

Draft prepared for presentation at  
OECD/CERF/CILN International  
Conference on Self-Employment

Preliminary and Incomplete  
Please Do Not Quote

Burlington, Ontario, Canada  
September 24-26, 1998

**EXPLAINING THE RECENT RISE IN SELF EMPLOYMENT:  
LIFECYCLE, COHORT AND AGGREGATE ECONOMY EFFECTS**

Danny Leung and Chris Robinson

University of Western Ontario

September 1998  
(Draft)

## I. INTRODUCTION

Trends in self employment have attracted considerable recent attention. For example, Blau (1987) noted a rise in aggregate self employment rates in the United States in the 1970's following several decades of decline. An increase in female self-employment rates has been examined in Devine (1994) and ethnic and racial patterns have been studied in Fairlie and Meyer (1996). Self-employment patterns have also been a major focus of recent interest in Europe and elsewhere. OECD (1992) notes a reversal of earlier declines in many countries and striking recent growth in some. A variety of explanations have been sought for the various observed trends, many of them based on changes in the relative attractiveness of self-employment due to a variety of policy changes and trends in the overall economy. In addition to interest in recent trends, the growing importance of self-employment in many countries led to more detailed examination and modeling of the basic determinants of self-employment. (See, for example, Evans and Leighton, 1989). Lifecycle patterns, however, have received relatively little attention. It has generally been noted that self-employment rates increase with age and that there is an unusually large increase for workers nearing retirement.<sup>1</sup> This paper presents a comparative analysis of aggregate and disaggregated trends in self-employment for Canada and the United States and discusses the identification of lifecycle, cohort and aggregate economy effects on these trends.

Self-employment measures are typically derived from class of worker questions. In section II the observed patterns of self employment over the 1967 to 1996 are presented for various groups of workers using various measures based on class of worker data. The aggregate patterns for nonagricultural workers show a general decline for both countries until the mid 1970's followed by an increase. For the United States the increase is relatively short lived, but for Canada the increase is more sustained with large changes towards the end of the period. The aggregate patterns, however, mask major differences by sex and occupation. For example, the rates for males in both countries show general declines, whereas the females show substantial increases. Given the relatively low female rates at the beginning of the period, this different growth pattern leads to strong convergence in the rates by sex. Similarly, a strong convergence pattern is observed in the United States data

between white and blue collar rates for males. At the start of the period white collar rates are much higher than blue collar, but by the end of the period the ranking is reversed.

Lifecycle profiles estimated from the cross sections are presented in Section III. Section IV estimates the lifecycle profiles based on a pseudo panel. In the absence of cohort and year effects, the lifecycle profiles from these two sources would be identical. To the extent they differ, there is evidence of cohort and/or year effects. The identification of these effects is discussed in section V, and preliminary estimates of some simple models are presented in section VI.

## **II. PATTERNS OF SELF-EMPLOYMENT, 1967-96**

### **1. *Measuring Self-Employment***

There are a variety of conceptual issues that are unresolved in the appropriate definition and measurement of self-employment. A legal definition, for example, that excludes individuals who “work for themselves” but who incorporate their businesses, thereby becoming “legally” employees, may not be an appropriate definition for an economic model of self-employment. That is, nothing “real” may have changed in the economy from an individual incorporating his/her business: the same relationship with customers, the same hours of work and the same prices for inputs and outputs may all be prevailing. Yet a legal definition of self-employment will record a change. Even restricting attention to unincorporated individuals and employees, it is not clear how the difference is to be conceptually determined. Presumably, a self employed individual could draw up a contract for a buyer of their services that mimics the contract between employers and employees, including provision for notice of termination of the contract etc. The same relation would then hold between the purchaser and seller of labour services whether the person was self-employed or employee and a meaningful economic distinction would not be obvious.

A distinction might be made on the basis of more freedom of contract in the relation between the self-employed person and the buyer (employer) of their services than exists between an employee

and an employer because of the legal restrictions on the latter. One source of variation in the amount of self-employment might then be variation in the legal restrictions and the recourse to self-employment for their avoidance. More generally, the importance of any changes in observed measures of self-employment will depend on the underlying concept being reflected in the measure. An increase in measured self-employment that is largely “hidden unemployment” of recently laid-off middle managers declaring themselves as “consultants” would be of more importance in the real economy than a tax induced cosmetic change.

The main data source for the analysis of trends for the United States is the Current Population Survey (CPS). Two basic approaches are used to measure self-employment. The method of the Bureau of Labor Statistics is based on the class of worker classification<sup>2</sup>. In the CPS surveys in the years 1967 to 1995 persons who are employed are asked to classify their employment by a question similar to the 1994 version as follows: “Were you employed by government, by a private company, a nonprofit organization, or were you self-employed...” Those who respond that they are self-employed (after 1966) are then asked “Is this business incorporated?” Those who say yes are classified as *wage and salary workers* and are treated as such for the rest of the survey.<sup>3</sup> These individuals are not separately identified in the CPS class of worker variables until 1976. An alternative approach is to make use of the “income from nonfarm self-employment” variable. This is available in the March CPS, but is also restricted to non-incorporated self-employed.

Canadian data are available from the monthly Labour Force Surveys. Unfortunately, however, the original data are not publicly available for most years and a long time series can only be obtained from relatively restrictive categories from published sources. The data used for this study come from several sources. The long time series come from two main sources. For 1967-1975 the November issues of the monthly publication *The Labour Force* are used and for 1976-1996 the data are from the Labour Force Historical Review CDROM. These are supplemented by data from four supplements to the usual labour force survey: the 1971 Union Affiliation Survey, the 1981 Work History Survey, the 1991 Survey of Consumer Finances and the 1996 Survey of Consumer Finances. The self-employment measures, as in the CPS, come from the class of worker variable. In the

Canadian questionnaire respondents are asked to describe their main job or business. The class of worker coding follows with no explicit question on the questionnaire itself, though the accompanying code sheet in recent years includes the question “in ...’s job, was he/she a paid worker, self employed or an unpaid family worker?” The structure is thus similar, but not identical to the CPS structure.<sup>4</sup>

## **2. *Trends in Aggregate Self-Employment Measures, 1967-96***

For the United States, self-employment measures based on class of worker information were examined in the March CPS files for 1967-95.<sup>5</sup> Initially two class of worker based measures were examined:

- (a) SEC1 is the fraction of employed workers in the reference week recorded as self employed on the class of worker variable for the main job that week. From 1967-1975 it only includes unincorporated self-employed and the incorporated self employed are included in wage and salary workers and not separately identified. From 1976-1995 it again only includes unincorporated self-employed, though incorporated individuals can now be identified. This measure corresponds to the Bureau of Labor’s measure.
- (b) SEC2 is the fraction of employed workers in the reference week recorded as incorporated or unincorporated self employed on the class of worker variable for the longest job last year.<sup>6</sup> It is measured only from 1976 onward. The universe is the same as SEC1 except that the class of worker on the longest job must have a valid code.<sup>7</sup>

While SEC1 measures the number of unincorporated self-employed, SEC2 does not necessarily measure the total of incorporated and unincorporated self-employed since, as argued by the CPS survey staff, an unknown number of incorporated individuals will have “correctly” (legally) identified themselves as wage and salary workers in the first question noted earlier and will never

be asked the question about incorporation. If the fraction doing this is constant, then the percentage changes in the two components of the total self-employed can both be calculated, but unless the fraction is known, the percentage change in total cannot be measured. If the fraction is small, then SEC2 could be taken as a measure of the total.

Two comparable class of worker based measures were also constructed using the Canadian data.

- (a) SE1 has the same definition as SEC1 for the United states.
- (b) SE2 is the fraction of employed workers in the reference week recorded as incorporated or unincorporated self employed on the class of worker variable for the main job in the reference week.

Although many official tabulations use the SE1 measure, on the same grounds as argued by the U.S. officials based on legal definitions of wage earners, there is a preference in the Canadian Departments for using the SE2 definition when discussing trends in self-employment as a labour market issue.<sup>8</sup> This is also a common definition in other countries, as noted in OECD (1992).

The structure of the Canadian Labour Force Survey Questionnaire is such that SE1 should measure the same concept as SEC1 does in the CPS. However, SE2 may include a larger number of incorporated individuals than its U.S. counterpart because of the absence of a clear two question sequence as in the U.S. where some incorporated individuals may be lost on the first question.

In both countries the self-employment ratio is much higher in agriculture (about 50%) than non-agricultural industries (about 10%), but these workers make up a relatively small percentage of the total self-employed. The analysis therefore focuses mainly on nonfarm self-employment. Figure 1 shows the trends in both measures of self-employment for both countries.

For the unincorporated rate there is a clear declining trend from the 1960's to the mid 1970's with the Canadian figure declining from a higher initial position. A downward pressure is exerted by the

decline in the share of agriculture in both countries. By the mid 1970's the rates are identical for both countries. However, thereafter there is a divergence. Both countries grow to the early 1980's, but thereafter the U.S. figure shows a slightly declining trends while the Canadian trend is definitely upward with an acceleration in the 1990's. From 1976 to 1995 the percentage increase in the rate is only 1.45% for the U.S. compared to 22.5% for Canada. The "total" self-employment measure, available only from 1976 onward, shows an increasing trend for both countries but at a faster rate for Canada, especially in the 1990's. The increase in rates for this measure over the 1976-1995 period are 14.5% for the U.S. and 40.68% for Canada.

To abstract from the effect of the decline in agricultural employment, Figure 2 shows the trends for both measures for the nonfarm sector. The unincorporated self-employment rates fall for both countries initially to a low in 1976; they both rise thereafter in a similar pattern until the 1990's when the Canadian rate accelerates rapidly. The total self-employment rates show a more consistent divergence. Both countries have growing rates from a point of close equality in 1976, but the Canadian rate of growth is much faster with an increase of 59.52% over the 1976-1995 period compared with 23.34% for the U.S. The U.S. growth is, in fact, all concentrated in the 1976-1983 period with a stagnant self-employment rate thereafter.

### **3. *Self-Employment Trends by Sex, and Occupation***

Overall, the picture for nonfarm self-employment shows a rise for both countries, but a relatively modest one for the U.S. This aggregate pattern, however, masks substantial differences at the disaggregated level. First, there is a substantial difference in the patterns by sex, as shown in Figures 3 and 4. In Figure 3 the unincorporated rates are plotted. These show several similarities across the countries. In both countries the male rates are higher than the female but there is substantial convergence. For the United States the female rate is only 41.52% of the male rate in 1967 but rises to 69.87% in 1995. The Canadian figures are 30.50% and 72.69% in 1967 and 1995 respectively. In addition both countries show broadly declining trends for the male self-employment

rates contrasting with increasing trends for the female rates. In Figure 4 the total self-employment rates for both sexes are generally increasing since 1976, although there is almost no change for U.S. males. The convergence in the sex ratios occurs here, as for the unincorporated rates. The female rate to male rate ratio for the U.S. increases from 33.21% in 1976 to 59.29% in 1995. The Canadian ratio goes from 41.33% to 58.17% in the same period. Thus increases in self-employment rates have come disproportionately from females.

Disaggregating by broad occupational groups - blue and white collar - also reveals marked differences in the trends. A full analysis is only available for the U.S. as this type of data is not available in Canada. Figure 5 shows the self-employment rates for blue and white collar males for both unincorporated and total self-employment.<sup>9</sup> There is a striking difference in the trend for the two occupational groups. While blue collar rates rise throughout the period, except for the last year, and the total and unincorporated rates are quite similar, the white collar rates are declining (unincorporated) or stagnant (total) and the gap between the total and the unincorporated is large. The unincorporated rate for white collar workers starts out very much higher than the blue collar rate (13.58% vs. 4.04%) but the different trends result in a strong convergence (8.76% vs. 8.19%). For the females (Figure 6) blue collar rates rise over the period with little difference between the total and unincorporated, as for the males. However, since 1976 the white collar rates have also been increasing, unlike the males. There is also an increasing gap between the unincorporated and total rates. Like the males, the female blue collar rates start out much lower than the white collar rates (1.95 vs. 5.39) and the rates converge over time (6.26% vs. 4.78%). Unfortunately, the Canadian data do not permit the same type of analysis by occupation.

### III. CROSS-SECTION LIFECYCLE PROFILES

The increases in self-employment from increased rates for male blue collar workers and females in general may be due to a variety of effects including cohort, aggregate economy and lifecycle effects. In this section cross section and cohort age profiles are presented and identification issues discussed. The U.S. data permit the construction of profiles with fine age gradations. However, the Canadian data are grouped. The detailed U.S. data are examined first; comparisons with Canada are then made following appropriate grouping.

Figures 7a-b and 8a-b plot the cross section age profiles for males and females for the years 1967, 1976, 1986, and 1995 for both measures of self-employment over the age range 16-63. The unincorporated rates, shown in Figures 7a-b, show a general increase in the self-employment rate with age. The same is true for the incorporated rates plotted in Figures 8a-b. Over time there is an upward shift in the female profiles; this is especially clear in the total rates (Figure 8b). By contrast, the male profiles show 1986 generally higher than the others. In the total self-employment rates for males there is a general crossing of profiles rather than the clear upward shift in the females.

Figure 9 compares the male and female profiles directly for 1976 and 1995 for total self-employment. The male profiles are steeper but do not increase in slope (except for the older ages) over the twenty year time period. The female profile does become steeper and by 1995 the slope is quite similar to the male profile over the early to mid age range.

Canadian data do not permit the same detailed analysis. However, age specific rates can be calculated for grouped data - generally ten year age groups. In Table 1, these age profiles for unincorporated self-employment rates are presented for the years 1971, 1981, 1991 and 1996 in the first four columns. Comparable U.S. data are presented in the next four columns for the same years except that 1995 is substituted for 1996. Tables 2 and 3 present the data separately by sex. Apart from the teenage years, the Canadian rates show a monotonic increase with age with especially large increases near retirement ages. The U.S. rates also show these patterns. The very high rates for

workers over 64 has been the subject of some discussion in the previous literature. Fuchs (1982)

The U.S. - Canadian comparisons can be most clearly seen by plotting the data in Tables 1-3. In Figure 10a-d the male unincorporated rates are shown for the four years. Figure 11a-d shows the female rates. For the males, the age patterns are very similar. In terms of levels, in 1981 the U.S. had higher rates for males at most ages but by 1995/6 there is a clear Canadian dominance that is largest at the older ages. Female rates showed very similar age patterns until 1995/96 when again the Canadian rates are higher, especially at the older ages.

#### **IV. COHORT AGE PROFILES**

The cross section age profiles are a mixture of lifecycle, cohort and “aggregate conditions” effects. If aggregate conditions were stable and there were no cohort effects, then these age profiles would accurately predict the actual experience of the true cohorts. To the extent that the cross section or synthetic cohort “predictions” deviate from the actual experience of the true cohorts, there is evidence of effects of changes in aggregate conditions or cohort effects.

The U.S. data permit detailed analysis of the lifecycle experience of many cohorts. Figure 12 plots the lifecycle profiles for the unincorporated self-employment rates of the cohorts born in 1926-27, 1935-36, 1945-46 and 1954-55. There is a clear tendency for the later cohorts to have lifecycle profiles above the earlier ones for the first half of the lifecycle. This contrasts with no clear pattern in cross section profiles over successive years. Figure 13 plots the same profiles for the total self-employment rates. Since this is only measured from 1976 the profiles do not overlap over the first half of the lifecycle for as many years as in Figure 12. Thus the relative positions of the cohorts is less clear. Where there is substantial overlap, latter cohorts appear to have higher profiles. The pattern is similar if the data are disaggregated by sex. The females, however, show a much clearer tendency for total self-employment rate lifecycle profiles of later cohorts to be higher (Figure 14). Comparing different cohorts at the same age, of course, involves a combination of cohort and year

effects, so these higher profiles do not necessarily reflect cohort effects.

It is instructive to compare the cohort and cross section age profiles directly. Relative to the 1976 cross section age profile, the cohorts always experience higher rates than are predicted by the cross section (Figures 15a-d). By the 1981 cross section this effect is mitigated (figures 16a-d). Overall it is clear that in many cases there is substantial divergence between the cross section and cohort lifecycle profiles. Thus there is clear evidence of cohort or year effects.

The cohort analysis for Canada is very limited relative to the United States because of the age grouping in the available data. The data are insufficient for useful graphical analysis and are presented instead in tabular form in Table 4. Reading across the table gives the cohort age profile; reading down yields the cross section profile. Thus for males, the cohort aged 15-24 in 1971 experienced increases in rates from a starting point of .0154 to .0619 by the time they were aged 25-34, .0968 by ages 35-44, and .1197 by ages 40-49. This compares with the synthetic cohort (cross section) which grows to .0714 by ages 25-34, and .1230 by 35-44. For the males, the cross section profiles generally over predict the actual experience of the cohorts in 1981, but reduce or reverse the over prediction by 1991. For example, the cohort aged 35-44 in 1971 experiences a rate of .0969 in 1981 compared with the synthetic “prediction” of .1258, but by 1991 the actually experienced rate of .1378 is slightly above the .1309 of the synthetic prediction. For the females there is no similar general tendency for synthetic over prediction in 1981, though it does occur for the 45-54 group.

## **V. IDENTIFYING COHORT, LIFE CYCLE AND AGGREGATE EFFECTS**

Many explanations for recent trends and patterns of self-employment have been advanced in the substantial self-employment literature of the last two decades. Some explanations are based on economy wide changes in technology or tax policy which in principle would have the same qualitative, and possibly, quantitative effect on all ages and cohorts. They may be thought of as aggregate effects. Some are based on the characteristics of individuals, such as their education, willingness to take risks, desire for independence, inherited wealth, etc. While these characteristics

will vary across individuals at a point in time, there may also be differences by cohorts in the average value of these characteristics, some of which may not be easily measurable. These will exert cohort effects. At the same time there are likely to be lifecycle or age effects on self-employment. The hypothesis of capital market constraints, for example, is likely to result in a lifecycle effect.<sup>10</sup> Distinguishing between lifecycle, cohort and aggregate effects and measuring their magnitude may therefore help to discriminate between various hypotheses regarding recent trends.

In general, for analyses in which age, cohort and year effects all play a role, there is a fundamental identification problem since, by definition:

$$\text{Year} = \text{Cohort Birth Year} + \text{Age}$$

Hence separate effects for all three obviously cannot be estimated. Moreover, if complicated interactions are permitted between these effects, e.g. if changes in “aggregate conditions” can have differential effects by age that are not necessarily even the same sign, then the usefulness of a decomposition would be limited.

The U.S. data set to be analysed consists of 28 years of cross section observations (1967 to 1995, excluding 1994) on 55 one year age groups (16-70). It can be split into various worker “categories” such as male/female, blue/white collar. For any category, the relevant data constitutes individuals falling into the  $N \times T$  age/year cells where  $N$  is the number of age groups and  $T$  is the number of years. The cell means represent the self-employment rate data at the most general level. In the previous sections various subsets of these cell means have been plotted. In a dummy variable regression framework this constitutes 1540 parameters for each worker category.. There is thus a problem of an unmanageably large number of parameters. Further, because of the identity relation between cohort, year and age, these  $N \times T$  age/year cells have exact counterparts as age/cohort cells or cohort/year cells, and hence create the basic identification problem discussed earlier. Finally, comparisons across worker categories are complicated if they are endogenous. That is, if workers with a “taste” for the independence of self-employment tend to choose, say, white collar occupations

to improve their chances of self-employment, comparison across blue/white-collar would not identify the effect of an exogenous (say via technological change) change in the proportion of blue-collar workers in the economy. The identification problem can be addressed if restrictions can be imposed on some of the effects. The problem of an unmanageable number of parameters will be addressed by imposing various functional form restrictions. Finally, endogeneity issues can be investigated using methods based on cohort grouped estimators.

For simplicity, consider the following model of self-employment with two exogenous worker categories, represented by a dummy variable indicating whether the worker is in a blue collar occupation:<sup>11</sup>

$$(1) \quad s_i(a,t,c) = \alpha(a,t,c)B_i(a,t,c) + \lambda_i(a,t,c)$$

$$= \alpha(a,t,c)B_i(a,t,c) + \mu(a,t,c) + \epsilon_i(a,t,c), \quad i=1,2,\dots,M$$

The subscript  $i$  indexes an individual identifier over the whole pooled sample. The dependent variable is a dummy variable equal to one if individual  $i$ , of age  $a$  when observed at time  $t$ , and belonging to cohort  $c$ , is a union member.  $B_i(a,t,c)$  is a dummy variable equal to one if this individual is in a blue collar occupation;  $\lambda_i(a,t,c)$  represents all other factors influencing the probability of self-employment. Since  $B_i$  is exogenous,

$$E \lambda_i(a,t,c)|B_i=1 = E \lambda_i(a,t,c)|B_i=0 = E \lambda_i(a,t,c) = \mu(a,t,c)$$

where  $\mu(a,t,c)$  is the mean over individuals in the population of age  $a$  at time  $t$  and belonging to cohort  $c$  of unobserved characteristics that affect self-employment. Individual  $i$ 's idiosyncratic deviation from this mean is  $\epsilon_i(a,t,c)$ , where  $E\epsilon_i(a,t,c)|B_i=1 = E\epsilon_i(a,t,c)|B_i=0 = E\epsilon_i(a,t,c) = 0$ . Since  $t = c + a$ ,  $c$  cannot vary given  $(a,t)$  equation (1) is equivalent to:

$$(2) \quad s_i(a,t) = \alpha(a,t)B_i(a,t) + \mu(a,t) + \epsilon_i(a,t) \quad i=1,2,\dots,M,$$

The expected age/year cell means for each worker category follow from taking conditional expectations of (2):

$$E s_i(a,t)|B_i=1 = \alpha(a,t) + \mu(a,t) \quad \text{and} \quad E s_i(a,t)|B_i=0 = \mu(a,t)$$

The actual age/year cell means for each category are:

$$(3) \quad \bar{s}(a,t)|B_i=1 = \alpha(a,t) + \mu(a,t) + \bar{\epsilon}(a,t)|B_i=1$$

$$\bar{s}(a,t)|B_i=0 = \mu(a,t) + \bar{\epsilon}(a,t)|B_i=0$$

where the bars indicate the sample means for the cells. From (3), cross section age profiles are plots of  $\alpha(a,t) + \mu(a,t) + \bar{\epsilon}(a,t)|B_i=1$  for blue collar workers and of  $\mu(a,t) + \bar{\epsilon}(a,t)|B_i=0$  for white collar workers, holding  $t$  constant and varying  $a$ . In large samples  $\bar{\epsilon}(a,t)|B_i=1$  and  $\bar{\epsilon}(a,t)|B_i=0$  should be approximately zero and the difference between the corresponding blue and white collar plots,  $\alpha(a,t)$ , represents the effect of an exogenous switch from white to blue collar occupation.

Consider the interpretation of the plotted profiles. The simplest interpretation follows from an additive specification, i.e. let:

$$\alpha(a,t,c) = \gamma^b + \delta^b(a) + \eta^b(t) + \theta^b(c) \quad \text{and} \quad \lambda_i(a,t,c) = \gamma_i + \delta_i(a) + \eta_i(t) + \theta_i(c), \quad \text{so that}$$

$$\mu(a,t,c) = \gamma + \delta(a) + \eta(t) + \theta(c)$$

where  $\delta(a)$  is the age effect,  $\eta(t)$  is the time effect and  $\theta(c)$  the cohort effect for white collar workers and  $\delta(a) + \delta^b(a)$ ,  $\eta(t) + \eta^b(t)$  and  $\theta(c) + \theta^b(c)$  are the corresponding effects for blue collar workers. Then in equation (2):

$$\mu(a,t) = \gamma + \delta(a) + \eta(t) + \theta(t-a), \quad \text{and}$$

$$\alpha(a,t) = \gamma^b + \delta^b(a) + \eta^b(t) + \theta^b(t-a)$$

The expected cell means are thus:

$$E s_i(a,t) | B_i=1 = \gamma + \delta(a) + \eta(t) + \theta(t-a) + \gamma^b + \delta^b(a) + \eta^b(t) + \theta^b(t-a)$$

and

$$E s_i(a,t) | B_i=0 = \gamma + \delta(a) + \eta(t) + \theta(t-a)$$

Ignoring the mean error which will be close to zero in large samples, the cross section age profile for white collar workers plots  $\mu(a,t) = \gamma + \delta(a) + \eta(t) + \theta(t-a)$ , holding  $t$  constant and varying  $a$ . This is a “true” age or lifecycle effect,  $\delta(a)$ , only if there are no cohort effects,  $\theta(c)$ , since varying  $a$ , for a given  $t$  automatically varies the cohort. Alternatively, consider the interpretation of holding  $a$  constant and varying  $t$ , i.e. comparing the cross section age profiles at any given age. This will only be a “true” aggregate conditions effect,  $\eta(t)$ , if there are no cohort effects,  $\theta(c)$ , since varying  $t$  holding  $a$  constant will automatically vary the cohort. To identify separate effects, either the aggregate conditions have to be measured directly and not be collinear with year, or some parametric restrictions have to be imposed on  $\delta(a)$ ,  $\eta(t)$  or  $\theta(c)$ . For example, there may be equality of aggregate conditions over various time periods, i.e. non-varying  $\eta(t)$  over some range of  $t$ , or constant age affects,  $\delta(a)$ , over some range of  $a$ , or similar restrictions on  $\theta(c)$ .

The cohort age profiles follow from plotting  $\mu(a,c) = \gamma + \delta(a) + \eta(c+a) + \theta(c)$ , holding  $c$  constant and varying  $a$ . Again, the result is a true lifecycle effect only if there are no aggregate conditions effects. The absence of both aggregate conditions effects and cohort effects would result in the cross section age profiles being identical to the cohort age profiles.

Thus far, occupation has been considered exogenous. Suppose instead it is correlated with self-employment status in the manner sketched above. In that case the model in equation (1) no longer has a zero correlation between  $B_i(a,t,c)$  and  $\lambda_i(a,t,c)$ . Instead, let

$$E[\lambda_i(a,t,c) | B_i=1] = \mu_b(a,t,c) \quad \text{and} \quad E[\lambda_i(a,t,c) | B_i=0] = \mu_w(a,t,c)$$

and rewrite (2) as:

$$(4) \quad s_i(a,t) = \alpha(a,t)B_i + \mu^w(a,t) + (\lambda_i(a,t) - \mu^w(a,t))$$

The expected cell means are then:

$$E s_i(a,t) | B_i=1 = \alpha(a,t) + \mu^b(a,t) \quad \text{and} \quad E s_i(a,t) | B_i=0 = \mu^w(a,t)$$

The difference in the blue and white collar profiles therefore now estimates  $\alpha(a,t) + \mu^b(a,t) - \mu^w(a,t)$  rather than  $\alpha(a,t)$ .

To produce a consistent estimate of  $\alpha(a,t,c)$ , the correlated error problem has to be addressed. One approach to this problem is to assume that it comes primarily at the cohort level - e.g. particular cohorts vary in their “taste” for independence. If the data are then grouped by cohort and this cohort is followed over time this “taste” will be held constant, and provided the proportion of blue collar workers changes over time,  $\alpha(a,t,c)$  can be estimated. Suppose that the  $\lambda_i(a,t,c)$  term in (1) can be written:  $\lambda_i(a,t,c) = \tau_i(c) + \phi_i(a,t,c)$ , where  $\tau_i(c)$  is an individual’s “taste” for independence that may be correlated with occupation and  $\phi_i(a,t,c)$  represents the remaining factors in  $\lambda_i(a,t,c)$  that are uncorrelated with occupation. Summing (1) over members of the same cohort for each (a,t) combination then yields:

$$(5) \quad \bar{s}(a,t,c) = \alpha(a,t,c)\bar{B}(a,t,c) + \bar{\tau}(c) + \bar{\Phi}(a,t,c),$$

or equivalently,

$$\bar{s}(a,t) = \alpha(a,t)\bar{B}(a,t) + \bar{\tau}(t-a) + \bar{\Phi}(a,t),$$

where  $\bar{s}(a,t,c) = (1/N(c,t))\sum s_i(a,t,c)$ , etc., are again the (a,t) cell means. The only source of endogeneity in (5) is the correlation between  $\bar{\tau}(c)$  and  $\bar{B}(a,t)$ . However, if  $\bar{\tau}(c)$  can be included in the regression in the form of cohort dummy variables, only the  $\bar{\phi}(a,t,c)$  term will remain in the error and  $\alpha(a,t,c)$  can be consistently estimated.

The grouping approach provides a possible solution for endogeneity problems. However, this comes at the cost of requiring extra restrictions. Estimating equation (2) in the most unrestricted form requires the estimation of  $2 \times (N \times T)$  parameters, where  $N$  is the number of age groups and  $T$  is the number of years. The individual level data contain  $N \times T \times I$  observations, where  $I$  is the average number of observations per age group. Given large enough  $I$ , these parameters could be estimated. In the grouped data there are only  $N \times T$  observations, hence the number of parameters has to be reduced. As a result, differences between a grouped estimate and an individual level estimate will reflect not only possible endogeneity, but also possible specification error at the grouped level.

## VI. PRELIMINARY EMPIRICAL ESTIMATES

In the first approach to identifying separate effects it is assumed that cohort characteristics change slowly relative to year effects. Specifically, it is assumed that grouping into 10 year age groups will result in possible cohort effects across these 10 year cohorts but not within them. In addition, the additive simplification of the previous section is imposed and differences by occupation are suppressed. Thus, the specification is:

$$E s_i(a,t) = \mu(a,t) = (\gamma + \theta_5) + \delta(a) + \eta(t) \\ + (\theta_1 - \theta_5)C_1 + (\theta_2 - \theta_5)C_2 + (\theta_3 - \theta_5)C_3 + (\theta_4 - \theta_5)C_4$$

where  $C_i$  is a dummy variable equal to one if the individual is in cohort group  $i$ . The five cohort groups are 1915-24, 1925-34, 1935-44, 1945-54 and 1955-64 with 1935-44 as the omitted

category.<sup>12</sup> At maximum, a cohort group has 280 (a,t) combinations; at minimum 165. Thus some restrictions on the number of parameters in the functions for a and t are desirable. Initial restrictions are a second order polynomial for  $\delta(a)$  and a full set of year dummies for  $\eta(t)$ . The omitted year is 1976. Linear probability estimates for this model, in total and separately by sex, are presented in Table 5. The age range was restricted to 16-59. The dependent variable is multiplied by 100 for convenience. The coefficients on the cohort dummy variables are not significantly different from zero for either males or females. Given the range of the point estimates, cohort differences amounted to a maximum of 0.67 of a percentage point on the self employment rate for males and 0.77 of a percentage point for females. This is roughly the magnitude of increase that would occur for a 2-3 year interval around age 30 as workers age.

The year effects can be more substantial. Relative to 1976, the 1980's showed year effects significantly higher for both males and females - by as much as 2.07 percentage points for the males and 2.67 percentage point for the females.

## FOOTNOTES

1. Fuchs (1982) examines the high self-employment rates in this age group.
2. See, for example, Bregger (1996).
3. The rationale is that they are legally employees of their own businesses (Bregger, 1996).
4. A disturbing feature of the Canadian data is that there is a very poor fit between the Labour Force Survey measures and that obtained by the census. In 1991 the census measured self-employment at 9.7% vs. 13.2% in the 1991 census, i.e. the Labour Force Survey gives a figure 36% higher than the census. This “bad fit” is discussed in a technical report to the census (Statistics Canada - Cat. No 92-338E) wherein the difference is attributed to “ (1) enumerator training in the LFS, and (2) specific manual and computer edits in the LFS and census” (p.32)
5. The source of the data is the UNICOM March CPS CDROM; the 1967 date was chosen rather than the available 1964 as the basic measure of US self-employment - unincorporated self employed - began in that year (Bregger, 1996).
6. The March files do not identify incorporated self employed in the class of worker variable for the main job in the reference week until the 1988 survey.
7. Approximately 98% satisfy this criterion.
8. Three Statistics Canada publications from the Labour and Household Surveys Analysis Division (*Self-Employment in Canada*, Cat 71-582, 1985; *Enterprising Canadians: The Self-Employed in Canada*, Cat 71-536, 1988; and *Labour Force Update The Self Employed* Cat 71-005-XPB, Autumn 1997) all use the SE2 definition for their analysis with the same

rationale: “However, for studying labour market behaviour....it is useful to set aside the distinction between incorporated and unincorporated businesses and treat all of these individuals as self-employed. This also has advantages in the study of trends in the number of self-employed individuals since changes in tax laws can prompt movements towards incorporation which impart a downward influence on the estimated number of self employed persons if those with incorporated businesses are classified as paid workers.” (*Self-Employment in Canada*, Cat 71-582, 1985, p.8) By contrast, articles in the U.S. Monthly Labor Review (e.g. Bregger, 1996) routinely use SEC1)

9. The measure for total self-employment is based on longest job last year rather than job last week. Theoretically the total must be at least as large as the unincorporated. In practice, if the incorporated numbers are very small, this condition may be violated.
10. See, for example, Holtz-Eakin et.al.(1994) for an analysis of the effects of liquidity constraints.
11. This section is based on Robinson (1998).
12. If the age range is restricted to 16-63, there are 1344 (a,t) observations (i.e. cell means) in total given the years 1967-95, with 1994 missing. Cohort birth years can thus range from 1904 to 1979. However, these extreme cohorts would only contribute one (a,t) observation each compared to a maximum possible of 28 (for cohorts 1932 - 1951). The cohort range was therefore restricted to 1915 to 1964.

## REFERENCES

- Blau, David M. "A Time Series Analysis of Self-Employment in the United States." *Journal of Political Economy* 95 (1987): 445-467.
- Bregger, John E. "Measuring Self-Employment in the United States." *Monthly Labor Review*, January/February 1996.
- Devine, Theresa. "Changes in Wage and Salary Returns to Skill and the Recent Rise in Female Self-Employment." *American Economic Review* 84 (1994): 108-113.
- Evans, David S., and Leighton, Linda S. "Some Empirical Aspects of Entrepreneurship." *American Economic Review* 79 (1989): 519-35.
- Fairlie, R., and Meyer, B. "Ethnic and Racial Self-Employed Differences and Possible Explanations." *Journal of Human Resources* 31 (1996): 757-93.
- Fuchs, Victor. "Self-Employment and Labor Force Participation of Older Males." *Journal of Human Resources* 18 (1982): 339-57.
- Holtz-Eakin, Douglas; Joulfaian, David; Rosen, Harvey S. "Sticking It Out: Entrepreneurial Survival and Liquidity Constraints." *Journal of Political Economy* 102 (1994): 53-75.
- OECD "Recent Developments in Self-Employment", Chapter 4 in OECD *Employment Outlook*, OECD Paris, July 1992.

TABLE 1

Cross-Section Age Profiles for Canada and the United States:  
Unincorporated Self Employed Nonagricultural Workers.

Age Group	CANADA				UNITED STATES				
	1971	1981	1991	1996	1971	1981	1991	1995	
15-16	.0133	.0295	.1661	.2822	.0468	.0469	.0270	.0506	
17-19	.0061	.0393	.0268	.0628	.0118	.0125	.0108	.0197	
20-24	.0159	.0251	.0358	.0353	.0184	.0244	.0232	.0201	
25-34	.0584	.0593	.0717	.0799	.0534	.0603	.0608	.0541	
35-44	.1015	.0705	.0840	.1014	.0736	.0892	.0893	.0816	
45-54	.1039	.0881	.1040	.1099	.0899	.0983	.1002	.0952	
55-64	.1194	.0805	.1184	.1555	.1084	.1040	.1161	.1138	
65-69	.2039	.1892	.2551	.2933	.1922	.1642	.1768	.1630	
70+	.3189	.2748	.1955	.3356	.2365	.2517	.2240	.2205	
15-64	.0720	.0652	.0805	.0959	.0663	.0696	.0747	.0707	
Total	.0751	.0775	.0823	.0987	.0716	.0734	.0782	.0740	

TABLE 2

Cross-Section Age Profiles for Canada and the United States:  
Unincorporated Self Employed Male Nonagricultural Workers.

CANADA

UNITED STATES

Age Group	1971	1981	1991	1996		1971	1981	1991	1995
15-16	.0135	.1537	.1166	.2294		.0738	.0653	.0430	.0589
17-19	.0084	.0192	.0174	.0463		.0134	.0138	.0127	.0248
20-24	.0180	.0262	.0465	.0363		.0207	.0313	.0297	.0228
25-34	.0714	.0619	.0787	.0859		.0566	.0705	.0708	.0576
35-44	.1230	.0767	.0968	.1077		.0923	.1094	.1060	.0949
45-54	.1258	.0969	.1213	.1269		.1111	.1213	.1203	.1096
55-64	.1309	.0809	.1378	.1785		.1370	.1248	.1415	.1348
65-69	.2128	.2023	.2860	.3128		.2356	.2052	.2274	.1976
70+	.3164	.3064	.2029	.3216		.2738	.2712	.2686	.2674
15-64	.0897	.0674	.0920	.1048		.0818	.0857	.0899	.0810
Total	.0929	.0704	.0943	.1082		.0885	.0902	.0944	.0853

TABLE 3

Cross-Section Age Profiles for Canada and the United States:

Unincorporated Self Employed Female Nonagricultural Workers.

CANADA

UNITED STATES

Age Group	1971	1981	1991	1996		1971	1981	1991	1995
15-16	.0131	.2659	.2188	.3281		.0172	.0269	.0123	.0428
17-19	.0037	.0586	.0365	.0787		.0100	.0111	.0089	.0147
20-24	.0133	.0238	.0259	.0341		.0157	.0168	.0163	.0171
25-34	.0305	.0554	.0637	.0731		.0471	.0470	.0489	.0498
35-44	.0525	.0615	.0695	.0941		.0418	.0626	.0702	.0666
45-54	.0572	.0740	.0816	.0895		.0567	.0676	.0772	.0793
55-64	.0913	.0796	.0870	.1201		.0635	.0742	.0833	.0890
65-69	.1826	.1684	.2017	.2561		.1235	.1083	.1151	.1221
70+	.3267	.1769	.1806	.3638		.1716	.2252	.1714	.1555
15-64	.0386	.0621	.0671	.0855		.0422	.0491	.0572	.0591
Total	.0414	.0634	.0682	.0876		.0454	.0520	.0594	.0611

TABLE 4

Cohort and Cross Section Age Profiles for Canada by Sex:  
Unincorporated Self Employed Nonagricultural Workers.

	MALES					FEMALES			
Age in 1971	1971	1981	1991	1996		1971	1981	1991	1996
*			.0465	.0559				.0486	.0484
**		.0361	.0787	.0976			.0578	.0637	.0850
15-24	.0154	.0619	.0968	.1197		.0105	.0554	.0695	.0915
25-34	.0714	.0767	.1213	.1449		.0305	.0615	.0816	.1021
35-44	.1230	.0969	.1378			.0525	.0740	.0870	
45-54	.1258	.0809				.0572	.0796		
55-64	.1309					.0913			

\*15-24 in 1991

\*\*15-24 in 1981

TABLE 5

Estimates of the Linear Probability Model for Unincorporated Self Employment  
(Nonagricultural Workers, 16-59)

	<u>Total</u>		<u>Males</u>		<u>Females</u>	
	Coefficient	t-statistic	Coefficient	t-statistic	Coefficient	t-statistic
C: 1915-24	0.1211	0.279	0.2415	0.386	-0.4301	-0.749
C: 1925-34	-0.1833	-0.698	-0.1213	-0.322	-0.5188	-1.486
C: 1945-54	0.1999	0.827	0.3142	0.899	0.2500	0.801
C: 1955-64	-0.4230	-1.052	-0.3639	-0.624	-0.4041	-0.768
Age	0.6471	16.624	0.7446	13.040	0.4752	9.456
Agesq	-0.0058	-12.580	-0.0063	-9.363	-0.0045	-7.520
Year 1967	0.7774	1.831	0.7622	1.264	0.6215	1.086
Year 1968	-0.0901	-0.192	-0.4491	-0.677	0.3779	0.597
Year 1969	-0.0754	-0.165	-0.2490	-0.386	0.0329	0.053
Year 1970	-0.1675	-0.371	-0.3002	-0.468	0.0271	0.045
Year 1971	0.1109	0.249	0.0501	0.079	0.1295	0.216
Year 1972	-0.1205	-0.272	0.1211	0.192	-0.4793	-0.806
Year 1973	0.3824	0.875	0.5468	0.881	0.1235	0.210
Year 1974	0.3356	0.771	0.1977	0.319	0.5030	0.862
Year 1975	-0.0496	-0.112	-0.1703	-0.270	0.2044	0.347
Year 1977	0.5031	1.216	0.4646	0.786	0.5787	1.051
Year 1978	1.1430	2.748	0.7793	1.306	1.7475	3.168
Year 1979	0.3725	0.899	0.0295	0.050	0.9526	1.736
Year 1980	1.2530	3.101	1.3200	2.262	1.4264	2.684
Year 1981	0.9216	2.249	0.8374	1.412	1.3566	2.523
Year 1982	1.3402	3.157	1.3579	2.211	1.5844	2.845
Year 1983	1.6567	3.829	1.8442	2.938	1.7459	3.085

Year 1984	1.6416	3.749	1.9961	3.135	1.6116	2.821
Year 1985	1.5276	3.444	1.4491	2.246	2.0101	3.476
Year 1986	0.9827	2.172	1.4727	2.238	0.8100	1.372
Year 1987	1.5494	3.339	2.0690	3.057	1.3836	2.294
Year 1988	1.5879	3.364	1.7374	2.538	1.7784	2.881
Year 1989	1.2251	2.498	0.7659	1.073	2.2154	3.469
Year 1990	1.3310	2.698	0.8277	1.152	2.3702	3.693
Year 1991	1.3826	2.719	0.7019	0.945	2.6746	4.052
Year 1992	1.3880	2.652	1.3521	1.774	1.9067	2.800
Year 1993	1.0343	1.924	1.3115	1.670	1.2609	1.808
Year 1995	0.5696	1.008	0.0939	0.114	1.6162	2.201
Constant	-8.7091	-9.063	-10.0504	-7.156	-6.5291	-5.244

R squared	.0108	.0134	.0082
Observations	229436	129616	99820

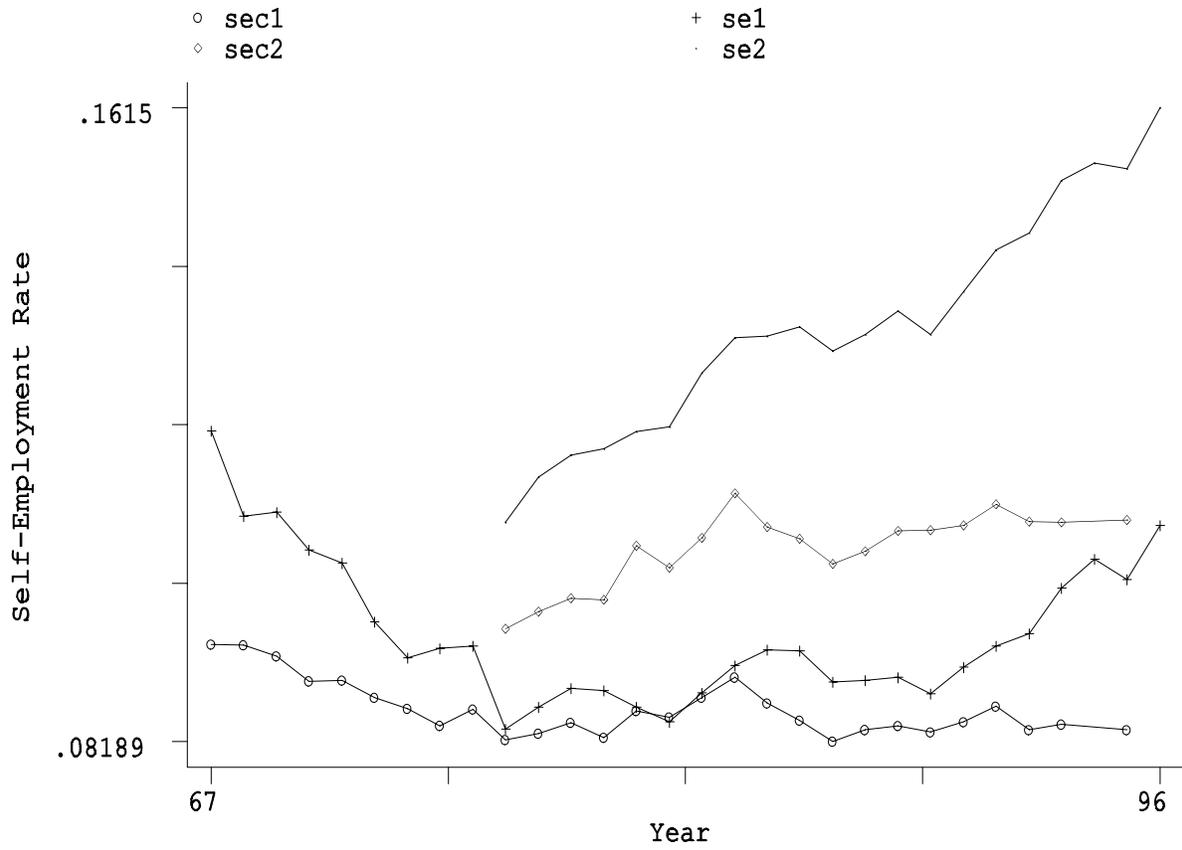


Figure 1 Self-Employment Rates by Measure: Canada & US

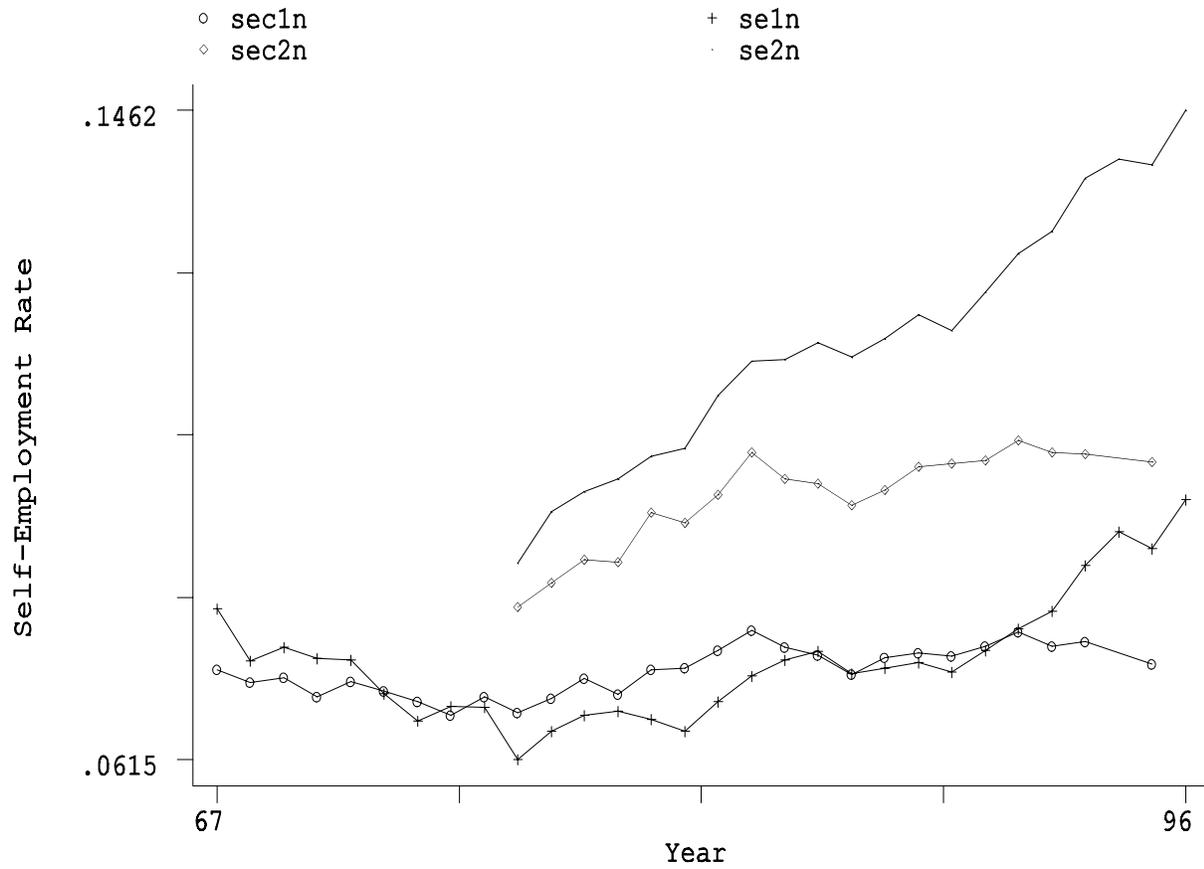


Figure 2 Nonfarm Self-Employment Rates by Measure: Canada & US

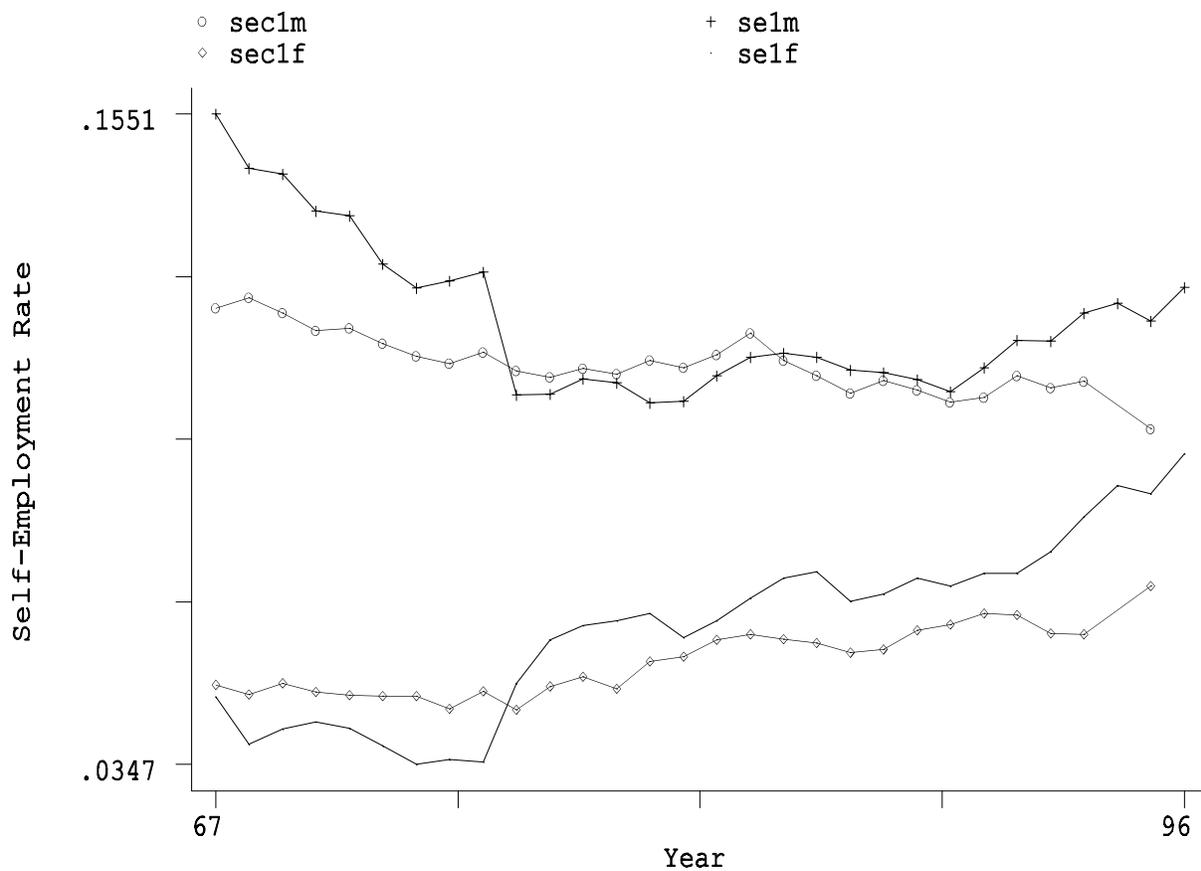


Figure 3 Unincorporated Self-Employment Rates by Sex:  
Canada & US

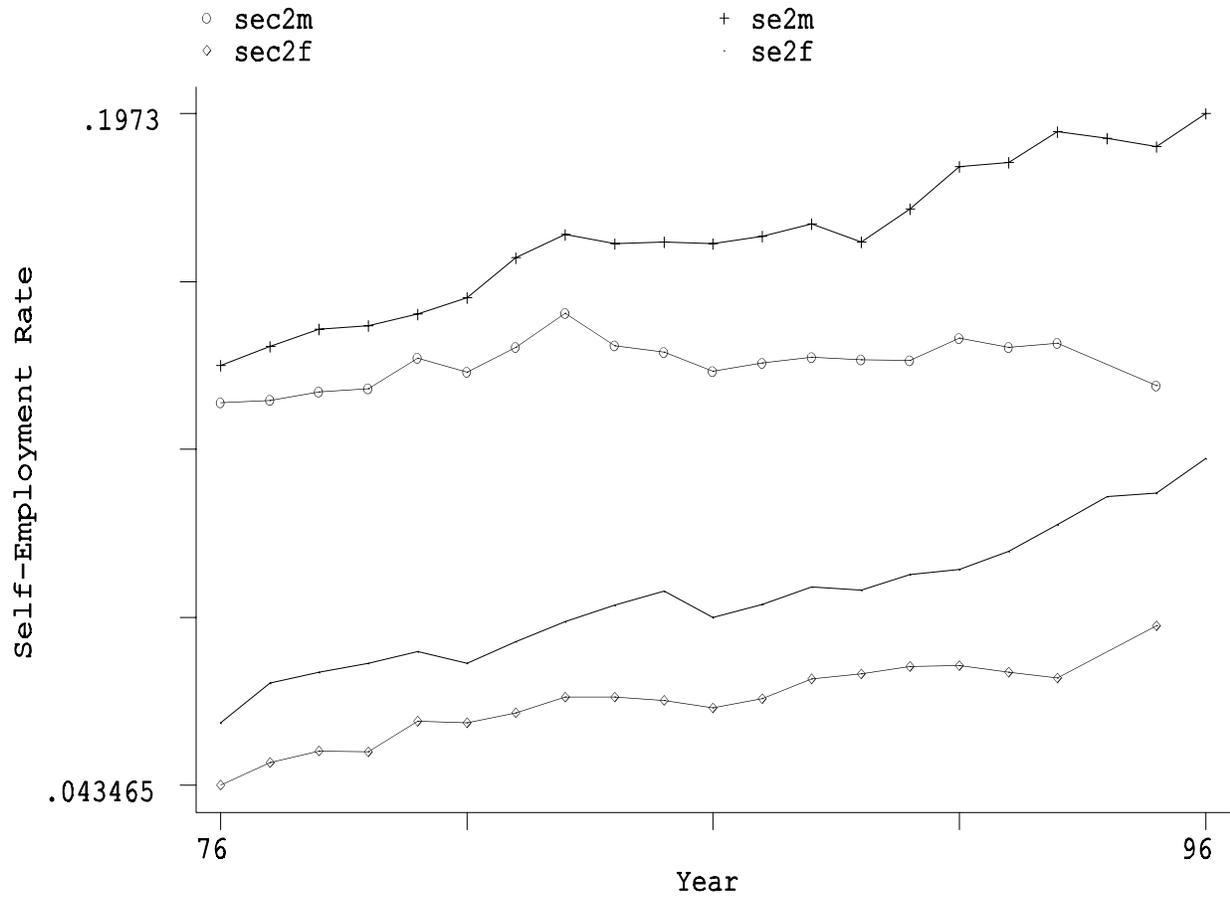
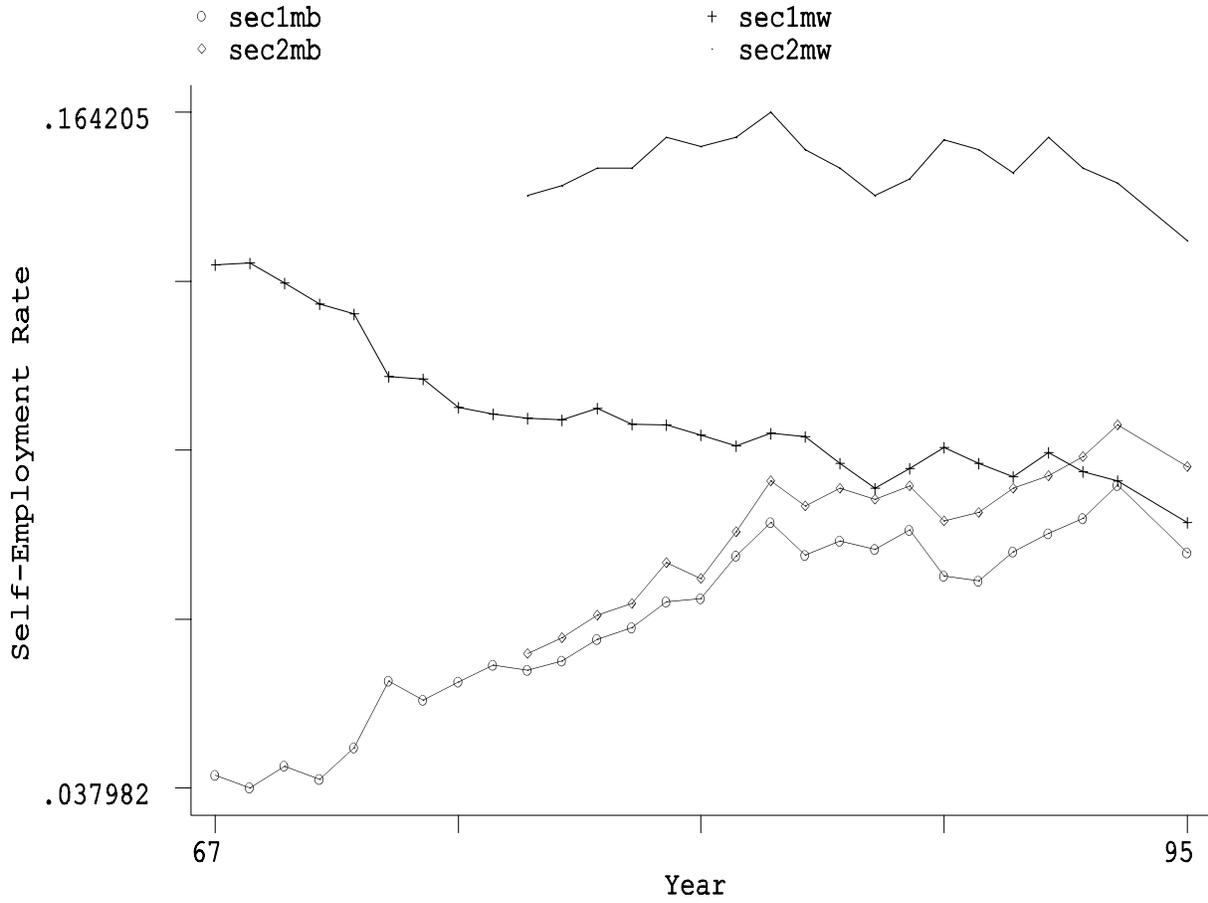


Figure 4 Total Self-Employment Rates by Sex: Canada & US



**Figure 5 US Male Nonfarm Self-Employment Rates by Occupation**

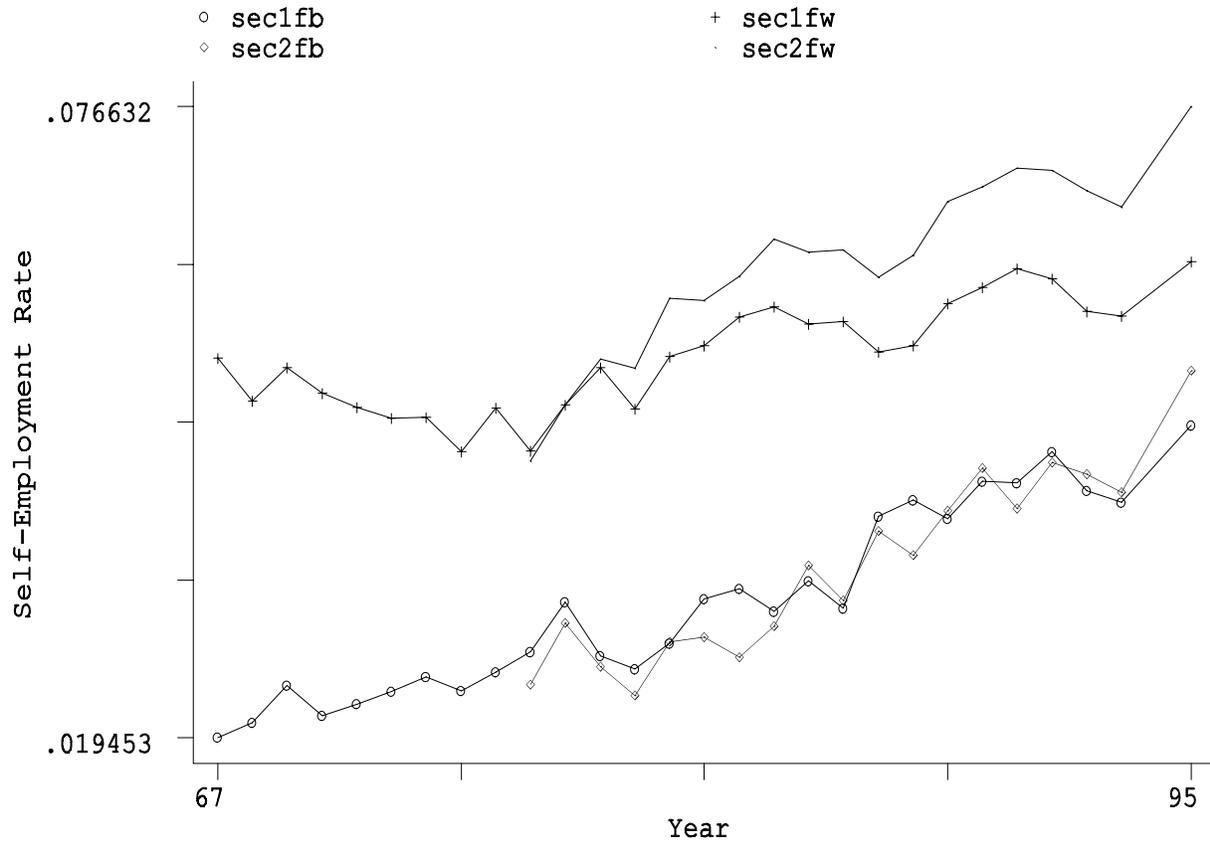


Figure 6 US Female Nonfarm Self-Employment Rates by Occupation

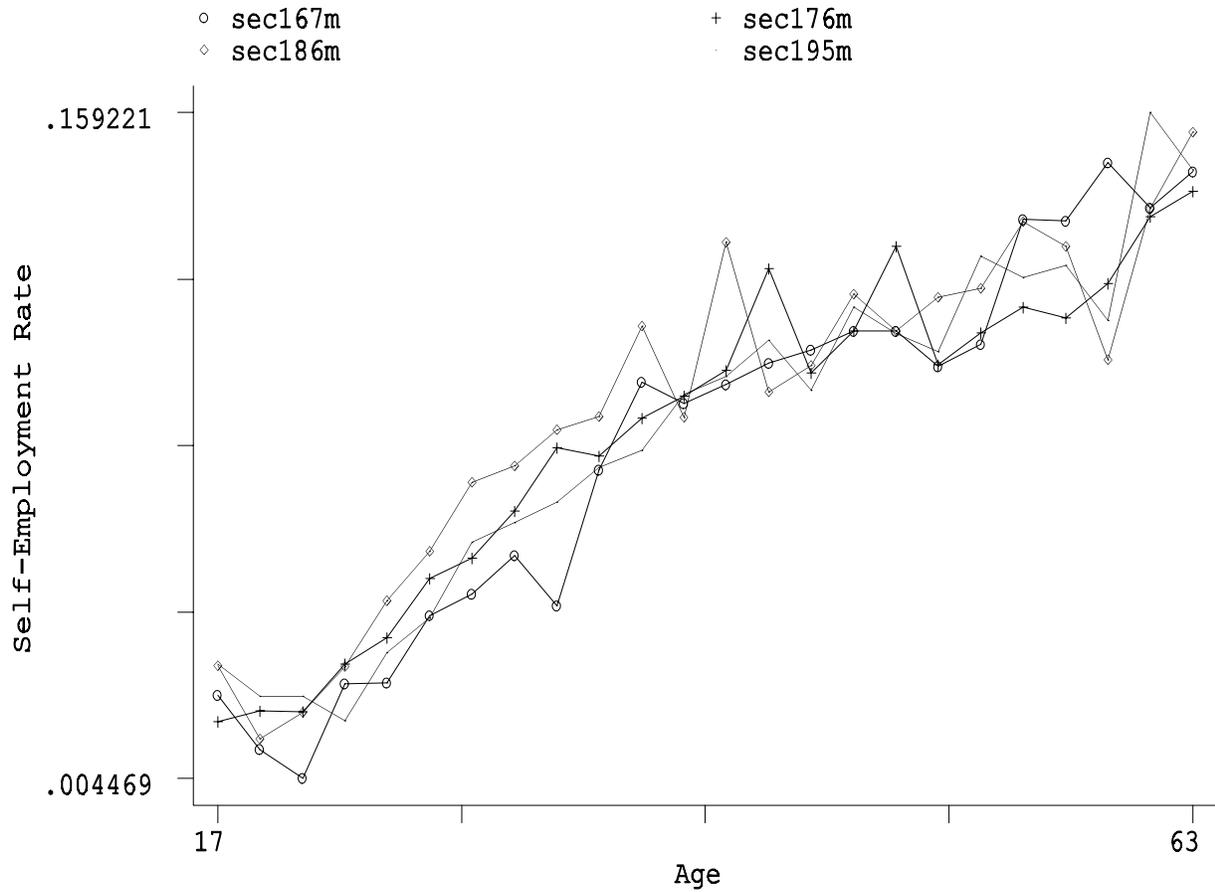


Figure 7a US Male Unincorporated Self-Employment Rates:  
Selected Years

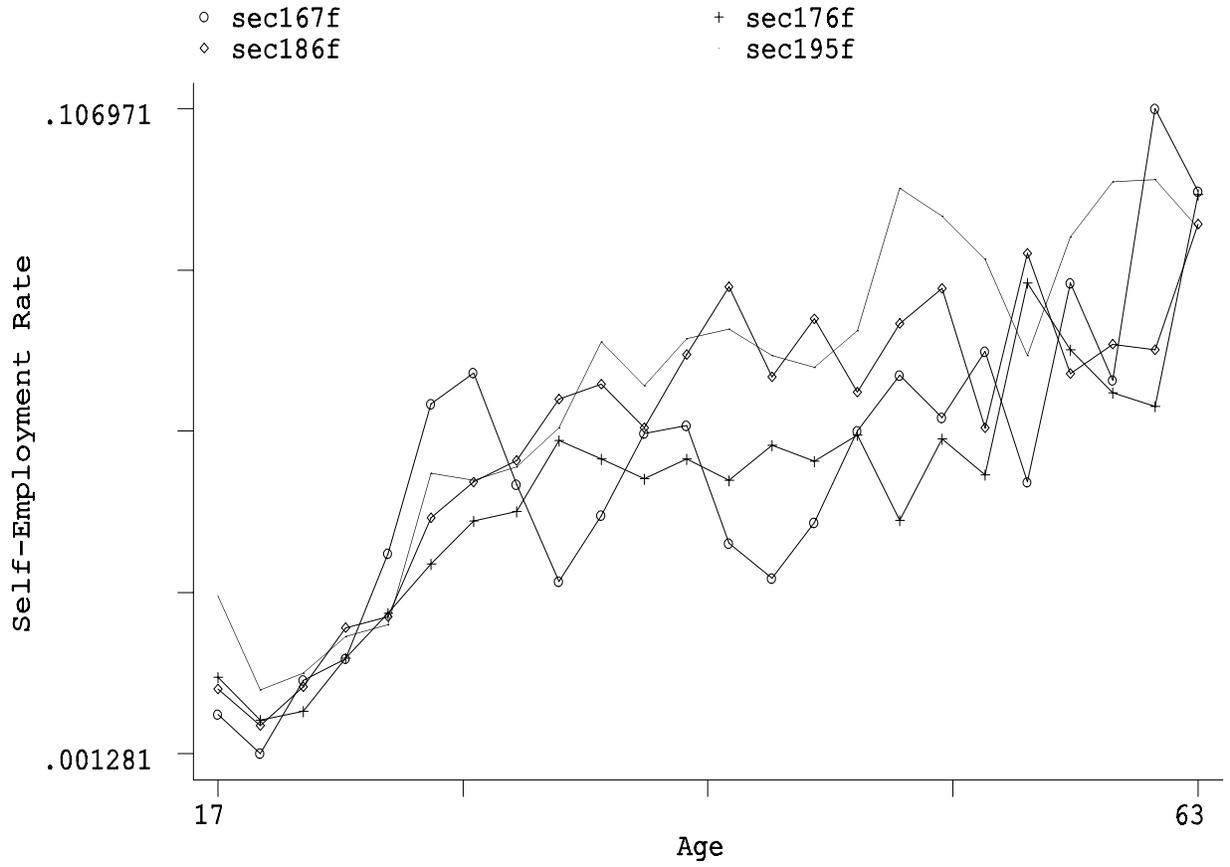


Figure 7b US Female Unincorporated Self-Employment Rates:  
Selected Years

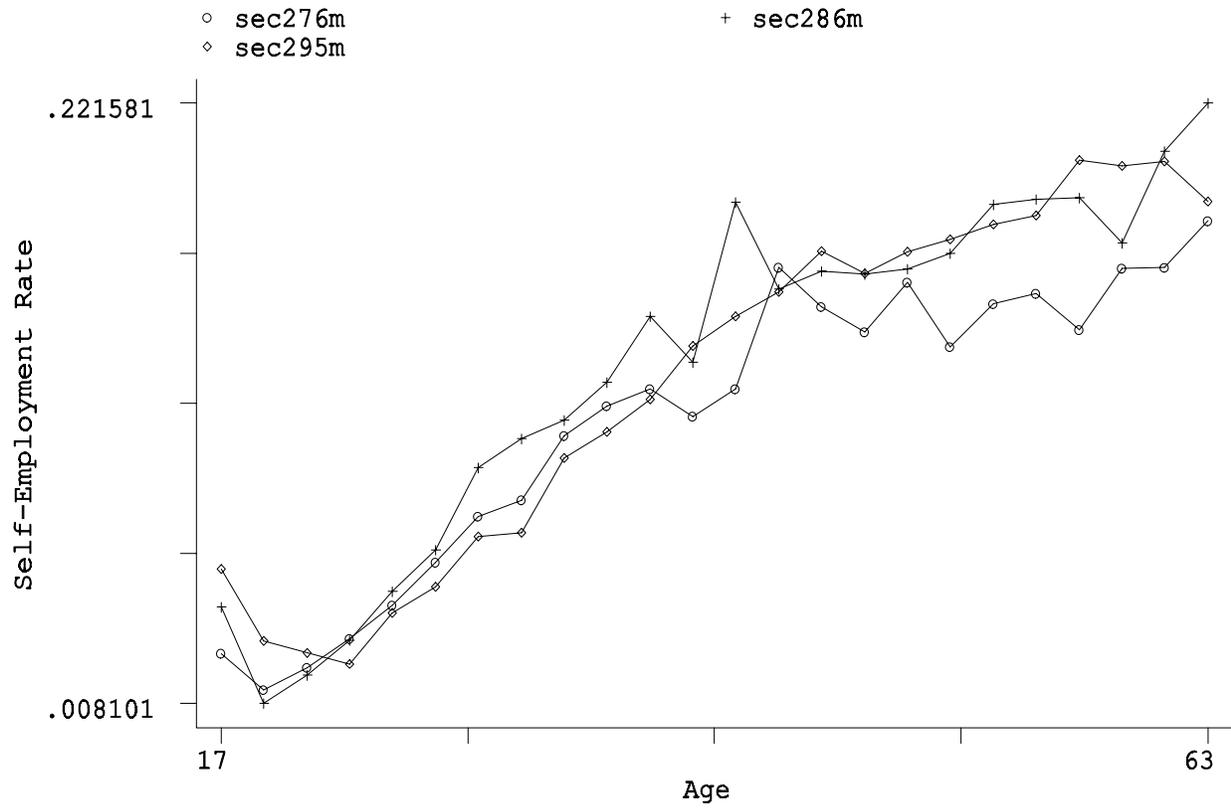


Figure 8a US Male Total Self-Employment Rates: Selected Years

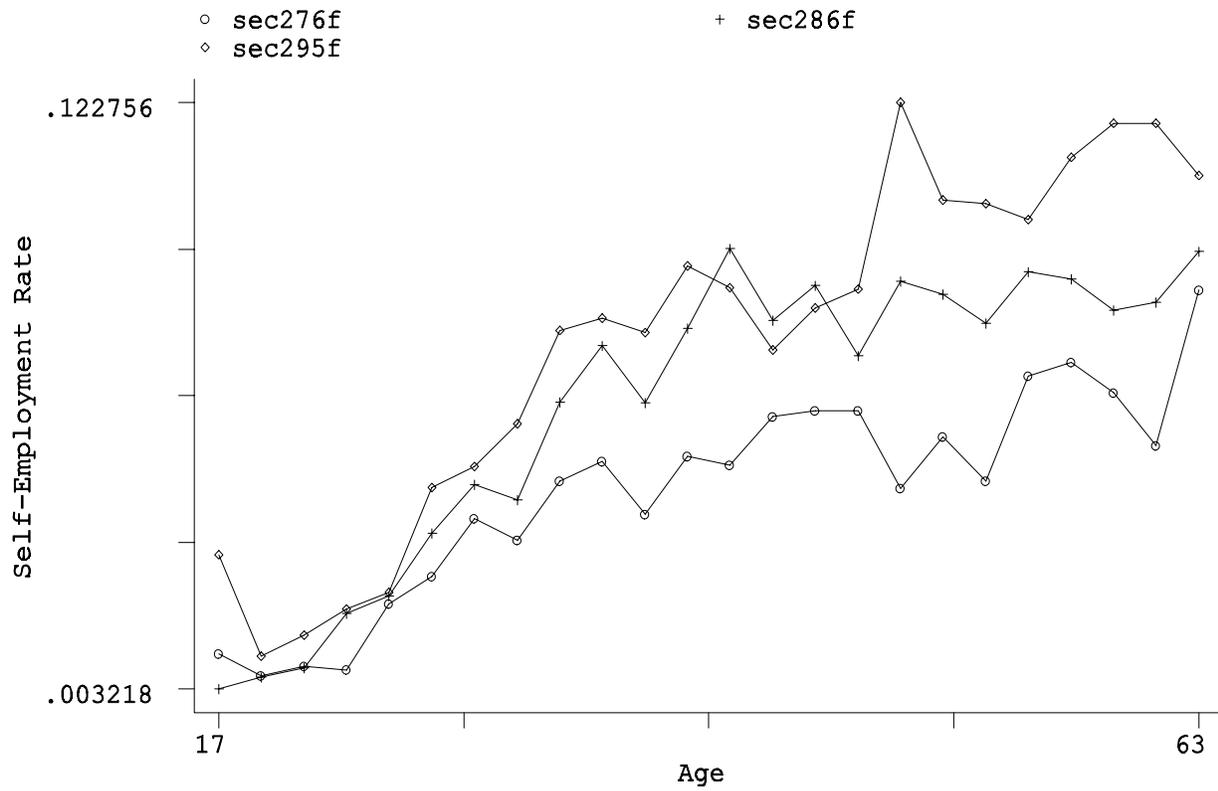


Figure 8b US Female Total Self-Employment Rates: Selected Years

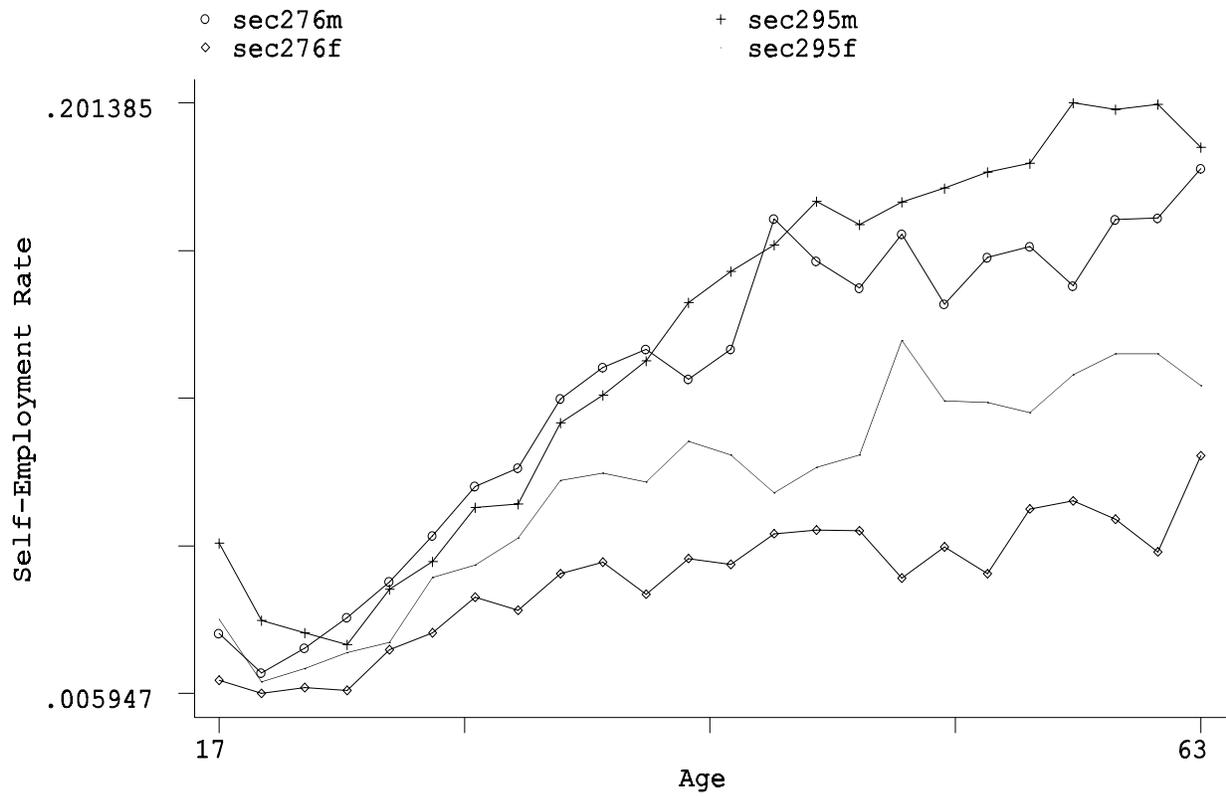


Figure 9 US Total Self-Employment Rates by Sex: Selected Years

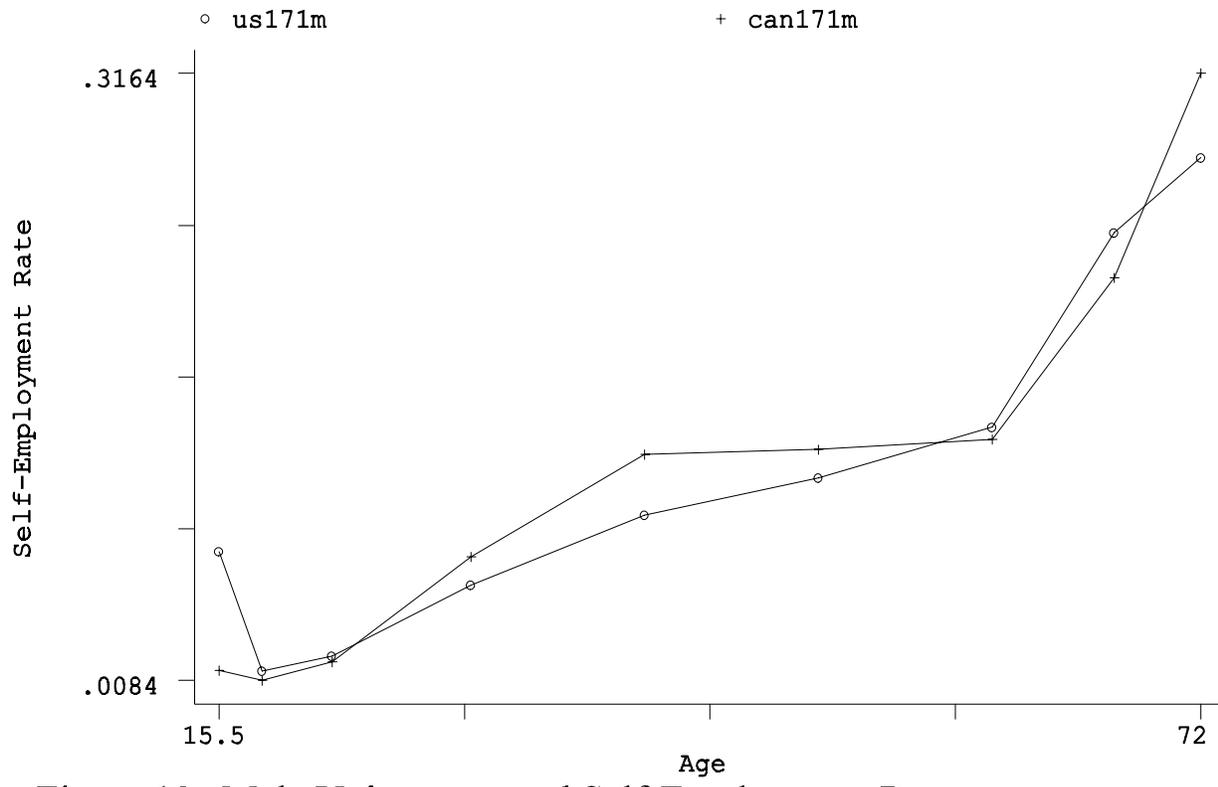


Figure 10a Male Unincorporated Self-Employment Rates:  
US & Canada, 1971

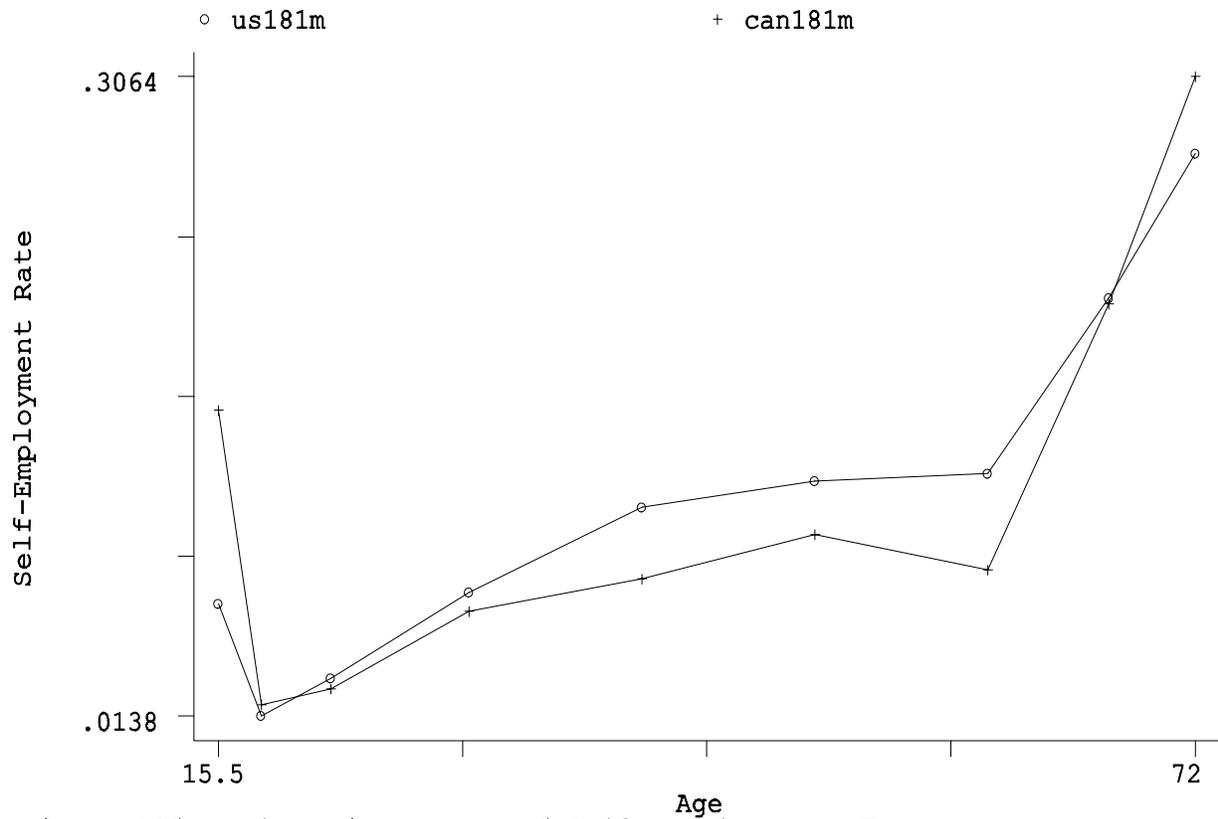


Figure 10b Male Unincorporated Self-Employment Rates:  
US & Canada, 1981

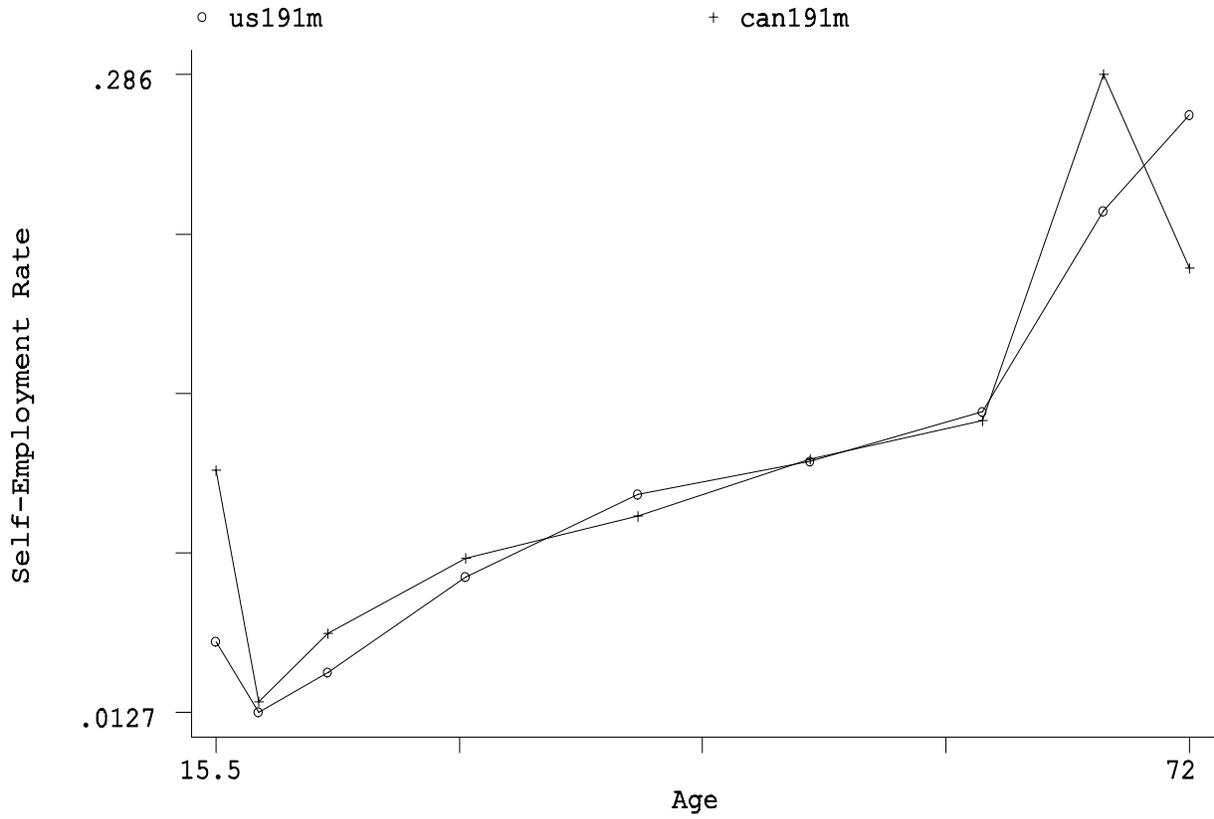
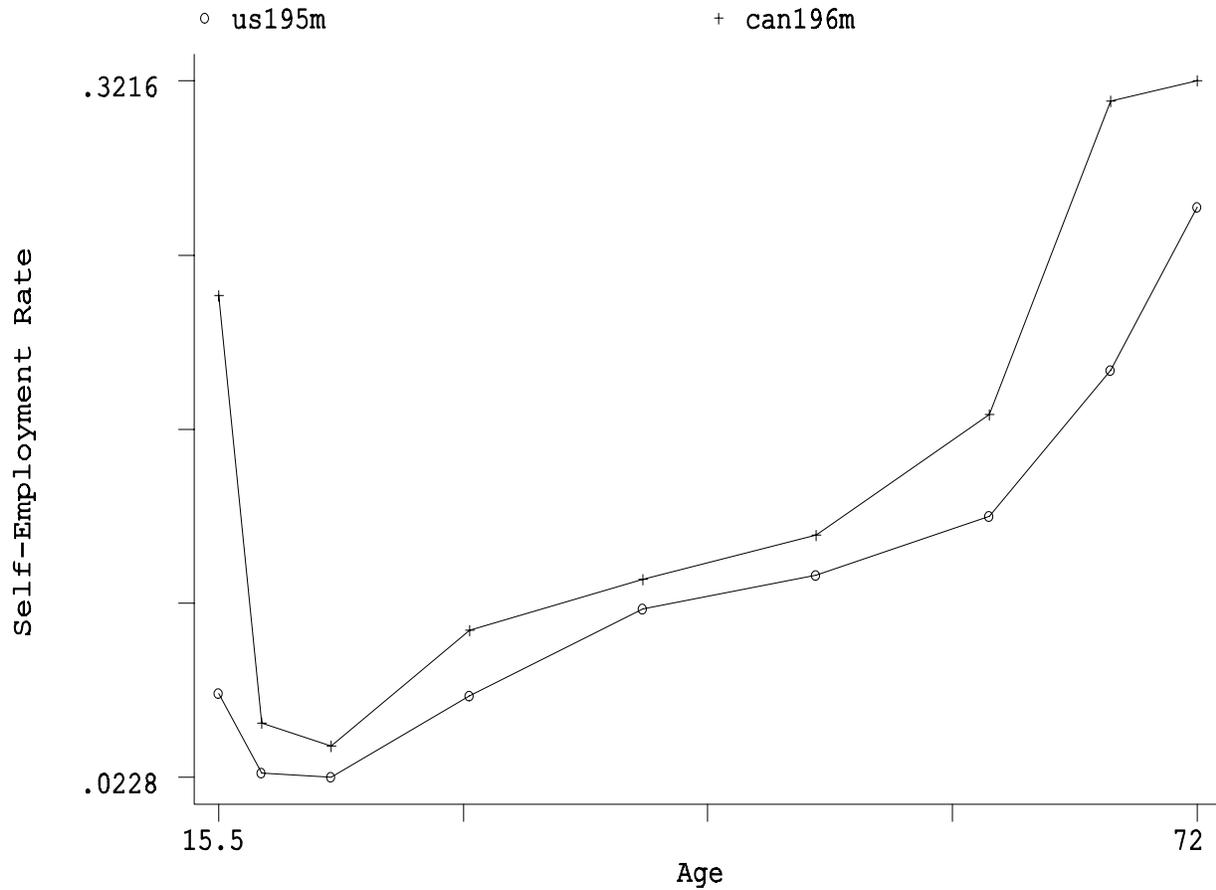


Figure 10c Male Unincorporated Self-Employment Rates:  
US & Canada, 1991



**Figure 10d Male Unincorporated Self-Employment Rates:  
US & Canada, 1995/1996**

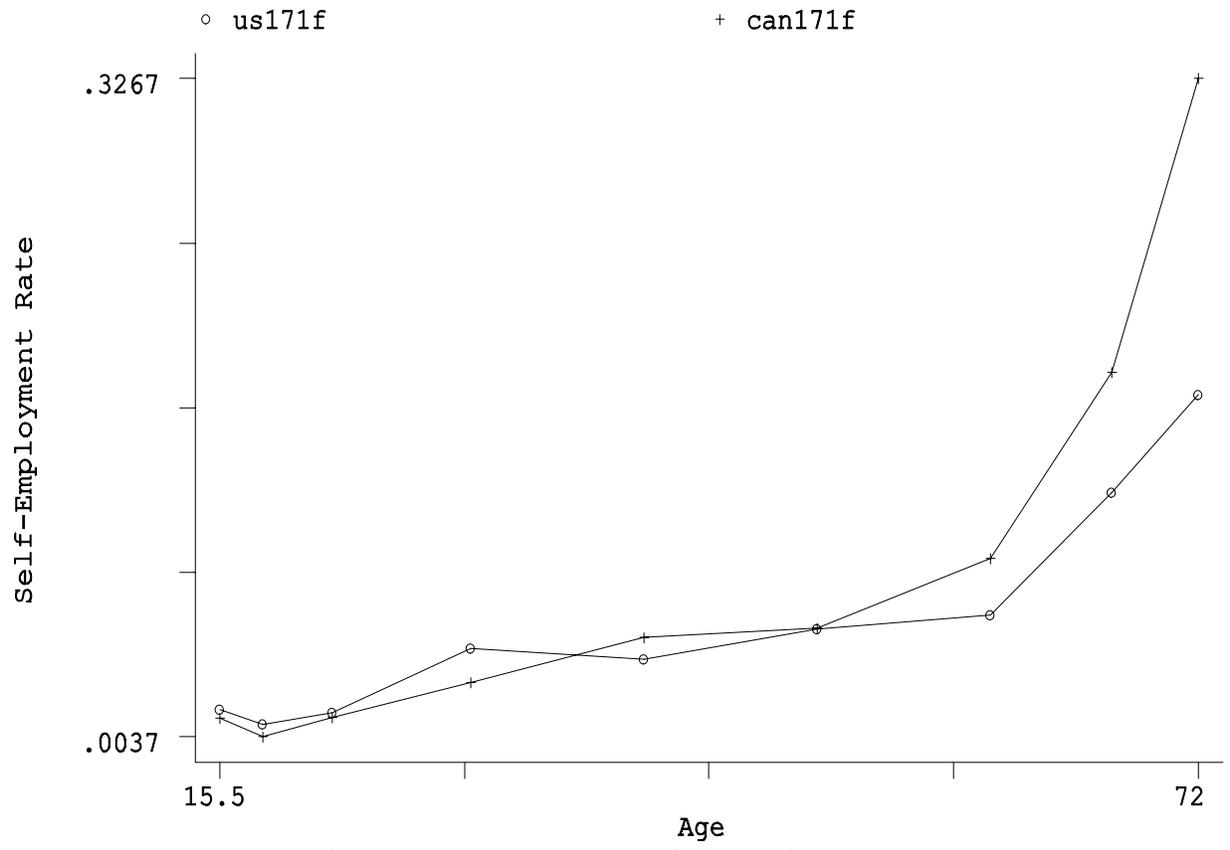


Figure 11a Female Unincorporated Self-Employment Rates:  
US & Canada, 1971

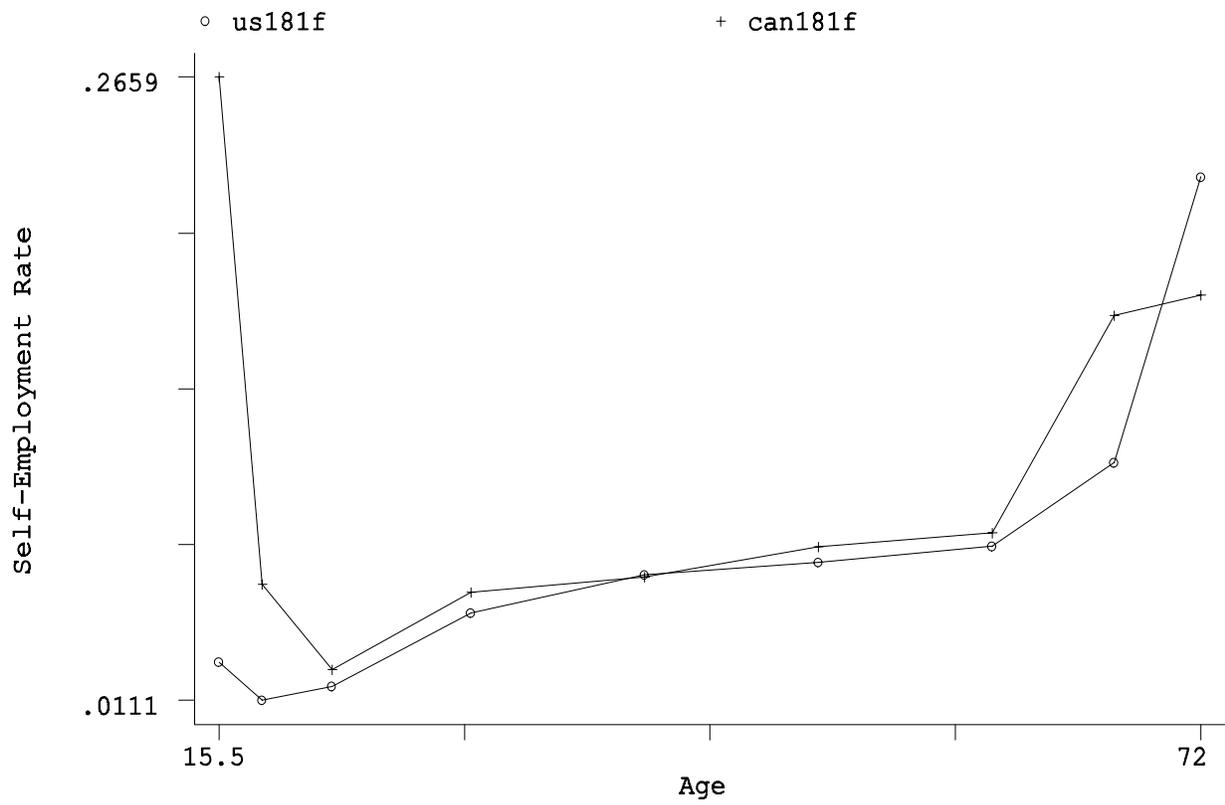


Figure 11b Female Unincorporated Self-Employment Rates:  
US & Canada, 1981

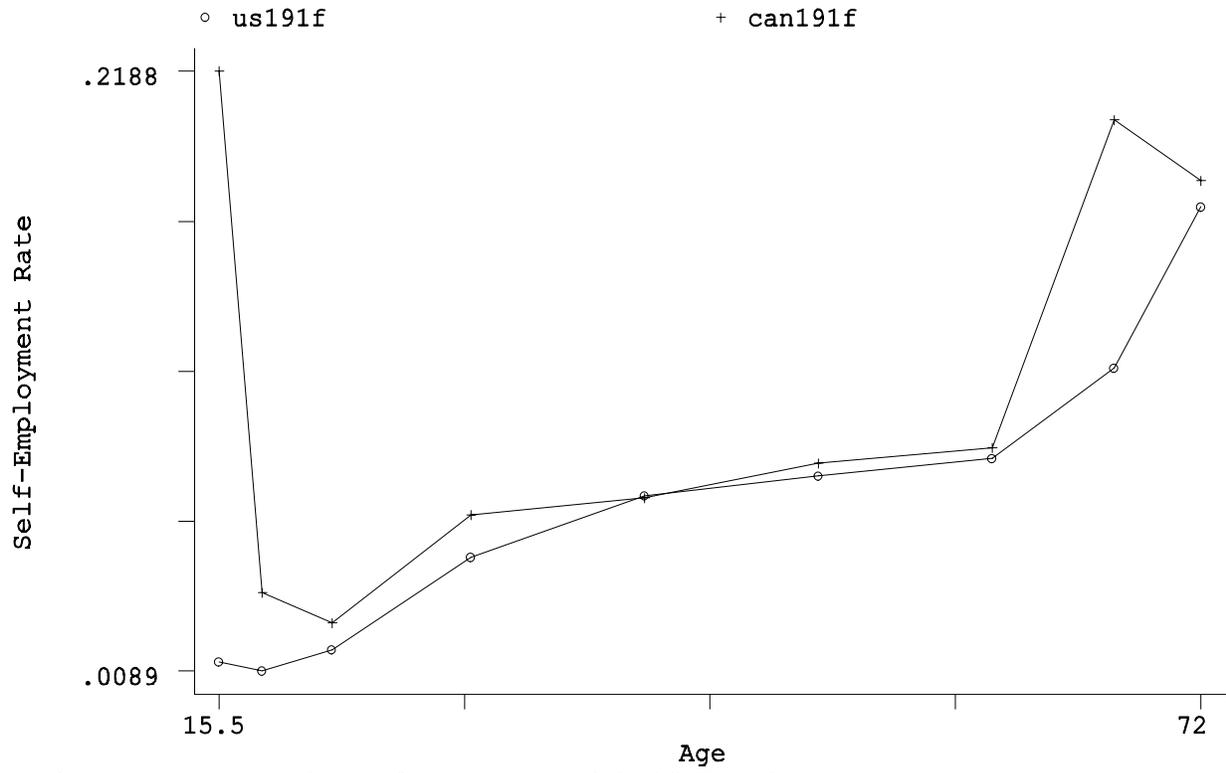
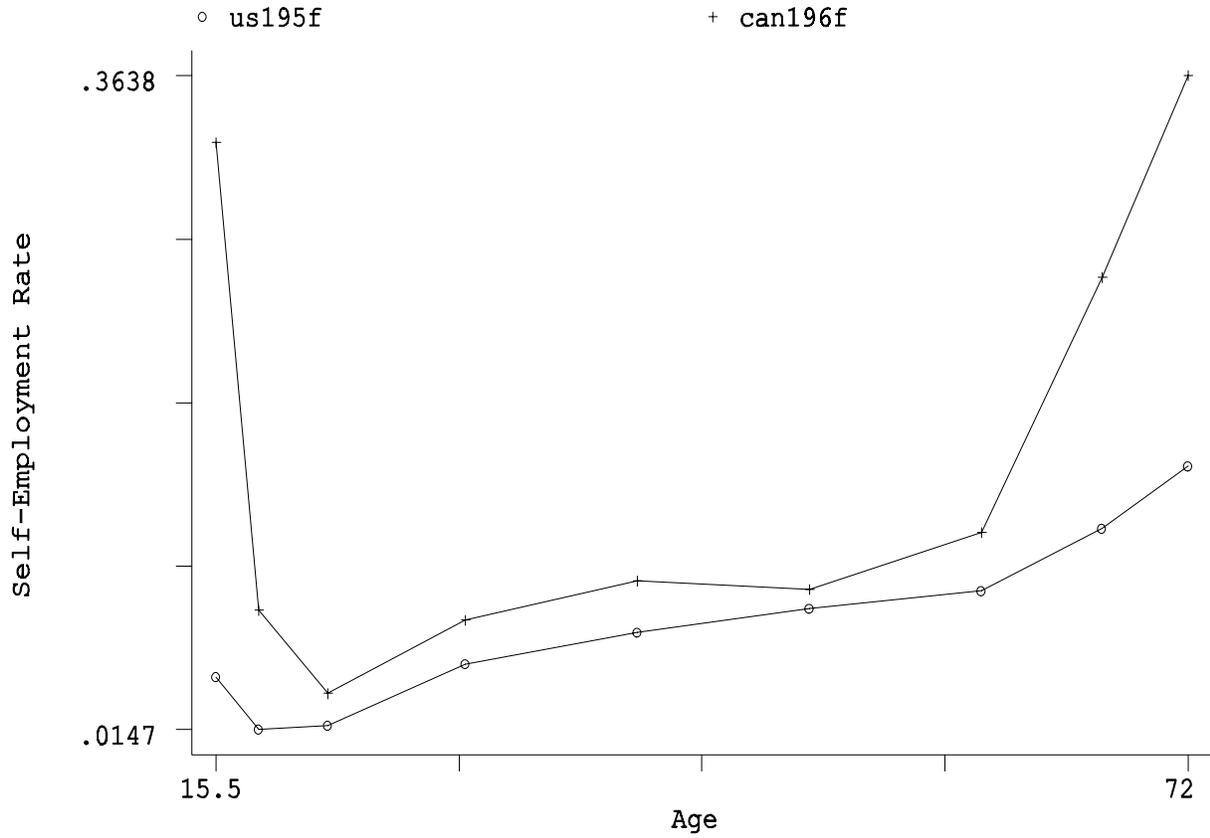


Figure 11c Female Unincorporated Self-Employment Rates:  
US & Canada, 1991



**Figure 11d Female Unincorporated Self-Employment Rates:  
US & Canada, 1995/1996**

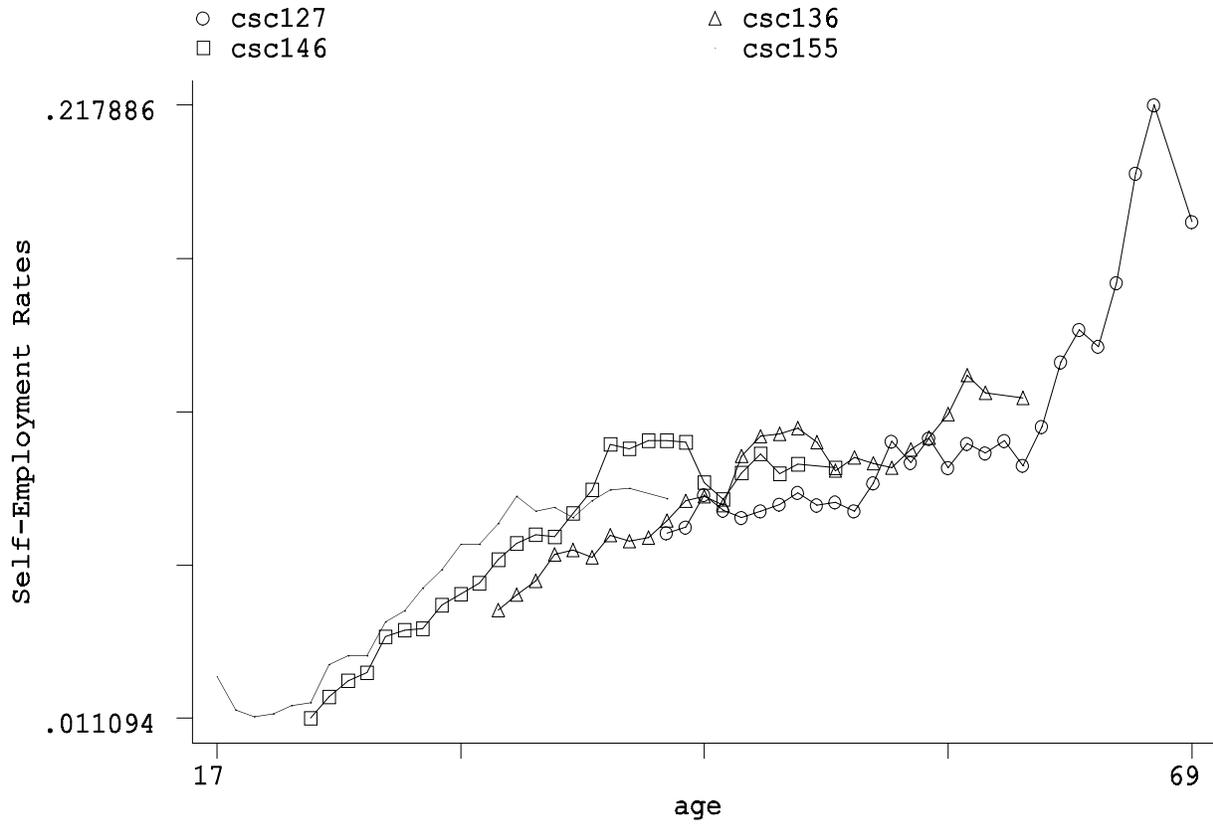


Figure 12 Lifecycle Profiles: Unincorporated Self-Employment Rates

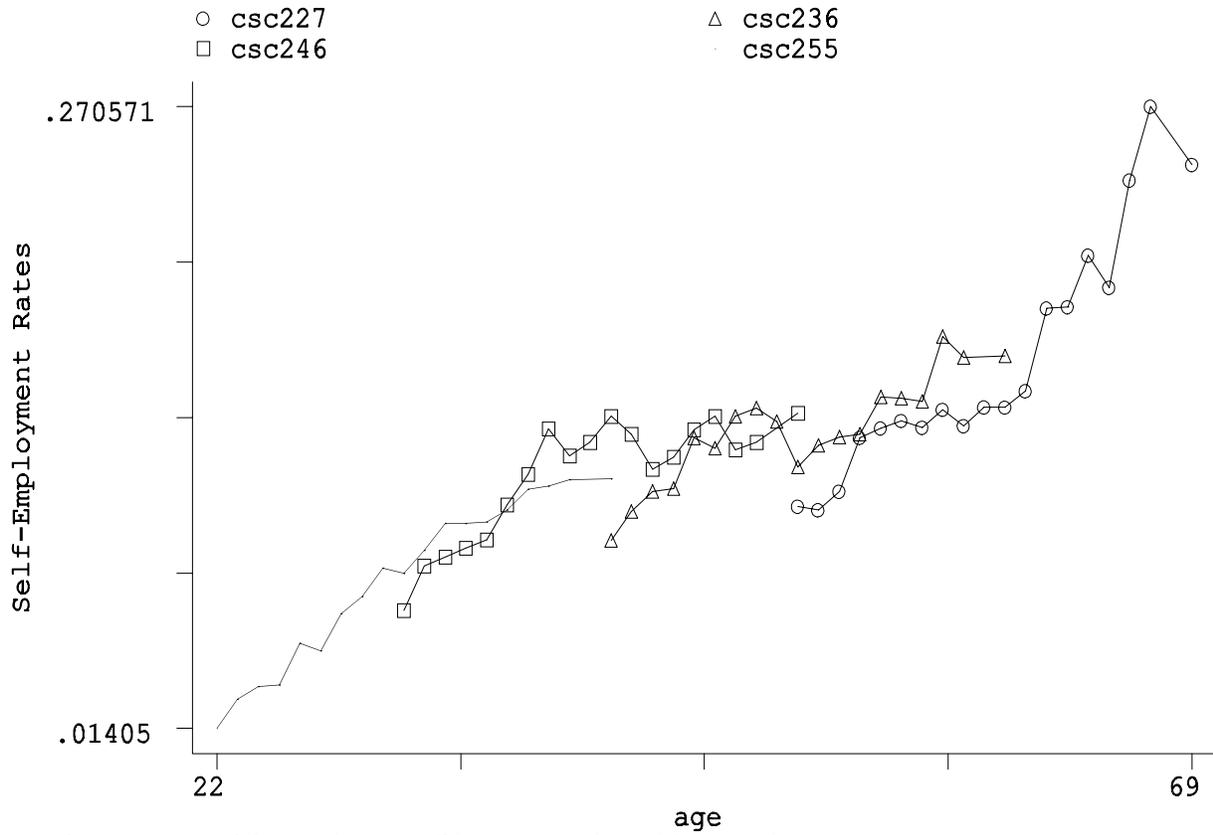


Figure 13 Lifecycle Profiles: Total Self-Employment Rates

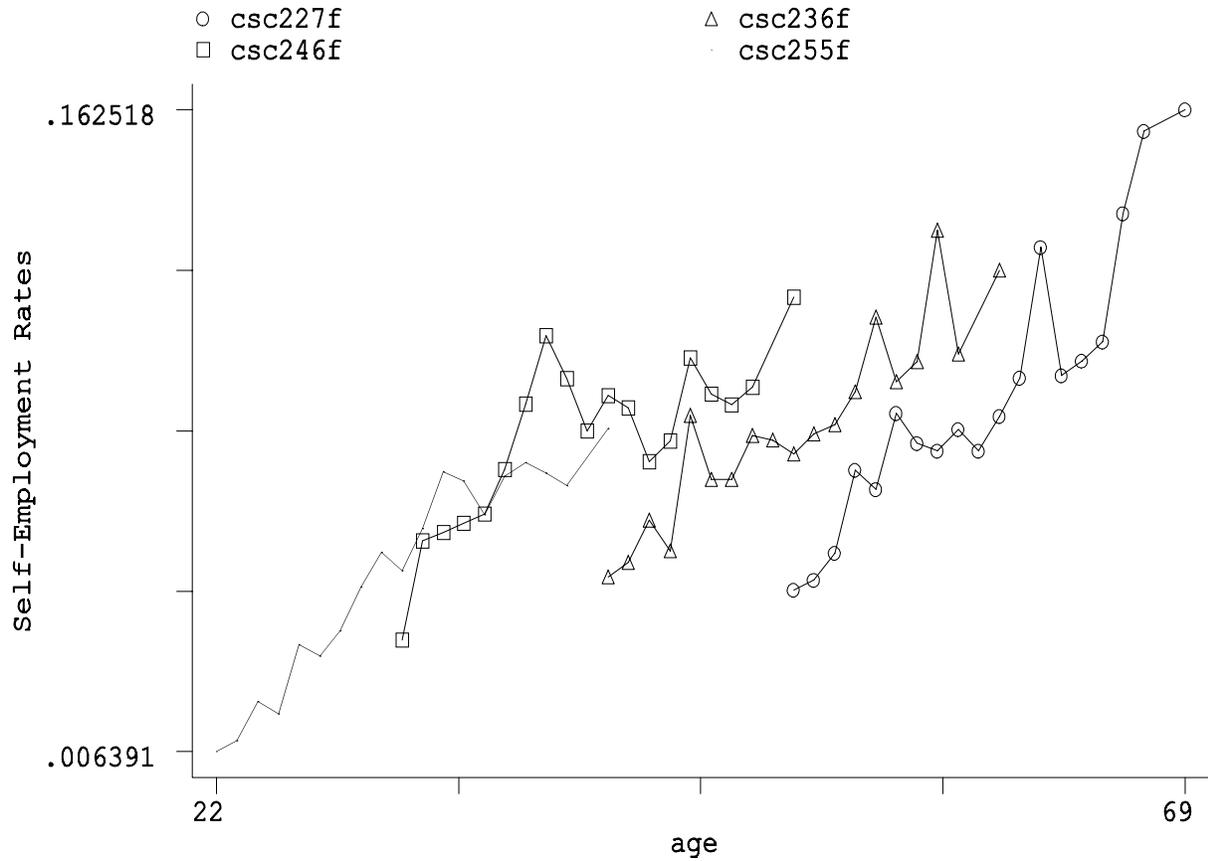


Figure 14 Lifecycle Profiles: Female Total Self-Employment Rates

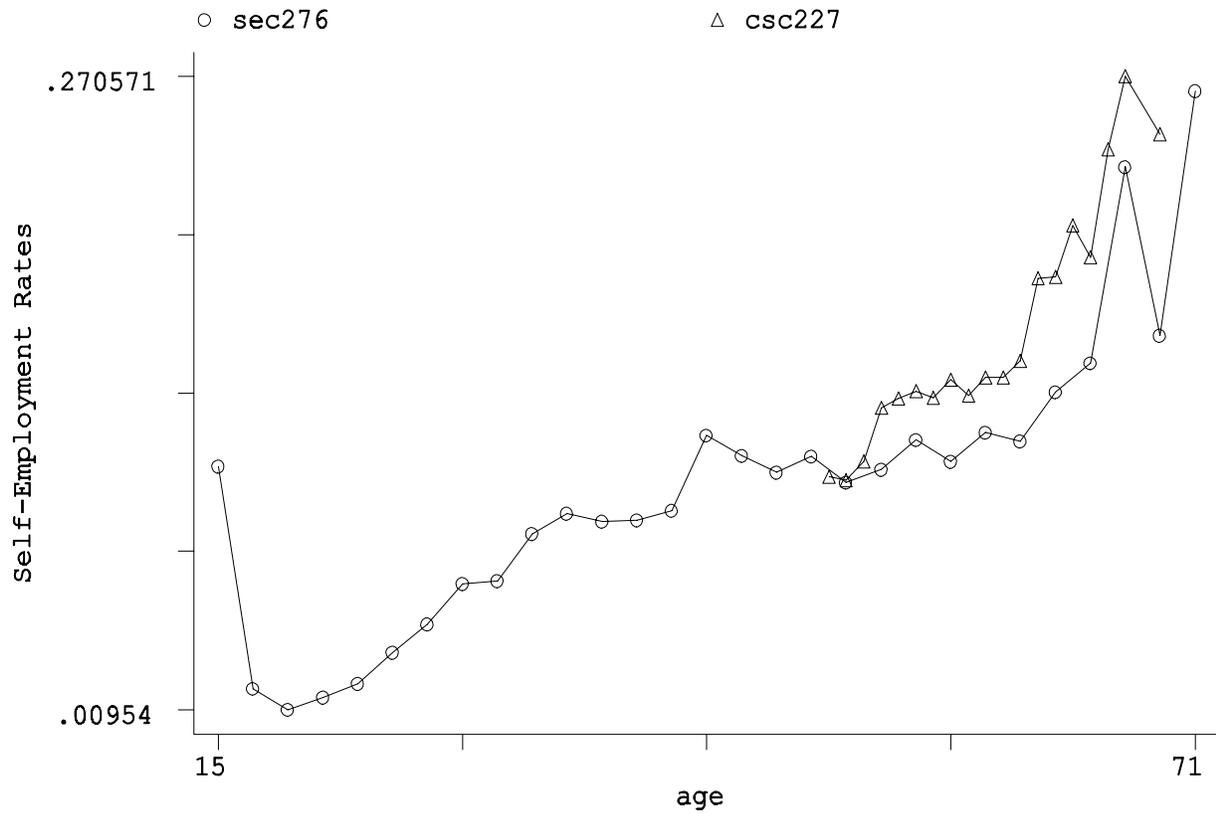


Figure 15a 1926-27 Cohort and 1976 Cross Section Age Profiles

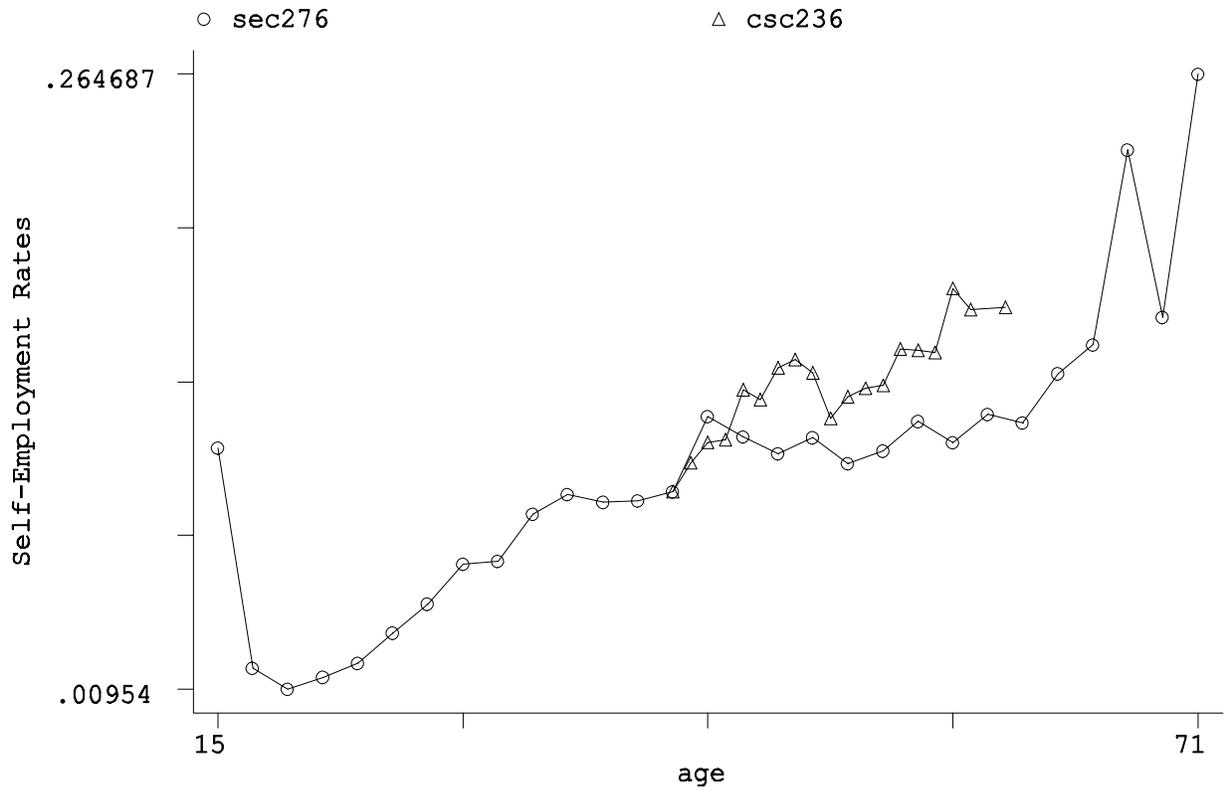


Figure 15b 1935-36 Cohort and 1976 Cross Section Age Profiles

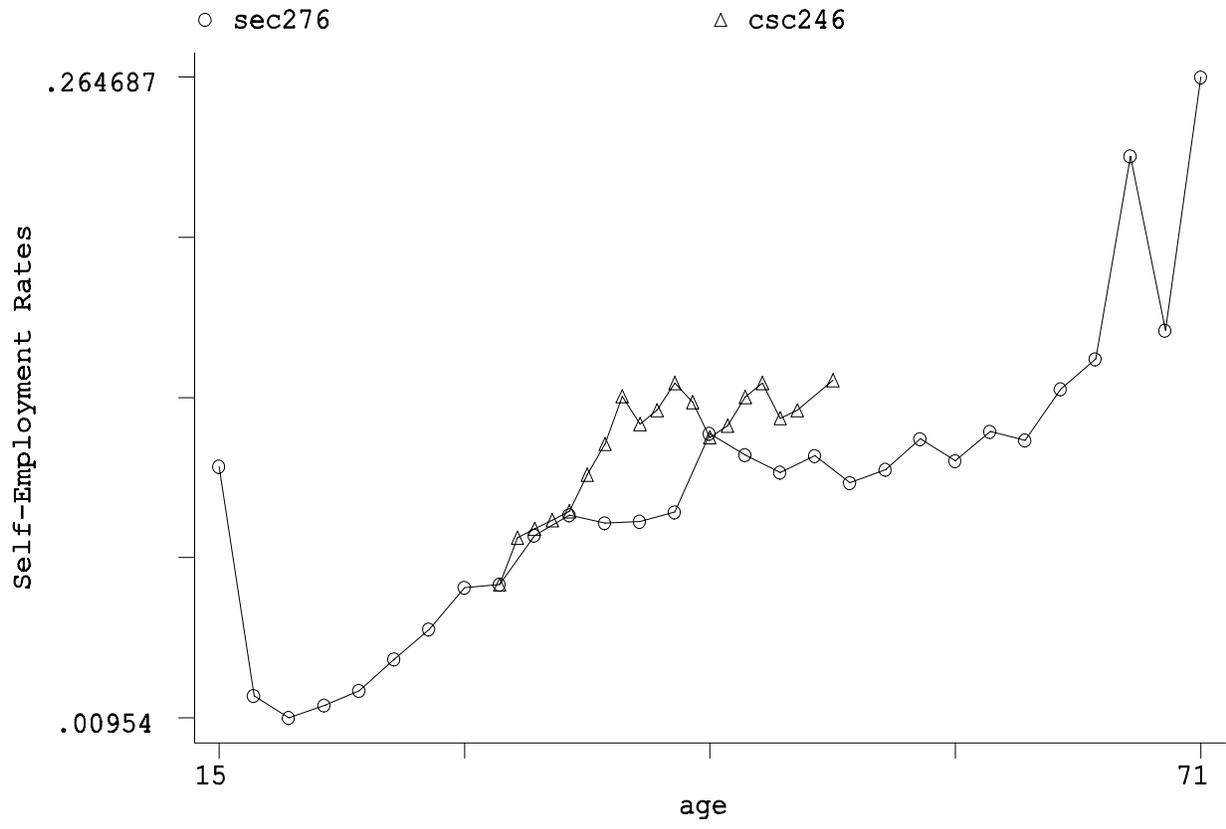


Figure 15c 1945-46 Cohort and 1976 Cross Section Age Profiles

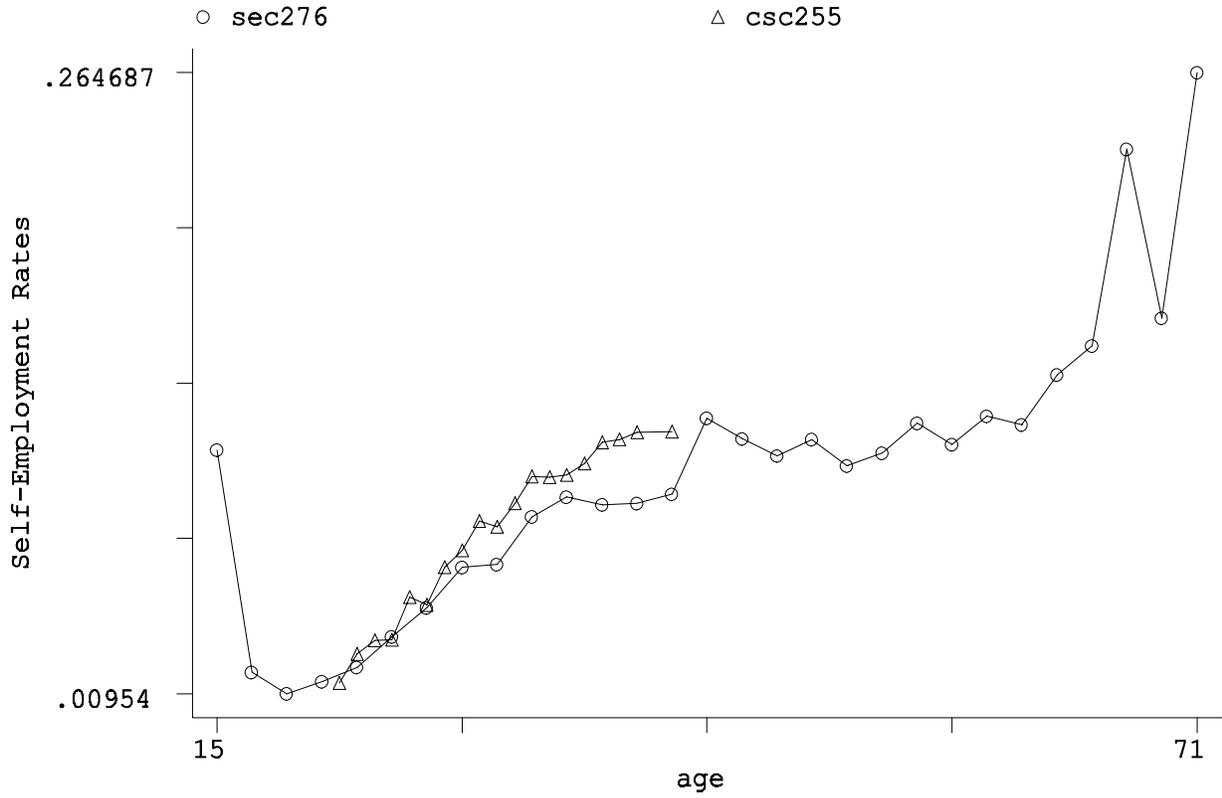


Figure 15d 1954-1955 Cohort and 1976 Cross Section Age Profiles

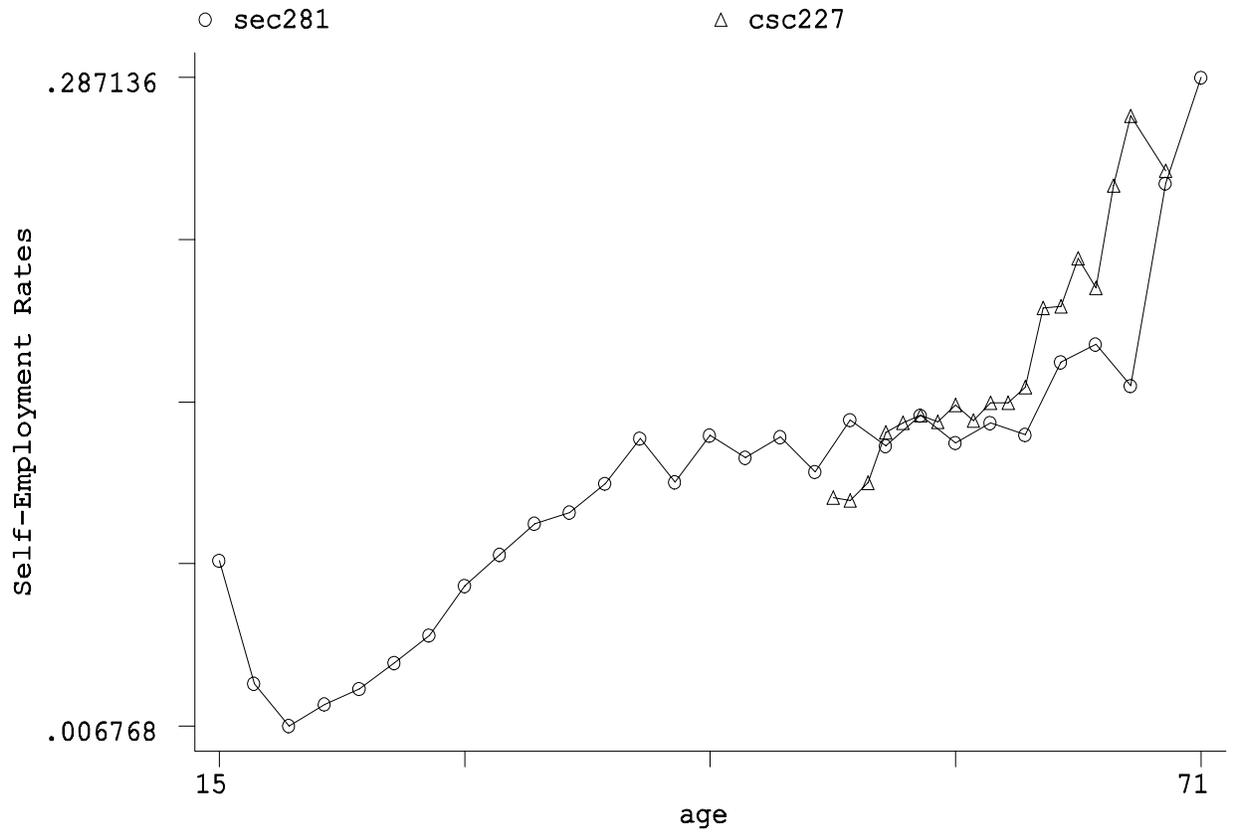


Figure 16a 1926-1927 Cohort and 1981 Cross Section Age Profiles

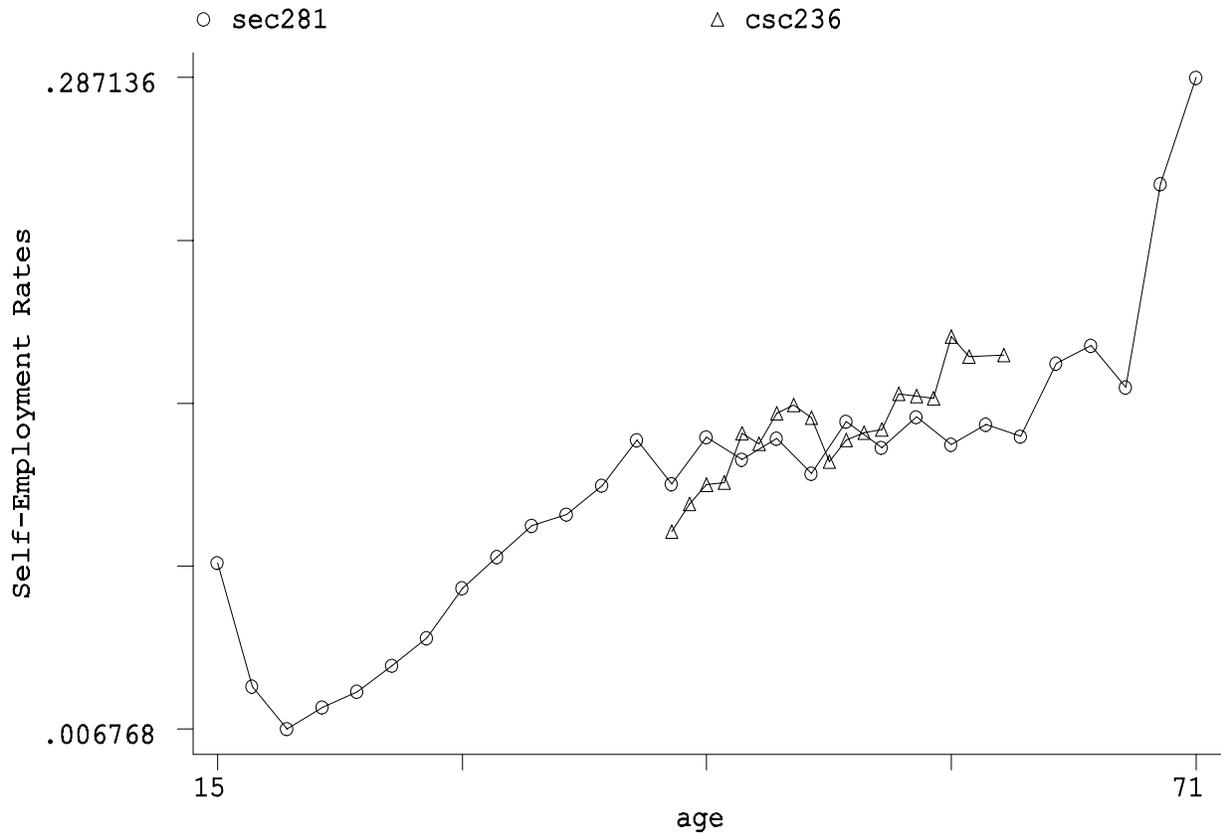


Figure 16b 1935-36 Cohort and 1981 Cross Section Age Profiles

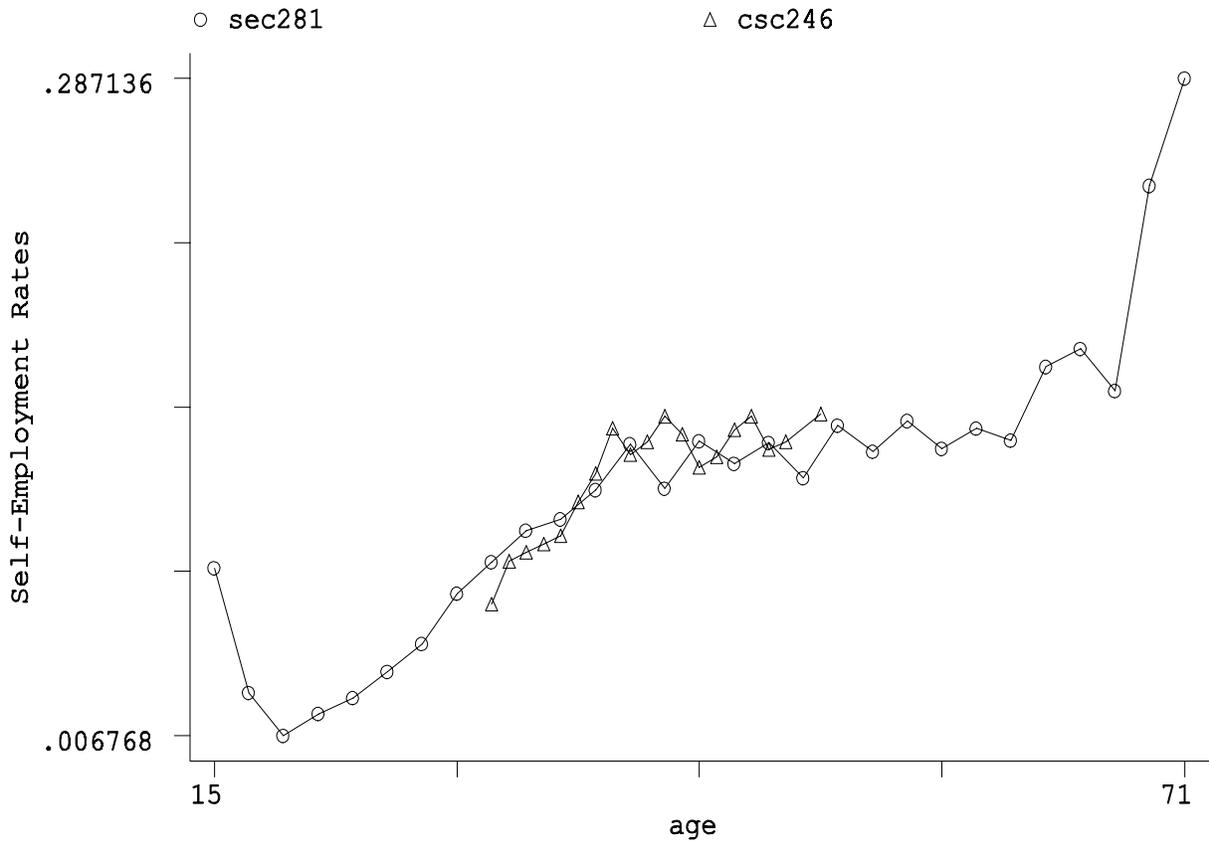


Figure 16c 1945-46 Cohort and 1981 Cross Section Age Profiles

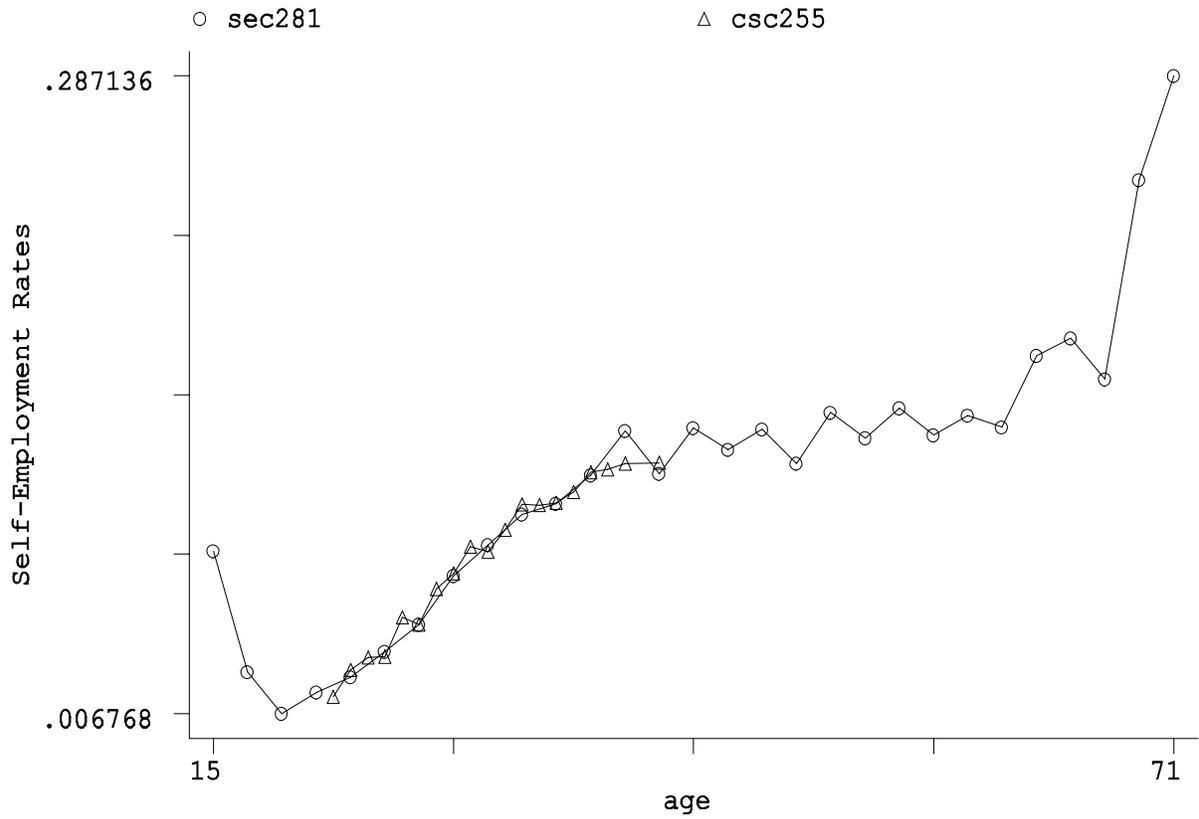


Figure 16d 1954-1955 Cohort and 1981 Cross Section Age Profiles