, *Statistische Nachrichten* Vol. 4/1997 0, p. 220-231.
- Hooimeijer, P. and H. Heida. 1995. Household projections and housing market behaviour. *Modelling Household Formation and Dissolution*. N. Keilman, A. Kuijsten and A. Vossen. Oxford, Clarendon Press.
- Jiang, L. and B. O'Neill. 2004. Towards a new model for probabilistic household forecasts. *International Statistical Review* 72(1): 51-64.

- Keilman, N., A. Kuijsten, A. Vossen. Eds. 1988. *Modelling Household Formation and Dissolution*. Oxford, Clarendon Press.
- Lee, R.D. and L.R. Carter 1992a. Modeling and Forecasting U.S. Mortality. *J. of American Statistical Association*, 87 (419): 659-671.
- Lee, R.D. and L.R. Carter. 1992b. Modeling and Forecasting U.S. Mortality: Rejoinder. *J. of American Statistical Association*, 87 (419): 674-675.
- Linke, W. 1988. The headship approach in modeling households: The case of the Federal Republic of Germany. In: *Modelling Household Formation and Dissolution*. N. Keilman, A. Kuijsten and A. Vossen. Oxford, Clarendon Press.
- Lutz, W. 1997. FAMSIM Austria: feasibility study for a dynamic microsimulation model for projections and the evaluation of family policies based on the European Family and Fertility Survey. Austrian Institute for Family Studies. Vienna.
- Mason, A. 1987. HOMES: A Household Model for Economic and Social Studies. East-west Center. Hawaii.
- McDonald, P. and R. Kippen. 1998. Household trends and Projections: Victoria 1986-2011, Victorian Department of Infrastructure, Melbourne.
- National Institute of Population and Social Security Research, 2006. Population Statistics of Japan 2006. Available at <http://www.ipss.go.jp/index-e.html>.
- Nelissen, J. H. M. 1991. Household and education projection by means of microsimulation model. *Economic Modelling* 8(4): 480-511.
- O'Leary, J. 1998. Estimating the Number of Households for Victoria as at 30th June 1996, Victorian Department of Infrastructure, Melbourne.
- O'Neill, B. and B. Chen. 2001. Demographic determinants of household energy use in the United States. In: *Population and Environment: Methods and Analysis*. W. Lutz, A. Prskawetz and W. Sanderson. A Supplement to *Population and Development Review* Vol. 28.
- Prskawetz, L. Jiang, and B. C. O'Neill. 2004. Demographic composition and projections of car use in Austria. *Vienna Yearbook of Population Research* 2004: 274-326.
- Registrar General and Census Commissioner of India. 2006. HH-16 Households by size and number of members seeking/available for work. Downloaded from the site <http://www.censusindia.net/index.html> on 30.08.2006.
- Statistics Austria. 1986a. Volkszählung 1981. Haushalte und Familien. Wien, Österreichische Statistisches Zentralamt, 1996. 630/26. Heft.
- Statistics Austria. 1996. Volkszählung 1991. Haushalte und Familien. Wien, Österreichische Statistisches Zentralamt, 1996. 1.030/26. Heft.
- Statistics Austria. 2005. Volkszählung. Haushalte und Familien. Wien, 2005.
- Statistics Austria. 2006. ISIS online database. Accessed at 20.09.06. at http://www.statistik.at/isis/current/isis_gui.shtml
- Statistics Canada. 2006. Downloaded from the site <http://www.statcan.ca/start.html> on 30.08.2006.
- U.S. Census Bureau. 2006. Historical time series of households by Size: 1960 to Present. Downloaded from the site <http://www.census.gov/> on 30.08.2006.
- Van Imhoff, E. and N. Keilman. 1992. LIPRO 2.0: An Implication of a Dynamic Demographic Projection Model to Household Structure in the Netherlands. Den Hague, Sweats and Zeitlinger.
- Van Imhoff, E., A. Kuijsten, P. Hooimeijer, L.J.G. van Wissen. Eds. 1995. *Household Demography and Household Modeling*. New York, Plenum Press.

- Zeng, Y., J. Vaupel, W. Zhenglian. 1998. Household projection using conventional demographic data/Frontiers of Population Forecasting, Supplement to Population and Development Review, Vol. 24. W. Lutz, J. Vaupel and D. A. Dhlburg.
- Zeng, Y., J. Vaupel, Z. Wang. 1997. A multidimensional model for projecting family households – with an illustrative numerical application. Mathematical Population Studies 15(3): 187-216.