Chapter 4

Deceleration of the trend in inequality of educational outcome in the Netherlands

4.1 Introduction

The association between family socioeconomic status and offspring's educational attainment has been studied long and intensely in social stratification and social mobility research as it is considered a major indicator of the openness of a society (for example Breen and Jonsson, 2005; Hout and DiPrete, 2006). In this chapter I will focus on one aspect of this research theme: the association between socioeconomic background and the highest achieved level of education, which will be called Inequality of Educational Outcome (IEOut), and in particular on how this IEOut has changed over time. The main motivation for studying how IEOut has changed over time is the following dilemma: previous research has found that for the Netherlands IEOut has decreased linearly over time (De Graaf and Ganzeboom, 1990, 1993; De Graaf and Luijkx, 1995; Ganzeboom, 1996; Sieben et al., 2001). Such a linear decrease in the association between socioeconomic background and highest achieved level of education is improbable. A linear trend would eventually lead to a negative association between socioeconomic background and highest achieved level of education, which would mean that having a high status background would become a hindrance instead of an asset for attaining education. This is implausible, and as a consequence the negative trend in IEOut will have to slow down. This leads to the main question that this chapter tries to answer: has there been a deceleration in the trend in IEOut, and if so, when did this deceleration take place? To answer this question, the effect of family background on highest achieved level of education (IEOut) is allowed to change over cohorts according to a smooth but flexible curve. This smooth curve is used to estimate the trend, which is the slope or first derivative of the curve, and the change in trend, which is the slope of the slope or second derivative of the curve. To assess whether and when the negative trend significantly decelerated, I will test whether the change in trend is significantly positive, since a slowing down of a negative trend means that the trend moves from more negative to less negative.

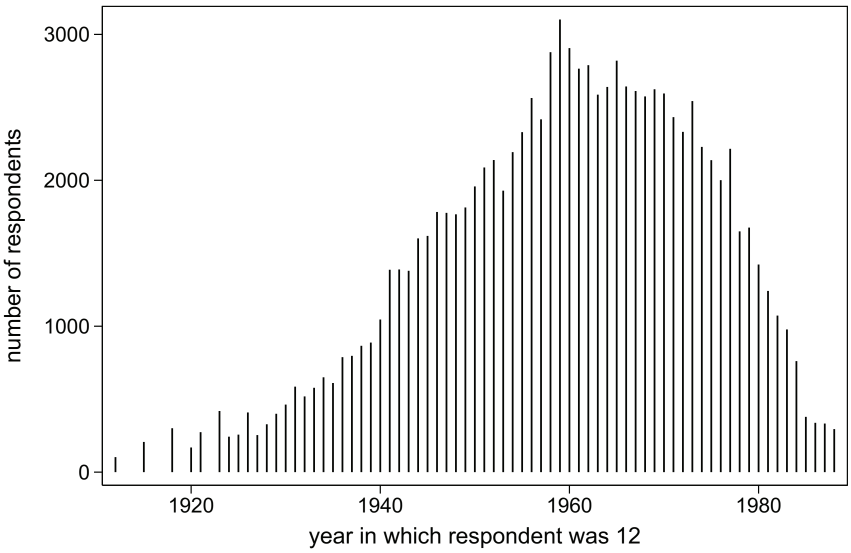
The secondary aim of this chapter is to assess the susceptibility of data assembled in the International Stratification and Mobility File (ISMF) (Ganzeboom and Treiman,

2009) to three potential sources of error. The first potential source of error is due to the fact that the ISMF consists of multiple surveys. The variables of these surveys have been post-harmonized and then stacked in order to create a single dataset. This could lead to a false trend if the quality of the surveys changed systematically over time. Such a systemic change in quality could for example occur because the response rate changed systematically over time. It is likely that the quality of a survey will influence the strength of the associations between the variables in that survey, as the associations in low-quality surveys will be contaminated by more noise than the associations in high-quality surveys. So, a false increasing (decreasing) trend in the association between family background and educational attainment can be expected if the quality of the surveys systematically increased (decreased) over time. The second potential source of error concerns the scale of education as used in the ISMF. This is an *a priori* scale based on the institutional years of education with an *ad hoc* correction of the level assigned to senior secondary vocational education (MBO). In Chapter 3 I proposed a scale with a stronger empirical foundation, which scales the levels such that education optimally predicts occupational status. The former scale will be referred to as the *a priori* scale while the latter will be referred to as the empirical scale. The most prominent difference between these two scales is that the *a priori* scale assigns too much value to lower vocational education (LBO). This can influence the estimated trend in IEOut as this difference in scaling means that the estimated trend is likely to respond differently to changes in the proportion of respondents with lower vocational education over time. The third potential source of error is missing data. This will lead to biased estimates if the likelihood of not answering a question is related to the value of the dependent variable (Allison, 2002). The dependent value in this chapter is the highest achieved level of education, and it is likely that the willingness and ability to finish a survey is associated with the respondent's highest achieved level of education. So, it is plausible that missing data could cause bias in the estimates of IEOut. The severity of this problem is influenced among other things by the proportion of observations that contain missing values. If the proportion of missing values changes over time, then the severity of this problem would change over time and thus also bias the estimated trend. The presence of these three potential sources of error is easier to detect when studying changes in the trend in IEOut, as this is a very subtle analysis. If the potential sources of error matter, then they will certainly show up in such an analysis.

4.2 Previous research

The challenge of studying the trend in IEOut is to cover a sufficient period of time such that the trend, and changes in the trend become visible. A common strategy is to take multiple surveys and compare respondents that are born in different years, that is, the time is captured by comparing so-called synthetic cohorts. By comparing synthetic cohorts, a single survey can cover a period of 40 years (when using respondents who are between 25 and 65 years old). This period can be further extended by adding surveys collected at different times. These cohorts are used as a measure of when the effect of social background on educational attainment took place. This is reasonable, given the strongly age-stratified nature of full-time education, which means that people born in the same year experience a very similar educational system. Within the Netherlands, this technique was first used for the study of the trend in IEOut by Peschar et al. (1986), and has been used in numerous other studies since (Peschar 1987; Ganzeboom and De Graaf 1989; De Graaf and Ganzeboom 1990; De Graaf and Luijkx 1992; De Graaf and Ganzeboom 1993; De Graaf and Luijkx 1995; Ganzeboom et al. 1995; Ganzeboom 1996; Rijken 1999; Korupp et al. 2000, 2002; Breen et al. 2009; and Chapter 2 of the current dissertation), and resulted in the International Stratification and Mobility File (Ganzeboom and Treiman, 2009). Six of these studies (De Graaf and Ganzeboom 1990; De Graaf and Luijkx 1992; De Graaf and Ganzeboom 1993; De Graaf and Luijkx 1995; Ganzeboom 1996; and Chapter 2 of the current dissertation) test whether the trend in IEOut in the Netherlands is linear or not, and none of these studies can reject the hypothesis that IEOut is linearly decreasing over time. In all cases, the tests for non-linearity of the trend were performed by comparing a model with a linear trend with a model with a non-linear trend. This non-linear trend was either a quadratic trend or a discrete trend, where the period was broken up into a series of cohorts, and separate IEOuts were estimated for each cohort. The use of these methods could explain why no non-linearity was found, as quadratic functions can easily be too rigid to adequately represent a non-linear trend, while a discrete trend is very flexible but expends a lot of statistical power, making it hard to find any significant evidence for non-linearity in the trend. The main aim of this chapter is to find out if any non-linearity in the trend can be found if one uses a model that is more flexible than a quadratic function but less flexible, and thus more powerful, than a discrete trend.

Figure 4.1: Number of observations per cohort



4.3 Data

The data consist of 54 surveys held in the Netherlands between 1958 and 2006 that were post-harmonized as part of the International Stratification and Mobility File (ISMF) (Ganzeboom and Treiman, 2009). Where available, survey weights have been used. The weighted number of respondents is 86,581. The number of respondents is unequally distributed over the cohorts, as is shown in Figure 4.1. Of these respondents, 9,416 lack information on father's occupational status, 651 on the respondent's highest achieved level of education, and 342 on both. If the proportion of missing information varies across cohorts, then this could bias the estimate of the trend. Figure 4.2 shows that the proportion of observations with missing information has changed systematically over time. The reason for these changes across cohorts could be in part an age-effect, as the older cohorts will consist mainly of people that were old at the time of the interview, and in part be a period effect, which can for example capture changes in a general attitude towards surveys and the introduction of computer-assisted interviewing, which makes it harder to skip questions.

Time is measured by the year in which the respondent was 12, which is the age at which most persons in the Netherlands make the most important decision in their educational career. Information is available for the cohorts aged 12 in 1912–1988. IEOut is measured by the strength of the metric association between highest achieved level of education of the respondents and their father's occupational status. Father's occupational status is measured in ISEI scores (Ganzeboom and Treiman, 2003). The original

Figure 4.2: Proportion of observations with missing values per cohort

ISEI score is a continuous variable ranging from 10 to 90, but it has been standardized to have an overall mean of 0 and a standard deviation of 1. The highest achieved level of education of the respondents is measured in either the original *a priori* scale from the ISMF or in the empirical scale estimated in Chapter 3. A description of the different levels of education and the two scalings have been reproduced in Table 4.1. The first three columns show the name of each level, their English translation, and their classification in the ISCED (UNESCO, 1997) scheme. The fourth column presents the institutional duration, or the number of years it would take a 'standard student' to finish that level of education. The final two columns present the two scales. The most striking difference between the two scalings is the value of lower vocational education (LBO), whose value is valued considerably higher in the *a priori* scale. Moreover, Table 4.1 shows that in the empirical scale the values of two educational categories, MAVO and HBO, changed in 1968. A major educational reform, the Mammoetwet or 'Mammoth Law', was implemented in that year. The metric of the scales in Table 4.1 is pseudo-years of education, but in the analysis both scales will be standardized to have a mean of zero and a standard deviation of one.

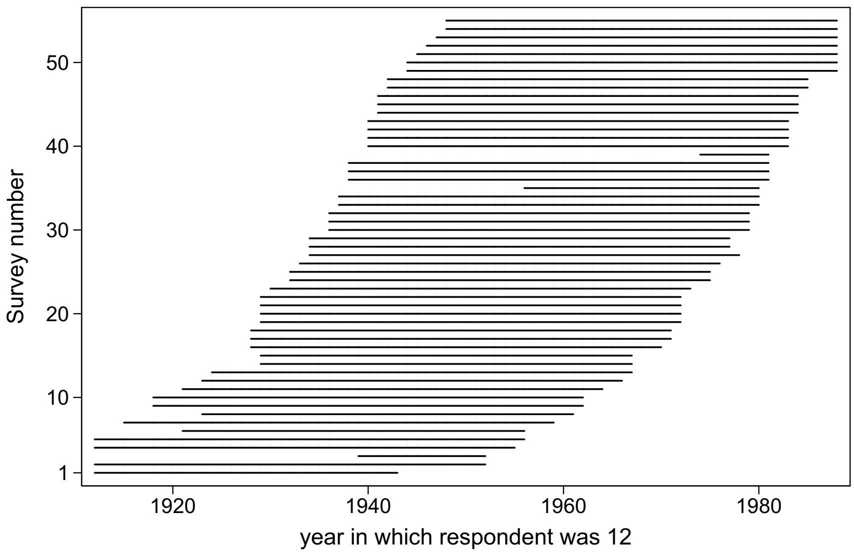
Table 4.1: Conversion of old educational levels into new educational levels and simplified educational levels

level	English translation	ISCED[a]	institutional duration	*a priori* scale	empirical scale
LO	primary	1	6	6.0	6.0
LBO	junior vocational	2C	10	9.0	7.0
MAVO	junior general secondary	2B[b]	9 / 10	10.0	10.5[c] / 9.5[d]
MBO	senior secondary vocational	3C	14	10.5	10.0
HAVO	senior general secondary	3B[b]	11	11.0	11.0
VWO	pre-university	3A	12	12.0	13.0
HBO	higher professional	5B	15	15.0	15.0[c] / 14.5[d]
WO	university	5A	17 / 16	17.0	17.0

[a] (UNESCO, 1997)
[b] These levels were originally intended to be terminal levels of education for most students (so 2C or 3C) but evolved into levels that primarily grant access to subsequent levels of education.
[c] before 1968
[d] after 1968

An important characteristic of this data is that it consists of different surveys. A major advantage of this approach is that this greatly increases the period that can be studied because these surveys were held at different times. This is illustrated in Figure 4.3, which indicates for each survey the cohorts to which it contributes observations. It shows that the oldest cohorts collect their observations from only four surveys, while other cohorts collect their observations from almost all surveys. So peculiarities of individual surveys are most likely to influence estimates in the earliest cohorts, since in these cohorts each survey is responsible for a sufficient proportion of the observations to have a noticeable influence. The characteristics of individual surveys are less likely to have an effect on the estimates in the middle cohorts, as no single survey is dominant in these cohorts. The appendix to this chapter shows that there are considerable variations among surveys in terms of response rate, proportion of missing cases, and the degree of detail in which the variables are measured.

Figure 4.3: Cohorts covered by different surveys (survey numbers correspond to the appendix and are ordered by the year in which the survey was held)



4.4 Method

For the estimation of the non-linear trend, a two-step process has been used. First, a new dataset is created containing, for each annual cohort, an estimate of IEOut for men and women, and their standard errors. The estimates are obtained by regressing the respondent's highest obtained level of education on father's occupational status, separately for men and women and each cohort. An annual cohort is combined with a neighboring cohort if it does not contain enough observations for a stable estimate. This resulted in the following combined cohorts: 1900/1901, 1902/1904, 1905/1907, 1910/1911. Second, a local polynomial curve (Cleveland, 1979; Loader, 1999; Fox, 2000) is fitted through these annual estimates of IEOut. This is done in such a way that estimates with small standard errors, that is, measured with great precision, receive more weight than estimates with large standard errors. These curves also provide estimates of the trend and the change in trend. These are the first and second derivatives of the smooth curve.

An attractive feature of the local polynomial smooth is that it uses information from nearby cohorts to create an improved estimate of IEOut at a cohort. This is illustrated in Figure 4.4. The point on the local polynomial curve for cohort 1938 is computed using the following four steps: First, the observations that will be used in the estimation are selected. This is typically done by selecting a fixed proportion, say 65%, of nearest observations. This is shown in Figure 4.4 in panel (a). Second, observations that are selected are weighted according to their distance from 1938. A common function used to create these weights is the tricube function[1]. The tricube function is shown in panel (b) in Figure 4.4. The tricube weights ensure that cohorts close to 1938 receive more weight when estimating the value of cohort 1938. Third, these weights are adjusted in such a way that they take into account that some cohorts are estimated with much more precision than others. The raw estimates of IEOut are regression parameters, so an estimate of the precision of the estimate is available in the form of the standard error. Weights based on the inverse of the square of the standard error would properly correct for the difference in precision between cohorts. The weights based on proximity to cohort 1938 and the weights based on the precision of the estimates of IEOut are combined by computing the product of these two. This is shown in panel (c) of Figure 4.4. Fourth, a regression of IEOut on cohort, cohort

[1] If the cohort of interest, the 'focal value' is represented by x_0 and the span (half the range that contains 65% of the observations) by h then the value x is assigned the weight

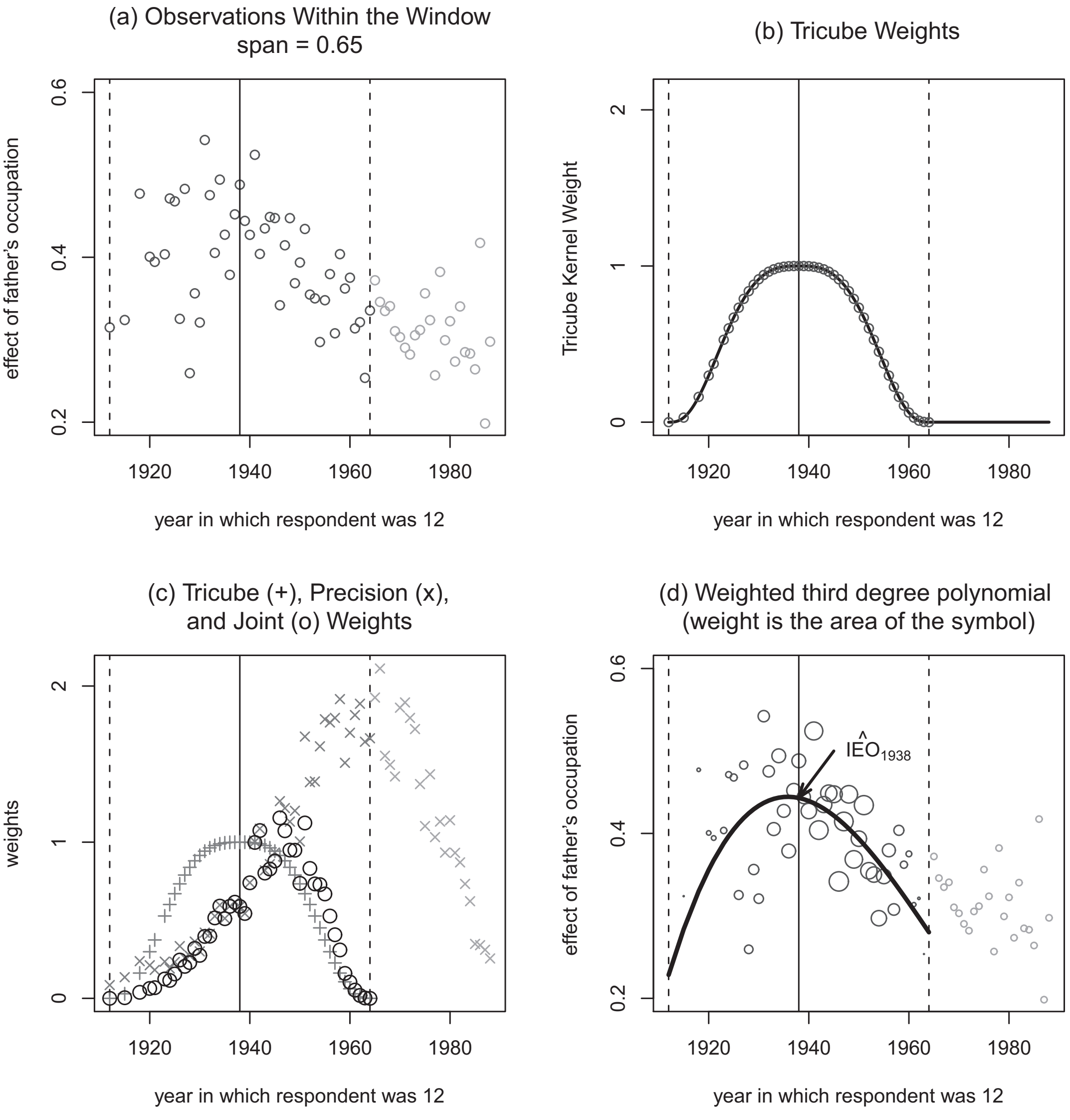
$$W = \begin{cases} \left[1 - \left(\frac{|x-x_0|}{h}\right)^3\right]^3 & \text{if } \frac{|x-x_0|}{h} < 1 \\ 0 & \text{if } \frac{|x-x_0|}{h} \geq 1 \end{cases}$$

squared, and cohort cubed, is estimated using these combined weights. The predicted IEOut from this regression for cohort 1938 is the local polynomial estimate. It uses most information from cohorts that are close by and are estimated with high precision, and less information from cohorts that are far away or are estimated with low precision. Furthermore, the slope of this regression line in 1938 is a local estimate of the trend, and the change in slope in cohort 1938 is a local estimate of the change in trend in 1938. These are obtained by evaluating the first and the second derivatives of the regression line at 1938. The entire local polynomial curve is obtained by repeating this process for all annual cohorts. This procedure is implemented in the `locfit` package (Loader, 2005) in the R statistical computing environment (R Development Core Team, 2005). This also provides confidence envelopes for the curve, the first and the second derivatives, using procedures discussed by Loader (1999).

What makes this method attractive is that it takes an intermediate position between two commonly used alternative methods of estimating a non-linear trend: a quadratic trend and a discrete trend. The first strategy is usually not flexible enough to adequately fit the data. The second strategy is too flexible, which means that too much statistical power is lost, making it hard to find any evidence for a non-linear trend.

The secondary aim of this chapter is to investigate sensitivity of the results to the three potential sources of error: The first potential source of error is the fact that the ISMF consists of multiple surveys that vary in quality. The survey effects are controlled for by adding dummies for surveys, and interacting these dummies with father's occupational status. The dummies are constructed in such a way that the main effect of father's occupational status represents the IEOut in an average survey, so differences between cohorts are no longer the result of differences across surveys. By adding survey dummies, and interactions between the survey dummies and father's occupational status, each survey has its own baseline IEOut, but the trend is constrained to be the same for all surveys. The reason for this choice was that there is good reason to expect that the quality of a survey can influence the effect of father's occupational status on the respondent's education, as the effect is likely to be smaller in more noisy data, but the effect of data quality on the estimated trend of the effect of father's occupational status is much less clear.

Figure 4.4: Obtaining local polynomial regression estimate for IEOut for cohort 1938, adapted from Figure 4.1 in (Fox, 2000, p. 24–25)



The second potential source of error is the fact that there are multiple ways in which the dependent variable — the respondent's education — can be scaled. The ISMF uses a common strategy by starting with the institutional years of education, the number of years a 'standard student' would need to finish that level of education, and applies an *ad hoc* correction to make sure that the ordering of levels corresponds with an *a priori* ordering. In Chapter 3 I proposed an alternative scale based on the idea that education predicts the occupational status of the respondent, and if education is better scaled then education should be better at predicting occupational status. This way one can estimate an optimal scale of education. These two scales were presented in Table 4.1. By comparing the estimated trend using the *a priori* scale with the estimated trend using the empirical scale, one can assess whether the differences between the scales actually matter.

The third potential source of error is the presence of missing data. The annual estimates of IEOut are controlled for missing data using Multiple Imputation (Little and Rubin, 2002). The idea behind Multiple Imputation is to create multiple 'complete' datasets by first estimating for each missing value a distribution of plausible values, and then drawing multiple values (in this case 5) from this distribution. This is done in Stata (StataCorp, 2007) using the `ice` (Royston, 2004, 2005a,b, 2007, 2009) module. The model of interest is estimated on each 'complete' dataset. The point estimate is the average of the point estimates from the different 'complete' datasets, and the variance of the sampling distribution (the standard error squared) is computed according to equations (4.1) to (4.3) (Little and Rubin, 2002).

$$se^2 = \overline{se^2} + (1 + 1/m)B \tag{4.1}$$

$$\overline{se^2} = \frac{\sum_{j=1}^{m} se_j^2}{m} \tag{4.2}$$

$$B = \frac{\sum_{j=1}^{m} (\beta_j - \overline{\beta})^2}{m-1} \tag{4.3}$$

Equation (4.1) shows that the variance of the sampling distribution (se^2) in the case of m 'complete' datasets consists of two elements: $\overline{se^2}$, and B. The first element is described in equation (4.2), and is the average of the variances of the sampling distributions in the different 'complete' datasets. This represents an estimate of the degree of uncertainty about a parameter which ignores the fact that some of the data is itself uncertain as it consists of imputations rather than real observations. The second element, in equation (4.1), and equation (4.3), corrects for this by using the differences in the parameters (β_j) between the different complete datasets as a measure of the uncertainty due to the imputations.

The key issue with multiple imputation is the model used for estimating the imputed values. This model must be at least as flexible as the model of substantive interest (Little and Rubin, 2002). For this reason separate imputation models are estimated for each combination of cohort and survey. Within each of these combinations, imputations for the missing values are created from a model using father's and respondent's occupational status and education, with interactions between whether the respondent is male or female and all these variables. The occupational status of the respondent and the level of education of the father are also used in the imputation model even though they will not be used in the final model of interest, in order to improve the imputations. However, the father's highest achieved level of education is only added when available, which was not the case in 10 surveys. Imputations were only carried out if the cohort-survey combination had at least 20 fully observed cases. As a result, not all missing values were imputed. There were 9,758 missing values for father's occupational status, of which 1,934 could not be imputed, and there were 993 missing values for the respondent's highest achieved level of education, of which 181 could not be imputed. Respondents with missing values that could not be imputed will still be ignored in the analysis.

4.5 Results

The results using estimates of IEOut while controlling for all the potential sources of error and using the empirical scale are shown in detail in Figure 4.5 for men and Figure 4.6 for women. Panels (a) show local polynomial curves fitted through the annual estimates of IEOut with their 95% confidence envelope. The confidence envelopes always remain above zero, indicating that the offspring of fathers with a higher status occupation did, on average, attain more education than the offspring of fathers with a lower status occupation. Panels (a) also show that the level of inequality appears to have changed over time for both men and women. This is tested in the panels (b), which show the trend in IEOut, that is, the first derivatives of the local polynomial curves in panels (a). A period of significant negative trend is found for both men (1941–1960) and women (1952–1977). The hypothesis that the trend is zero in the last period (after 1960 for men and 1977 for women) cannot be rejected, suggesting that the trend has indeed slowed down. Notice however that the confidence envelopes are very wide for both the youngest and the oldest cohorts, so the finding of zero trend in the most recent cohorts could also be due to lack of statistical power. The way to find out if the trend truly decelerated is to also estimate the changes in trends, the second derivatives, which are shown in panels (c). If the trend truly decelerated, then the second derivative should be significantly positive, indicating that the negative trend became less negative. Panels (c) show a significantly accelerating trend (negative second derivative) between 1935 and 1944 for men and 1949 and 1952 for women, but no significant deceleration. For men, the point estimates of the change in trend are positive before the trend became insignificant, providing some indication that the trend decelerated. For women, the point estimate of the change in trend is also briefly positive prior to the trend becoming non-significant, but this period is much shorter, and quickly becomes negative again, so the case for a decelerating trend for women is much less convincing. These results are summarized in panels (d). The curve represents the smooth estimates of IEOut from panel (a), while the shaded areas below that curve represent the periods of significant trend, and the shaded areas above the curve represents periods of significant change in trend.

Figures 4.7 and 4.8 show how controls for the three different potential sources of error influenced these results. Panels (a) and (b) use the *a priori* scale and the empirical scale of education respectively. Panels (c) show the trend using the empirical scale while controlling for survey effects. Panels (d) show the trend using the empirical scale while controlling for missing data. Comparing panels (a) and (b) shows that the scale of education does influence the estimated trend. A decelerating trend was found for men using the *a priori* scale, but this change in trend became insignificant when the empirical scale was used. For women, using the empirical scale leads to a significant positive estimate of the trend prior to the negative trend, and a significant transition between the positive and the negative trend. Neither of these characteristics was present when the *a priori* scale was used. The aspect of the trend that remains largely unaffected by the scale of education is the downward trend during the third quarter of the twentieth century. The panels (c) show the trend in the 'average survey', thus controlling for survey effects. This correction mainly affects the oldest cohorts, since these cohorts contain data from only a few surveys, as was shown in Figure 4.3. As a consequence, a problem in an individual survey could have an influence on the uncorrected results. The younger cohorts contain data from many surveys, so any problems with individual surveys is likely to be averaged out. One important way in which surveys differ from one another is the number of missing values, as is shown in the appendix to this chapter. If this is the main source of differences in the results between models that control and do not control for survey effects then the trends in panels (d), which control for missing data but not for survey effects, should closely correspond to the trends in panels (c). However, the trends in panels (d) closely correspond to the trends in panels (b), indicating that controlling for missing data hardly influences the results.

Figure 4.5: Trend in Inequality of Educational Outcomes and change in trend for men. (IEOut is measured as standardized regression coefficients. The local polynomial smooth has a span of .65 and uses weights proportional to the inverse of the variances.)

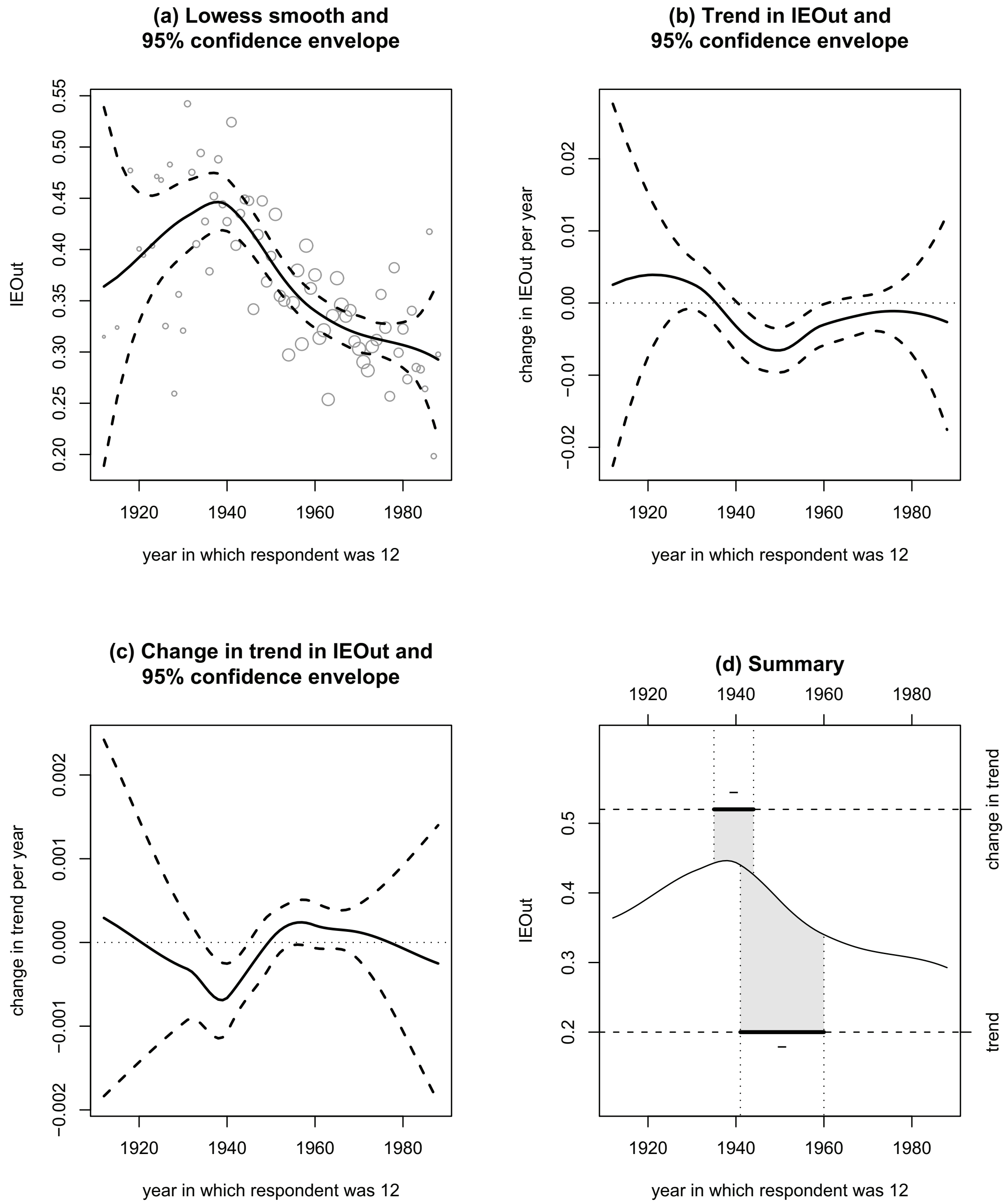


Figure 4.6: Trend in Inequality of Educational Outcomes and change in trend for women. (IEOut is measured as standardized regression coefficients. The local polynomial smooth has a span of .65 and uses weights proportional to the inverse of the variances.)

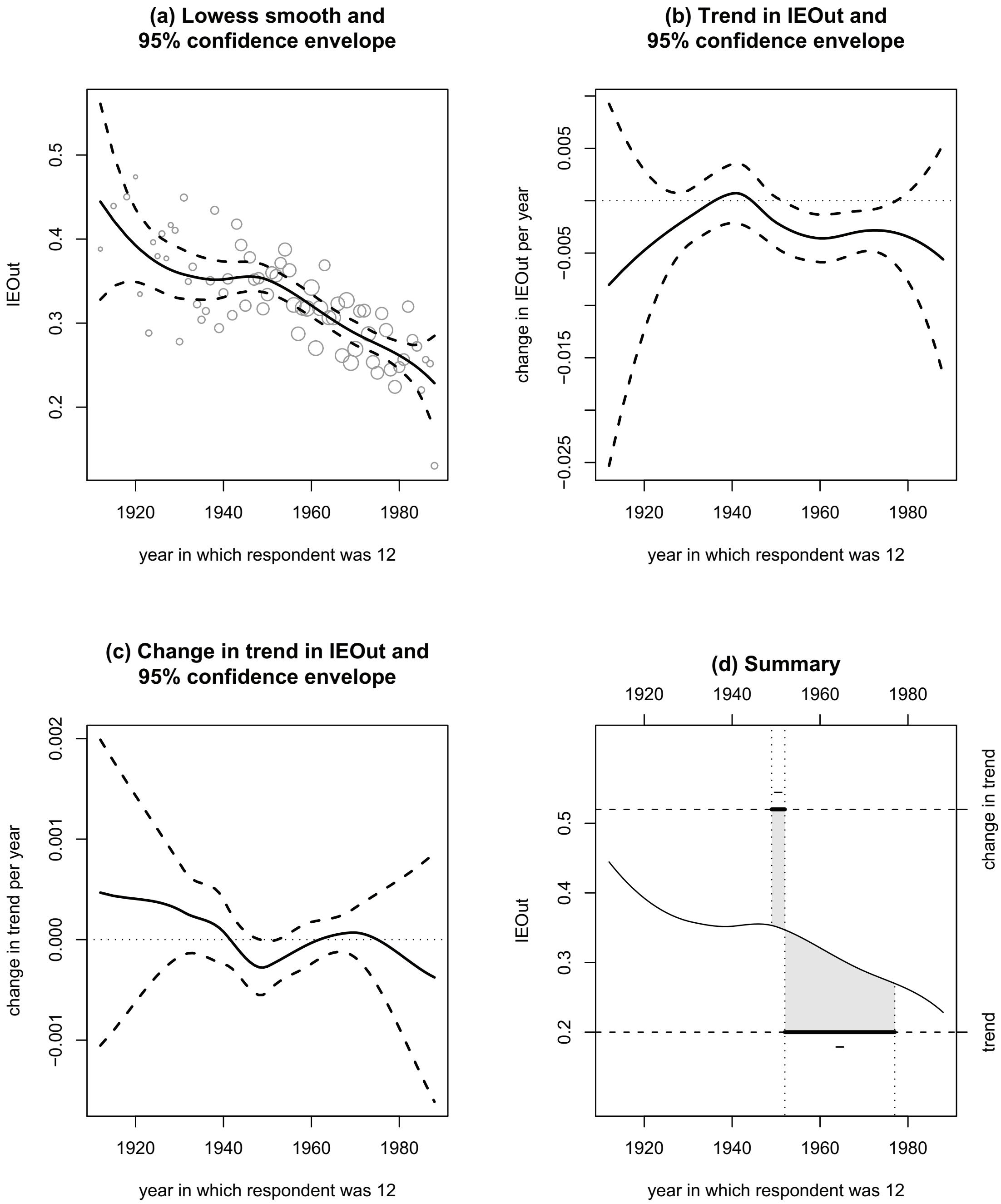


Figure 4.7: Trend in Inequality of Educational Outcomes and change in trend for men while using different scales of education and controlling for survey effects and missing data. (IEOut is measured as standardized regression coefficients. The local polynomial smooth has a span of .65 and uses weights proportional to the inverse of the variances)

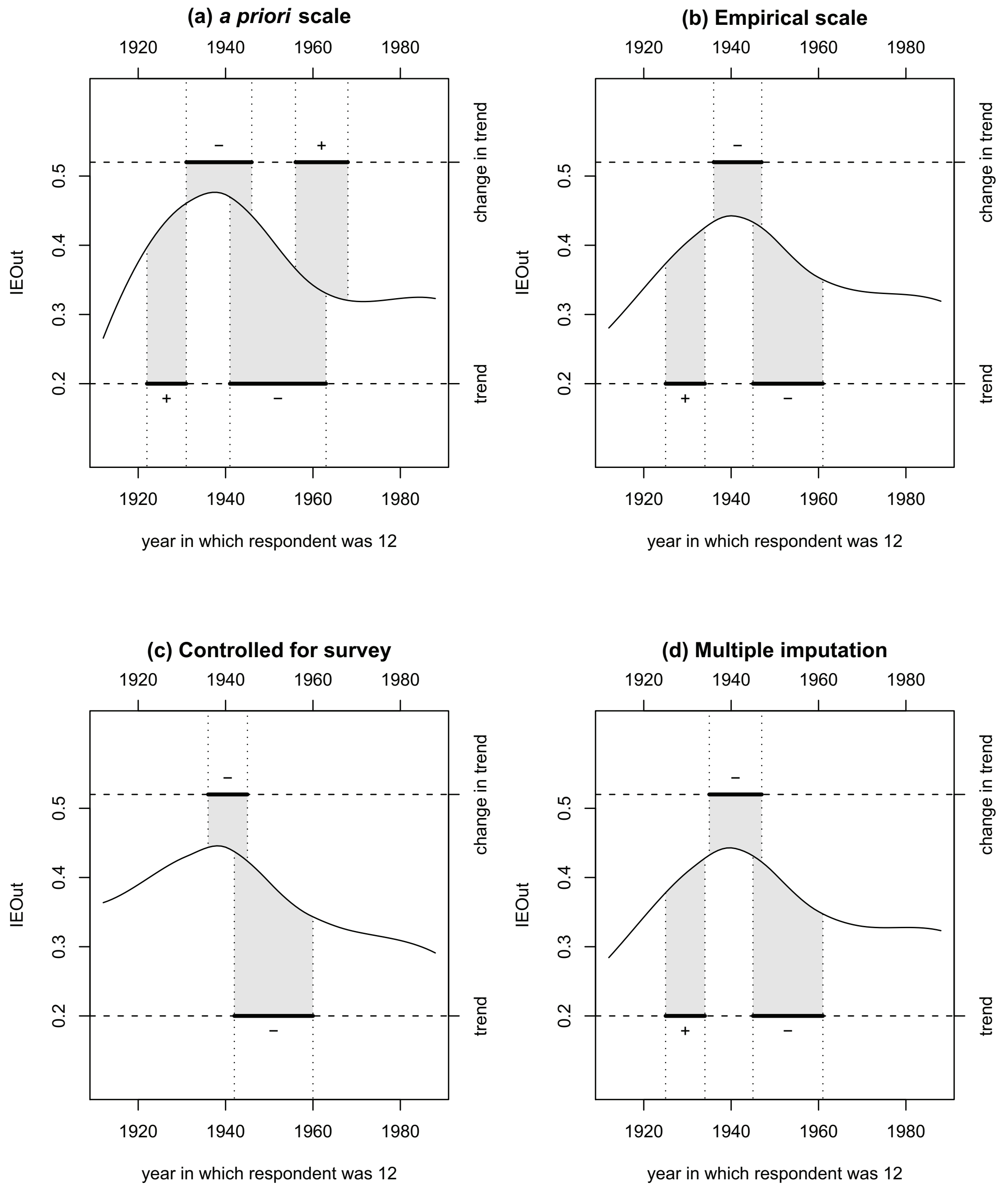
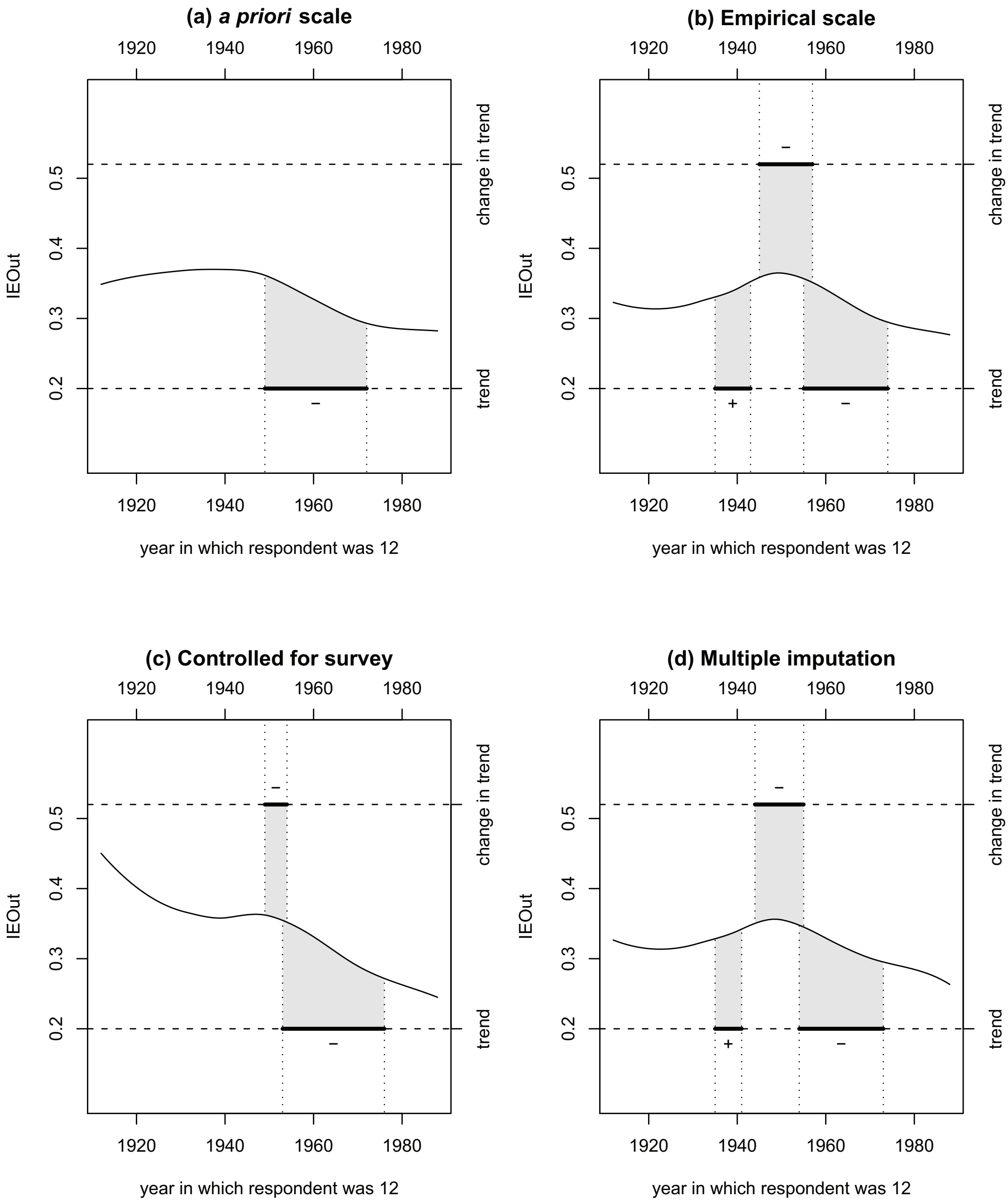


Figure 4.8: Trend in Inequality of Educational Outcomes and change in trend for women while using different scales of education and controlling for survey effects and missing data. (IEOut is measured as standardized regression coefficients. The local polynomial smooth has a span of .65 and uses weights proportional to the inverse of the variances)



4.6 Conclusion

This chapter had a primary and a secondary aim: The primary aim was to provide a detailed description of the trend in IEOut in the Netherlands between 1912 and 1988, and in particular whether the negative trend in IEOut has decelerated. Previous studies all found a positive IEOut and an overall negative trend in IEOut, but failed to find any evidence that this trend was non-linear. This chapter did find evidence that the trend has been non-linear, but has not found the deceleration in the decreasing trend in IEOut that was expected. The results are summarized in Table 4.2, which shows periods of significant trends and changes in trends while controlling for the different potential sources of error. The most robust findings in this chapter are a period of negative trend for both men and women, which was preceded by a period of significantly accelerating trend. The presence of the period of accelerating trend indicates that previously the trend was less negative, and for men there is a solid indication that the trend was even positive. There is some evidence that the negative trend decelerated prior to becoming insignificant for men, but this deceleration is not (yet) significant. There is no indication that the negative trend for women decelerated prior to becoming insignificant, indicating that the lack of significance of the negative trend in the youngest cohorts has more to do with lack of power than with a lack of negative trend.

Table 4.2: Periods of significant trend in IEOut and change in trend in IEOut

scale	corrected for	trend		change in trend	
		positive	negative	positive	negative
men					
a priori		1922–1931	1941–1963	1956–1968	1931–1946
empirical		1925–1934	1945–1961		1936–1947
empirical	survey		1942–1960		1936–1945
empirical	missing data	1925–1934	1945–1961		1935–1947
empirical	survey and missing data		1941–1960		1935–1944
women					
a priori			1949–1972		
empirical		1935–1943	1955–1974		1945–1957
empirical	survey		1953–1976		1949–1954
empirical	missing data	1935–1941	1954–1973		1944–1955
empirical	survey and missing data		1952–1977		1949–1952

The secondary aim of this chapter was to use this analysis to investigate the degree of susceptibility of the International Stratification and Mobility File [ISMF] (Ganzeboom and Treiman, 2009) to three potential sources of error: the scaling of education, survey effects, and missing data. Controls for missing data did not change the results, but both controls for survey effects and using different scales of education did moderately influence the estimated trend. Controls for survey effects primarily influenced the older cohorts, for both men and women, while different scalings of education primarily influenced the estimated trend in older cohorts for women.

Appendix: Surveys and indicators of their quality

Table 4.3: Indicators of data quality of Dutch surveys that were post-harmonized in the International Stratification and Mobility File (Ganzeboom and Treiman, 2009)

survey number[a]	year	birth cohorts	N	response rate	% missing	# categories respondent's education	# categories father's occupation
1	1958	1912–1943	783	94	0.4	9	23
2	1967	1912–1952	1125	68	2.0	4	20
3	1967	1939–1952	333	63.1	0.3	4	21
4	1970	1912–1955	1391	74	3.1	8	62
5	1971	1912–1956	1313	76	4.4	8	57
6	1971	1921–1956	1098		2.8	5	5
7	1974	1915–1959	739	67	6.0	12	49
8	1976	1923–1961	654	69	2.2	7	59
9	1977	1918–1962	2669	70	6.9	6	64
10	1977	1918–1962	1123	64	13.7	9	59
11	1979	1921–1964	1159	65	3.1	40	55
12	1981	1923–1966	1448	83	12.6	10	63
13	1982	1924–1967	1014	74	9.5	10	64
14	1982	1929–1967	1670		13.5	8	39
15	1982	1929–1967	590	60	7.4	16	61
16[b]	1985	1928–1970	3163	41	5.6	9	67
17	1986	1928–1971	1056	83	10.7	10	62
18	1986	1928–1971	2545	57	11.0	5	39
19	1987	1929–1972	1188	82	1.6	7	30
20	1987	1929–1972	639	60	1.2	8	57
21	1987	1929–1972	686	78	10.1	7	58
22[b]	1988	1930–1973	3482		7.2	9	66
23	1990	1932–1975	1765	48	5.9	7	65
24[b]	1990	1932–1975	3303		7.5	9	67
25	1991	1933–1976	787		47.5	6	56
26	1992	1934–1978	1579	43	4.0	10	67
27[b]	1992	1934–1977	3554		7.0	9	66
28	1992	1934–1977	1624		39.1	20	64
29	1994	1936–1979	1202	47.5	11.5	5	44
30	1994	1936–1979	845	58	3.2	9	64
31[b]	1994	1936–1979	3403		13.5	8	66
32	1995	1937–1980	1639	40	11.6	9	64
33	1995	1937–1980	1615	51.5	6.1	8	65
34	1995	1956–1980	948		7.2	9	62
35	1996	1938–1981	603	36	2.6	8	59
36	1996	1938–1981	1013	37	30.5	10	14
37[b]	1996	1938–1981	3680		7.4	8	68
38	1996	1974–1981	189	42	34.4	10	14

(Continued on next page)

Table 4.3 – continued from previous page

survey number[a]	year	birth cohorts	N	response rate	% missing	# categories respondent's education	father's occupation
39	1998	1940–1983	644	31	12.7	8	55
40	1998	1940–1983	1288	50	19.6	9	59
41	1998	1940–1983	1783	48	2.4	10	69
42[b]	1998	1940–1983	4039		6.1	8	67
43	1999	1941–1984	1889	43	6.5	8	64
44	1999	1941–1984	6674	66.4	23.1	8	68
45	1999	1941–1984	1062		17.4	7	61
46	2000	1942–1985	1381	40.6	2.1	10	62
47	2000	1942–1985	863	42.7	3.3	8	62
48	2002	1944–1988	1607	67.9	12.9	13	63
49	2003	1945–1988	1750		10.7	10	63
50	2003	1944–1988	6652	37.8	2.0	10	68
51	2004	1946–1988	1317	64	8.4	13	62
52	2005	1947–1988	1313	40	8.1	8	62
53	2006	1948–1988	1224	60	9.4	13	64
54	2006	1948–1988	1450	38	5.0	8	66

[a] Number refer to the data references.

[b] These are waves in a panel. Respondents in these waves are weighted to ensure that each respondent contributes only one observation.