Statistical Models for Time Use Data: An Application to Housework and Childcare Activities Using the Austrian Time Use Surveys from 2008 and 1992

Master Thesis in Statistics

University of Vienna

Bernhard Hammer

October 10, 2012

Abstract

Time use research analyses usually the total amount of time which is devoted to a certain activity on a given day. This variable is non-negative, there may be many persons not engaging in the activity on the survey day and the distribution is often right-skewed. The exact shape of the distribution is strongly dependent on the activity under investigation and characteristics of the analysed population. The statistical analysis requires models which account for the characteristics of time use data and which are flexible enough to accommodate distributions of different shapes. In this thesis I analyse changes in the time use of men and women for housework and childcare activities using the Austrian time use surveys from 1992 and 2008. Several statistical models are used and compared: The linear model, the Tobit model for censored data as well as generalized linear models with a negative binomial and a Poisson-gamma distribution as random components. The generalized linear model with a Poisson-gamma random component turns out to be a particular useful tool in time use research: It is flexible enough to provide a reasonable fit for most distributions and in particular also for activities with a large number of zero observations and a heavily skewed distribution. There has been a strong decline in the time which devoted to housework activities by women in Austria. Several characteristics can be identified which affect women's time use for housework in couple-households: the number of persons in the household, their age and the labour force participation of the partner. Higher educated women spend much less time on housework and women living in a city do less housework than those living in rural areas. None of these variables is significantly related to the housework of men. Their engagement is found to be independent of the year but also from household type, singles are doing with a bit more than one and half an hour the same amount of housework as men living in couple households with children. Average time devoted to childcare increased between 1992 and 2008 considerably for fathers and mothers, but the increase turns out to be insignificant after controlling for individual characteristics such as education.

Contents

1 Introduction **1**

1.1 Theory and Literature 2

2 The Austrian Time Use Surveys **10**

2.1 Comparability 10

2.2 Data and Descriptive Statistics 12

2.2.1 Household Structure and Sample Size 13

2.2.2 Labor Force Participation 14

2.2.3 Time Use for Housework 15

2.2.4 Time Use for Child Care 17

2.2.5 Time Use for Leisure and Personal Care 18

3 Statistical Models for Time-Use Data **20**

3.1 The Linear Model 21

3.2 The Tobit Model 22

3.3 Homoscedasticity 24

3.4 The Framework of Generalized Linear Models 25

3.4.1 Components of the Generalized Linear Model 26

3.4.2 Canonical Exponential Families and the Cumulant Function . 27

3.4.3 The Exponential Dispersion Family 28

3.5 The Negative Binomial Model 29

3.6 The Poisson-Gamma Model 30

3.6.1 Tweedie Distributions 33

4 Results **37**

4.1 Model Evaluation and Results for Housework 37

4.2 Model Evaluation and Results for Childcare 51

5 Conclusions **62**

List of Tables

2.1 Household Structure and Sample Size by Household Type 14
2.2 Labour Force Participation by Year, Sex and Household Type in Percent 15
2.3 Time Use for Housework and Paid Work by Sex, Household Type and Year in Minutes per Day 16
2.4 Time Use for Childcare by Sex and Year in Minutes per Day 17
2.5 Time Use for Leisure and Personal Activities by Sex, Household Type and Year in Minutes per Day 18
4.6 Housework: Results from the Linear Model 41
4.7 Housework: Results from the Tobit Model 43
4.8 Housework: Results from the Negative Binomial Model 46
4.9 Housework: Results from the Poisson-Gamma Model 49
4.10 Childcare: Results from the Linear Model 54
4.11 Childcare: Results from the Tobit Model 57
4.12 Childcare: Results from the Negative Binomial Model 58
4.13 Childcare: Results from the Poisson-Gamma Model 60

List of Figures

3.1 Empirical Density of Time Devoted Housework 21
3.2 Homoscedasticity Diagostic Plot: Logarithm of Variance vs. Logarithm of Group Means and Age 25
3.3 Housework of Women: Empirical Distribution and Fitted Negative Binomial Distribution 30
3.4 Distribution of the Number and Length of Housework Episodes: A Comparison with Fitted Poisson and Gamma Distribution 32
3.5 Empirical Density of Housework and Fitted Poisson-Gamma Distribution: Men 2008 36
3.6 Empirical Density of Housework and Fitted Poisson-Gamma Distribution: Women 2008 36
4.7 Time Devoted to Housework by Couples with Children: Empirical Density 40
4.8 Housework: Residual Analysis of the Linear Model 42

4.9 Housework: Evaluation of the Tobit Model - Comparison of Survey Data and Simulated Data from the Model 44
4.10 Housework: Evaluation of the Negative Binomial Model 47
4.11 Houswork: Evaluation of the Poisson-Gamma Model 50
4.12 Density Childcare 53
4.13 Childcare: Evaluation of the Linear Model 55
4.14 Childcare: Evaluation of the Tobit Model 56
4.15 Childcare: Quantile Residuals for the Negative Binomial Model . . . 59
4.16 Childcare: Quantile Plots for the Poisson-Gamma Model 61

1 Introduction

Data on the time use of individuals together with information on socio-economic background and household structure provide valuable information how individuals and subgroups of the population use and structure their time. The analysis of time-use data focuses usually on the total time which is devoted to a certain activity on a given day. This data is nonnegative, often right-skewed and may contain a large share of observations reporting zero time in the activity. These characteristics should be taken into account in the statistical analysis. Time use researchers use a wide range of different statistical models, including the linear model, the Tobit model for censored data and generalized linear models with a Poisson or negative binomial distribution as random component. The Tobit model is one of the most frequently used models. Many time use researchers regard it as suitable for time use data with zero observations, at the same time it is closely related to the linear model, the one most researchers are familiar with. There are however serious concerns about the use of the Tobit model for time use data. The main critique is that it does not account for the zero observations in an appropriate way, as the zero observations are not the result of censoring. We will see that the Tobit model does not fit for activities with a large share of zero observations, thus activities which are not carried out daily or almost daily. Examples are childcare and housework activities of men. Brown and Dunn (2011) suggest the use of a generalized linear model (GLM) with a Poisson-gamma distribution (also known as compound Poisson distribution) as random component. I show that a GLM with a Poisson-gamma distribution is indeed a valuable additional tool for time use researcher. It is flexible enough to provide a reasonable fit for different type of activities, in particular also for those with a high number of zero observations and a very skewed distribution.

In this thesis I compare the linear model, the Tobit model and generalized linear models with a negative binomial and compound Poisson distribution in an application to housework activities of men and women in Austria: Two diary based time use surveys conducted in 1992 and 2008 are used to analyse changes in the time use for household and childcare tasks in this period. In the first part of the thesis I motivate the research topic and provide an overview over the literature. Theory as well as empirical research from other countries suggests that we should expect an increase in the time which is devoted to housework by men and a strong decrease for women. Regarding the time used for childcare most of the studies find an increase for both, men and women. Higher educated men are found to contribute a higher share to housework and childcare. I analyse if these pattern for housework and childcare can be found also in Austrian data, the results are reported in part 4. Part 2 provides an overview over the Austrian time use surveys and Part 3 introduces and discusses the statistical models in detail.

I use the usual conceptualization of housework as the unpaid tasks carried out for the maintenance of the house or flat and to satisfy the basic needs of the household members. It includes tasks such as preparing meals, washing the dishes, cleaning house and garden, doing the laundry and repair work as well as the production of durable goods used in the household, for example carpenter work, sewing and knitting. For these activities I will use the terms *housework* or *chores*. Care activities are not included in this category: Care for adults is not included because the categorization of these activities is not comparable in the two surveys; and childcare is treated as a distinct activity because the nature is different from other household work. One important difference is that contrary to housework people enjoy most of the childcare activities. With the terms *household tasks* and *household work* I refer to both, housework and childcare activities.

1.1 Theory and Literature

Housework and childcare activities are usually distributed among all household members who are able to carry out these activities. Gender plays and has ever played a particularly important role in the allocation of household tasks; studying the engagement in household work requires therefore studying the allocation of household work between men and women. Since the industrial revolution and the emergence of gainful employment it is the norm that women are mainly responsible for the household and men as main breadwinner for generating income in paid work. The decades after World War II saw an extreme gender-polarization in the distribution of production activities, with men doing very little household tasks and women specializing as housewives. As Esping-Andersen (2009) points out, this period was rather exceptional. In former times household work was more labour intensive (e.g. doing the laundry by hand...) and often included the production of goods, for example through farming activities. A sole devotion to household work in a narrow meaning was for the broad population for economic reasons not feasible and a privilege for those who were economically better off. The 1950s and 1960s with a high growth of income allowed a broader class of the population to take over the male-breadwinner female-housewife family model. More recent decades saw (again) an increase in the labour force participation of women: In Austria the labour force participation rate of women in the age from 25-54 increased from 52.6 percent in 1971 to 67.1 percent in 1992 and 81.5 in 2008. An similar increase could also be observed in other countries: For example in Sweden from 58.2 percent in 1970 to 87.6 in 2008 and in the US from 49.1 in 1970 to 75.8 percent in 2008. However, in the US as well as in Sweden and other Nordic countries the participation rate is rather constant since the 1990s; on a rather low level in the US; on a high level between 85 and 90 percent in the Scandinavian countries[1].

The increase in labour force participation and consequently in women's time devoted to paid work has been accompanied by a steep decline in the time devoted to housework. Gimenez-Nadal and Sevilla (2012) use data from the Multinational Time Use Study (MTUS) to analyse changes in the time use between the 1970s and the first decade of the current century for Australia, Canada, Finland, France, Norway and the US. The MTUS contains a collection harmonized time use studies, currently more than 60 datasets from 22 countries[2]. For each of the mentioned countries there are at least two surveys available from which one has been conducted in the 1970s. They find a strong decrease in the time women devote to unpaid work and a strong increase in the average time devoted to paid work in all the analysed countries. The time use for childcare requires a more sophisticated analysis than a simple comparison of averages, as it also depends on fertility and the number of children. Using data from the US Sayer et al. (2004) find an increase in the time

[1] *Source:* International Labour Organization, LABORSTA database.
[2] www.timeuse.org/mtus

devoted to childcare between 1965 and 1998 for mothers and fathers with an initial decline between 1965 and 1975, which she explains with the decline in the number of children per household. Gauthier et al. (2004) uses time use surveys from 16 countries conducted since the 1960s and confirms the trend, on average are both parents devoting more time to childcare with a particular strong increase from a low level for fathers. Analysing Canadian data in greater depth she finds a significantly stronger involvement in childcare of persons with higher education.
It seems that the reasons for the shift in the time use of women from unpaid household work to paid work are not only economic factors, it is found to be more pronounced for higher educated (and better off) couples (Esping-Andersen, 2009). The changes in time use can be rather summarized as a change of women's preferences, giving paid work with its advantages such as career, economic independence and social security a much higher value. This resulted also in a change of household composition (fewer children) and in the design of the life-course (later marriage, postponement of childbearing...). The trend of using less time for chores is also supported by the availability of new technology and new products, (e.g. dishwasher, convenience food) and a more widespread outsourcing of housework, for example by eating in a restaurant. The opposite development in the amount of time used for chores suggest that explanations for the increase in men's time devoted to household tasks have to be found in the division of housework within the households.

The question why men and women structure their activities so differently is a very central question in research on gender inequality. There are not only differences in total time devoted to household work, there are also remarkable differences regarding type and structure of these tasks. Women do the housework tasks which are scheduled and less time flexible such as preparing the meals, while men choose the tasks which are more discretionary. Using the Australian time use survey from 1997 Craig (2006) finds that mothers spend not only around double the time with and for their children even after accounting for the labour force status, but also that the more demanding tasks (physical care) as well as activities bound to a timetable (transporting children, cooking, putting children to bed,...) fall overproportionally on the mother. The fathers care is proportionally higher in the more enjoyable tasks such as talking and playing with the child. Furthermore, fathers spend time with children often when mothers are present, therefore not relieving women from responsibility. It is usually the women who carry the responsibility for managing care.

As Hook (2010) points out, the gender-polarized distribution of production puts women at economic risk and men at social risk. The responsibility of women for the home and the children limits their employment and career opportunities, affecting the earnings over the whole life-course. A break up of the relationship leaves them therefore in a very difficult economic situation, reflected in the high share of single parents which are at risk of poverty. The responsibility of men for breadwinning in turn limits their relationship with their children. An indicator that the argument - polarization of household and childcare tasks as social risk for men - has to be taken very seriously is the Austrian legislation which still discriminates men regarding child custody. The idea that the mother is the best caregiver for children is also embedded in law. Only recently the constitutional court repealed a law which assigned child custody for children born out of wedlock automatically and solely to the mother - after a decision of the European Court of Human Rights that this constitutes a violation of Article 14 (prohibition of discrimination) taken together with Article 8 (right to respect for family life) of the European Convention on Human Rights[1].

One of the most influential contribution to the research on family and households was certainly Gary Becker's *A Treatise on the Family* (Becker, 1981). Becker tries to explain decisions made within the households by applying neoclassical microeconomic methods. Based on economic reasons (contrary to norms, sentiments, ideology...) households decide rationally about the division of household labour. Given better earnings prospects of women it maximizes household utility if women specialize in household work and men in market production. This type of economic reasoning

[1] See http://www.humanrightseurope.org/2011/02/austria-loses-child-custody-human-rights-dispute/. The decision of the Constitutional Court of Austria can be found on http://www.menschenrechte.ac.at/dokumentation/2012/VfGH/VfGH_28_06_2012.pdf and the decision of the European Court of Human Rights on http://www.menschenrechte.ac.at/orig/11_1/Sporer.pdf

plays without doubt a fundamental role in the decisions of households and families, but there are several other dimensions of at least the same importance. Recent theories try to include further aspects in the explanations for the gendered distribution of household work. While Becker assumes a unitary utility function for the household emphasizes the relative resource hypothesis that there are players with conflicting interests. The division of production activities requires a lot of cooperation and bargaining among the household members. The relative resource hypothesis states that the partner with more external resources such as income has a greater power to bargain her-/himself out of the unpopular household tasks. This hypothesis is supported by empirical findings that the relative earnings of the women in relation to that of men are related to the division of household labour (Cunningham, 2007). However, Gupta (2007) finds that not the relative income matters but the absolute level. Having outside options and not being economically dependent clearly increases the bargaining power. She also points out that the relative resource hypothesis fails to explain why women with about the same income as their partner still do a much larger share of household work. Gender ideology as explanation for the division of household work recognizes the importance of social norms in these decisions. The division of household work is regarded as the performance of gender roles; gender-identity is expressed through the performance or non-performance of certain household work. The ideology explanation for the gendered division of housework is supported by the findings that couples with more gender-egalitarian attitudes tend to share housework more equally (Knudsen and Wærness, 2008). This explanation is not at odds with the basic idea of the bargaining power hypothesis, as the support or disapprovement of the personal environment is hard currency in the bargaining within the household.

Research in the last decade focused strongly how the national context and the welfare state are influencing the division of labour. The macro-level environment shapes social norms (normative expectations about behaviour) and influences attitudes and gender role-models. Using data from the International Social Survey Program Geist (2005) shows that in conservative welfare state regimes (Austria, Germany, Mediterranean Countries) it is more rare for couples to share housework equally than in social-democratic regimes (Scandinavian countries), which explicitly promote gender equity. The national context also influences more pragmatic decision makers by providing or denying access to resources and opportunities, thereby influencing the relative resources of spouses and providing incentives for adapting certain family models. Hook (2010) finds that long parental leaves are positively related with gender specialization and lower contributions of men to household work. She suggests that paternity leave boosts not only the short term involvement in housework and childcare, but also in the long run as fathers acquire skills as caretaker and the paternity leave fosters the relation between the father and children.

There are several indicators why we should expect changes in the amount and the gender division of time devoted to housework and childcare in Austria. First there is the higher labour force participation of women. Although women work often part-time (around 43% of all employed women in 2010[1]), own income decreases their economic dependency and increases their bargaining power within the household. And although Austria is still a rather conservative country regarding gender roles, attitudes became much more egalitarian in the last decades, in particular those of younger age-groups (Wernhart and Neuwirth, 2005). Because also better educated persons usually hold more egalitarian views (e.g. Brooks and Bolzendahl (2004)), it should translate into a more equal division of household labour among young and better educated couples.

In Austria also some public policies are promoting an increased participation of men in housework and childcare activities. A in this respect outstanding campaign was initiated in 1996 by the then minister for women, Helga Konrad. She planned to obligate spouses to an equal division of household work by law. It would have given women the right to divorce if men do not contribute to household work. Her campaign with the catchy title *Ganze Männer machen Halbe-Halbe* (full men share half-half) met strong resistance. However, an alleviated version of the law, containing the order for a balance of the contributions to the relationship and to housekeeping as well as the duty to contribute to housekeeping according to own possibilities, has

[1] *Source:* EUROSTAT, Employment and Unemployment (Labour Force Survey)

been enacted in 2000 (For a description of this campaign as well as the reactions in public and politics, see Steger-Mauerhofer (2007)). Another initiative of Helga Konrad was the expansion of public child care facilities through resources provided by the federal government (Kindergartenmilliarde). In 1997 and 1999 the government invested 600 million Schilling (€43.6 Mill.) to extend the coverage of public childcare. Public provided childcare has been expanded also in the years thereafter: While the enrolment in public child care of children below the age of 3 is still very small in Austria, there has been an increase in the pre-primary school enrolment rate of 3 to 6 year old: For 3 year old from 29 percent in 1992 to 53 percent in 2008 and for the 4 and 5 year old from 75 percent in 1992 to 90 percent in 2008[1]. Since the year 1990 fathers have the right to go on parental leave and draw child care benefit (Karenzgeld). This benefit compensates the caregiver to small children for the loss of labour income. In 1996 the total length of the period during which child care benefits were paid was linked to the participation of both parents: It was paid 18 months to one parent and for the full 24 months if both parents used it. In 2002 a new form of child care benefit (Kinderbetreuungsgeld) has been introduced: Until 2008 it was paid for 30 months to one parent and for 36 months if both parents were using it. For an more detailed overview over changes in the Austrian family policy until 2004 see Cizek et al. (2004)). As has been emphasized by Hook (2010), these changes should have an effect on the allocation of chores and in particular childcare duties. However, the share of men on paternal leave increased only slowly from 0.18 in 1990 to around 2 percent in 2001 and is still below 5 percent.

There are three hypotheses I want to test on Austrian time use data:

- Men in 2008 devote more time to housework and childcare than in 1992. This holds after controlling for other influences such as household structure and the number of children.

- Higher educated men devote more time to household work and women less time.

- The decrease for women and increase for men in the time used for chores and

[1] *Source:* OECD, http://stats.oecd.org

childcare is more pronounced for younger persons.

The following sections provide an overview over the data and introduce and discuss the statistical models to test these hypotheses.

2 The Austrian Time Use Surveys

In Austria time use surveys have been conducted in 1981, 1992 and 2008 as special programmes of the Austrian microcensus. Because of methodical differences I use only the two newer surveys: The survey conducted in 1981 was based on interviews while the surveys from 1992 and 2008 are based on time diaries, which is seen as the most reliable method to collect data on time use. The sample population of the microcensus are the private households; in selected households all household members older than 10 were asked to fill in time diaries. These diaries have time slots of fifteen minutes for the time from 5:00 to 23:00 and half an hour slots from 23:00 to 5:00. In each time slot the respondent had to fill in a range of information: The main activity she/he was carrying out; the secondary activity in case there were several activities carried out parallel; if the activity was carried out at home our outside the home; who else was present; and if the activity was carried out also for another household (e.g. preparing a meal or doing the laundry). The activities could be described in own words and were later coded into more than 200 categories in 1992 and more than 300 in 2008. The surveys contain beside the information from the time diary also variables from the microcensus. While in the basic programme of the microcensus it is obligatory to participate, the participation in special programmes is voluntary. In 1992 around 47 percent of the sampled persons participated in the survey (Österreichisches Statistisches Zentralamt, 1995). In 2008 it was 38.3 percent of the originally sampled households which took part in the survey (Statistik Austria, 2011), but there is a number of households in which not all adult members filled out a time diary. As Statistics Austria points out in the documentation to the survey, persons with a very high personal time-burden are likely to be underrepresented.

2.1 Comparability

There are some methodical differences also between the 2008 survey and the one conducted in and 1992. The survey period in 2008 was a full year from March 2008 to March 2009, while the 1992 survey was conducted only in March and September. These differences could affect the time use measures because of seasonal differences in activities and because of the exclusion of holiday periods. With testing for differences in women's housework and childcare activities over months no significant differences can be found for housework, neither for 2008 nor for 1992. There are significant but small differences in the time use for childcare; in March 1992 women spent around 15 minutes more on childcare than in September. For men there are differences in housework, they did on average 16 minutes less housework in March than in September 1992. That is, we expect not huge seasonal effects, but smaller differences in the time use estimates can be due to the different survey periods.

Another difference between the two surveys is the beginning and the end of the reported time-span: While in 1992 the diary was from 0:00 to 24:00, the reported time-span in 2008 was from 5:00 to 5:00. Affected by this difference are the travel times. I assigned travel times to an outside-home activity always to the next outside-home activity in the diary. The travel times from an outside-home activity to home are assigned to the last outside activity. At the beginning of the diary it might not be clear where the person is coming from and at the end of the diary where the way is leading to. These ways are categorized as unknown activity. As the frequency and purpose of traveling is usually different at midnight and 5:00 in the morning this might cause some differences, but not very large one: The average time spent on traveling with unknown purpose is less than 3 minutes and the total average time which could not be classified is about 4 minutes in 1992 and 7 minutes in 2008.

The categorization of activities differs in the two surveys, being somewhat more detailed in 2008. But the sub-categories housework and childcare allow the construction of a comparable measure which contains the same activities in both years. There can be small differences in the measure of paid work due to a different treatment of education related to work, the bulk of activities counting as paid work however is the same in both years; it includes any kind of paid work, preparation for work including the way to work as well as unpaid or paid help in a family business.

Some variables in the basic programme of the microcensus were also subject to a change. Of relevance in this analysis is the change in the concept of labour force status because the employment status of the partner will be included as control variable in the models. In the 2008 the employment status has been determined by the labor force concept of the International Labour Organization, according to which a person counts as employed if she/he was working at least one hour in the reference week or has been sick, on holiday, on parental leave, doing military service or had similar reasons to sustain from work. In 1992 a livelihood-concept was used according to which the person themselves could assign them to a group (employed, unemployed, pensioner, housewife, parental leave...) with the restriction on employment that they are usually working at least 12 hours a week. To make these two concepts more comparable I reclassified in 2008 recruits, persons who reported to work less than 12 hours a week and persons who report they did not work because they are on parental leave as not in the labour force.

2.2 Data and Descriptive Statistics

The time use survey in 2008 includes 8,234 individuals from 4,757 households and the survey from 1992 contains 25,233 individuals from 12,169 households. However, of interest in this analysis are those persons who face a trade-off between market and household work and in particular couples, as we are interested in the division of the tasks between the spouses. I therefore restrict the analysis to a subpopulation, respectively a subsample. One of the groups which I exclude is the pensioners: Many pensioners have an abundance of time; its use differs considerably from those in the labour force. Unpaid activities which are for persons in the labour force often an annoying duty become in retirement more a form of leisure activity, for example gardening or cooking. Pensioners use therefore a much higher share of their time for housework activities (247 minutes per day in the age-group 61-70 vs. 165 minutes in the age-group 31-50). Because in the 2008 data there is only information about employment but it is not possible to distinguish between housewives/househusbands and retirees, I exclude all persons aged 55 and older. The age 55 is chosen because women exit the labour force between 55 and 60, including these age groups would affect the comparability between men and women.

Time use for household work is also strongly affected by household structure. I focus in the analysis on 2 types of households: Couples with children and couples without children who are not living with another adult person. For the descriptive statistics I include also single parents and single households as well as a category for the "rest".

Although children often contribute a considerable amount of household work, their share is usually different from the one of their parents. I therefore exclude those household members which are the child of the household head or of her/his spouse and those below 15 years of age. The exclusion of those younger than 15 affects only households in the "rest"-category as only in three generation households and households with non-core-family members there are children which are neither those of the household head nor her/his spouse.

2.2.1 Household Structure and Sample Size

Table 2.1 gives an overview over household structure and the sample size by household type. There has been a remarkable change in the household structure between 1992 and 2008: The most remarkable development is the increase in the share of single households from 14.5 to 27.8 percent of all households and the decrease in the share of couple-households with children from 49.1 to 43.2 percent. The average size of the households in the sample decreased from 3.2 persons per household in 1992 to 2.65 persons in 2008. These changes also affect the total time use averages without necessarily requiring large changes by household type. Single women for example use much more time for paid work and less time for chores than those living in couple households. The change in household structure leads therefore to an increase of the average time devoted to paid work and a decrease in the average time devoted to chores.

Table 2.1: Household Structure and Sample Size by Household Type

	Share in Percent	Avg. Household Size	No. Households	No. Persons
1992				
Single Households	14.5	1.00	907	907
Couples with Children	49.1	3.87	3789	6673
Couples, no Children	13.5	2.00	972	1586
Single Parents	7.0	2.47	505	505
Other Households	15.9	4.50	1153	2062
Total	**100.0**	**3.20**	**7326**	**11733**
2008				
Single Households	27.8	1.00	696	696
Couples with Children	43.2	3.76	1238	2110
Couples, no Children	15.4	2.00	494	787
Single Parents	6.4	2.41	260	260
Other Households	7.2	4.00	182	302
Total	**100.0**	**2.65**	**2870**	**4155**

Notes: The percentages and the household size are calculated using survey weights. The category "Other Households" includes more generation households as well as couples and single parents living with another (possibly non-family) adult person and non-family households.

2.2.2 Labor Force Participation

The labor force participation of women has been increasing considerably between 1992 and 2008. As it has been accompanied by an increase of single households with high participation rates the increase by household type was rather moderate: The participation of women living in couple households with children increased from 49 to 60 percent and for single mothers from 70 to 81 percent, but stayed constant for all other household types (Table 2.2).

Table 2.2: Labour Force Participation by Year, Sex and Household Type in Percent

	Men				Women			
	1992		2008		1992		2008	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Labour Force Participation								
Single Household	90.2	1.5	82.1	2.1	87.7	1.5	86.3	1.9
Couples with children	95.5	0.4	94.3	0.8	49.1	0.8	59.7	1.5
Couples without children	90.2	1.1	94.0	1.3	77.9	1.4	79.4	2.0
Single Parents	86.2	4.4			70.2	2.2	81	2.6

2.2.3 Time Use for Housework

Table 2.3 provides an overview over the time use for paid work and household work by year, family type and sex. Austria follows what has been observed for other countries: There is a pronounced decline in the amount of time women devote to housework. The average time women devote to household work declined from around five hours in 1992 to three and half an hour in 2008. The largest decline is observed for women living with a partner and children from almost six hours on housework in 1992 to a bit more than four hours in 2008. Women living with a spouse and without children devote with two hours housework per day in 2008 around one hour less to housework than in 1992. For single mothers the decline was also about one hour to 2:20 hours of daily housework. The contribution of men in turn is found to be rather stable; there is a small and insignificant increase to around 100 minutes per day. Regarding men's engagement there is surprisingly little difference between household types. Single men use around the same time for household tasks as men living in a core family with children or as a couple without children. The additional housework arising through children is obviously carried out by the mother.

For analysing paid work I additionally exclude students, as they are often living in single households but do not participate in the labour market. The changes in the time use for paid work are less pronounced than those for childcare: The total time of women in paid work increased by a bit more than half an hour. A puzzle is the seemingly strong increase of paid work of women living in single households from 312 to 367 minutes a day. A closer inspection shows that in both years about the same share of single women are taking part in the labour market (87.7 percent in 1992 vs. 86.3 percent in 2008), and they have about the same number of average working hours (39). The working hours of those who were working on the survey day were also quite similar (516 in 1992 vs. 517 in 2008). However, in 1992 around 65 percent of them were doing paid work on the survey day, but in 2008 it was 73 percent. This higher share is found for all weekdays. I cannot explain this increase, but it could represent a methodical difference or difference in the sample more than a real increase of paid work in the population. The average for men is a bit lower in 2008, but because of the large variance in the time allocated to paid work there can be rather huge differences due to the sample.

Table 2.3: Time Use for Housework and Paid Work by Sex, Household Type and Year in Minutes per Day

	Men				Women			
	1992		2008		1992		2008	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Housework Total								
Single Households	92	5.7	102	5.5	161	6.3	132	6.6
Couples with Children	93	2.3	102	4.1	347	2.8	250	4.3
Couples without Children	99	4.9	107	6.0	230	5.4	177	6.5
Single Parents	160	14.7			262	6.9	202	8.8
Other Households	87	4.6	83	13.1	299	5.4	193	11.0
Total	**93**	**1.8**	**102**	**2.9**	**299**	**2.2**	**209**	**3.0**
Paid Work								
Single Households	382	14.8	359	15.9	312	12.2	367	14.8
Couples with Children	385	5.0	375	9.7	140	3.5	168	6.7
Couples without Children	364	10.9	381	15.7	278	9.3	288	13.3
Single Parents	365	35.0			279	12.1	289	16.1
Other Households	435	9.6	451	30.3	207	7.1	219	19.9
Total	**389**	**3.9**	**377**	**7.1**	**194**	**3.0**	**231**	**5.4**

Notes: The values reported for paid work are the averages over all weekdays and does not include persons still in education. For the standard error the approximation *standard deviation*/$\sqrt{\textit{number of observations}}$ is used. In 2008 there are too few observations in the category men - single parents to report reliable results.

2.2.4 Time Use for Child Care

Table 2.4 shows the averages of time devoted to childcare by sex and year for couples and single mothers with children below the age of 15. There are too few observations to report the values for single fathers as they usually do not have child custody for smaller children. We see an increase in time devoted to childcare of about 20 minutes for both, men and women. Remarkable is in any case the increase in the time of men devoted to basic childcare such as washing, feeding or bringing the child to bed from six to twenty three minutes. It also includes activities such as breast feeding, so it is "natural" that women devote more time to this activity. And although also mothers devote around 20 minutes more to these activities, the large relative increase for men is an indicator for a higher engagement in childcare, albeit on a low level.

Table 2.4: Time Use for Childcare by Sex and Year in Minutes per Day

	Men				Women			
	1992		2008		1992		2008	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Childcare								
Couples	38	1.4	70	3.5	114	2.4	137	4.5
Single Mothers					96	7.0	116	9.5
Basic Childcare								
Couples	6	0.4	23	1.5	41	1.1	58	2.8
Single Mothers					31	3.1	36	3.1

Notes: Basic childcare includes activities such as feeding, changing nappies, bringing the child to bed, waking the child up, dressing the child and preparing the child for school.

What is usually mentioned in time use research but in the analysis finally ignored is the secondary activity. Childcare is sometimes carried out parallel with other tasks: Women in couple households with children below the age of fifteen used in 1992 on average 114 minutes for childcare as main activity and around 29 minutes for childcare as secondary activity (of which 17 during housework). In 2008 they used 137 minutes for childcare as main activity and 28 minutes for childcare as secondary activity (of which 13 during housework). Men carried out childcare as secondary activity for 5 minutes in 1992 and 11 minutes in 2008; but it is not carried out while doing housework but mostly during time which is mainly used for leisure and personal care. I also focus solely on the main activity, but it may be worth studying other measures such as the secondary activity or total time spent with the children (see e.g. Craig (2006)).

2.2.5 Time Use for Leisure and Personal Care

The most remarkable change in time use between 1992 and 2008 is certainly the rather large drop in the time which women devote to housework. Since a day has 1440 minutes this raises the question what happened to the time which is not use for housework anymore? We saw already that the part of the reduction in housework time has been used for an increase in paid work and an increase of the time devoted to childcare. Another part has been used to increase the time devoted to leisure activities; the women in the subsample (younger than 55, not child of household head or his spouse) increased the time devoted to leisure activities by about 30 minutes to 3 hours in 2008 (Table 2.5).

Table 2.5: Time Use for Leisure and Personal Activities by Sex, Household Type and Year in Minutes per Day

	Men				Women			
	1992		2008		1992		2008	
	Mean	SE	Mean	SE	Mean	SE	Mean	SE
Leisure								
Single Household	293	10.6	332	11.3	269	8.6	261	10.2
Couples with children	258	3.2	253	6.0	204	2.4	229	4.3
Couples without children	283	6.9	281	10.2	242	5.6	264	8.2
Single Parents	298	26.3			210	6.9	224	11.0
Other Households	231	5.9	235	19.9	186	4.5	249	12.8
Total	**261**	**2.5**	**272**	**4.6**	**212**	**1.9**	**241**	**3.3**
Personal Activities								
Single Household	642	7.5	629	8.8	652	5.4	645	8.0
Couples with children	655	2.2	647	4.9	652	1.6	663	3.7
Couples without children	651	5.4	638	6.7	653	4.4	659	7.5
Single Parents	674	15.3			659	6.3	645	8.1
Other Households	652	4.2	639	16.0	658	3.3	663	10.3
Total	**653**	**1.8**	**641**	**3.6**	**653**	**1.4**	**658**	**2.8**

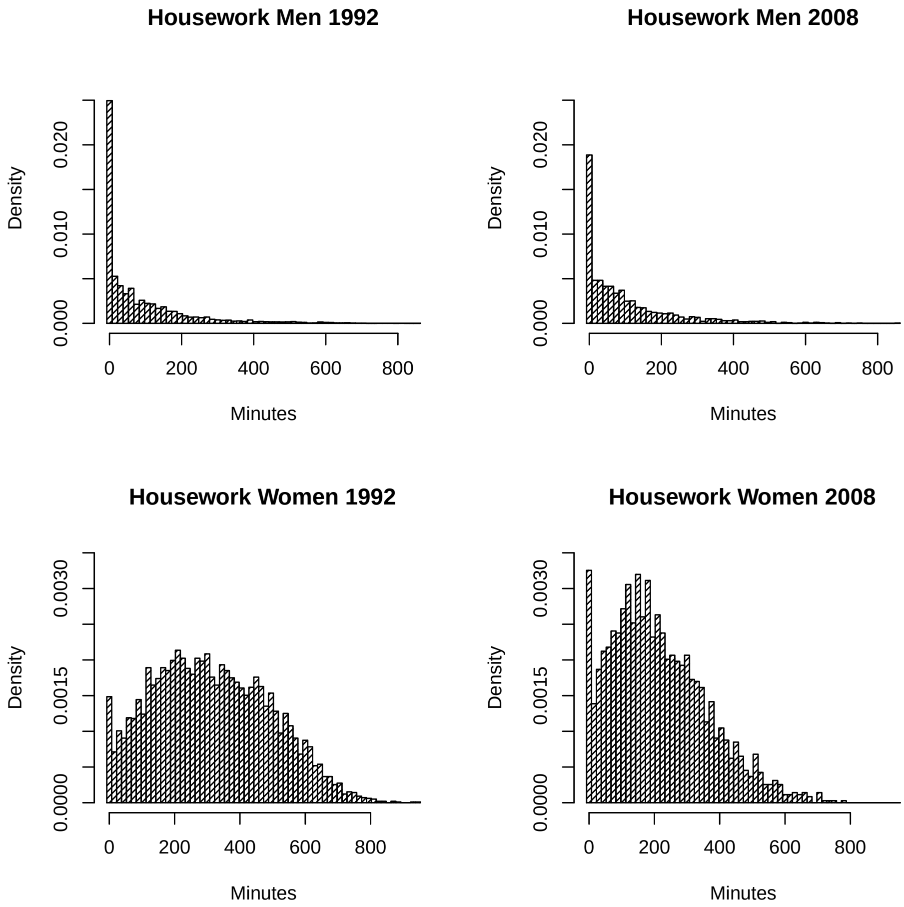
Notes: For the standard error the approximation $\textit{standard deviation}/\sqrt{\textit{number of observations}}$ is used. In 2008 there are too few observations in the category men - single parents to report reliable results.

The category of activities in which people spend the most time are personal activities. This category includes activities which serve the "maintenance" of the body such as sleeping, eating, personal hygiene or having a rest. There are surprisingly little differences between gender, household types and years: For all these groups the mean lies between ten and a half and eleven hours a day.

3 Statistical Models for Time-Use Data

Characteristic for time use data is the frequent observation of zeros, thus many persons do not engage in the analysed activity on the survey day. Brown and Dunn (2011) distinguish zero observations into two types: *Structural zeros* arise because a person does not engage in a certain activity at all; e.g. many adults without children never engage in childcare. *Sampling zeros* arise because the person does not engage in the activity on the survey day, but might otherwise carry out this activity frequently. Couples for example might divide housework and childcare duties according to the day of the week, or carry out certain housework tasks only on a specific day. Possible models for data with many structural zeros are hurdle models which have a component to model the zero observations, for example a binomial model, and a model for the duration given the hurdle is exceeded (e.g. Flood and Gråsjö (1998)). In our analysis structural zeros are less of a problem because it is very rare that a person does not engage in any kind of housework at all, and certainly also rare that a parent never engages in any kind of childcare.
Our task is to find an appropriate model for a variable which takes on only non-negative values, whose distribution has a point mass at zero and which can be heavily right-skewed. In the survey it appears as a discrete variable in form of the number of 15-minute and 30-minute time slots in which the activity has been carried out. Figure 3.1 shows as example the density of the total time devoted to housework activities for men and women in 1992 and 2008. Well visible is the left-shift of the distribution for women. In 2008 a larger share of women does not engage in housework at all on the survey day, and those who do use much less time.

Figure 3.1: Empirical Density of Time Devoted Housework



3.1 The Linear Model

The linear model (LM) used to be the most frequently used model in time use research. Regarding the distribution of the time which men devote to housework it would probably not be the first choice. But we will see that the LM can fit to activities with a low number of zeros, such as the time which mothers in couple households devote to housework. The linear regression model is of the form

$$y_i = \sum_{j=1}^{k} x_{ij}\beta_j + \epsilon_i. \tag{3.1}$$

where y_i is the variable of interest, thus the time spent in a certain activity, x_i is a vector of individual characteristics and β_j are the coefficients which describe the (linear) relation between the characteristics x_i and the response variable y_i. The ϵ_i capture influences which are not explained by $\sum_{j=1}^{k} x_{ij}\beta_j$; they are assumed to be independent and identically (iid) normal distributed with mean zero and variance σ^2. In matrix form the model can be written as

$$y_i = \beta' x_i + \epsilon_i, \tag{3.2}$$

The advantage of the linear model is its robustness and the straightforward interpretation of the regression coefficients. But the use of the linear model in time use research raises several concerns, I will point out only some of them: Because of the skewness and the zero-observations are the residuals often non-normal distributed. The LM might produce reasonable results even when the normality assumption is violated, but the usual tests for testing hypotheses (t-test, F-test) rely on this assumption. The same is true for the iid-assumption: We will see in Section 3.3 that the variance of the time devoted to housework and childcare is related to some of the explanatory variables. That is, the ϵ_i are not identically distributed but vary with some of the explanatory variables x_i (heteroscedasticity). This can lead to biased estimates of the standard errors and the significance levels. It is not only the violation of assumptions required for hypothesis testing which raises concerns, but the appropriateness of the model in general: There are no boundaries for possible values of $\beta' x_i$ in the model; it can therefore hardly be appropriate since the data is nonnegative and there can be many zero observations.
The frequent observation of zero time for an activity is a characteristic of the data which most researchers are aware of and try to account for in their modelling approach. An alternative to the linear model which is frequently used is the Tobit model for censored data.

3.2 The Tobit Model

The Tobit model is used in applications when the true value of a variable is not observed above or below some value. This is for example the case when survey data is top coded, thus if for certain variables (e.g. income) only the information that it lies above a certain value is available. But, as Wooldridge (2002) describes, such situations also appear when people have to choose corner solutions. Thus, they face a maximation problem but cannot choose outcomes below or above a certain value. Wooldridge (2002) motivates the use of the Tobit model with an example on charity spending, which clearly cannot lie below zero. The idea can be applied to time use as well: Assume that an individual faces the problem of choosing the right amount y of time spent in housework activities. The amount is dependent on a vector of individual characteristics x, such as employment status or the presence of children. The relation between the optimal time in housework and the characteristics x is described by the coefficient vector β. To determine the optimal amount of time which should be spent on housework the individual has to solve the optimization problem

$$y_i = \max(0, \beta' x_i). \tag{3.3}$$

Assuming that this model describes the process of choosing the amount of time spent in housework, we can translate it into a stochastic model of the following form:

$$y_i^* = \beta' x_i + u_i, \tag{3.4}$$

where y_i^* is an unobserved, latent variable. The u_i is an error term which captures individual preferences, characteristics not captured in the x_i and measurement errors. The observed variable y_i is defined by

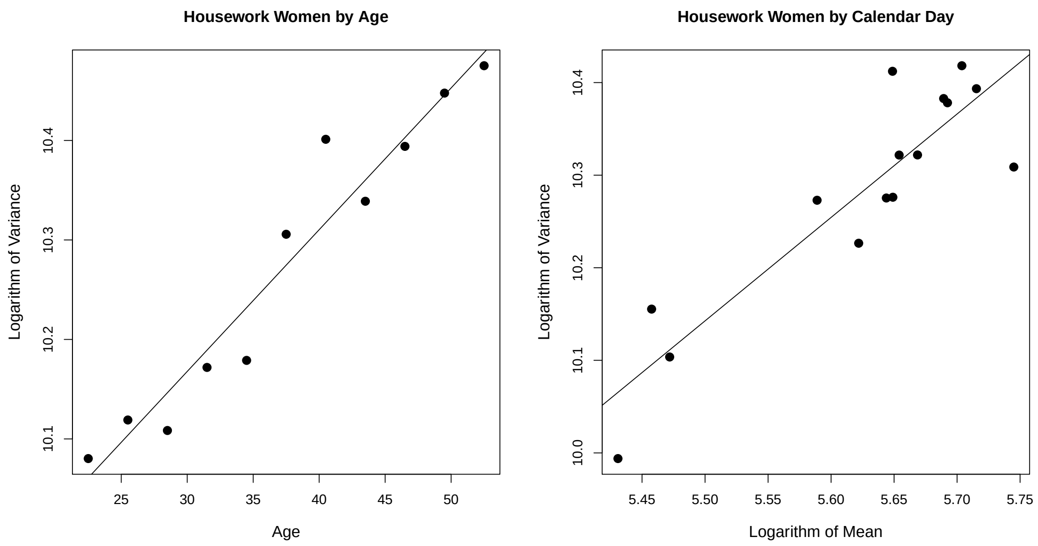
$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > 0 \\ 0 & \text{if } y_i^* \leq 0 \end{cases} \tag{3.5}$$

In the Tobit model the errors u_i are assumed to be normal distributed with mean zero and variance σ^2. This assumption together with the equation 3.4 and 3.5 constitute the standard censored Tobit model. The latent variable is interpreted as the propensity to perform a given activity. Once it reaches a given threshold, the activity is realized. Some of the time use researchers think that the Tobit model fits the data better than the linear model as it somehow accounts for the zero observations (e.g. Bonke (2010)). But the application of the Tobit model in time use research give rise to at least as severe concerns as the linear model: The zero observations are to a large degree sampling zeros, thus neither censored nor the corner solutions of an optimization process. The Tobit model requires the normality of the latent variable and as Brown and Dunn (2011) emphasize, the physical interpretation of the regression coefficients is rather difficult as they describe the influence of the explanatory variables on the latent variable and not on the observed variable.

3.3 Homoscedasticity

An assumption of the linear and the Tobit model is homoscedasticity. Recall that in the linear model we assume that the dependent variable Y_i (the time which observation i devotes to the activity) is normal distributed with mean $\mu_i = \beta' x_i$ and variance σ. That is, the variance is assumed to be independent of the mean μ_i, respectively the explanatory variables x_i. In the Tobit model we have a similar assumption for the distribution of the latent variable. This assumption is untenable for time use data. To illustrate this I first show that the variance is strongly related to age, an important variable in the regression models; older persons devote on average more time to housework and less to childcare. I divide the female observations into age groups (<24 and 24-54 in 3-year age-groups) and calculate the variance as well as the mean for each group. The logarithm of the variance is plotted against age in the first plot of Figure 3.2. We see a clear relation between age and the variance of the housework variable. For the second plot I formed groups according to the calendar day using two consecutive days per group and calculated for each group mean and variance. The logarithm of the group means and the logarithm of the variance are then plotted against each other. There is a clear positive relation between the mean and the variance, indicating that the normal distribution is not optimal to model this data. Brown and Dunn (2011) argue that the relation between the logarithm of the mean M and the logarithm of the variance V is of the form $log(V) = c + p * log(M)$ which would suggest the relation between mean and variance to be of the form $V = c' * \mu^p$ for some constant c'. This type of variance function with $p \in (1, 2)$ corresponds to Poisson-gamma models. In such cases where the range of the response variable Y is restricted and where the variance depends on the mean the use of a generalized linear model (GLM) can be more appropriate.

Figure 3.2: Homoscedasticity Diagostic Plot: Logarithm of Variance vs. Logarithm of Group Means and Age



3.4 The Framework of Generalized Linear Models

Two types of GLMs which have been used in time use research and which are also used in this thesis are GLMs with a negative binomial and a compound Poisson distribution as random components. This section gives a brief introduction to the framework of generalized linear models based on Lindsey (1997) and Altman (2009). In the linear regression model we are used to the idea of modelling the observations directly by $y_i = \beta' x_i + \epsilon_i$. Equivalently, we can think of modelling the expectation of each observation, $\mathrm{E[Y|X_i]} = \mu_i$ and assuming a conditional distribution for Y given μ_i. This distribution is assumed to to be the same for all μ_i and come from the exponential dispersion family (this family is described below). For example, we can write the linear regression model as

$$Y_i \sim N(\mu_i, \sigma^2), \tag{3.6}$$

where

$$\mu_i = \mathrm{E}[Y|x_i] = \beta' x_i. \tag{3.7}$$

In other words: We assume a distribution of the Y_i and model the mean of this distribution. The advantage of this approach is its flexibility: We can use this idea to model non-normal distributed data, for example discrete data. The observations Y_i may be discrete, while μ_i typically is a continuous variable.

3.4.1 Components of the Generalized Linear Model

The flexibility of the GLM framework lies not only in a broader choice of distributions for the errors; GLMs have 3 components:

1. The distribution of Y given μ_i, which is called the *random component.* This distribution is assumed to belong to the exponential dispersion family with a constant scale parameter.

2. The second component is the linear predictor, thus explanatory variables and regression coefficients that produce a linear prediction η_i:

$$\eta_i = \sum_{j=1}^{k} x_{ij}\beta_j \tag{3.8}$$

3. The third component is the link function g, which links the expectation of the i^{th} observation μ_i to its linear predictor η_i:

$$g(\mu_i) = \eta_i. \tag{3.9}$$

The link function g is required to be monotonic and differentiable over the range of μ_i. It determines how the linear predictor influences the mean. Two common choices are the identity link if an additive influence of the explanatory variables on the dependent variable is assumed, and a logarithmic link if the influence is assumed to be multiplicative. A particular choice is the canonical link, which transforms the mean into a canonical parameter. The advantage of the canonical link is, that there exist sufficient statistics for the coefficient vector β.

3.4.2 Canonical Exponential Families and the Cumulant Function

The random component in generalized linear models is required to be a member of the exponential dispersion family, which is introduced in this section. The introduction and overview is based on Jørgensen (1997), who provides a comprehensive and detailed study of these distributions. To a (one-dimensional) exponential family belong distributions whose density can be written in the following form:

$$f(y;\eta) = \exp\{a(y)b(\eta) - g(\eta) + h(y)\}, \tag{3.10}$$

where $a(\cdot)$, $b(\cdot)$, $g(\cdot)$ and $h(\cdot)$ are known functions and η is a parameter vector. If $a(y) = y$ and the parameterization so that $\theta = b(\eta)$, then the density is called *naturally parameterized* or in *canonical form* with the *canonical parameter* θ; y is called the *canonical statistic*. The density can then be written as

$$f(y;\theta) = c(y)\exp\{y\theta + \kappa(\theta)\}. \tag{3.11}$$

The function $\kappa(\theta)$ for a fixed θ has the role of a normalizing constant, so that the integral of $f(y;\theta)$ over the domain of y is one. It has therefore the form $\kappa(\theta) = \log\int e^{\theta y}\nu(dy)$ with $\nu(dy) = c(y)dy$. The function $\kappa(\theta)$ is denoted as *cumulant function*. The cumulant function is closely related with the moment generating function and like moments are cumulants a sequence of numbers which can be used to describe a distribution. The first cumulant is the mean of the distribution and the second cumulant the variance. However, cumulants of higher order however are not the same as the corresponding central moments. The reader is referred to Kolassa (2009) for the definition of cumulants and a discussion of their properties. The mean μ and the variance $var(Y)$ of the distribution can be derived from $\kappa(\theta)$ by taking the first and second derivate, respectively:

$$\mu = \kappa'(\theta) = \tau(\theta) \tag{3.12}$$

$$var(Y) = \kappa''(\theta) = \tau'(\theta) \tag{3.13}$$

The function $\kappa'(\theta)$ respectively $\tau(\theta)$ is called the mean value mapping and describes the relationship between the parameter θ and the mean μ of the distribution. Since τ has an inverse (it is monotone increasing and continuous) the variance can be expressed as function of the mean by

$$V(\mu) = \tau'(\tau^{-1}(\mu)). \tag{3.14}$$

The function V is called the variance function and characterizes a distribution within the canonical exponential family. That is, the variance function is unique for a distribution within the class of natural exponential distributions (a proof can be found in Jørgensen (1997), p.51). The distributions corresponding to a variance function of the form $V(\mu) = \mu^p$ with $p \notin (0,1)$ is called the Tweedie family. The density of this family will be derived from the variance function in Section 3.6.1.

3.4.3 The Exponential Dispersion Family

The *exponential dispersion family* is a generalization of the exponential family, distributions in this family include a scale parameter ϕ (also called dispersion parameter). The dispersion (e.g. variance) of a distribution usually depends on its mean. The scale parameter allows adjusting the dispersion without changing the mean. Densities of this type of distributions can be written as

$$f(y;\theta,\phi) = c(y,\phi)\exp\left\{\frac{y\theta - \kappa(\theta)}{\phi}\right\}. \tag{3.15}$$

This model is called the *reproductive exponential dispersion model*. The transformation $Z = \frac{Y}{\phi}$ is referred to as duality transformation; the density of the transformed variable has the form

$$f(z;\theta,\lambda) = c(z,\lambda)\exp\{z\theta - \lambda\kappa(\theta)\}, \tag{3.16}$$

where $\lambda = \frac{1}{\phi}$. This model is called *additive dispersion model*; Depending on the distribution one or the other form might be slightly easier to derive. For the reproductive exponential dispersion family the mean for Y is $\kappa'(\theta)$ and the variance function is $\phi\kappa''(\theta)$. The mean for Z in the additive dispersion model is $\lambda\kappa'(\theta)$ and the variance function is $\frac{1}{\lambda}\kappa''(\theta)$.

Example: Negative Binomial

The negative binomial distribution is known to model the number of of successes in a sequence of Bernoulli trials before the r^{th} failure. It has two parameters; the probability of a success p and the number of failures r. Let Z be negative binomial distributed with the parameter p and r. The probability that there are z successes before the r^{th} failure is

$$P(Z = z) = \binom{z+r-1}{z} p^z (1-p)^r \tag{3.17}$$

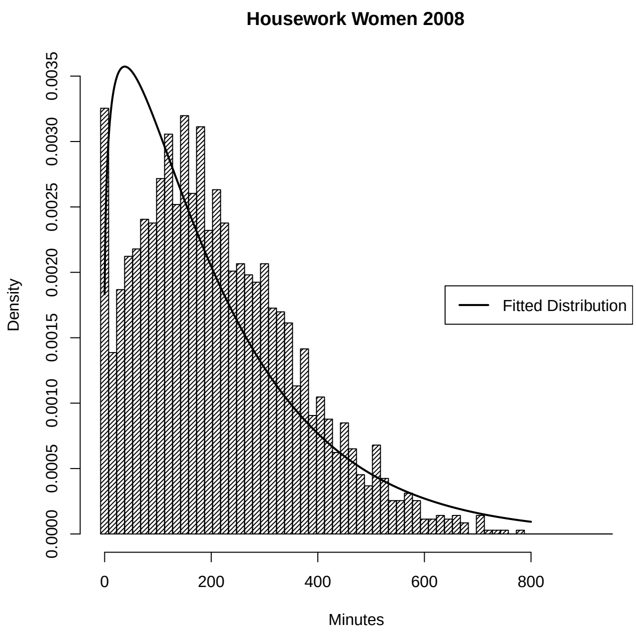
$$= \binom{z+r-1}{z} \exp\{z\theta + r\log(1-e^\theta\}). \tag{3.18}$$

The negative binomial distribution belongs therefore to the exponential dispersion family, the canonical parameter is $\theta = \log(p)$, the cumulant function $-\log(1-e^\theta)$ and the dispersion $\frac{1}{r}$. The mean μ of the distribution is $\mu = r\kappa'(\theta) = r\frac{e^\theta}{1-e^\theta} = \frac{rp}{1-p}$ and the variance $Var(Z) = \frac{1}{r}\kappa''(\theta) = \frac{1}{r}(\frac{p}{1-p} + \frac{p^2}{(1-p)^2}) = \mu + \frac{1}{r}\mu^2$.

3.5 The Negative Binomial Model

Hook (2010) for example uses a GLM with a negative binomial distribution as random component in her analysis. This model is often used for overdispersed count data. The GLM with a negative binomial distribution might suit the time use data better than the linear and Tobit model as the negative binomial distribution is non-negative, it can account for skewness and it allows zeros. In the application to time use data we interpret it as success if the activity is carried out in a fifteen minute time slot and as failure if this is not the case. For the negative binomial to be an appropriate model we have to assume that the probability to carry out the activity is a constant p for each time slot and in particular that this probability is independent of the activities carried out 15 minutes before. The problem in the application to time use data can be seen immediately: If an activity is carried out at a certain time it is quite probable that the same activity is carried out also in the next fifteen minutes, the activities in each time-slots are certainly not independent of each other. To fit the data it would require a high probability that there is no success before the r^{th} failure, thus that the person does not engage in the activity; But once there is a success there should be a high probability that there are more. The negative binomial cannot account for a high ("excessive") number of zeros. This is shown in Figure 3.3 for the distribution of time spent women spend in housework activities. The negative binomial distribution cannot accommodate the peak at zero and the second peak at around 180 minutes. One could argue that there might be a variable which "explains" the zero observations, resulting in a good fit of a more complex model. This is however unlikely, as the zero observation emerge largely as result of the sample, thus by coincidence.

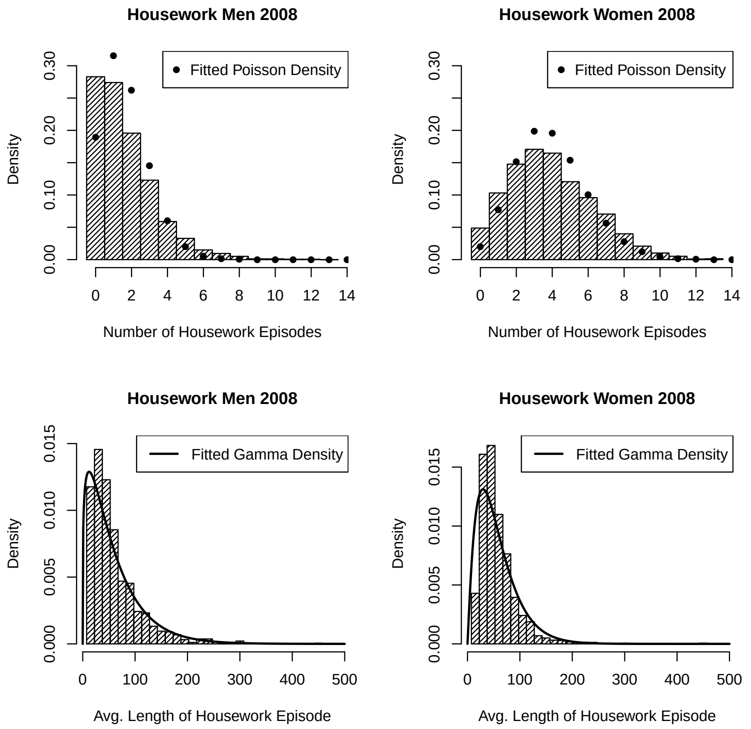
Figure 3.3: Housework of Women: Empirical Distribution and Fitted Negative Binomial Distribution



3.6 The Poisson-Gamma Model

The suggestion of Brown and Dunn (2011) to use a GLM with Poisson-gamma distribution (also compound Poisson distribution) is interesting as it recognizes the deeper structure of the data: The total time on a day which is devoted to a certain activity consists usually of several episodes with a different length. For example there are usually several episodes a day spent on preparing meals: A short period preparing breakfast, a longer period for lunch and dinner. Brown and Dunn (2011) argue that the number of times a person engages in a certain activity might be modelled with a Poisson process, allowing also the occurrence of zeros. Once a person engages in an activity, the duration can be modelled by a gamma distribution which can account for the skewness to the right. Assume the random variable N is Poisson distributed and Z_i follows a gamma distribution. Then the sum $Y = \sum_{i=1}^{N} Z_i$ follows a Poisson-gamma distribution. The sum Y represents the total time spent in this activity on a certain day. Figure 3.4 shows the distribution of the number of housework periods as well as their length and compares it with a Poisson and gamma distribution, respectively. Admittedly there is neither the Poisson model for the number of housework episodes nor the gamma distribution as model for their length completely convincing. However, the process which generates time use data is certainly a complex one and differs from activity to activity as well as across individuals and subpopulations. We cannot expect a perfect fit of relative simple models; the advantage of the Poisson-gamma is its flexibility and we will see that it provides a reasonable fit for very different distributions. The Poisson-gamma is also used by actuars as insurances encounter a similar problem of modelling the frequency of claims as well as their size (Kaas (2001)).

Figure 3.4: Distribution of the Number and Length of Housework Episodes: A Comparison with Fitted Poisson and Gamma Distribution



3.6.1 Tweedie Distributions

Those exponential dispersion models with a variance function of the form $V(\mu) = \mu^p$ with a real number p not lying between 0 and 1 are called *Tweedie distributions* and p is called the *shape* parameter. To the Tweedie family belong the normal distribution ($p = 0$), the Poisson distribution ($p = 1$ and $\phi = 1$) as well as the gamma distribution ($p = 2$).

Example: Poisson distribution

A Poisson distributed random variable Y has the density

$$f(y; \lambda) = \frac{\lambda^y}{y!} e^{\lambda} = \frac{1}{y!} e^{y \log \lambda - \lambda} \tag{3.19}$$

which can be written in canoniocal form as

$$f(y; \theta) = \frac{1}{y!} e^{y\theta - e^{\theta}}, \tag{3.20}$$

with the canonical parameter $\theta = \log \lambda$ and the cumulant generating function $\kappa(\theta) = e^{\theta}$. The mean and the variance are $\kappa'(\theta) = \kappa''(\theta) = e^{\theta} = \lambda$ and the variance function is $V(\mu) = \mu$; the Poisson distribution belongs therefore to the Tweedie family with $p = 1$.

Example: Gamma Distribution

The density of a gamma distribution is

$$f(y; a, b) = \frac{b^a}{\Gamma(a)} y^{a-1} e^{-by}. \tag{3.21}$$

Using the mean parametrization $\mu = \frac{a}{b}$ and $\phi = \frac{1}{a}$ the distribution can be written as

$$f(y, \mu, \phi) = c(y, \phi) \exp\left\{ \frac{-\frac{1}{\mu} y - \log \mu}{\phi} \right\} \tag{3.22}$$

with $c(\mu, \phi) = \frac{a^a}{\Gamma(a)} y^{a-1}$. The canonical parameter is $\theta = -\frac{1}{\mu}$ and the cumulant generating function is $\kappa(\theta) = -\log(-\theta)$. The mean μ of the distribution of Y is therefore $\kappa'(\theta) = \frac{1}{-\theta} = \mu$ and the variance function is $\kappa''(\theta) = \frac{1}{(-\theta)^2} = \mu^2$. So, also the gamma distribution belongs to the Tweedie family with $p = 2$.

Jørgensen (1997) shows that Tweedie distributions with $1 < p < 2$ correspond to compound Poisson distributions. The way to derive the densities for the Tweedie class from the variance function is also described in Jørgensen (1997) and in Swan (2006). First we find the corresponding cumulant function to the variance function $V(\mu) = \mu^p$. We know that $\kappa'(\theta) = \mu$ and $\kappa''(\theta) = \mu^p$.

$$\begin{aligned} \mu^p &= \kappa''(\theta) \\ &= \frac{\partial \kappa'(\theta)}{\partial \theta} \\ &= \frac{\partial \mu}{\partial \theta} \end{aligned} \tag{3.23}$$

By taking reciprocals we get

$$\mu^{-p} = \frac{\partial \theta}{\partial \mu.} \tag{3.24}$$

Integrating with repect to μ gives us

$$\theta = \begin{cases} \frac{1}{1-p}\mu^{1-p} & \text{if } p \neq 1 \\ \log(\mu) & \text{if } p = 1 \end{cases} \tag{3.25}$$

Because $\mu = \kappa'(\theta)$ we get for θ the expression $\theta = \frac{1}{1-p}\kappa'(\theta)^{1-p}$. After rearranging the terms we get again a differential equation:

$$\kappa'(\theta) = [(1-p)\theta]^{\frac{1}{1-p}} \tag{3.26}$$

Since we can ignore the constants without affecting the results we get for the cumulant function

$$\kappa(\theta) = \begin{cases} \frac{1}{2-p}[(1-p)\theta]^{\frac{2-p}{1-p}} & \text{if } p \neq 1, 2 \\ e^{\theta} & \text{if } p = 1 \\ -\log(-\theta) & \text{if } p = 2 \end{cases} \tag{3.27}$$

And as function of μ:

$$\kappa(\theta) = \begin{cases} \frac{1}{2-p}\mu^{2-p} & \text{if } p \neq 1, 2 \\ \log(\mu) & \text{if } p = 1 \\ -\frac{1}{\mu} & \text{if } p = 2 \end{cases} \tag{3.28}$$

The cases $p = 1$ and $p = 2$ correspond to the Poisson and gamma distribution, respectively. For p fixed and $p \neq 1, 2$ the density has the form

$$f(y; \mu, \phi) = c_p(y, \phi) \exp\left\{\frac{1}{\phi}\Big(y\frac{\mu^{1-p}}{1-p} - \frac{\mu^{2-p}}{2-p}\Big)\right\} \tag{3.29}$$

Jørgensen (1997) shows that the function $c_p(y, \phi)$ is given by

$$c_p(y, \phi) = \begin{cases} \frac{1}{y}\sum\limits_{n=1}^{\infty} \frac{\kappa_p^n(-\frac{\phi}{y})}{\phi^n \Gamma(n\frac{2-p}{p-1})n!} & \text{if } y \geq 0 \\ 1 & \text{if } y = 0 \end{cases} \tag{3.30}$$

The exponential dispersion model is in the following way related to the underlying Poisson and gamma distribution : $\lambda = \frac{\mu^{2-p}}{\phi(2-p)}$;$a = \frac{2-p}{p-1}$, $b = \frac{\mu^{1-p}}{\phi(p-1)}$ when they are parameterized as in 3.19 and 3.21, respectively.

Figure 3.5 and 3.6 show a compound Poisson distribution fitted to the empirical distribution of the time spent in housework activities. The fit is considerably better than for the negative binomial distribution. The huge advantage of the compound Poisson distribution is that it can have a point mass at zero.

For the estimation of the GLM model with the compound Poisson distribution I use the R-packages *tweedie* (Dunn, 2010) and *statmod* (Smyth et al., 2012). The power p of the Poisson-gamma distribution is estimated with the function `tweedie.profile()` from the tweedie package. After having a value for p specified, the GLM with a distribution from the Tweedie family can be estimated with the function `tweedie()` from the statmod package.

Figure 3.5: Empirical Density of Housework and Fitted Poisson-Gamma Distribution: Men 2008

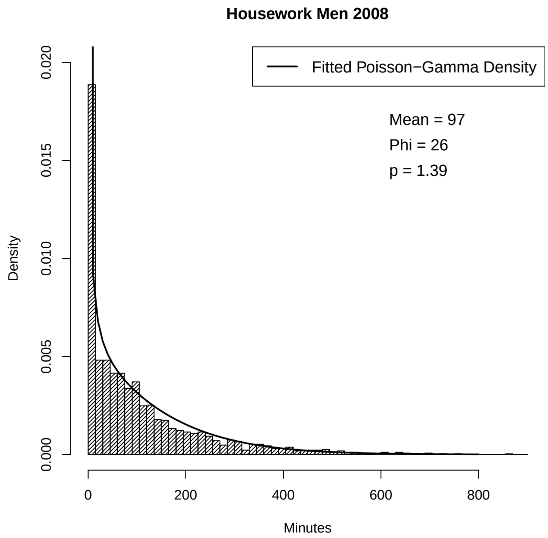
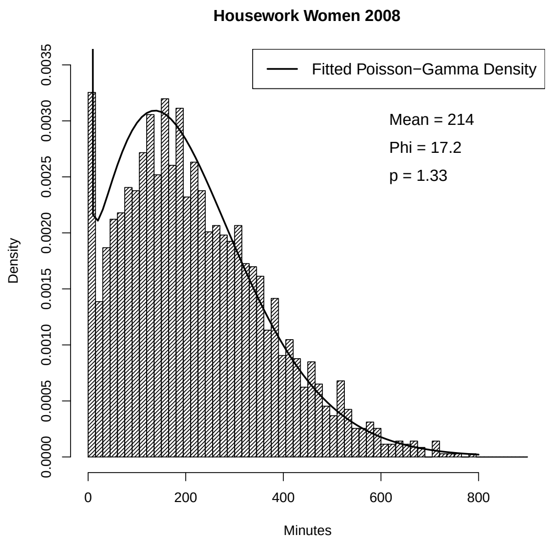


Figure 3.6: Empirical Density of Housework and Fitted Poisson-Gamma Distribution: Women 2008



4 Results

In this section the linear model, the Tobit model and the GLMs with the negative binomial and Poisson-gamma distribution are applied to housework and childcare activities of men and women. That is, I evaluate how well these models describe the data and test for the assumed relationship between time devoted to housework/childcare activities and year, age and education. To analyse changes in time use over years it is necessary to compare persons with similar characteristics. Since the focus is on people living in couple households it has to be kept in mind that this group in 2008 could be different from the group in 1992 in characteristics which cannot be observed. A hint in this direction is the strongly increasing share of single households. In particular women who do not want to devote several hours a day to housework and childcare activities might decide to live alone. A change in the time which is devoted to housework and childcare by a certain group might be (also) a selection effect rather than a real increase in the average time devoted to these tasks.

4.1 Model Evaluation and Results for Housework

I try to control for some of the characteristics of the individuals in the data which might affect their time use for housework but are not of primary interest. The time use of persons in rural areas might differ from those living in cities, due to differences in attitudes and the differences in infrastructure. I expect the time use for housework to be lower in cities, as the dwellings usually lack a labour intensive garden and there is a better supply of support services for the household (e.g. food deliveries, laundry and cleaning services). To capture these differences a dummy variable is included which indicates if the person lives in a city larger than 100.000 inhabitants. Another important factor influencing the total amount of housework is the total number of persons in the household and possibly the age of children. I also include the number of persons in the household and the number of children below the age of six as control variables. And because spouses often specialize in either household work or paid work, the partner's engagement in paid work is expected to affect the amount of household work. A variable indicating if the partner is working is therefore included.

The variables which are of primary interest regarding the hypotheses are year, age and education. Year is included as a dummy variable indicating that an observation is from the year 2008. If the first hypothesis that men used more time for housework in 2008 is true, then the corresponding coefficient should be significant and positive. For women we saw the huge decline in time spent for housework already in the descriptive statistics, we can therefore expect a large and negative coefficient for the 2008 indicator variable. The second hypothesis states that education affects the time spent for housework, with higher educated women doing less and higher educated men doing more housework. Two dummies are included for education, indicating that the highest completed education level is upper secondary with the school leaving exam (and qualification for university entrance) Matura and tertiary education, respectively. An education level below Matura is termed as basic education. It is not possible to distinguish between persons having only compulsory education and those with a completed lower secondary education such as an apprenticeship or a vocational middle school (Lehre, Berufsbildende Mittlere Schule), because there was no distinction between these type of education in the 1992 survey. The third hypothesis states that young people have much more liberal attitudes in 2008, which should lead to a higher participation in housework activities for men and less housework for women. In other words, the relation between age and housework is expected to change between 1992 and 2008. A mixed term of age and year is included to allow for a different relation of housework activities and age in the two years. If young people do more housework in 2008, the coefficient should be negative. Using the framework of GLM, the linear predictor for housework can be written as:

$$\begin{aligned}\eta_{housework} = &\ const. + \beta_1 * \textit{year 2008} + \beta_2 * \textit{upper secondary} + \beta_3 * \textit{tertiary} + \\ &\beta_4 * \textit{age} + \beta_5 * \textit{year 2008*age} + \beta_6 * \textit{household size} + \\ &+ \beta_7 * \textit{no. of children below 6} + \beta_8 * \textit{living in city} + \beta_9 * \textit{partner is working}\end{aligned} \tag{4.31}$$

In the linear model the linear predictor is related to the mean of the response variable through the identity function. In the Tobit model it is also the identity function, but it describes the relation between the linear predictor and the mean of the latent variable. For the GLM models with the negative binomial and Poisson-gamma distribution a log-link is used. This means, that the influence of the covariates is assumed to be multiplicative rather than linear.

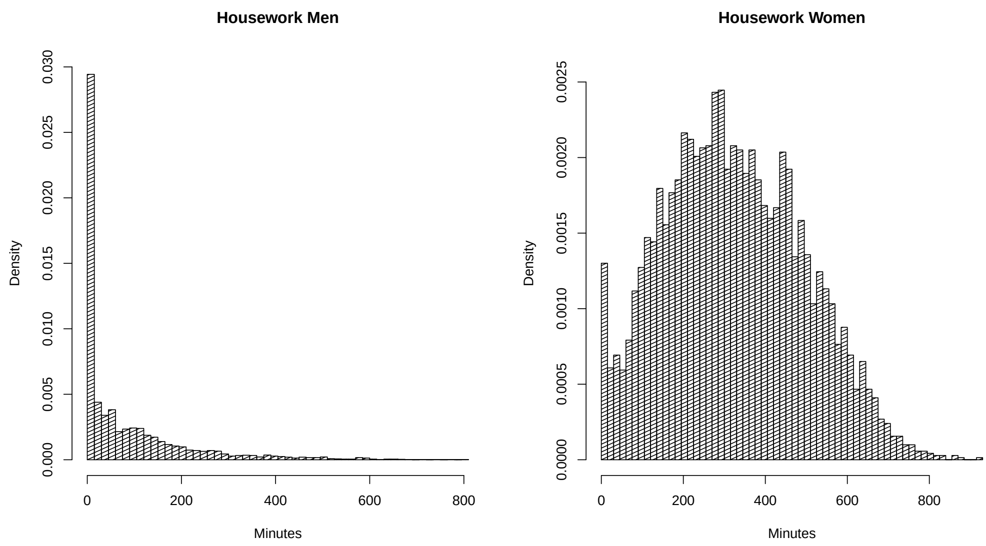
In a first step only persons who live in couple-households with children are included. The household type and in particular the presence of children affects the time use in a quite fundamental way; it would be not appropriate to assume the same relation between response variable and the explanatory variables for different types of households.

The density of housework activities for couples with children is plotted in Figure 4.7. For women the distribution of the daily time in housework activities looks like a censored normal distribution, we would therefore expect a reasonable fit of the Tobit model. Also the linear model might fit well as there are only few zero observations. But neither of these two models can be expected to fit for men, as the distribution is (almost) monotone decreasing and characterized by a high number of zero observations. This distribution can hardly be interpreted as a censored normal distribution. It turns out only the GLM with a compound Poisson distribution produces a reasonable fit for housework of men and an at least acceptable fit for women. The Poisson-gamma GLM is then used to estimate the model also for couples without children.

Housework: Results from the Linear Model

The results for the linear model are reported in Table 4.6. We have to forget about the results for men: As the R^2 and the residual plot (Figure 4.8) show, the model has an explanatory power of zero and residuals are far away from being normal distributed. For women the fit is much better: An R^2 of 0.14 is for time use data rather high and the residuals follow approximately a normal distribution. We know already that the variance of sample distribution increases with the time which is devoted to housework. This can be seen also in the plot of the predicted values vs. the residuals. The heteroscedasticity could affect the confidence intervals, but

Figure 4.7: Time Devoted to Housework by Couples with Children: Empirical Density



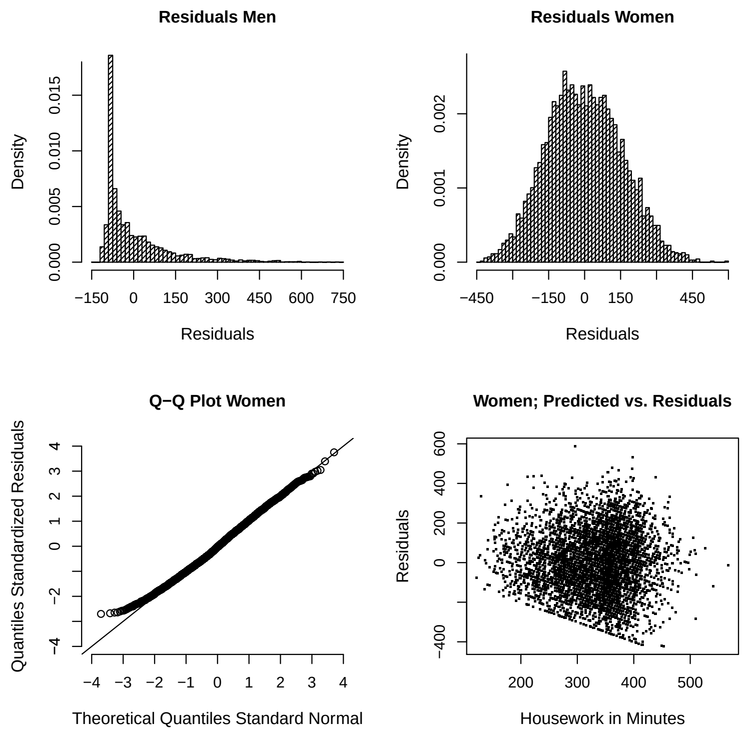
it should not bias the estimates. With the exception of the mixed term of age and the 2008-indicator are all of the included variables significant and have the expected sign. After controlling for changes in other characteristics the reduction in time devoted for housework was 1:48 hours from 1992 to 2008. Surprising are the rather large coefficients for education: Women with tertiary education devote more than an hour less to housework than those with only basic education; those with Matura but without tertiary education do still around 45 minutes less housework than those with basic education. This reduction of housework for higher educated women can be also partly an income effect, something we cannot control for due to a lack of income data. Higher educated persons usually also have higher income and can afford to outsource some of the household work, for example by engaging a cleaning service. The higher participation rates in paid work for better educated suggest that the lower time use for housework is not only an income effect, but that the time use preferences of higher educated women favour paid work (and as we will see, also childcare) over chores. Age has also a quite large effect, each year increases the average daily time devoted to housework by 2.8 minutes. Children (household size) increase housework by half an hour per child and by about 40 minutes when they are young (below 6). Also the city-effect is quite large: Persons living in a city do on average about half an hour less housework than those living in less urban areas. The indicator if the partner is employed is significant, but with 16 minutes comparatively low.

Table 4.6: Housework: Results from the Linear Model

	(Housework Men)	(Housework Women)
Year 2008	-11.57	-96.57***
	(25.73)	(27.82)
Upper Secondary (Matura)	-8.249	-45.36***
	(6.095)	(7.674)
Tertiary Education	-1.509	-72.41***
	(6.976)	(9.258)
Age	0.431	2.801***
	(0.329)	(0.391)
Age*Year 2008	0.686	0.157
	(0.621)	(0.695)
Household Size	0.483	32.92***
	(2.349)	(2.766)
Number of Children below 6	-3.550	10.20**
	(3.597)	(4.419)
Living in City	-1.027	-27.96***
	(5.764)	(7.013)
Partner is Employed	-0.0914	16.52**
	(3.983)	(7.548)
Constant	65.71***	112.2***
	(16.23)	(20.39)
Observations	4,068	4,715
R-squared	0.006	0.139

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Figure 4.8: Housework: Residual Analysis of the Linear Model



Housework: Results from the Tobit Model

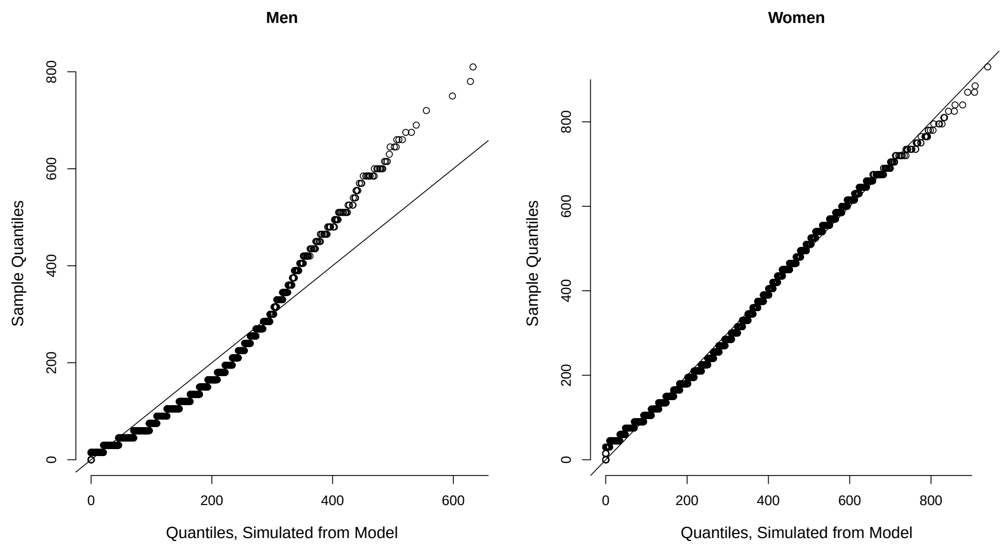
The results from the Tobit model are for women not much different from those of the linear model, because there are only a few zero observations. But it does also not deliver more reliable results for men. To evaluate the fit of the model I compare the sample distribution of time used for housework with a distribution which is simulated using the estimated model. That is, a normal distributed error term with mean zero and the estimated variance is added to the fitted values for the latent variable. Negative values are then censored and set to zero. Figure 4.9 shows the corresponding quantile plots. While the distribution of the sample and the simulated data is similar for women, it differs considerably for men: The Tobit model predicts about the same number of zero observations but underestimates the number of observations who spend a positive but small amount of time in housework activities as well as the heavy right tail of the distribution. This is exactly what we expect: The Tobit model fits a normal distribution with its left-tail censored a zero but, because less than half of the observations are censored, with strictly positive mean. It therefore underestimates the many observations which are positive but close to zero in the sample; overestimates the number of observations where the model mean is expected; and underestimates again the heavy tail of the sample distribution. The Tobit model is not appropriate to describe the distribution of the time which men devote to housework activities.

Table 4.7: Housework: Results from the Tobit Model

	(Housework Men)	(Housework Women)
Year 2008	8.306	-98.48***
	(38.01)	(28.17)
Upper Secondary (Matura)	-8.644	-45.13***
	(9.143)	(7.760)
Tertiary	7.128	-72.02***
	(10.33)	(9.362)
Age	0.334	2.820***
	(0.494)	(0.395)
Age*Year 2008	0.510	0.174
	(0.917)	(0.704)
Household Size	-0.0951	33.03***
	(3.520)	(2.796)
Number of Children below 6	-4.136	10.56**
	(5.405)	(4.469)
Living in City	-0.393	-28.08***
	(8.623)	(7.092)
Partner is Employed	8.497	16.86**
	(5.974)	(7.636)
Constant	19.12	110.0***
	(24.42)	(20.62)
Standard Error	170.8	158.6
Observations	4,068	4,715
Censored	1,489	61

Standard errors in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Figure 4.9: Housework: Evaluation of the Tobit Model - Comparison of Survey Data and Simulated Data from the Model



Housework: Results from the Negative Binomial Model

The results from the GLM with a negative binomial random component are shown in Table 4.8. Because the GLM with a negative binomial does not fit very well to our data and the results as well as the interpretation are similar to those of the Poisson-gamma model, I defer the interpretation of the results to the next section. The fit of the model is evaluated using two approaches: Like for the Tobit model I compare the empirical distribution with a distribution of the housework variable which is simulated using the estimated model. And additionally randomized quantile residuals are plotted, which is just another way to compare model and empirical distribution. Quantile residuals are defined in the following way (Dunn and Smyth, 1996): Let $F(y;\mu,\phi)$ be the continuous cumulative distribution function of an exponential dispersion model $P(\mu,\phi)$. If y has the distribution $P(\mu,\phi)$, then $F(y;\mu,\phi)$ is uniformly distributed on $[0,1]$. The quantile residual for observation i is defined by

$$r_i = \Phi^{-1}(F(y_i;\hat{\mu}_i,\hat{\phi}_i)), \tag{4.32}$$

with $\Phi()$ denoting the cumulative distribution function of the standard normal. Apart from the sampling variability of $\hat{\mu}$ and $\hat{\phi}$ are the quantile residuals standard normal distributed. A large deviation from the standard normal distribution indicates that the model $P(\mu, \phi)$ for y is not correct. Since in our case $F()$ is not continuous a more general definition is applied: Let $a_i = \lim_{y \uparrow y_i} F(y_i; \hat{\mu}_i, \hat{\phi}_i)$ and $b_i = F(y_i; \hat{\mu}_i, \hat{\phi}_i)$. The randomized quantile residual is defined by

$$r_i = \Phi^{-1}(u_i) \tag{4.33}$$

where u_i is a uniform random variable on the interval $(a_i, b_i]$. The randomization strategy is used to avoid overlapping points in the plots and is similar to the strategy of jittering. The time spent in a certain activity is clustered at multiples of fifteen and there are mainly discrete explanatory variables in the model, this would result also in a clustering of the quantile residuals. Without some form of jittering or randomization it would be not visible how many of the quantile residuals actually lie at a certain point. The quantile residuals are estimated using the function `qres()` from the statmod package.

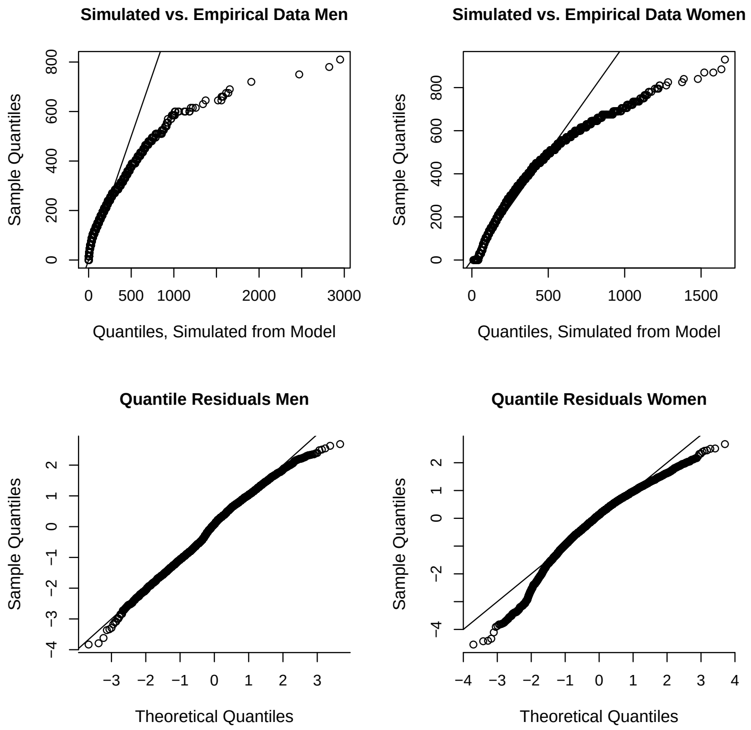
The plots with the simulated distribution and the quantile residuals for the negative binomial model show, that the model-distribution has a much stronger tail to the right than the empirical distribution. Another discrepancy between the model and the empirical distribution is at the left end: In particular for women there are more observations that use zero and little time for chores than the model distribution would predict, causing the quantile residuals of these observations lying far outside the range in which we would expect realizations of a standard normal. We can conclude that the negative binomial model is not appropriate to model the time use for housework, although for men it fits certainly better than the linear or the Tobit model.

Table 4.8: Housework: Results from the Negative Binomial Model

	(Housework Men)	(Housework Women)
Year 2008	-0.0139	-0.455***
	(0.450)	(0.108)
Upper Secondary (Matura)	-0.0998	-0.137***
	(0.110)	(0.0299)
Tertiary Education	-0.0152	-0.246***
	(0.125)	(0.0361)
Age	0.00530	0.00795***
	(0.00591)	(0.00149)
Age*Year 2008	0.00473	0.00389
	(0.0109)	(0.00270)
Household Size	0.00372	0.0954***
	(0.0415)	(0.0108)
Number of Children below 6	-0.0416	0.0346**
	(0.0649)	(0.0171)
Living in City	-0.0200	-0.0831***
	(0.104)	(0.0272)
Partner is Employed	0.0112	0.0603**
	(0.0718)	(0.0293)
Constant	4.206***	5.155***
	(0.300)	(0.0787)
Deviance	4.74	0.36
Observations	4,068	4,715
chi2	10.28	475.1
p-Value	0.328	0

Standard errors in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Figure 4.10: Housework: Evaluation of the Negative Binomial Model



Housework: Results from the Poisson-Gamma Model

In terms of significance and the sign of the coefficients delivers the Poisson-gamma more or less the same results than the other models (Table 4.9). Although the model fit for men is much better (Figure 4.11), there is none of the included variables significantly related to the time men devote to housework, the coefficients are small and insignificant. There is in particular no significant difference between the years. For women we see again the already familiar pattern with all of the coefficients significant except the age-year mixed term. The interpretation of the coefficients however is different from the linear and Tobit model: They do not describe the additive influence of the explanatory variables, but they describe it as multiplicative. The estimate for the year 2008 indicator for example mean, that women do $e^{-0.443} = 0.64$ times the housework in 2008 compared to 1992. For the mean of about 4 hours this corresponds to a reduction of a bit more than one and half an hour, which is similar to the estimates from the linear model. Women with tertiary education do 33 percent less housework than women with basic education. For the mean of about 4 hours this is again consistent with the linear model which estimates the housework time for tertiary educated a bit more than one hour lower. Those women with upper secondary education (Matura) but without tertiary education do about 13 percent less housework than those with basic education. The age effect is also found significant and positive with women at the age of 50 doing 17 percent more housework than those at age of 30. An additional household member increases housework by about 10 percent with a higher effect as long as the additional member is below the age of 6. The city effect is as expected; women who live in a city do about 10 percent less housework. And women whose partner is employed do slightly more housework than those with an inactive partner.

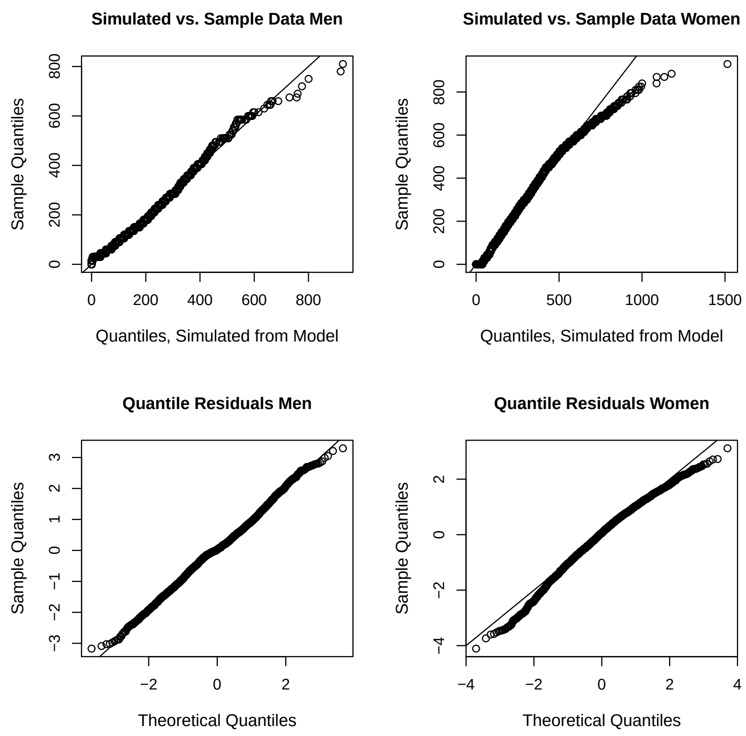
For couples without children the results are a bit different: Men as well as women do around 20 percent less housework if they live in cities. For women the age-effect is stronger than for the women with children; the results indicate that women at the age of 55 do 1.6 times more housework than those at the age of 30. No significant effect can be found for education and the indicator if the partner is employed or not.

Table 4.9: Housework: Results from the Poisson-Gamma Model

	Couples with Children		Couples without Children	
	(Men)	(Women)	(Men)	(Women)
(Intercept)	4.205***	5.171***	5.094***	5.395***
	(0.191)	(0.061)	(0.958)	(0.464)
Year 2008	-0.039	-0.443***	-0.175	-0.457**
	(0.295)	(0.093)	(0.354)	(0.165)
Upper Secondary (Matura)	-0.100	-0.145***	0.134	-0.048
	(0.073)	(0.025)	(0.119)	(0.059)
Tertiary	-0.016	-0.253***	0.019	-0.134
	(0.081)	(0.031)	(0.148)	(0.081)
Age	0.005	0.008***	0.007	0.019***
	(0.004)	(0.001)	(0.005)	(0.002)
Age*Year 2008	0.005	0.004	0.006	0.004
	(0.007)	(0.002)	(0.008)	(0.004)
Household Size	0.005	0.093***		
	(0.027)	(0.008)		
Number of Children below 6	-0.043	0.031*		
	(0.042)	(0.013)		
City	-0.016	-0.089***	-0.222*	-0.198***
	(0.068)	(0.022)	(0.097)	(0.044)
Partner is Employed	0.004	0.055*	-0.0631	-0.012
	(0.046)	(0.024)	(0.096)	(0.044)
phi	27.500	17.889	25.174	25.570
Likelihood-ratio	688.731	14091.089	342.738	6575.990
p-Value LR-Test	0.000	0.000	0.000	0.000
Number of Observations	4068	4715	1037	1336

Standard errors in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Figure 4.11: Houswork: Evaluation of the Poisson-Gamma Model



4.2 Model Evaluation and Results for Childcare

In the application and evaluation of the models with regard to childcare only couple households with at least one child below the age of 10 are included. Older children need no or only a rather small amount of childcare with the result that there would be a large share of zero observations. It could be seen already in the application to housework that in this case the Poisson-gamma would be the only model which might fit to the data.

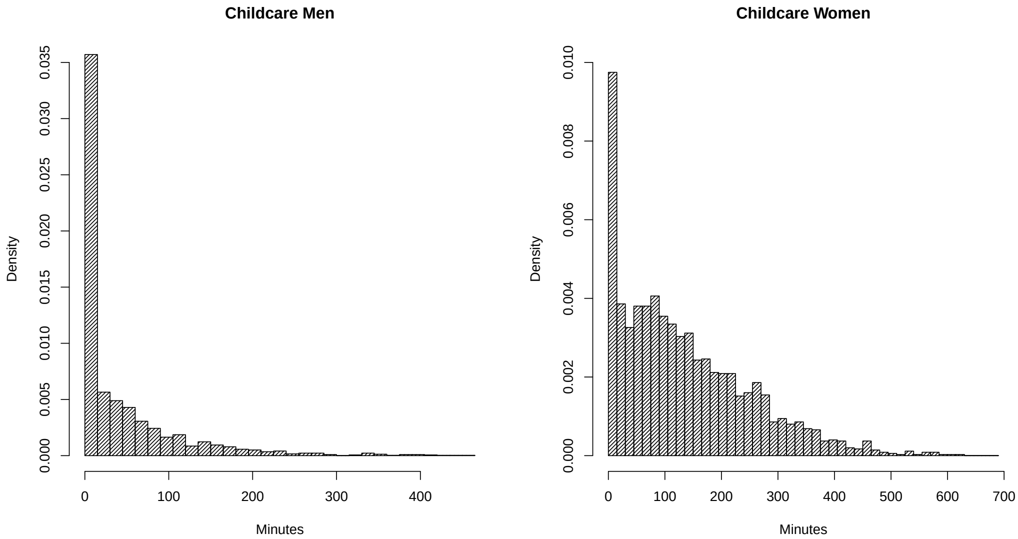
Most of the arguments used for the specific form of the linear predictor for housework are valid also for childcare. But childcare activities depend more on the age of the children than it is the case for housework, therefore I try to better account for the age structure of children: Household size is dropped, all households in the sample consist of a couple and at least one child below the age of ten. Included is instead the total number of children below 10, the number of children aged 0-2 and the number of children aged 3-5. The coefficient for the number of children below ten can be interpreted as additional time devoted to childcare for an additional child between the age of six and the age of ten. The coefficients for the number of children aged 0-2 and those aged 3-5 should be interpreted as the additional time which is devoted to childcare if there is a child in this age group; while time devoted to childcare increases with the number of children, we cannot assume that this happens in a linear fashion. Because is already rare that there are two children in one of these age-groups, the coefficients mainly capture the effect of having one child in these ages. The city indicator captures possible differences between urban dwellers and those in rural areas, but it has to be interpreted in a different way than for housework: The larger supply of child-care facilities in cities would point in the direction that parents in cities use less time for childcare. On the other hand, the lack of "natural" playground for children in cities (gardens, fields, forests) require more supervision of children and consequently a larger amount of time devoted to childcare activities.

For childcare the linear predictor has the following form:

$$\eta_{housework} = \textit{const.} + \beta_1 * \textit{year 2008} + \beta_2 * \textit{upper secondary} + \beta_3 * \textit{tertiary}+$$
$$\beta_4 * \textit{age} + \beta_5 * \textit{year 2008*age} + \beta_6 * \textit{no. children below 10} + \beta_7 * \textit{no. of children aged 0-2}+$$
$$+ \beta_8 * \textit{no. of children aged 3-5} + \beta_9 * \textit{living in city} + \beta_{10} * \textit{partner is working} \tag{4.34}$$

A histogram with the empirical density of the time devoted to childcare is shown in Figure 4.12. For men the shape of the empirical density of the time devoted to childcare is similar as the density for housework activities: There is high share of zero observations, and the density is slowly and monotone decreasing. For women there is, compared to housework, a larger share of zero observations. The density is rather constant with slight peak at about 100 minutes and is then decreasing. After analysing the time used for housework activities we have already a feeling which models could fit the data: For a good fit of the linear model there are certainly too many zeros. The distribution for women however looks like as if the Tobit model could deliver good results. No good fit should be expected from the negative binomial for the childcare activities of women as there is the high number of zeros and the second peak around 100 minutes, a pattern which the negative binomial cannot accommodate.

Figure 4.12: Density Childcare



Childcare: Results from the Linear Model

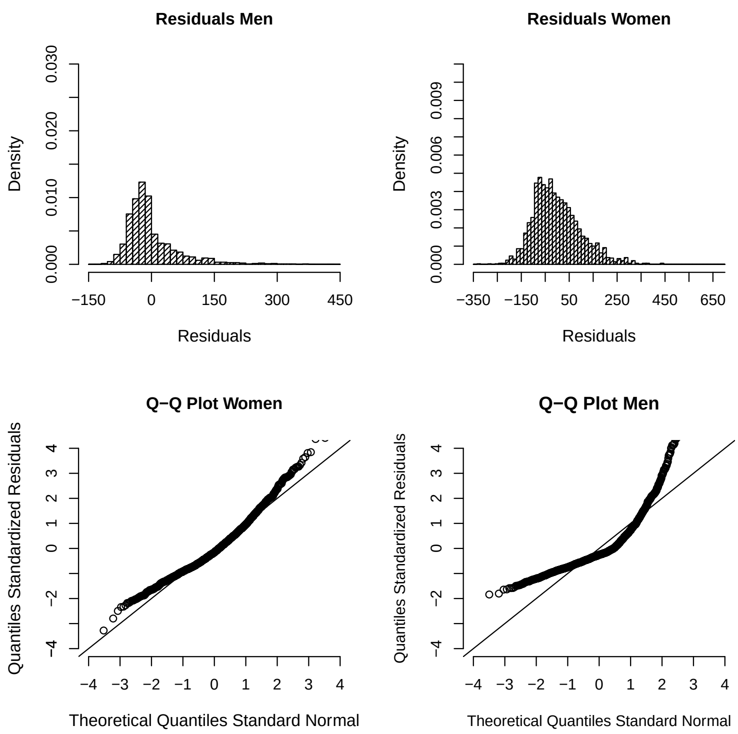
The results of the linear model are interesting as we get a significant and positive relation between the year 2008 indicator and the time spent in childcare activities for men. But, as we could already suspected by looking at the distribution of the dependent variable, the model does not fit well, neither for men nor for women. The residuals are not normal distributed but have a shorter tail to the left and a heavier tail to the right than would be expected from a normal distribution (Figure 4.13).

Table 4.10: Childcare: Results from the Linear Model

VARIABLES	Childcare Men	Childcare Women
Year 2008	38.27*	-19.12
	(19.65)	(27.65)
Upper Secondary (Matura)	0.194	14.05**
	(4.327)	(6.390)
Tertiary Education	16.51***	24.81***
	(4.829)	(7.360)
Age	-0.814***	-2.065***
	(0.276)	(0.458)
Age*Year 2008	-0.185	1.151
	(0.526)	(0.794)
Number of Children below 10	-0.422	10.39***
	(2.671)	(3.872)
No. Children aged 0-2	29.69***	92.03***
	(3.124)	(4.534)
No. Children aged 3-5	12.76***	27.14***
	(2.961)	(4.332)
Living in City	17.36***	8.984
	(4.267)	(6.291)
Partner is Working	6.978**	21.49***
	(3.045)	(7.719)
Constant	42.89***	116.3***
	(11.21)	(18.52)
Observations	2,112	2,332
R-squared	0.133	0.278

Standard errors in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Figure 4.13: Childcare: Evaluation of the Linear Model



Childcare: Results from the Tobit Model

The Tobit model fits, as suspected, quite well for the time in women devote to childcare activities. The distribution simulated from the model is quite similar to the empirical distribution of childcare activities. Education has a significant positive effect; having upper secondary education with Matura increases the propensity of childcare 13 minutes, tertiary education by about 27 minutes. The age of children is clearly related to the time devoted to childcare: A child in the age from 6 to 10 increases the propensity for childcare by 12 minutes, if it is in the age-group 3-5 by an additional half an hour and by additional 100 minutes if it is aged 0-2. Women whose partner is working are using more time for childcare activities. For men we have a similar problem as in the application to housework; the model underestimates the number of values which are close to zero and the heavy tail of the empirical distribution.

Figure 4.14: Childcare: Evaluation of the Tobit Model

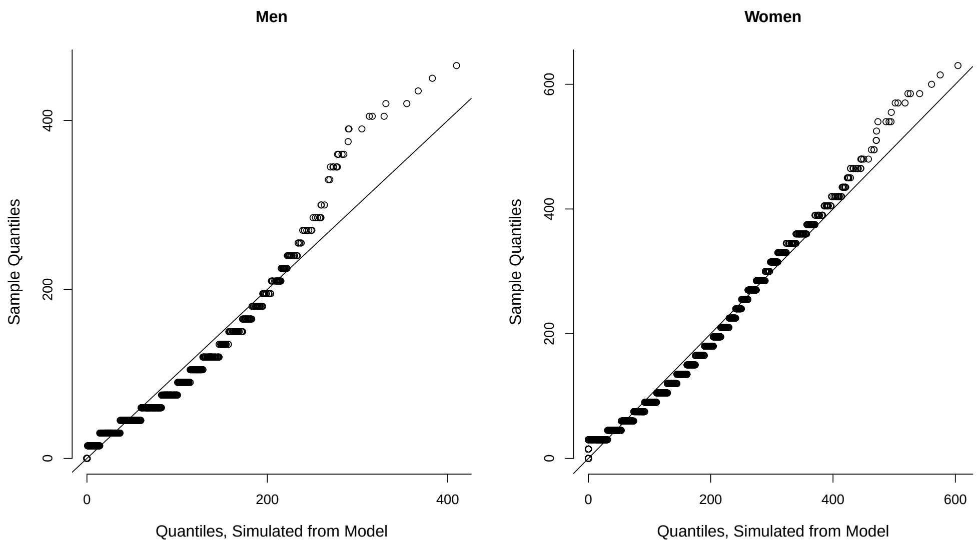


Table 4.11: Childcare: Results from the Tobit Model

VARIABLES	Childcare Men	Childcare Women
Year 2008	23.52	-24.97
	(34.02)	(30.81)
Upper Secondary (Matura)	8.522	15.24**
	(7.647)	(7.112)
Tertiary Education	28.05***	27.49***
	(8.431)	(8.181)
Age	-2.306***	-2.474***
	(0.521)	(0.515)
Age*Year 2008	1.053	1.464*
	(0.921)	(0.887)
No. of Children below 10	-4.225	11.70***
	(4.785)	(4.301)
No. of Children aged 0-2	59.45***	99.23***
	(5.616)	(5.036)
No. of Children aged 3-5	30.62***	31.51***
	(5.365)	(4.827)
Living in City	33.43***	10.08
	(7.386)	(7.007)
Partner is Working	15.49***	25.12***
	(5.540)	(8.641)
Constant	21.22	110.3***
	(20.63)	(20.72)
Standard Error	104.5	109.4
Censored	1,026	279
Observations	2,112	2,332

Standard errors in parentheses; *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Childcare: Results from the Negative Binomial Model

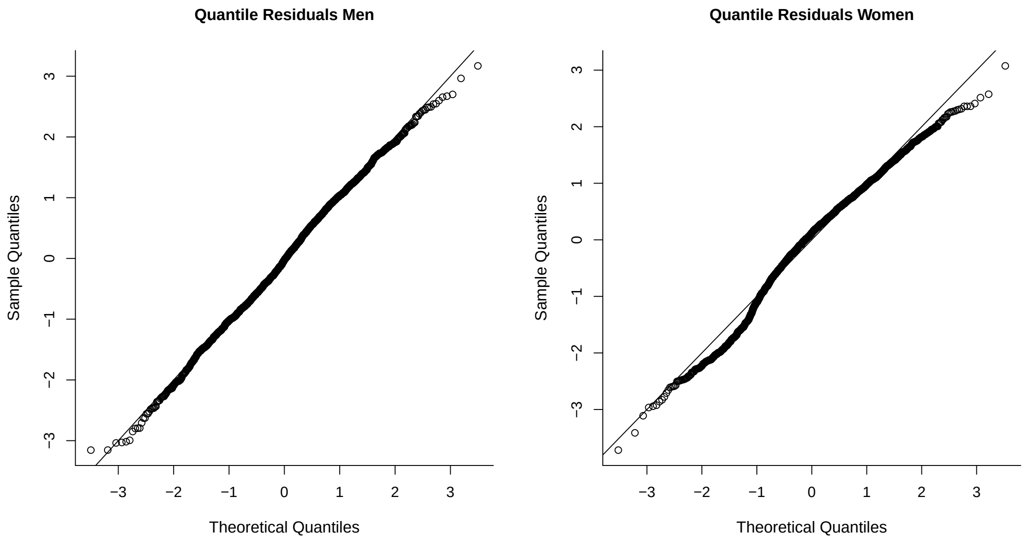
The GLM with the negative binomial distribution fits quite well for the childcare activities of men but as expected, not so well to the childcare activities of women. Also for men there is a strong relation between the age of children and the time devoted to childcare activities: Men devote double the time to childcare if there is a child aged 0-2 and 44 percent more if the child is aged 3-5 than for a child between 6 and 10. Surprising is the rather large effect for age: The results indicate that fathers aged 40 spend 35 percent less time to childcare activities than those aged 25 after controlling for the age of the children. Also living in a city affects the childcare activities of men, indicating that men in cities devote 1.5 times more time to childcare than those outside cities. As already pointed out, responsible for this large effect could be the lack of space for children which necessitates in a round the clock supervision.

Table 4.12: Childcare: Results from the Negative Binomial Model

VARIABLES	Childcare Men	Childcare Women
Year 2008	0.220	-0.211
	(0.728)	(0.319)
Upper Secondary (Matura)	0.0617	0.0875
	(0.165)	(0.0725)
Tertiary Education	0.351*	0.174**
	(0.183)	(0.0835)
Age	-0.0295***	-0.0191***
	(0.0103)	(0.00510)
Age*Year 2008	0.0128	0.0118
	(0.0195)	(0.00913)
No. Children below 10	0.0260	0.0936**
	(0.0987)	(0.0432)
No. Children aged 0-2	0.702***	0.608***
	(0.124)	(0.0515)
No. Children aged 3-5	0.367***	0.241***
	(0.110)	(0.0491)
Living in City	0.478***	0.0702
	(0.164)	(0.0721)
Partner is Working	0.199*	0.195**
	(0.119)	(0.0879)
Constant	3.857***	4.744***
	(0.423)	(0.209)
Logarithm of Over-Dispersion Parameter	1.812	0.232
Observations	2,112	2,332
p-Value χ^2 - Test	0	0

Standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Figure 4.15: Childcare: Quantile Residuals for the Negative Binomial Model



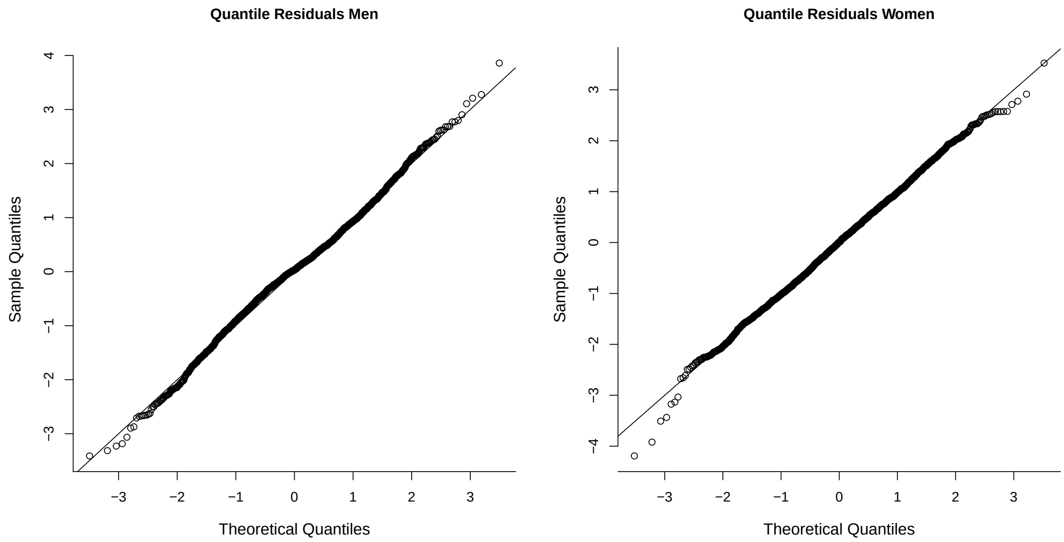
Childcare: Results from the Poisson-Gamma Model

The results from the GLM with a Poisson-gamma distribution confirm the estimates from the Tobit and negative binomial model for women and men, respectively (Table 4.13). These results are somewhat surprising: Contrary to what we would expect from the descriptive statistics there is no significant effect of the year. The effect of education is found to be large; men with tertiary education devote around 34 percent more time to childcare than those with only basic education, women with tertiary education about 18 percent more. The strong increase of average time devoted to child care and the insignificant effect of year in the model can be partly explained by an increase of average education levels: The share of women with in couple-households with children who hold a university degree increased from 5 percent in 1992 to 13 percent in 2008; the share of those with Matura but without tertiary education increased from 9 percent to 15 percent. The share of men with tertiary education increased from 7.2 to 12.7 percent, and the share of those with Matura but without tertiary education from 10.9 to 12.5 percent. Not surprising is the relation between time devoted to childcare and the age of the children: Men devote 2.5 times more time to childcare if the child is younger than 3 years as compared to a child aged 6-9, women still more than twice the time. A child aged 3-5 increases childcare time by about 20 percent for women and 30 percent for men compared to a child from 6-9. The labour force status of the partner affects both, men and women, childcare time increases by about 20 percent if the partner is working.

Table 4.13: Childcare: Results from the Poisson-Gamma Model

	(Childcare Men)	(Childcare Women)
(Intercept)	1.389***	2.216***
	(0.268)	(0.136)
Year 2008	0.013	-0.273
	(0.405)	(0.197)
Upper Secondary (Matura)	0.018	0.094*
	(0.097)	(0.045)
Tertiary	0.297**	0.169***
	(0.100)	(0.051)
Age	-0.031***	-0.021***
	(0.007)	(0.004)
Age*Year 2008	0.018	0.013*
	(0.011)	(0.006)
No. Children below 10	-0.028	0.054
	(0.061)	(0.028)
No. Children aged 0-2	0.950***	0.775***
	(0.118)	(0.054)
No. Children aged 3-5	0.323***	0.213***
	(0.069)	(0.032)
Living in City	0.368***	0.064
	(0.087)	(0.044)
Partner is Working	0.193**	0.169**
	(0.070)	(0.057)
p-Value LR-Test	0.000	0.000
N	2112	2332

Figure 4.16: Childcare: Quantile Plots for the Poisson-Gamma Model



5 Conclusions

The distribution of the total time spent in a certain activity depends strongly on the activity and the characteristics of the individuals under investigation. All the discussed statistical models can be appropriate for some activities and a certain subpopulation. The linear and the Tobit model fit quite well to the housework of women with children as there are only few women who did not engage in housework at the survey day and the distribution can be approximated with a normal distribution. The Tobit model fits also for data with a moderate share of zero observations, such as the childcare of mothers with small children. However, both the linear and the Tobit model assume homoscedasticity, an assumption which is untenable for time use data. Heteroskedasticity can lead to inconsistent estimates of the coefficients and the confidence intervals. The linear and the Tobit model are inappropriate for housework and childcare activities of men: The corresponding distributions are characterized by a large share of zero observations and a monotone decreasing density, a shape which cannot be described by these models as they are based on a normal distribution. The GLMs provide a much better fit for these distributions, in particular the Poisson-gamma model. The negative binomial model can fit to time use data but should be used with care: It is unsuitable for a frequent characteristic of time use data with a large share of zero observations and a density with a heavy right-tail or another peak at a strict positive value. This shape arises because if an activity is carried out, it is usually carried out for a longer period. A better choice is the generalized linear model with a Poisson-gamma random component, which turns out to be a very useful tool in time use research: It is flexible enough to provide a reasonable fit for most distributions and in particular also for activities with a large number of zero observations and a heavily skewed distribution. There is however not a generally preferable model for time use data, the model has to be selected according to the specific activity and subpopulation of interest.

We find for Austria a strong decline in the time which devoted to housework activities by women. But we clearly have to reject the first hypothesis that there has been an increase in the time men use for household tasks: The strong decline in time devoted to housework by women has not been accompanied by an increase in the contribution of men. For childcare we find an increase between 1992 and 2008 in the descriptive statistics, but no significant difference between the years after controlling for other characteristics. The increase in the average education level can explain part of the increase in the average time devoted to childcare, as higher educated persons are found to spend considerable more time on childcare activities and the share of persons with tertiary education and those with Matura is much higher in 2008 than in 1992.

For the second hypothesis that higher educated women spend less time for household work and higher educated men more, evidence is mixed. The idea behind the hypothesis was that better educated persons hold more gender egalitarian views and share household work more equally. With regard to housework the results do not support this argument. Higher educated women spend much less time on housework, but there is no evidence that higher educated men do more housework. But fathers with higher education are found to devote much more time to childcare than those with only basic education. The finding that both, men and women spend more time on childcare when they have higher education is an indicator for different time use preferences of higher educated persons general. But in relative terms is the effect of education on the childcare activity much larger for men, resulting in a more equal distribution of childcare activities for better educated couples.

There is no support for the third hypothesis that the engagement in household work increased more for younger men or decreased more for younger women as no change in the relationship between age and housework and age and childcare can be found.

References

Altman, R. (2009). Lecture Notes: Generalized Linear and Nonlinear Modelling. http://people.stat.sfu.ca/∼ raltman/stat402/ (accessed July 7, 2012).

Becker, G. (1981). *A Treatise on the Family.* Harvard Univ Pr.

Bonke, J. (2010). Childrens Housework–Are girls more active than boys? *Electronic International Journal of Time Use Research*, 7(1):1–16.

Brooks, C. and Bolzendahl, C. (2004). The transformation of US gender role attitudes: Cohort replacement, social-structural change, and ideological learning. *Social Science Research*, 33(1):106–133.

Brown, J. and Dunn, P. (2011). Comparisons of Tobit, Linear, and Poisson-Gamma Regression Models. *Sociological Methods & Research*, 40(3):511–535.

Cizek, B., Gysi, S., and Österreichisches Institut für Familienforschung (2004). *Internationales Jahr der Familie 1994 + 10: Bericht über 10 Jahre Österreichische Familienpolitik*. Bundesministerium für soziale Sicherheit, Generationen und Konsumentenschutz.

Craig, L. (2006). Does Father Care Mean Fathers Share? *Gender & Society*, 20(2):259–281.

Cunningham, M. (2007). Influences of Womens Employment on the Gendered Division of Household Labor over the Life Course. *Journal of Family Issues*, 28(3):422–444.

Dunn, P. (2010). R-package `tweedy`: Tweedy exponential family models. Version 2.0.7.

Dunn, P. and Smyth, G. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, pages 236–244.

Esping-Andersen, G. (2009). *The Incomplete Revolution: Adapting to Women's New Roles.* Polity Press.

Flood, L. and Gråsjö, U. (1998). Regression Analysis and Time Use Data A Comparison of Microeconometric Approaches with Data from the Swedish Time Use Survey (HUS). Working Papers in Economics, Göteborg University, Department of Economics.

Gauthier, A., Smeeding, T., and Furstenberg Jr, F. (2004). Are Parents Investing Less Time in Children? Trends in Selected Industrialized Countries. *Population and development review*, 30(4):647–672.

Geist, C. (2005). The welfare state and the home: Regime differences in the domestic division of labour. *European Sociological Review*, 21(1):23–41.

Gimenez-Nadal, J. and Sevilla, A. (2012). Trends in time allocation: A cross-country analysis. *European Economic Review*.

Gupta, S. (2007). Autonomy, Dependence, or Display? The Relationship Between Married Womens Earnings and Housework. *Journal of Marriage and Family*, 69(2):399–417.

Hook, J. (2010). Gender Inequality in the Welfare State: Sex Segregation in Housework, 1965–2003. *American Journal of Sociology*, 115(5):1480–1523.

Jørgensen, B. (1997). *The Theory of Dispersion Models*. Chapman & Hall/CRC.

Kaas, R. (2001). Compound Poisson Distributions And GLMs Tweedies Distribution. In *3rd Actuarial and Financial Mathematics Day*.

Knudsen, K. and Wærness, K. (2008). National context and spouses housework in 34 countries. *European Sociological Review*, 24(1):97–113.

Kolassa, P. McCullagh, J. (2009). Cumulants. *Scholarpedia*, 4(3):4699.

Lindsey, J. (1997). *Applying generalized linear models*. Springer Verlag.

Statistik Austria (2011). Standard-dokumentation: Zeitverwendungserhebung 2008 - 2009. Technical report.

Österreichisches Statistisches Zentralamt (1995). *Zeitverwendung 1992/1981: Ergebnisse des Mikrozensus, März/September 1992 und September 1981*. Österreichische Staatsdruckerei.

Sayer, L., Bianchi, S., and Robinson, J. (2004). Are Parents Investing Less in Children? Trends in Mothers and Fathers Time with Children1. *American Journal of Sociology*, 110(1):1–43.

Smyth, G., Hu, Y., Dunn, P., and Phipson, B. (2012). R-package `statmod`: Statistical modeling. r package version 1.4.14.

Steger-Mauerhofer, H. (2007). *Halbe/Halbe: Utopie Geschlechterdemokratie? Zur partnerschaftlichen Teilung der Versorgungsarbeit in Österreich*. Milena.

Swan, T. (2006). *Generalized estimating equations when the response variable has a Tweedie distribution: An application for multi-site rainfall modelling*. PhD thesis, University of Southern Queensland.

Wernhart, G. and Neuwirth, N. (2005). Geschlechterrollenwandel und Familienwerte (1988–2002). Austrian Institute for Family Studies, Working Paper Nr. 54.

Wooldridge, J. (2002). *Econometric Analysis of Cross Section and Panel Data*. The MIT press.