Inequality Trends in Sweden 1978 – 2004*

David Domeij[§] and Martin Flodén[¶]

SSE/EFI Working Paper Series in Economics and Finance
No 720
May, 2009

Abstract

We document a clear increase in Swedish earnings inequality in the early 1990s. Inequality in disposable income and earnings net of taxes and transfers also increased, but much less than the increased inequality in pre-government earnings. These different developments are most likely explained by the generous Swedish welfare system. Consistent with these observations, we see no clear trend in consumption inequality.

We also estimate stochastic processes for household earnings. A simple random-walk process captures much of the life-cycle dynamics. But we find clear evidence that the true earnings process is not a random walk. We demonstrate that some estimation methods result in severe upward bias in the estimated volatility of permanent shocks if serial correlation in temporary shocks is ignored.

Our estimation results show that the increase in earnings inequality is almost entirely driven by an increase in *residual* earnings inequality. Moreover, this increase was mostly generated by an increased volatility of persistent shocks.

JEL classification: **D31, D33, E24, J31**

Keywords: Income inequality, consumption inequality, stochastic earnings process

---

*We thank the editors, two anonymous referees, Paul Klein, Roine Vestman, and participants at the PIER IGIER conference 2007 for helpful comments. We also thank the Swedish Retailing Institute for providing data. Financial support from the Jan Wallander and Tom Hedelius Foundation at Svenska Handelsbanken and the Swedish Research Council is gratefully acknowledged.

[§]Department of Economics, Stockholm School of Economics, Box 6501, SE-113 83 Stockholm, Sweden, david.domeij@hhs.se.

[¶]Department of Economics, Stockholm University, SE-106 91 Stockholm, Sweden, and CEPR, martin.floden@hhs.se.

1 Introduction

This paper documents trends in Swedish earnings, income and consumption inequality between 1978 and 2004. We document a clear increase in Swedish earnings inequality in the early 1990s. Inequality in disposable income and earnings net of taxes and transfers also increased, but much less than the increased inequality in pre-government earnings. We see no clear trend in consumption inequality.

From a welfare perspective, it is important to understand why earnings inequality rose. For example, the policy consequences may be very different if the inequality was generated by changes in the returns to observable characteristics, such as education, rather than by changes in unobservables, such as luck. The welfare implications also depend on changes in the life-cycle dynamics of earnings. In particular, increased volatility of temporary earnings shocks can more easily be insured than increased volatility of persistent shocks, and therefore have less important welfare implications.

We first document that the increased inequality was almost entirely driven by an increase in unobservables, i.e. in *residual* earnings inequality. To further investigate the causes behind the increased inequality, we estimate stochastic earnings processes. Based on these estimates, we document that the rise in residual inequality was mostly generated by an increased volatility of persistent shocks.[1] Our finding that inequality in earnings net of taxes and transfers has increased much less than inequality in pre-government earnings suggests that the Swedish welfare state has been effective in providing insurance. Further support for this is that although the persistent shocks are relatively difficult to insure against at the household level, consumption inequality has *not* increased.

The paper also makes a methodological contribution. Household earnings is often conveniently assumed to follow a random-walk process. That process provides a parsimonious specification of the income process, and is therefore useful in applied work. We show that the random-walk specification captures much of the Swedish earnings process when we also allow for serial correlation in the temporary shocks. The stochastic income processes are typically estimated using some minimum distance method. Parameters in the earnings process are then chosen so that some moments implied by the process are close to the corresponding empirical moments. If the true earnings process is a random walk, the choice of moments to focus on is of little importance. But we demonstrate that the choice of moments is of crucial importance for the resulting parameter estimates if the true earnings process deviates slightly from a random walk. More specifically, our results indicate that it is important to use moment conditions that explicitly describe how earnings inequality evolves over the life cycle and over time.

A number of previous papers have studied trends in Swedish inequality. There is ample evidence that Swedish income inequality fell substantially between 1970 and 1990. Most of

[1] Gustavsson (2007) estimates a similar earnings process but focuses on male earnings and uses only data for 1991-1999. In line with our results, he finds that most of the rising inequality is explained by an increased volatility of persistent shocks. But Gustavsson's results are not identical to ours. Whereas we find that the variance of persistent shocks increased from the 1980s to the 1990s, we do not find an increase between 1991 and 1999. This difference may be explained by Gustavsson's short time period, and his assumption that shock prior 1991 had the same variance as those in 1991.

the Swedish wage compression occurred in the 1970s and was mainly a result of narrowing age, education and gender wage differentials.[2]

Our finding that Swedish earnings inequality increased in the 1990s confirms previous findings in Johansson (2006) and Björklund and Freeman (2008). The rise in inequality during the 1990s did, however, not fully offset the falling inequality during the 1970s. Johansson (2006) shows that the Gini coefficient for family income in 2002 was clearly below its 1970 value and much below its 1950 value. Nordström Skans et al. (2006) finds that much of the rising wage inequality during the 1990s can be attributed to rising wage dispersion between firms, and Gustavsson (2006) finds evidence that the returns to observable individual-specific qualities contributed to higher inequality. Domeij (2008) finds that much of the rise in residual earnings inequality was generated by changes in the industry composition and educational patterns.

A number of developments in the Swedish economy may explain why the falling inequality during the 1970s was reversed during the 1990s. First, the Swedish wage-setting process was reformed in several steps. The system with centralized wage bargaining broke down in the mid 1980s and was replace by a higher degree of industry-wide bargaining. In the early 1990s, the public sector went from grade-based wage setting to individual wage setting. Starting in the mid 1980s and accelerating in the early 1990s, a number of markets were deregulated and became more competitive, which made wage compression more difficult to achieve. Possibly, these trends were reversed in 1997 after the Agreement of Industry Development and Wage Formation was signed. This agreement stipulated common guidelines for wage setting across industries.

Second, several macroeconomic imbalances were accumulated during the 1980s. A number of devaluations of the krona in the late 1970s and early 1980s was followed by a decade with a fixed exchange rate but high inflation. The rapid expansion of the public sector in the 1970s had arguably also resulted in artificially low unemployment.[3] The second half of the 1980s was characterized by a rapid credit expansion and rising asset prices as a consequence of deregulated domestic credit markets and a tax system that resulted in low or negative real interest rates after tax. This process cumulated with a major crisis in the early 1990s. The most direct causes of the crisis were a deregulation of foreign exchange markets in 1989 in combination with rising world interest rates and a Swedish tax reform that further raised the after-tax real interest rate. Between 1990 and 1993, the unemployment rate increased from 2 to 11 percent, the employment rate fell from 83 to 73 percent, GDP per capita fell by more than 6 percent, and the public debt increased from 43 to 72 percent of GDP. The second half of the 1990s was characterized by high GDP growth and fiscal consolidation. In spite of the thriving economy, the unemployment and employment rates did not return to their pre-crisis levels. For example, at the next peek in 2001 the unemployment rate was 6 percent and the employment rate was 75 percent.

Finally, following the crisis Sweden let the currency float and implemented inflation targeting in 1992-1993, joined the European Union in 1995, and reformed the fiscal framework starting in the mid 1990s. These reforms had major impact on the Swedish economy, although it is unclear to what extent they contributed to changes in income inequality.[4]

[2] See for example Edin and Holmlund (1995).

[3] Lindbeck (1997) suggests that the public sector acted as "employer of last resort".

[4] The fiscal consolidation affected the generosity the welfare systems. For example, the replacement

Our study uses several different data sources. A common feature of these sources is that all income data is based on tax registers. We consequently avoid problems associated with top coding, incomplete recall, and biased and low response rates. We instead face another potential problem. A major tax reform ("the tax reform of the century", see Agell et al., 1996), was implemented in 1990-1991. The tax reform aimed at broadening tax bases and reducing marginal tax rates. Following this reform, some items in the tax registers were reclassified. This led some researchers to, in particular Edin and Fredriksson (2000), to challenge the comparability of pre- and post-reform income data. Böhlmark and Lindqvist (2006) on the other hand demonstrate that although comparability in principle could be questioned, the problems are minor in practice. We carefully try to compose income definitions that are comparable over time, and we do not see important breaks in our time series around the tax reform.

The paper is structured as follows. The next section describes the data, our sample selection procedures, and compares the implied aggregate time series to the national accounts. Section 3 documents facts for inequality in earnings, income, and consumption over time. Section 4 documents similar facts for inequality over the life cycle. In Section 5, we estimates stochastic earnings processes. Section 6 concludes.

2 Data and sample selection

We mainly use data from three different data sets. Our main source of income and earnings data is LINDA (Longitudinal Individual Data for Sweden), while we use HINK (Household Survey on Income) for hours, wages, wealth, and HUT (Household Expenditure Survey) for consumption.

2.1 The LINDA database

LINDA is a register-based longitudinal data set compiled by Statistics Sweden from the Income Register based on filed tax reports, the Census (in 5 year intervals from 1960 to 1990), and other registers. It consists of a large panel of individuals and their family members (as defined for tax purposes), resulting in approximately $300,000$ individuals per year.[5] We use data from 1978 to 2004. As household head, we use the oldest adult male in the household.[6] In households without adult males, the oldest female is defined as the head. For each household we record the sex and age of the head together with the number of adults, children, and consumption equivalents in the household.[7] From 1990 and on, we also have some information about education. We construct three education classes: less than high-school, high-school graduates, and some college education.

rates in most social security systems were reduced from 90 to 80 percent in the early 1990s. This may have affected inequality measures for disposable income and post-government earnings.

[5] See Edin and Fredriksson (2000) for a further description of the LINDA data set.

[6] Men between 17 and 24 are classified as adults if the oldest woman in the household is less than 18 years older than the man.

[7] We use the OECD consumption equivalence scale. The number of consumption equivalents is then 1 for the first adult, plus 0.7 times the number of additional adults, plus 0.5 times the number of children (aged 0-16).

We calculate pre-government earnings (y) and disposable income (y^D) for 1985-2004 as the measure of labor income suggested by Statistics Sweden (2006) to be comparable between years in LINDA. Pre-government earnings consist of wages and salaries, the part of business income reported as labor income, and taxable compensation for sick leave and parental leave, while disposable income consists of the sum of factor income and positive transfers minus taxes and negative transfers. We construct comparable measures for 1978-1984.[8] We calculate financial capital income (y^A) as the sum of three components, each restricted to be non-negative. The first component is the sum of net interest income, net dividends, and net realized capital gains. The second component is income from roomers and boarders, and the third component is 35 percent of net business income. Finally, we calculate post-government earnings (y^T) as the difference between disposable income and capital income. Since taxed paid on capital income constitute a small part of total tax payments, and since we cannot separate capital taxes earnings taxes, we assume that all taxes are labor income taxes. Nominal values were converted to 2004 SEK using the CPI.

We use three broad sample selection criteria for the LINDA data. First, when comparing the aggregates in LINDA to data in the national accounts, we include all sampled individuals but drop other household members. We refer to this as the *NIPA sample*. Second, when analyzing inequality trends over time and life-cycles, we only keep households where the head is between 25 and 59 years old and where the head was sampled by LINDA. We refer to this as the *broad sample*. Third, when estimating earnings processes we focus on households with a strong attachment to the labor market. In particular, there is no information about hours worked in LINDA. Consequently, we cannot condition the sample on labor supply. We instead remove households with low earnings. The effective hourly minimum wage in Sweden in 2004 was around SEK 75 (Skedinger, 2007), and we exclude households with earnings less than half this minimum wage multiplied by 160 hours and 12 months.[9] For other years, we adjust the minimum wage by calculating the mean real earnings for each year, estimating a linear time trend for these means and removing that time trend from the SEK 75 minimum wage.[10] Moreover, the estimation approach requires that we can follow households over time, and we have a maximum of 26 earnings differences on a single household. We include households where we have information on n earnings differences. We refer to this as the *estimation sample*. In the benchmark estimation sample we require $n = 2$ differences but we also consider $n = 20$. The first columns in Table 1 show how earnings and income has developed over time in our broad sample.

We are often interested in measures of residual inequality. When using the LINDA data, we calculate residual income (ε^y) by running year-specific regressions of log income against a complete set of age dummies (D^a) together with the number of consumption equivalents (E) and family dummies (D^f). There are 15 family dummies, indicating single men, single women, and couples with 0, 1, 2, 3, and 4+ children. The regression we run is then

$$\ln y_{i,t} = D^y_t + \beta_{1,t} D^a_{i,t} + \beta_{2,t} D^f_{i,t} + \beta_{3,t} E_{i,t} + \varepsilon^y_{i,t} \tag{1}$$

where D^y are year dummies. In some specifications we also control for the education levels

[8] STATA code is [will be] available on our web pages.

[9] Note that when estimating processes for post-government earnings we use post-government (rather than pre-government) earnings for the selection criteria.

[10] This method implies a minimum wage around SEK 60 in the 1980s, consistently with values reported by Skedinger and used in our analysis of wages in HINK below.

reported between 1990 and 2004. We then run the regression

$$\ln y_{i,t} = D_t^y + \beta_{1,t} D_{i,t}^a + \beta_{2,t} D_{i,t}^f + \beta_{3,t} E_{i,t} + \beta_{4,t} D_{i,t}^e + \varepsilon_{i,t}^y \tag{2}$$

where D^e are dummies indicating less than high-school, high-school graduates, or some college education.

2.2 The HINK database

The HINK data set is a revolving panel with $10,000 - 20,000$ households per year, based both on registers and telephone surveys conducted by Statistics Sweden. Sampling occurs at the individual level, all household members of the sampled individual are included, and the sampled individual stays in the sample for two consecutive years.[11] We have data for 1975, 1978, and 1980 to 1992. To construct the equivalent to the broad sample in LINDA, we define the household head as in LINDA, and we include households where the household head is between 25 and 59 years old. To obtain a representative sample of households we weight households by household size.[12]

Pre-government earnings, post-government earnings, and disposable household income are defined in accordance with the definitions in LINDA, except that the earnings measures now include two thirds of business income instead of the reported labor part of business income. We use two measures of wealth. Total net wealth (a^+) is calculated by Statistics Sweden using estimated real estate values together with information on other assets. Financial net wealth (a) is calculated as total net wealth minus the estimated real estate values net of mortgages.

Annual hours (l) are the reported number of hours worked per week times the reported number of weeks worked. The hourly wage (w) are then calculated as the respective individual's pre-government earnings divided by the number of hours worked. For calculations based on wage data, we only include individuals where the hourly wage is above half the minimum wage. This sample, which we refer to as the *wage sample*, roughly corresponds to the LINDA estimation sample. According to Skedinger (2007) the minimum wage in the service sector was around SEK 60 (in 2004 SEKs) throughout the 1980s. To examine the robustness of the results, we also consider a more narrow wage sample where we require 900 annual hours and exclude students, self-employed, farmers and non-classified individuals. The final columns in Table 1 show how average wages, hours, and wealth have developed over time in our sample.

When using the HINK data, we calculate residual wages by running year-specific regressions of wages against a quartic polynomial in age, the number of consumption equivalents, and family dummies. There are 5 family dummies, indicating single men without children, single women without children, couples without children, singles with children, and couples with children.

[11] Non-married cohabitants without children form a household in HINK but two separate households in LINDA. Furthermore, a maximum of two adults are included in a HINK household, but LINDA does not have that restriction.

[12] The weights are N/n where N is the number of households in the sample and n is the number of household members that could have been sampled.

2.3 The HUT database

The HUT data set is a cross-sectional survey carried out by Statistics Sweden. A representative cross-sectional sample of Swedish individuals between ages 0 and 74 years is chosen, and all household members of the sampled individual are included in the survey. We use data from the surveys in 1985, 1988, 1992, 1996, and 1999. To construct the broad sample, we weight households as in the HINK data, we use the same definition of the household head, and we drop households with reported food consumption less than 10 SEK (2004 prices) per consumption equivalent and day. This results in around $2,000$ remaining households in the first years, and about half as many in the later years when the sample size was reduced.

The data collection in HUT consists of several steps. An initial interview is first conducted about the household composition. Second, the household is asked to keep book of all expenditure during a two-week period. During this period the household records all expenditure of all household members.[13] In connection with the bookkeeping a questionnaire is filled out concerning the expenditure during the previous year for: housing, holiday cottages, petrol, insurance, traveling abroad, and the purchase of certain durable goods. Expenditure on clothing, shoes and traveling within Sweden is collected in the same way for the previous two-month period.

We calculate non-durable consumption (c) as the measure of total consumption expenditure minus rent paid for tenants, mortgage payments, repairs, and vehicle purchases. The measure of consumption including services from housing (c^+) is calculated as non-durable consumption plus rents for tenants, mortgage payments, and repairs. Disposable income was calculated in HUT by linking the person to the Income Register based on filed tax reports. Disposable income concerns the year of the survey and consists of the sum of factor income and all transfer income (e.g. pension payments, unemployment benefits, paid sick-leave, housing assistance, etc.) net of taxes. Table 1 reports average consumption according to these different measures for each of the sample years.

When using the HUT data, we calculate residual consumption by running year-specific regressions of consumption against the number of consumption equivalents and age and family dummies. We use dummies for the age groups $25-33$, $34-42$, $43-51$, and $52-59$.[14] The family dummies are the same as in the HINK data.

2.4 Comparison with other sources

Figures 1-3 show the aggregate implications from our data together with the aggregates reported in the national accounts.[15] Figure 1 shows that per capita labor income in the LINDA database closely tracks the wage sum reported in the national account. The income measures are however consistently higher in LINDA than in the national accounts. The

[13] An equal number of households start the bookkeeping every week during the survey year.

[14] We use different methods to control for age in the three datasets because of the difference in sample sizes.

[15] For these comparisons, we include all households or individuals in the datasets. That is, we do not focus on household heads or households with heads in working age.

largest difference (8 percent) is recorded in 1991, which is the year after the "tax reform of the century" where we would anticipate some measurement problems and inconsistencies.[16] For recent years the difference is between 1 and 3 percent.

Figure 2 shows that the consumption data in HUT differ more substantially from the values reported in the national accounts. Per capita consumption in HUT is consistently lower than in the national accounts, and the difference varies between 5 and 10 percent. One possible explanation for the different averages between the national accounts and HUT is that HUT only samples individuals aged $0 - 74$, so that the consumption of old individuals is underrepresented. It is however unlikely that the consumption of these old individuals would raise the average substantially. Another possible explanation is that housing related expenditure is defined differently in the national accounts and HUT, and they are consistently reported to be smaller in HUT. Excluding these housing related expenditures, the difference between average consumption in the national accounts and HUT falls to between 1 and 5 percent.

The total hours worked relative to the population aged 25 to 59 according to the HINK database and the labor force surveys are reported in Figure 3. The overall pattern, with increasing hours in the 1980s and sharply falling hours in the early 1990s, is similar for the two series. But the series for hours in HINK is less reliable than in the labor force survey, which is reflected by a more volatile series in the figure. One problem with the HINK survey is that it asks about normal hours worked, disregarding vacation and sick leave. Many individuals consequently report that they work 40 hours per week and 52 weeks per year although all workers have at least four weeks of vacation and more than five percent of workers were on sick leave for the full week in a typical week in this time period. This generates an upward bias in average hours in HINK. The fraction of individuals reporting exactly 52 work weeks increased in the mid 1980s, which possibly explains why the average in HINK rose above that in the labor force survey in these years.

Another problem with the HINK data that presumably generates a downward bias is missing values. In some years missing values were set to zero in the survey. To be consistent we recode all missing values to zeros. It is clear that most of these missing values are actually from individuals that did not work (the fraction of missing values is for example much higher for children and retirees than for prime-aged individuals), but there will be a downward bias if true missing values exist and are treated as zeros. Between 10 and 16 percent of individuals aged 25 to 59 report positive labor income but either zero hours worked or have a missing value, indicating that such true missing values actually do exist. This fraction of missing values also varies between the years, and unusually many missing values in 1983 explains much of the fall observed fall in average hours in HINK by around 10 percent between 1982 and 1983. Reassuringly, these problems with missing values seem to mostly relate to groups of individuals who are not included in our analysis of wage inequality below. When we exclude students, retirees, conscript soldiers, self employed, farmers, and non-classified individuals the series for hours is smooth.

Table A1 in the Appendix compares the time series for equivalized earnings and disposable income from the different data sets. The values are mostly consistent. Table A2 further shows that our three data sets, LINDA, HUT, and HINK, are similar in terms of average

[16] Agell et al. (1996) summarize the important aspects of the tax reform.

age of the household head, household size, and number of consumption equivalents.

3 Time series facts

This section documents how inequality in wages, hours, income, consumption and wealth has developed over the last three decades. We first analyze how inequality has changed at the individual level by focusing on hours and wages. Figure 4 shows a clear fall in *inequality in hourly wages* between 1975 and 1992 in the wage sample where we include all individuals aged 25 to 59 with positive hours and a sufficiently high wage. Both panel (a) and (d) show a fall in overall wage inequality. The fall in inequality occurred both at the top of the distribution (panel b) and at the bottom of the distribution (panel c). All these developments are robust to further restrictions on the sample, for example by requiring 900 annual hours and excluding students, self-employed, farmers and non-classified individuals, although the levels of inequality are lower in the more narrow sample.

This fall in inequality is consistent with the development found by Edin and Holmlund (1995) up until the mid 1980s. Based on other selection criteria (for example focusing on a broader age group) they however find that wage dispersion according to all these measures increased somewhat during the latter half of the 1980s. Although we find larger levels of inequality than Edin and Holmlund, inequality around 1990 was low in an international comparison.[17]

To understand why the wage inequality fell, we examine how the gender premium and the returns to education and experience has evolved. Panel (a) in Figure 5 shows that the fall in wage inequality between 1975 and 1992 partly can be explained by the substantial fall in the education premium. This development is in line with Domeij and Ljungqvist (2009). Using census data and tax records covering the total population, they show that the skill premium fell by almost 30 percent between 1970 and 1990.[18] Panels (b) and (c) however show that there were no clear trends for the gender or experience premia in the wage sample. But in the narrow wage sample, where we focus on those working at least 900 hours per year, we see a clear fall in the gender premium from around 40 percent in the late 1970s to just below 30 percent from the early 1980s and on. That is, the gender premium appears to have fallen for those with a strong labor-market attachment. Panel (d) reports the evolution of residual wage inequality, i.e. not explained by education, gender, experience or family composition. This plot shows that also changes in this residual component contributed to the falling inequality. Panels (a) to (c) in Figure 5 also report the evolution of the wage premia along the education, gender and experience dimensions according to LOUISE data during the 1990s.[19] The education premium appears to have

[17] See for example the other papers in this volume.

[18] Edin and Holmlund (1995) also document a fall in the skill premium until the early 1980s using another dataset. Their data indicate that the skill premium was relatively flat, or possibly increasing, in the second half of the 1980s.

[19] LOUISE (Longitudinal Education, Income, and Employment Data for Sweden) is a register-based longitudinal data set compiled by Statistics Sweden from the Income Register based on filed tax reports for 1990 and on, the 1990 Census, and other registers. It basically covers the full population aged 16 and above. We however only have LOUISE data on an individual level and consequently cannot use this database for the household variables that are the main focus of this paper. A shortcoming of the LOUISE

increased by around 10 percent in the 1990s, which is in line with Gustavsson (2006) and Domeij and Ljungqvist (2009). The gender premium appears to have fallen somewhat during the 1990s. The experience premium on the other hand first increased during the crisis in the early 1990s, but then fell.

We next examine trends in wages and hours worked for men and women separately. Figure 6 first shows that the fall in wage inequality during the 1980s documented in Figures 4 and 5 applies both to men and women, but that the fall was slightly larger for men than for women. There is also a clear fall in the *dispersion of hours* among women but not among men. The level of dispersion is, not surprisingly, much lower when focusing on the narrow wage sample that only includes those who worked at least 900 annual hours. There is still a fall in the variance of log wages among women, but then from 0.08 to 0.06 rather than from 0.6 to 0.2. The dramatic fall in the dispersion of hours worked among women seen in panel (b) in Figure 6 is most likely explained by the trend increase in the fraction of women working close to full time. In particular, among women with positive hours, the fraction working at least 900 hours increased from 82 percent in 1980 to 92 percent in the early 1990s.

Panels (c) and (d) in Figure 6 show that there are no trends in the correlations between wages and hours. Table 2 shows that this correlation on average was -0.24 for men and -0.23 for women between 1978 and 1992. This table also shows a strong positive correlation between head and spouse hours. The correlations are much smaller, around -0.03, in the narrow sample but there is still no time trend. This indicates that the clearly negative correlation in the benchmark sample was generated by some individuals reporting few hours of work but with relatively high wages.

We now turn to analyzing trends in inequality at the household level. Figure 7 documents a clear increase in *earnings inequality* in the first half of the 1990s. Panel (a) shows similar patterns and magnitudes for the development of the variance of raw, equivalized, and residual earnings. This indicates that most of the increased inequality is driven by an increase in residual inequality. Panel (b) consequently shows that the inequality explained by the age and family components has been relatively stable. Controlling for education in addition to age and family composition has little impact. Only a small part of inequality is explained by education, and there are no clear trends in the educational component. During the Swedish macroeconomic and banking crisis in the early 1990s, unemployment increased from 2 to 11 percent and the employment rate fell from 83 to 71 percent. Panel (a) shows that changes in inequality largely coincide with changes in unemployment. Consequently the increase in unemployment and lower participation resulted in a dramatic increase in the dispersion of labor income in the early 1990s.

Although changes in households' labor market status are of key importance for understanding the overall inequality trends, it is also interesting to understand the trends among those who work. Figure 8 shows that inequality increased across the whole distribution, and thus that the higher unemployment does not explain all of the higher inequality. In-

data is that while it contains data on hours worked for 1990, we have to extrapolate that information for the years 1991-2002. Specifically, we assume that the average hours of work for detailed demographic groups (age×gender×education×industry) is given by its 1990 value multiplied by a year×age×gender specific factor, such that average hours worked in age×gender groups are consistent with Statistics Sweden's estimates for the period 1991 to 2002.

equality at the top shows a clear increase during the last 15 years as displayed in panel (b). Moreover, the unemployment rate and most measures of inequality fell in the late 1990s, but inequality at the very top has continued increasing as indicated by the p99/p90 ratio.[20] Another way to explore the development of inequality among those who work is to exclude households with little labor income. As a result of the economic crisis, a large fraction of the households had no labor income. For example, panel (a) in Figure 9 shows that earnings for the 10th percentile was at or close to zero from the mid 1990s and on. This explains the development for the p50/p10 ratio in the bottom-left panel in Figure 8. Furthermore, panel (b) in Figure 9 shows that an increasing fraction of households do not have enough earnings to satisfy the selection criteria for inclusion in the benchmark estimation sample. The bottom panels in Figure 9 report inequality measures based on this benchmark estimation sample. Not surprisingly, the cross-sectional variance of earnings is much lower when focusing on households with a strong attachment to the labor market. The magnitude of changes in the variance is also much lower, although the time trend displays a similar pattern as for the broad sample in Figure 8.[21] Interestingly, however, both the levels and trends in the p99/p90 and p90/p50 ratios in panel (d) in Figure 9 are remarkably similar to those in the broad sample displayed in panel (b) in Figure 8. This suggests that measures of inequality at the top are relatively insensitive to selection criteria, and possibly that the top tail of the earnings distribution is well approximated by a Pareto distribution.

Figure 10 and Table 3 show that the Swedish welfare system has moderated the effects on inequality throughout the period, and in particular during the turbulent 1990s. The variances of post-government earnings, i.e. earnings after taxes and transfers, and disposable income are much lower than the variance of pre-government earnings. Furthermore, inequality in post-government earnings remained remarkably stable when unemployment and pre-government earnings inequality increased dramatically in the 1990s. Inequality in disposable income has however increased. The different developments for post-government earnings and disposable income indicate an increased dispersion of capital income. Table 3 indeed shows that inequality in capital income was higher in the more recent time period according to all measures. This is also in line with evidence in Roine and Waldenström (2009). Although there is an upward trend in disposable income inequality, the early 1990s and the early 2000s stand out as periods of more rapid change. These periods coincide with falling asset prices. A possible explanation for the increased dispersion of capital income is therefore that some households have chosen to realize (big) losses in these years, an issue that we however have not explored in this paper.

Figures 11-13 show various measures of *consumption inequality*. There are no clear trends in these figures but a small increase in consumption inequality cannot be ruled out. This trend is however less pronounced than the increasing trend in inequality in disposable income. A possible explanation for these different trends is that savings increased among high-income households. Table 4 indeed shows that the correlation between disposable income and savings has increased over time. A closer look at the data reveals that the increased correlation is driven by high-income households. The savings ratio for households

[20] In a more detailed examination of top incomes in Sweden, Roine and Waldenström (2009) document a clear increase in inequality at the top in the 1990s.

[21] One difference is the development of the variance of log earnings during the 1980s. This variance increased in the broad sample in Figure 8 but falls in the narrow sample in Figure 9. This fall is consistent with the falling dispersion of wages and hours in Figure 6, which also builds on a narrow sample.

with disposable income above the median increased from 27 to 37 percent from the 1980s to the 1990s. If we include housing-related expenditures in the consumption measure, the savings ratio instead increased from 6 to 16 percent. That the increase is ten percentage points in both cases indicates that non-housing related savings increased among high-income households. The savings ratio was relatively constant for low-income households.

We finally examine the time trends in *wealth inequality*. Panels (a) and (b) in Figure 14 indicate that both the level of financial wealth and total wealth fell relative to disposable income up until 1987. This development has previously been documented by Statistics Sweden (2000) but is still remarkable considering that the stock market increased by almost than 600 percent in nominal terms between 1980 and 1987 (including the stock-market crash in October 1987). Equity however only accounted for around 10 percent of households' financial wealth in the 1980s (Werin, 1993). The falling wealth ratio can instead partly be explained by house prices falling in relative terms in the first half of the 1980s, as can be seen in panel (b) in Figure 14. Moreover financial liberalization during the 1980s contributed to a credit expansion that resulted in a low private savings rate. The households' net savings ratio was on average 8 percent during the 1970s and for the first half of the 1980s, but then fell sharply to an average less than three percent during the latter half of the 1980s. The credit expansion however contributed to strongly appreciated asset prices in 1988-1990. Residential house prices for example rose by around 30 percent in real terms, and commercial property prices rose even faster. This explains why the wealth-to-income ratios increased in spite of the low savings ratios. This upward trend came to a halt with the economic crisis in the early 1990s. Calculations from Statistics Sweden indicate that the wealth-to-income ratio has risen in recent years.

The lower panels in Figure 14 show the Gini coefficients for net financial wealth and net total wealth. Inequality increased for both measures during the 1980s. The level of inequality as measured by the Gini coefficient is high in an international comparison. The high Gini coefficients are mostly generated by a large fraction of households reporting no or negative net private wealth (see Table 3). To some extent this is a result of measurement errors. In particular condominiums tend to be systematically undervalued. But these measurement problems are not quantitatively very important, and the high fraction of households with little private wealth is indeed a reality in the Swedish data.[22] A likely explanation for this is that the comprehensive pension system and generous social insurance systems reduce the need to accumulate private wealth for life-cycle and precautionary reasons (Domeij and Klein, 2002). Further supporting the importance of the Swedish welfare system is the observation that the correlations between wealth and various income and consumption measures are remarkably low, as shown in Table 4.

[22] From 1997 and on, Statistics Sweden adjust for most of these measurement problems. The Gini coefficients reported for that time period in panel (d) in Figure 14 are still high, indicating that these measurement problems were not a severe problem.

4 Life-cycle profiles

This section explores how inequality evolves over an individual's or household's life cycle. To examine the inequality for some variable x over the life cycle, we calculate

$$\sigma^2_{h,s} = \mathrm{var}\left(\ln x_{h,s}\right)$$

where h denotes an age group and s a year or cohort. We then regress this variance against age and year or cohort dummies,

$$\sigma^2_{h,s} = \beta_0 + \beta_1 D^h_{h,s} + \beta_2 D^s_{h,s} + \varepsilon_{h,s}. \tag{3}$$

Figure 15 and 16 report the life-cycle profiles $\beta_1 D^h_{h,s}$ for wages, earnings, and consumption when we control for year and cohort effects, respectively.[23] The life-cycle profiles for *wage inequality* are completely different in these two figures. Inequality increases over the life-cycle when we control for time effects but falls when we control for cohort effects. A closer examination of the life-cycle profiles of inequality for each cohort clarifies these apparently conflicting observations. This examination reveals that for each cohort, except the very oldest, the variance falls over the life-cycle, but also that inequality is lower for younger cohorts.[24] Moreover, inequality within the old age groups was substantially higher than within the young age groups in the late 1970s, but then fell during the 1980s so that no difference between age groups remained in the early 1990s. This fall in inequality among the old explains the downward trend in wage inequality in Figure 4. A possible explanation for this development is that egalitarian wage policies mostly affect new entrants on the job market, and that they because of wage rigidities have little effect on older workers. These policies may have compressed wages among young workers already in the 1970s. But the impact on inequality among older age groups was delayed until the younger workers with compressed wages aged. This discussion indicates that different cohorts were affected by wage compression at different points in time, and thus that controls both for time and cohort effects would be relevant. Although we see no obvious approach to sort out which controls result in the most representative life-cycle profile, we note that the explanatory power of the cohort dummies is clearly better, possibly indicating that the downward-sloping profile is most representative.

Also the life-cycle profiles for *earnings inequality* in panels (b) and (c) differ when we control for time and cohort effects. Contrary to the previous discussion about wage inequality, we argue that several observations indicate that time effects rather than cohort effects are important for capturing the earnings inequality. First, and most important, we now consider a longer time period including the major crisis that the Swedish economy went through in the early 1990s. This crisis simultaneously affected all cohorts as seen in panel (a) in Figure 17, which suggest the importance of allowing for time effects when studying this longer time period. Second, the explanatory power of the time dummies

[23] For the wage regressions in HINK, we use dummies for age groups $25-29$, $30-34$, ..., $55-59$, and cohorts $1916-1929$, $1930-1934$, $1935-1939$, ..., $1950-1954$, $1955-1967$. For the earnings regressions in LINDA, we use a complete set of age and cohort dummies but we exclude cohorts with five or fewer observations (i.e. $1919-1923$, and $1975-1979$). For the consumption regressions in HUT, we use dummies for age groups $25-33$, $34-42$, $43-51$, $52-59$, and cohorts $1926-1939$, $1940-1944$, $1945-1949$, $1950-1954$, $1955-1959$, $1960-1974$.

[24] See Domeij and Floden (2009) for further information on these observations.

is much better than that of the cohort effects in our regression. Third, the time effects estimated from (3) closely follow the Swedish unemployment rate, as is shown in panel (b) in Figure 17. The correlation between the time effects and the unemployment rate is 0.92. We have also added the unemployment rate u when controlling for cohort effects. We then estimate the life-cycle profile from

$$\sigma^2_{h,k} = \beta_0 + \beta_1 D^h_{h,k} + \beta_2 D^k_{h,k} + \beta_3 u_{k+h} + \varepsilon_{h,k}$$

where k denotes the birth year of a cohort. The results are reported in panels (b) and (c) in Figure 16, and show that the life-cycle profiles then become flatter and somewhat more similar to those obtained when controlling for time effects.

There are however some indications that also cohort effects might have been important. Panel (a) in Figure 17 for example shows that inequality increased particularly at young ages during and after the crisis. That all cohorts were affected suggests that time effects are important, but the larger increase for young households indicates the presence also of cohort effects. The importance of the crisis for understanding the life-cycle profiles of inequality is further seen when splitting the sample into pre and post crisis periods. Whether we control for time or cohort effects then has little impact on the estimated life-cycle profiles of inequality (see panels (b) and (c) in Figure 16).

This analysis clearly indicates that the crisis in the early 1990s had major impact on the overall inequality trends in Sweden. Indeed, the increase in unemployment and transitions out of the labor force in the early 1990s are crucial for understanding the inequality trends. The lower panels in Figure 17 show inequality measures corresponding to panels (b) and (c) in Figures 15 and 16, but for the estimation sample that focuses on households with a strong attachment to the labor market. For this sample, it is of little importance whether we control for time or cohort effects. There are three further observations that stand out from these graphs. First, the magnitude of inequality is considerably smaller. The major part of inequality in the broad sample is thus explained by households with low attachment to the labor market, and not by wage inequality among those who work. Second, the difference between the life-cycle profiles for raw and equivalized earnings inequality is much more evident when focusing on the estimation sample. Inequality increases more over the life cycle when considering raw rather than equivalized earnings. We have not explored the sources of this difference, but a candidate explanation is that in households with two adults, it is more likely that both of them work at older ages when they do not have young to take care of.

5 Earnings dynamics

We now turn to estimating stochastic processes for earnings over the life-cycle. There are two main reasons for our interest in these processes. First, parsimonious specifications of such processes are important inputs in many macroeconomic models. Second, these processes are an important tool for understanding the dynamics of inequality over the life cycle. By allowing us to identify the persistence of shocks they help us analyzing the welfare implications of the rising earnings inequality that we documented in the previous sections.

The LINDA data set follows households over time and allows us to examine the dynamics of household earnings and income over time and over life cycles. We are in particular interested in the dynamics of residual earnings. We first consider the earnings process

$$\ln y_{i,h,t} = \mathbf{x}'_{i,h,t}\psi_t + \alpha_{i,h,t} + \beta_i + \varepsilon_{i,t}, \tag{4}$$

$$\alpha_{i,h,t} = \alpha_{i,h-1,t-1} + \eta_{i,t} \text{ if } h > 1 \tag{5}$$

$$\alpha_{i,1,t} = \eta_{i,t} \tag{6}$$

where $\mathbf{x}$ are household observables, and subscripts i, t, and h denote individuals, time, and age (starting at $h = 1$ at age 25), respectively.[25] Moreover, α is a permanent earnings component, β_i is a mean-zero individual-specific fixed effect with variance σ^2_β, $\varepsilon_{i,t}$ is a mean-zero temporary earnings shock with variance $\sigma^2_{\varepsilon,t}$, and $\eta_{i,t}$ is a mean-zero permanent earnings shock with variance $\sigma^2_{\eta,t}$.

We first estimate this process with GMM using moments based on log differences of the earnings residual. Let $g_{i,t} = \Delta\left(\ln y_{i,h,t} - \mathbf{x}'_{i,h,t}\psi_t\right)$ and note that the process (4)-(6) then implies that

$$var(g_t) = \sigma^2_{\eta,t} + \sigma^2_{\varepsilon,t} + \sigma^2_{\varepsilon,t-1} \tag{7}$$

and

$$cov(g_t, g_{t+s}) = \begin{cases} -\sigma^2_{\varepsilon,t} & \text{if } s = 1 \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

When using these moments based on log differences, estimation of the permanent shocks η in the process (4)-(6) hinges crucially on the assumption of no serial correlation in the temporary shocks ε. To see this, consider the process estimated by Blundell et al. (2008) on U.S. data. They estimate the process

$$\ln y_{i,t} = \mathbf{x}'_{i,t}\psi_t + \alpha_{i,t} + \varepsilon_{i,t} + \theta\varepsilon_{i,t} \tag{9}$$

$$\alpha_{i,t} = \alpha_{i,t-1} + \eta_{i,t} \tag{10}$$

and find $\sigma^2_\varepsilon = 0.0415$, $\sigma^2_\eta = 0.0102$, and $\theta = 0.1132$.[26] Suppose that this is the true data generating process. An econometrician assuming $\theta = 0$ and using equations (7)-(8) to estimate the process will then find $\hat{\sigma}^2_\eta = 0.0196$ rather than the true value $\sigma^2_\eta = 0.0102$.[27] Ignoring the moving average term thus results in a substantial upward bias of the variance of permanent shocks even when the moving error parameter is small.

It is often important to use parsimonious specifications of the income process when calibrating macroeconomic models. Rather than allowing for moving-average terms in the

[25] The household observables in $\mathbf{x}$ are age and family dummies, and consumption equivalents as described at the end of Section 2.1.

[26] These estimates are for 1979. Blundell et al. report year-specific estimates for $1979-1992$. To simplify the following discussion we assume that variances are constant over time.

[27] This process implies that $var(g_t) = \sigma^2_\eta + 2\left(1-\theta+\theta^2\right)\sigma^2_\varepsilon$ and $cov(g_t, g_{t+1}) = -(1-\theta)^2\sigma^2_\varepsilon$. The econometrician would therefore set $\hat{\sigma}^2_\varepsilon = -cov(g_t, g_{t+1}) = (1-\theta)^2\sigma^2_\varepsilon$, and $\hat{\sigma}^2_\eta = var(g_t) - 2\hat{\sigma}^2_\varepsilon = \sigma^2_\eta + 2\theta\sigma^2_\varepsilon$. If the econometrician instead use the level moments (11)-(12), but still ignore the moving average term, the variances would approximately be estimated as $\hat{\sigma}^2_\varepsilon = \left(1+\theta^2\right)\sigma^2_\varepsilon$, and $\hat{\sigma}^2_\eta = \sigma^2_\eta$. Note that the estimated variance of the temporary shock then is (approximately) $\hat{\sigma}^2_\varepsilon = var(\varepsilon_t + \theta\varepsilon_{t-1})$. That is the variance of the permanent shock is not biased while what is actually temporary over two periods will be estimated as being temporary over only one period.

process, we therefore first consider alternative moment conditions for estimating the process (7)-(8).[28] More precisely, we use moment conditions similar to those in Guvenen (2009) and Heathcote et al. (2004, 2008). Let $e_{i,h,t} = \ln y_{i,h,t} - \mathbf{x}'_{i,h,t}\psi_t$. The information used to identify the process is then described by

$$var(e_{h,t}) = var(\alpha_{h,t}) + \sigma_\beta^2 + \sigma_{\varepsilon,t}^2, \tag{11}$$

and

$$cov(e_{h,t}, e_{h+s,t+s}) = var(\alpha_{h,t}) + \sigma_\beta^2 \tag{12}$$

where the variance for the persistent component α is

$$\text{var}(\alpha_{1,t}) = \sigma_{\eta,t}^2$$

$$var(\alpha_{h,1}) = h\sigma_{\eta,1}^2 \text{ if } h > 1$$

and

$$var(\alpha_{h,t}) = var(\alpha_{h-1,t-1}) + \sigma_{\eta,t}^2 \text{ if } h > 1, \text{ and } t > 1.$$

We describe in the appendix how we aggregate these moment conditions across individuals and how we use data to calculate the corresponding empirical moments. We use three alternative aggregation methods. The first method, suggested by Krueger et al. (2008) and used widely in the literature is summarized in Appendix A.1.[29] This method aggregates moments across individuals of different ages resulting in time series of average variances and covariances between the different years. Since we have data for 27 years this results in 378 moments. The second method is summarized in Appendix A.2. This method instead aggregates moments across years, resulting in life-cycle profiles of average variances and covariances between different ages. With 35 ages in the data, this results in 594 moment. The third method, used by Heathcote et al. (2008), maintains all time-series and life-cycle information and does not aggregate in any dimension. This results in $9,946$ moments.

Under all three methods, the moment conditions (11)-(12) use more explicit information about how inequality evolves over the life cycle than the moment conditions (7)-(8). One may therefore suspect that estimation based on (11)-(12) is more robust to serial correlation in the temporary shocks than (7)-(8).[30]

To highlight the different implications of using these different moment conditions and aggregation methods we initially constrain the variance of shocks to be constant over time. Allowing for time-variation in the variances does not change the results or conclusions

[28] We however consider richer processes below.

[29] Guvenen (2007) and Blundell et al. (2008) are recent examples.

[30] Indeed, if the econometrician use the level moments (11)-(12), but ignores the moving average term, the variances will approximately be estimated to $\hat{\sigma}_\varepsilon^2 = (1+\theta^2)\,\sigma_\varepsilon^2$, and $\hat{\sigma}_\eta^2 = \sigma_\eta^2$. Note that the estimated variance of the temporary shock then is (approximately) $\hat{\sigma}_\varepsilon^2 = var(\varepsilon_t + \theta\varepsilon_{t-1})$. That is the variance of the permanent shock is not biased while what is actually temporary over two periods will be estimated as being temporary over only one period. This also shows that the bias when using moments based on earnings levels will be small as long as the moving average term θ is small.

For example, when using Blundell et al.'s process (i.e. $\theta = 0.1132$) the bias in the estimate of σ_ε^2 will be just above one percent. When estimating their process on artificial data under the restriction that $\theta = 0$, we find $\sigma_\varepsilon^2 = 0.042$ and $\sigma_\eta^2 = 0.010$, which cannot be distinguished from the true process, $\sigma_\varepsilon^2 = 0.0415$ and $\sigma_\eta^2 = 0.0102$.

of the following analysis. Table 5 shows that the different estimation methods result in notable differences. In particular, estimates based on moments from earnings differences reported in the first column result in a large variance of permanent shocks and a small variance of temporary shocks, whereas the estimates based on moments from earnings levels result in the opposite. We will now examine the implications of these estimates and compare them to what we see in the data. From (11) we note that the cross-sectional variance of earnings across all households aged h is $var\left(e_h\right) = h\sigma_\eta^2 + \sigma_\beta^2 + \sigma_\varepsilon^2$. The cross-sectional variance thus increases linearly at rate σ_η^2 over the life cycle. The different estimates of the variance of permanent earnings shocks therefore have dramatically different implications for how earnings inequality develops over the life cycle. The estimates from the difference specification implies that the cross-sectional variance increases from 0.11 at age 25 to 1.47 at age 59 while the estimates based on levels in column IV imply an increase from 0.12 to 0.25. As displayed in panel (a) in Figure 18, the life-cycle profiles implied by the estimates based on earnings differences are counterfactual.[31] According to the data, the cross-sectional variance increases from 0.11 to 0.23 over the life cycle. This is instead very close to the implications of the process estimated from level moments, as reported in panel (b). We demonstrated above that serially correlated temporary shocks would result in an upward bias in the estimated variance of permanent shocks when using moments based on earnings differences. The existence of such serial correlation is therefore a plausible explanation for why the method based on earnings differences result in unrealistically high estimates of the permanent shock.

Because the estimation method that based on earnings differences selects parameters in the earnings process to match the dynamics of these earnings differences, it naturally also matches the empirical counterparts better than the estimation methods based on earnings levels. The variance and first covariance of earnings growth rates are on average 0.06 and −0.01 in the data. These averages are also matched perfectly by the difference method while the level method implies a variance of 0.14 and a covariance of −0.07 as is shown in column II in Table A3 in the Appendix. This upward bias in the variance of annual earnings growth is possibly generated by neglected serial correlation in the temporary shocks. As is shown in footnote 30, ignoring this serial correlation will result in an annual variance term that also contains the variance of the serially correlated terms. If matching the moments for annual earnings growth rates is important, one may give weight also to the difference moments in the estimation process. But we will see in the next subsection that a richer stochastic process fits better to the difference moments even when those moments are not used to estimate the process.

That the different estimation methods result in different results can, as our previous discussion indicates, only be explained by a misspecification of the earnings process. That is, the random-walk process is too restrictive to capture all earnings dynamics. We argue that when calibrating a typical macroeconomic model, it is more important to capture the cross-sectional variance of earnings (either over the life cycle or across the population at a particular point in time) rather than for yearly growth rates. If using a parsimonious random-walk process is important, we thus argue that the process typically should be

[31] The high variance of permanent shocks does not only imply an urealistic life-cycle profile of earnings inequality but also that the cross-sectional variance of earnings across all households at any point in time will be unrealistically high. The estimates from column I for example imply a cross-sectional variance of 0.78 rather than 0.18 as in the data.

estimated from earnings levels. In some cases it may however be important to allow for a richer specification of the earnings process. An example of such a case is when analyzing economic developments directly from the data rather than with the help of a calibrated model. Before examining such processes based on Swedish data, we examine how the estimation of the parsimonious process varies with the choice of earnings measure, sample selection criteria, and time periods. These results are presented in Table 6.

Overall, the results are consistent with those in Sections 3 and 4 where we presented measures of inequality over time and over the life cycle. First, the variances are substantially lower when focusing on households with a strong attachment to the labor market. In particular, excluding households with low income reduces the variance, but also requiring long valid time series of earnings data contributes to a lower variance. Second, the variances are also substantially lower for post-government earnings than for pre-government earnings, presumably because of the generous Swedish welfare state. Third, the variances of both the temporary and persistent shocks increased after 1990 for pre-government earnings. This development is consistent with the rising inequality that we documented in Section 3. For post-government earnings however only the volatility of permanent shocks increased after 1990. We return to interpreting this observation after estimating richer earnings processes that allow us to identify the sources of uncertainty and inequality with more precision.

5.1 Alternative stochastic processes

We now consider a more general earnings process where we allow for moving average terms and persistent shocks that are not necessarily permanent[32],

$$\ln y_{i,h,t} = \mathbf{x}'_{i,h,t}\psi_t + \alpha_{i,h,t} + \beta_i + \varepsilon_{i,t} + \theta_1\varepsilon_{i,t-1} + \theta_2\varepsilon_{i,t-2}, \tag{13}$$

$$\alpha_{i,h,t} = \rho\alpha_{i,h,t-1} + \eta_{i,t} \text{ if } h > 1 \tag{14}$$

$$\alpha_{i,1,t} = \eta_{i,t}. \tag{15}$$

So far we have only focused on estimation of this process under the restriction that it is a random walk ($\rho = 1$) with serially uncorrelated temporary shocks ($\theta = 0$). But we have seen indications that the random-walk process is too restrictive to capture all aspects of earnings dynamics. In particular, the random walk process forces the variances and covariances to be linear in age, but we see concavity in the empirical life-cycle profiles (see e.g. the empirical moments in Figure 18). We also see that the covariances $cov\,(h, h+j)$ fall as the horizon j increases, whereas the random walk process implies that the covariance does not vary with the horizon.

To investigate the importance of allowing for a richer specification of the process, we now allow for persistent but not necessarily permanent shocks ($\rho < 1$) and moving-average

[32] We estimate this process using level moments corresponding to (11)-(12) as described in Appendix A.3, using both time-series and life-cycle moments.

terms ($\theta \neq 0$). Table 7 reports estimates of the process (13)-(15) for pre- and post-government earnings with different restrictions on under the ARMA parameters. Overall, the results are similar to those reported in columns I and V in Table 6. The variances of the permanent shocks are similar to those in the corresponding columns in Table 6, and the autoregressive parameter is estimated to be around 0.96. Because the autoregressive parameter is lower than unity, the life-cycle profiles of earnings inequality will have some curvature. Figure 19 shows that the richer ARMA(1,2) process therefore captures more of the empirical life-cycle and time series profiles of earnings inequality, but that the improvement over the random-walk process is not dramatic.[33]

The moving-average components reported in Table 7 are large. As predicted by our analysis in footnote 30, allowing for moving-average terms results in smaller estimates of σ_ε^2 but does not affect the total variance of the temporary shocks ($\left(1+\theta_1^2+\theta_2^2\right)\sigma_\varepsilon^2$). The smaller estimates of σ_ε^2 imply that the empirical moments based on annual growth rates of earnings, (7)-(8), are captured much better by the ARMA(1,2) process than by the random-walk process as seen in Column III in Table A3.

A closer look at the data reveals that the empirical variance of earnings growth residuals, $var\left(g_t\right)$, falls somewhat over the life cycle, from around 0.08 at age 26 to 0.05 at age 59 in the benchmark estimation sample. This feature is not captured by the ARMA(1,2) process since it implies that the variance of growth rates is constant over the life cycle. We have therefore considered two further modifications of the estimated process to allow for richer life-cycle dynamics. First, we followed Guvenen (2009) and allowed for heterogeneous income profiles. Equation (13) was then replaced by

$$\ln y_{i,h,t} = \mathbf{x}'_{i,h,t}\psi_t + \alpha_{i,h,t} + \beta_i + \gamma_i h + \varepsilon_{i,t} + \theta_1\varepsilon_{i,t-1} + \theta_2\varepsilon_{i,t-2}, \tag{16}$$

where γ_i is the individual component in the life-cycle profile of income. This component has zero mean, variance σ_γ^2 and correlation $\rho_{\beta\gamma}$ with the individual fixed effect β_i. Estimation of (16) together with (14)-(15) thus involves estimation of the two additional parameters σ_γ^2 and $\rho_{\beta\gamma}$. The Swedish data give no support to this specification. We found $\sigma_\gamma^2 = 2\times10^{-7}$ and that the other parameter estimates were unaffected.

Second, we allowed for age-specific components in the volatility of temporary shocks and replaced (13) by

$$\ln y_{i,h,t} = \mathbf{x}'_{i,h,t}\psi_t + \alpha_{i,h,t} + \beta_i + \varepsilon_{i,h,t} + \theta_1\varepsilon_{i,h,t-1} + \theta_2\varepsilon_{i,h,t-2}, \tag{17}$$

where $\varepsilon_{i,h,t}\tilde{}N\left(0,\sigma_{\varepsilon,h,t}^2\right)$. Our estimates for $\sigma_{\varepsilon,h,t}^2$ indeed show a clear life-cycle profile. The variance of the temporary shocks falls by 45 percent between ages 25 and 35, and then remains roughly constant. Averaged over the life-cycle, however, the temporary variance is the same as in Table 7, and allowing for age-specific variances did not affect the other parameter estimates. The final columns in Table A3 moreover show that the average levels of both the variances and covariances are only captured slightly more accurately when allowing for age-specific variances.

[33] Recall that we have now allowed for year-specific variances when estimating the random-walk process. This explains why there is variation in the time-series profiles, and why the theoretical covariances in panel (c) have different slopes.

Figure 20 presents the full time-series of estimated variances of temporary and persistent shocks, estimated with the ARMA(1,2) process for pre- and post-government earnings as in columns IV and VIII in Table 7.[34] For *pre-government earnings*, the volatility of both temporary and persistent shocks was relatively constant during the 1980s, but the volatility of both shocks increased in the early 1990s. This increased volatility coincides with the macroeconomic crisis that hit the Swedish economy in that period (see e.g. the unemployment series in Figure 7). Although the volatility of both shocks increased, the impact of higher volatility of persistent shocks was a much more important force behind the rise in Swedish earnings inequality following the crisis. Table 8 shows how the changed volatility of temporary and persistent shocks, respectively, contributed to the rise in earnings inequality. In 2004, the variance of residual pre-government earnings was 0.201. The table shows that the variance would have been 0.164 if both the volatility of temporary and persistent shocks had remained at their pre-1990 levels. Holding the variance of persistent shocks constant at the pre-1990 level, the variance in 2004 would have risen to 0.170 because of changes in the variance of temporary shocks. If instead the variance of temporary shocks had remained at its pre-1990 level, the variance in 2004 would have risen to 0.194 as a consequence of the rising volatility of persistent shocks.

The lower panels of Figure 20 show that the volatility of persistent shocks to *post-government earnings* was relatively constant during the 1980s, rose substantially during the crisis in the early 1990s, and then returned to levels just slightly above those seen prior to the crisis. The volatility of temporary shocks fell somewhat during the 1980s and then remained relatively constant throughout both the crisis and the recovery in the 1990s. The variance of post-government earnings would have been 0.104 instead of 0.121 if the volatility of the shocks had remained at their pre-crisis levels. Table 8 shows that permanent shocks explain all the increase, and that changes in the volatility of temporary shocks actually tended to reduce the overall inequality.

These estimates, both for pre- and post-government earnings, thus show that the rising Swedish earnings inequality during the 1990s was generated by an increased volatility of persistent shocks. A candidate explanation for the higher inequality *after* the crisis is that the crisis had long-lasting effects on households that were hit by shocks during the crisis. Such effects need not only take the form of the shocks being persistent but could possibly also generate larger volatility for these households even after the crisis, and thus explain the higher volatility of shocks during the last decade. An alternative explanation is that there has been a trend increase in inequality that need not be directly related to the crisis. In support of this latter hypothesis, Figure 21 reports estimates of cohort-specific variances of the individual fixed effects $\sigma^2_{\beta,t-h}$.[35] The figure shows that there has been a trend increase in this variance for both pre- and post-government earnings. That is, inequality has increased also for households that entered the labor market after the crisis. A more thorough way to examine the plausibility of these hypotheses would be to closely track households' transitions in and out of the labor market and in the earnings distribution over time. This is left for future work.

[34] As we noted above, allowing for the richer ARMA(1,2) specification rather than the more restrictive random-walk specification had no dramatic impact on the overall properties of the dynamic process. But the ARMA(1,2) specification identifies yearly variances of the temporary and persistent shocks that are much smoother, and presumably more accurate, than those identified by the random-walk specification.

[35] We have previously restricted σ^2_β to be constant over time. Allowing for cohort-specific $\sigma^2_{\beta,t-h}$ has little impact on estimates of the other parameters in the process.

6 Concluding remarks

The broad picture of the development of income, wealth, and consumption inequality that we document in this paper is consistent with previous studies. We clearly see that income inequality has increased during the last decades. Previous studies also document a dramatic reduction in income inequality during the 1970s. We do not date the turning point, but we find some evidence that in particular earnings inequality increased permanently during the macroeconomic crisis in the early 1990s. Much of this increased inequality was generated by movements in and out of the labor market. In spite of the large rise in pre-government earnings inequality, there was only a small increase in inequality in earnings net of taxes and transfers or disposable income. These different developments are most likely explained by the generous Swedish welfare system. Consistent with these observations, we see no clear trend in consumption inequality.

The increase in earnings inequality is almost entirely driven by an increase in *residual* earnings inequality. But we only control for age, family composition and the level of education. Changes in the households' *type* of education and their sectorial belonging therefore show up in the residual. Domeij (2008) shows that such compositional changes contributed to the increased inequality in Sweden during the 1990s.

We also estimate stochastic processes for household earnings. A simple random-walk process can capture much of the life-cycle dynamics. But we also find clear evidence that the true earnings process is not a random walk. In particular, we find evidence of serial correlation in the temporary shocks. We point out that some estimation methods result in severe upward bias in the estimated volatility of permanent shocks if the presence of such serial correlation is ignored. The earnings process is better described by a highly persistent ARMA(1,2) process. In addition to allowing for serial correlation in the temporary shocks, this process also implies that the cross-sectional variance increases less than linearly over the life cycle, just as in the data.

We find a clear increased volatility of earnings shocks during the severe macroeconomic crisis in the early 1990s. For pre-government earnings, the volatility of both temporary and persistent shocks remained high after the crisis. Abstracting for the crisis period, the volatility of persistent shocks in post-government earnings was only slightly higher in the recent time-period relative to the 1980s, and there is a small downward trend in the volatility of temporary shocks. We show however that the trends in the variance of temporary shocks have been of little importance for the trends in pre- and post-government earnings inequality, and that the rise in Swedish earnings inequality was almost entirely driven by larger persistent shocks.

Appendix A Estimation of the stochastic process for earnings

Appendix A.1 Using time-series information

The construction of moments in this subsection follows the approach outlined in Krueger et al. (2008). It is also similar to Guvenen (2009) and Heathcote et al. (2008). To simplify, this presentation ignores the fact that our panel is unbalanced.[36] See Krueger et al. for details with unbalanced panels.

The empirical moments are constructed as follows. Let

$$g_{i,h,t} = \ln y_{i,h,t} - \mathbf{x}'_{i,h,t}\boldsymbol{\psi}_t$$

denote residual earnings for individual i of age h. We include ages $25 - 59$, and thus let h range from 1 to $H = 35$. In our baseline sample we have data for $1978 - 2004$. Time t thus ranges from 1 to $T = 27$.

Let further

$$\mathbf{g}_i = \begin{pmatrix} g_{i,h,1} \\ g_{i,h+1,2} \\ \vdots \\ g_{i,h+T-1,T} \end{pmatrix}$$

denote the vector of relevant information for an individual i aged h in the first period, and let

$$\mathbf{m}_i = \mathbf{g}_i\mathbf{g}'_i$$

Our $T(T+1)/2 = 378$ empirical moments $\mathbf{m}$, are then

$$\mathbf{m} = \mathbf{vec}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{m}_i\right) = \begin{pmatrix} var_i\left(g_{t=1}\right) \\ cov_i\left(g_{t=1}, g_{t=2}\right) \\ \vdots \\ cov\left(g_{t=1}, g_T\right) \\ var\left(g_{t=2}\right) \\ \vdots \end{pmatrix}$$

Appendix A.2 Using life-cycle information

The construction of moments in this subsection uses an approach similar to that above. The difference is that we now collect information on the variances and covariances of residual earnings between specific ages whereas we in the previous subsection collected information on variances and covariances between specific years.

[36] By "unbalanced" we mean not only that observations may be missing in the usual sense, but also that some moments do not exist. For example, a 58 year old in 1978 will be excluded from the sample in 1980. We can then only calculate the variance and first covariance for that individual.

Define residual earnings $g_{i,h,t}$ as above. Let

$$\mathbf{g}_{i,t} = \begin{pmatrix} g_{i,1,t} \\ g_{i,2,t+1} \\ \vdots \\ g_{i,H,t+H-1} \end{pmatrix}$$

denote the vector of relevant information for an individual i aged 25 ($h = 1$) in period t, and let

$$\mathbf{m}_i = \mathbf{g}_i \mathbf{g}_i'$$

Our $H\left(H+1\right)/2 - \left(H-T\right)\left(H-T+1\right)/2 = 594$ empirical moments $\mathbf{m}$, are then

$$\mathbf{m} = \mathbf{vec}\left(\frac{1}{N}\sum_{i=1}^{N}\mathbf{m}_i\right) = \begin{pmatrix} var_i\left(g_{h=1}\right) \\ cov_i\left(g_{h=1}, g_{h=2}\right) \\ \vdots \\ cov\left(g_{h=1}, g_H\right) \\ var\left(g_{h=2}\right) \\ \vdots \end{pmatrix}$$

Appendix A.3 Moment conditions for the ARMA process

The information used to identify the ARMA(1,2) process is described by

$$var(e_{h,t}) = var\left(\alpha_{h,t}\right) + \sigma_\beta^2 + \sigma_{\varepsilon,t}^2 + \theta_1^2\sigma_{\varepsilon,t-1}^2 + \theta_2^2\sigma_{\varepsilon,t-2}^2, \tag{A.18}$$

and

$$cov(e_{h,t}, e_{h+s,t+s}) = \begin{cases} \rho var\left(\alpha_{h,t}\right) + \sigma_\beta^2 + \theta_1\sigma_{\varepsilon,t}^2 + \theta_1\theta_2\sigma_{\varepsilon,t-1}^2 & \text{if } s = 1 \\ \rho^2 var\left(\alpha_{h,t}\right) + \sigma_\beta^2 + \theta_2\sigma_{\varepsilon,t}^2 & \text{if } s = 2 \\ \rho^s var\left(\alpha_{h,t}\right) + \sigma_\beta^2 & \text{otherwise} \end{cases} \tag{A.19}$$

where the variance for the persistent component α is

$$\mathrm{var}\left(\alpha_{1,t}\right) = \sigma_{\eta,t}^2,$$

$$var\left(\alpha_{h,1}\right) = \rho^{2(h-1)} var\left(\alpha_{1,1}\right) + \sigma_{\eta,1}^2 \sum_{j=1}^{h-1} \rho^{2(j-1)} \text{ if } h > 1,$$

and

$$var\left(\alpha_{h,t}\right) = \rho^2 var\left(\alpha_{h-1,t-1}\right) + \sigma_{\eta,t}^2 \text{ if } h > 1, \text{ and } t > 1.$$

References

Agell, Jonas, Peter Englund, and Jan Södersten (1996), "Tax reform of the century – the Swedish experiment", *National Tax Journal* 49(4), 643-664.

Björklund, Anders, and Richard Freeman (2008), "Searching for optimal inequality/incentives", in: R. Freeman, B. Swedenborg, and R. Topel (eds.), *Reforming the Welfare State: Recovery and Beyond in Sweden*, Chicago University Press

Blundell, Richard, Luigi Pistaferri, and Ian Preston (2008), "Consumption inequality and partial insurance", *American Economic Review*, forthcoming

Böhlmark, Anders, and Matthew Lindqvist (2006), "Life-cycle variations in the association between current and lifetime income: Replication and extension for Sweden", *Journal of Labor Economics* 24(4), 879-96

Domeij, David (2008), "Rising earnings inequality in Sweden: the role of composition and prices", *Scandinavian Journal of Economics* 110(3), 609-634

Domeij, David, and Paul Klein (2002), "Public pensions: to what extent do they account for Swedish wealth inequality?", *Review of Economic Dynamics*, 5, 503-534

Domeij, David, and Lars Ljungqvist (2008), "The missing Swedish skill premium: Sweden versus the United States 1970-2002", manuscript, Stockholm School of Economics

Edin, Per-Anders, and Peter Fredriksson (2000), "LINDA – Longitudinal INdividual DAta for Sweden", manuscript, Uppsala University

Edin, Per-Anders, and Bertil Holmlund (1995), "The Swedish wage structure: the rise and fall of solidarity wage policy", in: R. Freeman and L. Katz (eds.), *Differences and Changes in Wage Structures*, Chicago: The University of Chicago Press, 307-344

Gottschalk, Peter, and Sheldon Danzinger, (2003), "Wage inequality, earnings inequality and poverty in the U.S. over the last quarter of the twentieth century", manuscript, Boston College

Gustavsson, Magnus (2006), "The evolution of the Swedish wage structure: new evidence for 1992-2001", *Applied Economics Letters*, 13(5), 279-286

Gustavsson, Magnus, (2007), "The 1990s rise in Swedish earnings inequality - persistent or transitory?", *Applied Economics,* 39(1), 25-30

Guvenen, Fatih (2009), "An empirical investigation of the labor income process", *Review of Economic Dynamics*, 12, 58-79

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni Violante (2004), "The cross-sectional implications of rising wage inequality in the United States", manuscript, New York University

Heathcote, Jonathan, Kjetil Storesletten, and Giovanni Violante (2008), "The macroeconomic implications of rising wage inequality in the United States", manuscript, New York University

Johansson, Mats, (2006), "Inkomst och ojämlikhet i Sverige 1951-2002", Working Paper 2006:3, Institutet för Framtisstudier (in Swedish)

Krueger, Dirk, Fabrizio Perri, Luigi Pistaferri, and Gianluca Violante (2008), "Guidelines to Achieve Cross-country Data Comparability", manuscript, University of Pennsylvania

Lindbeck, Assar (1998), "The Swedish Experiment", *Journal of Economic Literature*, 35, 1273-1319

MaCurdy, Thomas (1982), "The use of time-series processes to model the error structure of earnings in a longitudinal data analysis", *Journal of Econometrics* 18, 83-114

Roine,Jesper and Daniel Waldenström (2009), "Top incomes in Sweden over the twentieth century", in: A. B. Atkinson and T. Piketty (eds.), *Top Incomes: A Global Perspective*, Volume II, Oxford University Press

Skedinger, Per (2007), "The design and effects of collectively agreed minimum wages: evidence from Sweden", IFN Working Paper No. 700

Statistics Sweden (2000), "Förmögenhetsfördelningen i Sverige 1997, med tillbakablick till 1975", Rapport 2000:1 [in Swedish]

Statistics Sweden (2006), "Aggregerade variabler i LINDA", manuscript [older version available at: linda.nek.uu.se/F%F6rslag%20aggregerade%20variabler.xls]

Statistics Sweden (2006b), "Wealth Statistics 2004"

Werin, Lars (1993), "Det finansiella systemet i omvandling", in: L. Werin (ed.), *Från räntereglering till inflationsnorm*, SNS, Stockholm [in Swedish]

Table 1: Means (2004 SEK; thousands, except hours and wages)

	LINDA				HUT				HINK					
	Earnings	Earnings	Capital	Disp.	Consumption		Cons. plus housing		Wages		Hours		Net wealth	
	(pre gvt.)	(post gvt.)	income	income	househ.	equiv.	househ.	equiv.	(men)	(women)	(men)	(female)	financial	total
1978	245.0	160.4	8.6	168.6					124.7	103.0	2138	1516	141.3	224.8
1979	247.1	169.1	8.7	177.6										
1980	245.6	167.3	9.7	176.8					116.6	103.4	2096	1513		
1981	234.7	162.0	10.6	172.4					113.2	97.7	2081	1531		
1982	236.9	165.5	10.5	175.7					106.9	98.8	2080	1546	221.2	289.6
1983	234.1	168.6	9.9	178.2					104.8	96.3	2073	1556	169.8	238.3
1984	235.8	175.6	9.7	185.1					104.2	89.6	2126	1587	186.9	252.5
1985	237.7	178.6	9.7	188.0	177.6	91.2	226.6	116.7	105.4	90.0	2157	1629	228.1	301.8
1986	247.1	183.1	8.9	191.8					104.6	91.1	2151	1640	213.2	282.1
1987	257.7	190.4	9.5	199.5					105.7	94.0	2093	1643	191.7	279.8
1988	263.2	196.1	9.2	205.0	186.6	100.3	240.3	129.0	109.9	98.6	2096	1678	237.8	355.0
1989	275.1	202.9	9.6	212.1					116.4	96.8	2128	1737	324.7	464.4
1990	278.4	208.3	9.7	217.6					116.0	99.2	2099	1745	305.9	456.6
1991	286.0	236.2	20.9	255.9					118.6	97.0	2051	1730	216.3	423.8
1992	274.9	234.2	13.7	247.4	180.7	97.7	243.5	132.1	122.4	100.6	2045	1727	237.2	394.7
1993	255.1	223.1	13.0	235.4										
1994	254.2	219.8	16.2	235.1										
1995	255.7	212.0	13.4	224.7										
1996	263.7	211.0	16.3	226.3	160.3	92.0	218.8	127.6						
1997	268.2	209.3	19.3	227.3										
1998	280.9	215.7	21.2	235.1										
1999	294.2	222.5	28.1	248.1	198.7	108.6	252.7	140.1						
2000	309.4	235.3	33.5	265.7										
2001	317.9	245.8	21.7	265.8										
2002	319.5	251.2	19.9	269.3										
2003	318.3	251.0	16.5	266.3										
2004	322.5	253.5	18.5	270.6										

Notes: The household definition in LINDA changed between 1990 and 1991. As a result the average household size in our sample increased by 12.1 percent (see Table A2). The LINDA data are based on all households with head aged 25-59, and the data are not equivalized. The HUT data are based on all households with head aged 25-59 and consumption above the minimum threshold. The HINK wealth data are based on all households with head aged 25-59. Wages and hours data are based on individuals aged 25-59 with positive hours, with real wage above the minimum threshold and living in a household with head aged 25-59.

Table 2: Correlations between hours and wages

		Wages		Hours	
		head	spouse	head	spouse
Wages	head		0.13	-0.21	0.01
	spouse			-0.04	-0.23
Hours	head	-0.24^{a}			0.43
	spouse		-0.23^{b}		

Notes: All data is from HINK for 1978 to 1992. $^{a)}$ 'Head' here refers to men. $^{b)}$ 'Spouse' here refers to women.

Table 3: Cross-sectional dispersion

	LINDA					HUT	HINK		
	Earnings (head)	Earnings (pre gvt.)	Earnings (post gvt.)	Capital income	Disp. income	Consumption househ.	Wages	Net wealth financial	Net wealth total
Variance of log	0.873	0.772	0.303	4.131	0.304	0.183	0.208	3.063	2.548
1978-1990	0.629	0.556	0.297	2.254	0.240	0.180			
1991-2004	1.100	0.972	0.308	5.875	0.363	0.186			
p90/p50	2.151	1.739	1.534	1568.464	1.544	1.691	1.684	17.376	6.749
1978-1990	1.959	1.633	1.472	58.072	1.462	1.658			
1991-2004	2.330	1.837	1.593	2197.795	1.620	1.714			
p50/p10	61.165	54.756	1.995	n.a.	1.789	1.691	1.842	n.a.	n.a.
1978-1990	5.552	4.380	2.107	n.a.	1.821	1.659			
1991-2004	302.154	105.133	1.892	n.a.	1.760	1.712			
Gini coefficient	0.340	0.297	0.223	0.776	0.226	0.238	0.255	1.029	0.850
1978-1990	0.309	0.272	0.216	0.682	0.204	0.236			
1991-2004	0.368	0.321	0.229	0.864	0.247	0.239			
Coefficient of variation	0.627	0.666	0.508	11.321	0.861	0.478	0.755	5.344	3.500
1978-1990	0.620	0.587	0.467	3.463	0.449	0.501			
1991-2004	0.634	0.739	0.546	18.618	1.243	0.464			
Share ≤ 0	0.090	0.074	0.021	0.531	0.016	0.000	0.000	0.333	0.262
1978-1990	0.068	0.058	0.020	0.511	0.017	0.000			
1991-2004	0.108	0.087	0.022	0.474	0.015	0.000			

Notes: All variables except wages are equivalized. The average p50/p10 ratios for head and pre-government earnings exclude 1997 where more than 10 percent of households report zero earnings.

Table 4: Correlations

	LINDA					HUT		HINK	
	Earnings (head)	Earnings (pre gvt.)	Earnings (post gvt.)	Capital income	Disp. income	Consumption	Saving	Net wealth financial	Net wealth total
Earnings (pre gvt.)	0.91		0.77	0.08	0.61			0.03	0.08
1978-1990	0.90		0.75	0.09	0.77				
1991-2004	0.92		0.79	0.07	0.45				
Disposable income	0.54	0.61	0.71	0.52		0.47	0.59	0.02	0.06
1978-1990	0.68	0.77	0.95	0.17		0.44	0.44		
1991-2004	0.41	0.45	0.49	0.84		0.49	0.68		

Note: All variables are equivalized.

Table 5: Estimated earnings processes

	I	II	III	IV
σ^2_β	0.062	0.000	0.034	0.044
	(0.001)	(0.000)	(0.001)	(0.001)
σ^2_ε	0.010	0.060	0.075	0.070
	(0.000)	(0.000)	(0.000)	(0.000)
σ^2_η	0.040	0.007	0.004	0.004
	(0.000)	(0.000)	(0.000)	(0.000)
Moments	differences	levels	levels	levels
	(7)-(8)	(11)-(12)	(11)-(12)	(11)-(12)
Aggregation	ages	ages	years	none

Notes: Estimates of the random walk process for pre-government earnings based on different moment conditions and aggregation methods. The variances σ^2_ε and σ^2_η were restricted to be constant over time. Standard errors in parenthesis are based on estimates from 100 Monte Carlo simulations. First stage controls are consumption equivalents, and age and family dummies. The sample consists of 149,057 households, resulting in 2,019,704 and 1,838,294 observations on incomes and income differences, respectively.

Table 6: Estimated earnings processes, various sample selection criteria

	pre-government earnings				post-government earnings			
	I	II	III	IV	V	VI	VII	VIII
σ^2_β	0.045	0.155	0.032	0.058	0.030	0.074	0.023	0.058
	(0.000)	(0.001)	(0.001)	(0.002)	(0.000)	(0.001)	(0.001)	(0.001)
σ^2_ε	0.061	0.359	0.044	0.170	0.036	0.129	0.030	0.089
	(0.001)	(0.003)	(0.001)	(0.002)	(0.000)	(0.001)	(0.000)	(0.001)
σ^2_η	0.006	0.014	0.005	0.011	0.003	0.005	0.003	0.004
	(0.001)	(0.002)	(0.001)	(0.002)	(0.000)	(0.001)	(0.000)	(0.001)
required observations	2	2	20	20	2	2	20	20
exclude low income	yes	no	yes	no	yes	no	yes	no
# households	149,057	158,817	35,914	40,450	160,098	165,405	40,691	44,476
# differences	1,838,294	2,010,714	853,074	964,485	2,036,251	2,145,049	966,172	1,061,537
$\sigma^2_{\varepsilon,1978-1990}$	0.059	0.259	0.046	0.143	0.040	0.140	0.035	0.106
$\sigma^2_{\varepsilon,1991-2004}$	0.063	0.451	0.043	0.196	0.032	0.119	0.025	0.073
$\sigma^2_{\eta,1978-1990}$	0.003	0.003	0.003	0.006	0.002	0.001	0.002	0.002
$\sigma^2_{\eta,1991-2004}$	0.008	0.024	0.007	0.016	0.005	0.009	0.004	0.006

Notes: Estimates of the random walk process based on (11) and (12), using both life-cycle and time-series moments. The rows for σ^2_ε and σ^2_η report averages over years. Column I is identical to column IV in Table 5, except that we now allow for year-specific variances. See also Table 5.

Table 7: Estimated earnings processes, various process specifications

	pre-government earnings				post-government earnings			
	I	II	III	IV	V	VI	VII	VIII
σ^2_β	0.043	0.042	0.027	0.029	0.029	0.028	0.019	0.020
	(0.001)	(0.001)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
σ^2_ε	0.050	0.053	0.043	0.047	0.028	0.030	0.025	0.027
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)
σ^2_η	0.005	0.004	0.015	0.011	0.003	0.003	0.008	0.007
	(0.001)	(0.001)	(0.001)	(0.001)	(0.000)	(0.001)	(0.000)	(0.001)
θ_1	0.602	0.476		0.367	0.622	0.477		0.352
	(0.004)	(0.002)		(0.004)	(0.005)	(0.002)		(0.004)
θ_2		0.368		0.175		0.395		0.173
		(0.004)		(0.006)		(0.005)		(0.006)
ρ			0.956	0.965			0.957	0.965
			(0.001)	(0.001)			(0.001)	(0.001)
$\sigma^2_{\varepsilon,1978-1990}$	0.046	0.047	0.039	0.040	0.030	0.031	0.027	0.027
$\sigma^2_{\varepsilon,1991-2004}$	0.053	0.058	0.047	0.053	0.026	0.029	0.023	0.026
$\sigma^2_{\eta,1978-1990}$	0.003	0.003	0.011	0.009	0.002	0.002	0.006	0.005
$\sigma^2_{\eta,1991-2004}$	0.006	0.005	0.019	0.013	0.004	0.004	0.010	0.008

Notes: Estimates of the process (13)-(15) using both life-cycle and time-series moments. See Tables 5-6 for further details.

Table 8: Cross-sectional variance of log residual earnings in 2004

	pre gvt.	post gvt.
with fixed σ_{ε}^2 and σ_{η}^2	0.164	0.104
with fixed σ_{η}^2	0.170	0.098
with fixed σ_{ε}^2	0.194	0.126
with new σ_{ε}^2 and σ_{η}^2	0.201	0.121

Note: The table reports the cross-sectional variance of log residual earnings in 2004 based on simulations with different variances of the income shocks. For the period $1978 - 1990$, we use the variances from the estimates behind columns IV and VIII in Table 7. For $1991 - 2004$, we either fix the variances at their $1978 - 1990$ average, or use the actual estimated values.

Table A1: Comparing data sets

	Earnings				Disposable income						Correlation	
	mean		var(log)		mean			var(log)			Earnings & Disp. inc.	
	LINDA	HINK	LINDA	HINK	LINDA	HUT	HINK	LINDA	HUT	HINK	LINDA	HINK
1978	147.1	144.7	0.540	0.899	102.1		117.1	0.336		0.256	0.842	0.797
1979	149.3		0.511		107.9			0.295			0.845	
1980	149.2	145.6	0.540	0.995	107.8		122.6	0.304		0.292	0.732	0.831
1981	143.5	137.6	0.571	0.998	105.8		119.0	0.275		0.287	0.809	0.823
1982	142.8	129.0	0.540	0.924	106.4		112.2	0.230		0.286	0.750	0.828
1983	142.1	127.2	0.526	1.010	107.9		108.1	0.197		0.303	0.656	0.774
1984	144.1	129.6	0.583	1.295	113.0		111.5	0.199		0.321	0.807	0.816
1985	146.1	133.8	0.592	1.233	115.5	110.8	118.0	0.200	0.105	0.292	0.843	0.668
1986	153.1	141.4	0.547	1.076	119.0		119.9	0.195		0.286	0.820	0.663
1987	161.1	148.1	0.581	1.144	124.5		124.6	0.209		0.224	0.812	0.632
1988	165.7	151.6	0.562	1.155	128.5	126.6	128.8	0.215	0.123	0.248	0.684	0.665
1989	172.6	167.6	0.555	1.073	133.0		136.4	0.224		0.229	0.683	0.669
1990	172.8	163.7	0.582	1.292	134.9		135.0	0.244		0.257	0.753	0.444
1991	162.4	158.7	0.698	0.974	144.6		145.6	0.353		0.255	0.233	0.648
1992	156.6	163.3	0.823	0.939	140.7	141.0	151.5	0.361	0.148	0.269	0.629	0.840
1993	145.8		0.983		134.7			0.300			0.673	
1994	146.0		1.054		135.4			0.332			0.448	
1995	147.8		1.040		130.0			0.303			0.497	
1996	153.6		1.125		132.0	133.9		0.322	0.158		0.599	
1997	157.7		1.168		133.9			0.342			0.528	
1998	166.2		1.120		139.3			0.332			0.347	
1999	174.7		1.049		147.3	143.2		0.350	0.225		0.461	
2000	184.8		0.947		158.4			0.369			0.419	
2001	190.8		0.887		159.8			0.381			0.257	
2002	191.8		0.887		162.5			0.475			0.139	
2003	190.4		0.899		159.6			0.463			0.579	
2004	192.2		0.924		161.9			0.398			0.540	

Note: Variables are equivalized and in 2004 SEK.

Table A2: Descriptive statistics

	LINDA				HUT				HINK				
	# househ.	age	househ. size	cons. equiv.	# househ.	age	househ. size	cons. equiv.	# ind.	# househ.	age	househ. size	cons. equiv.
1978	82,491	40.2	2.3	1.7					8,699	7,467	42.7	2.4	1.9
1979	83,495	40.2	2.2	1.7									
1980	84,238	40.0	2.2	1.7					8,145	7,287	43.2	2.2	1.7
1981	84,903	40.0	2.2	1.7					7,861	7,134	43.0	2.2	1.7
1982	82,555	40.2	2.2	1.7					8,268	7,578	42.9	2.2	1.7
1983	83,230	40.2	2.2	1.7					7,501	7,088	42.4	2.2	1.7
1984	84,029	40.2	2.1	1.7					7,674	6,819	42.4	2.2	1.7
1985	84,830	40.2	2.1	1.7	3,359	40.3	2.7	2.0	7,722	6,610	42.5	2.2	1.7
1986	85,725	40.2	2.1	1.7					7,692	6,514	42.0	2.1	1.6
1987	86,543	40.3	2.1	1.6					7,504	6,462	41.8	2.0	1.6
1988	88,122	40.3	2.1	1.6	2,889	40.6	2.6	2.0	7,751	6,590	41.7	1.9	1.6
1989	88,826	40.4	2.1	1.6					8,994	7,252	42.1	2.0	1.6
1990	90,354	40.4	2.1	1.7					7,540	6,439	41.6	1.9	1.6
1991	87,048	40.8	2.4	1.8					7,593	6,414	41.7	1.9	1.6
1992	88,421	40.9	2.3	1.8	2,949	40.7	2.6	1.9	9,802	8,538	42.3	1.9	1.6
1993	89,937	41.0	2.3	1.8									
1994	91,580	41.1	2.3	1.8									
1995	92,575	41.3	2.3	1.8									
1996	93,721	41.4	2.3	1.8	811	41.8	2.3	1.8					
1997	95,263	41.4	2.3	1.7									
1998	96,203	41.5	2.2	1.7									
1999	96,852	41.6	2.2	1.7	1,253	42.0	2.5	1.9					
2000	97,565	41.7	2.2	1.7									
2001	98,205	41.9	2.2	1.7									
2002	98,275	41.9	2.2	1.7									
2003	98,136	42.0	2.2	1.7									
2004	97,678	42.0	2.2	1.7									

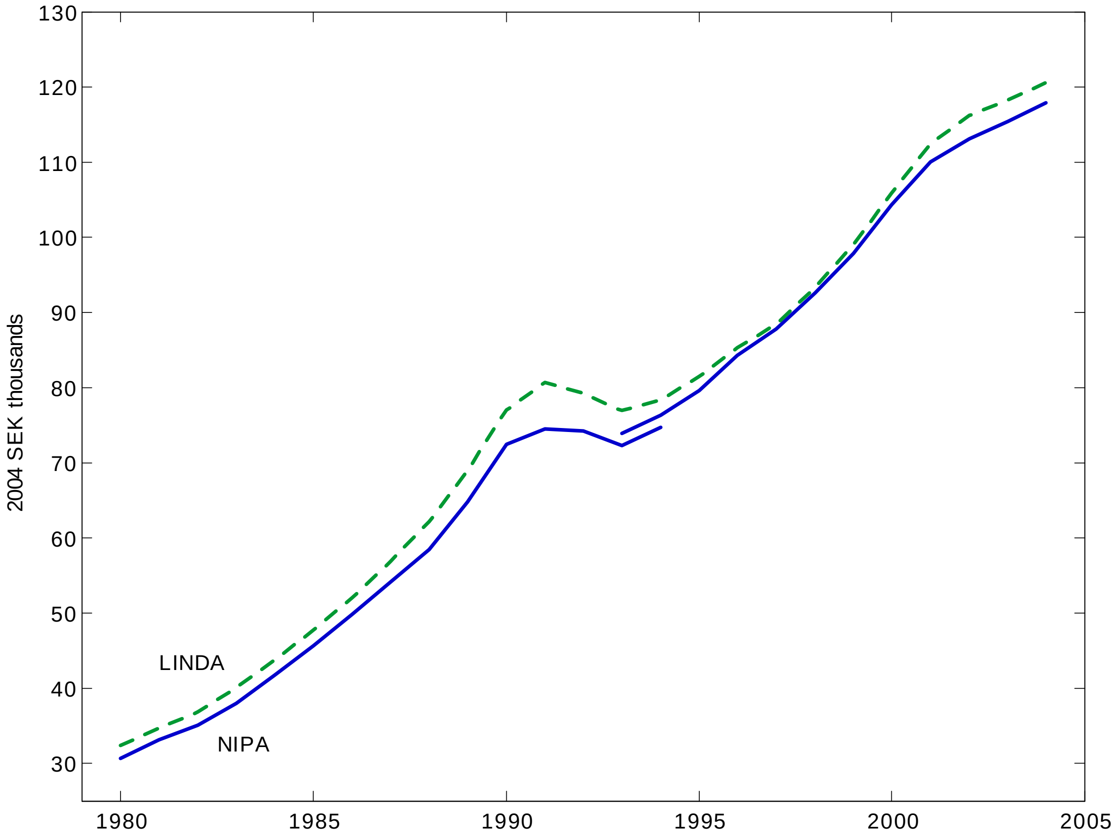
Notes: The household definition in LINDA changed between 1990 and 1991.

Table A3: Implications for moments based on earnings differences g_t

				age-specific variances	
	data	random walk	ARMA(1,2)	random walk	ARMA(1,2)
	I	II	III	IV	V
$var\,(g_t)$	0.060	0.144	0.079	0.127	0.077
$cov\,(g_t, g_{t+1})$	−0.010	−0.070	−0.022	−0.061	−0.021

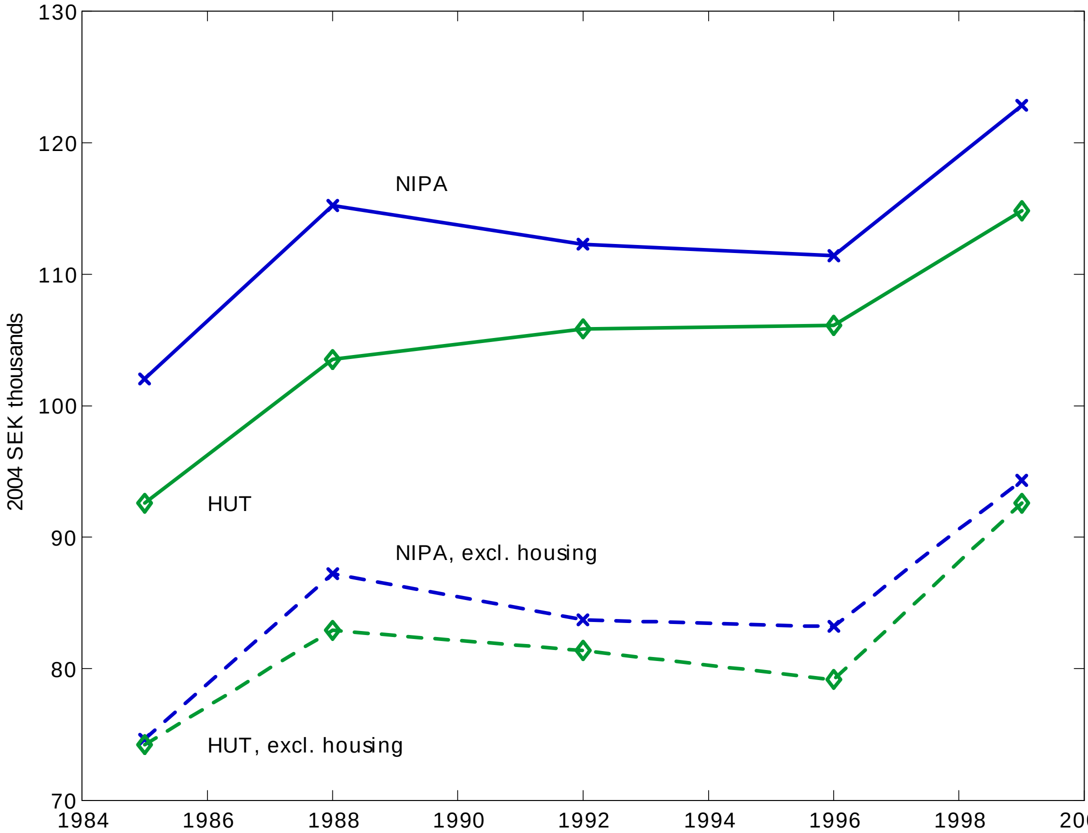
Notes: The table shows the average moments (7)-(8) from the data and implied by the processes estimated from the moments (11)-(12). The process in column II is identical to column IV in Table 5, the process in column III is identical to column IV in Table 7. The estimated processes in columns IV-V also allow for age-specific variances of the temporary shock, $\sigma^2_{\varepsilon,h,t}$.

Figure 1: Comparison of income data from LINDA with NIPA



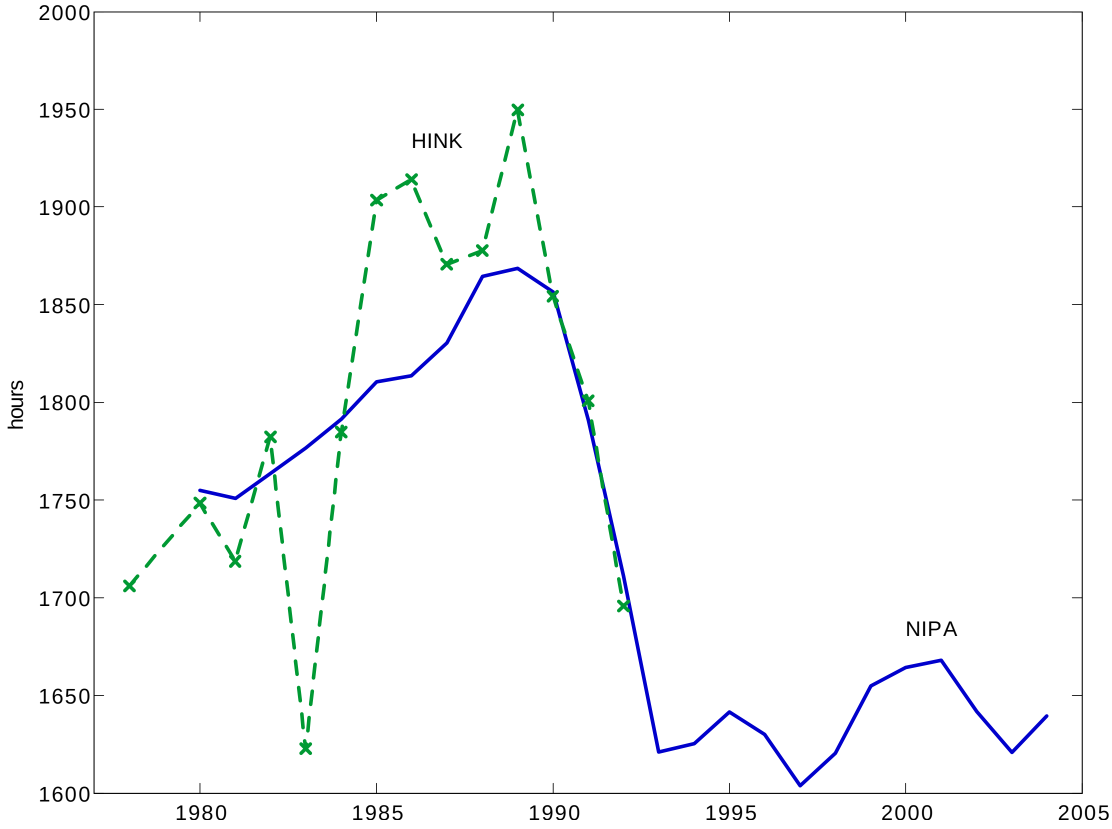
Note: The figure shows mean labor income per capita in current prices. NIPA data are per capita wage sums from the national accounts. Data for 1980-1994 are from Statistics Sweden N10 SM 9501. Data for 1993-2004 are from Statistics Sweden's online national accounts database.

Figure 2: Comparison of consumption data from HUT with NIPA



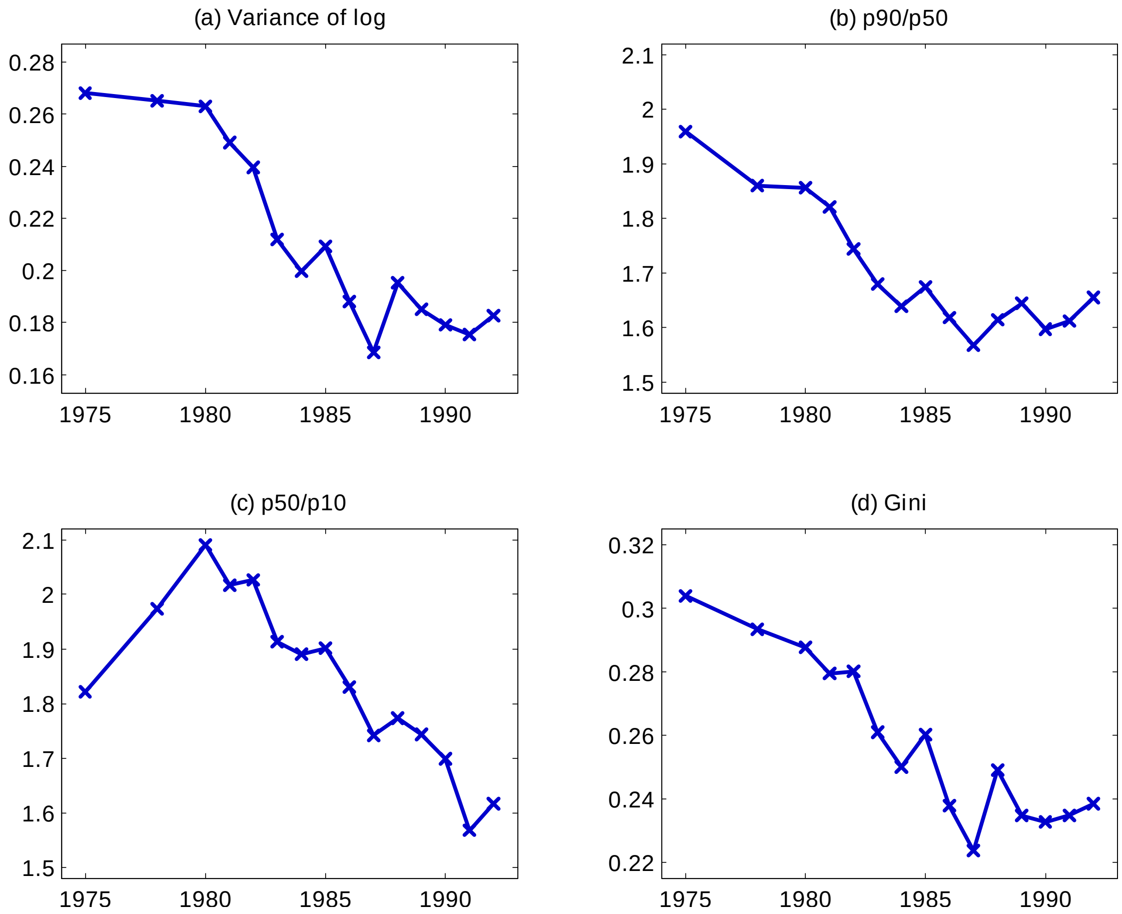
Note: The figure shows mean consumption per capita in current prices. NIPA data are household consumption expenditure, excluding non-profit organizations, based on national accounts data reported by the National Institute of Economic Research.

Figure 3: Comparison of hours data from HINK with "NIPA"



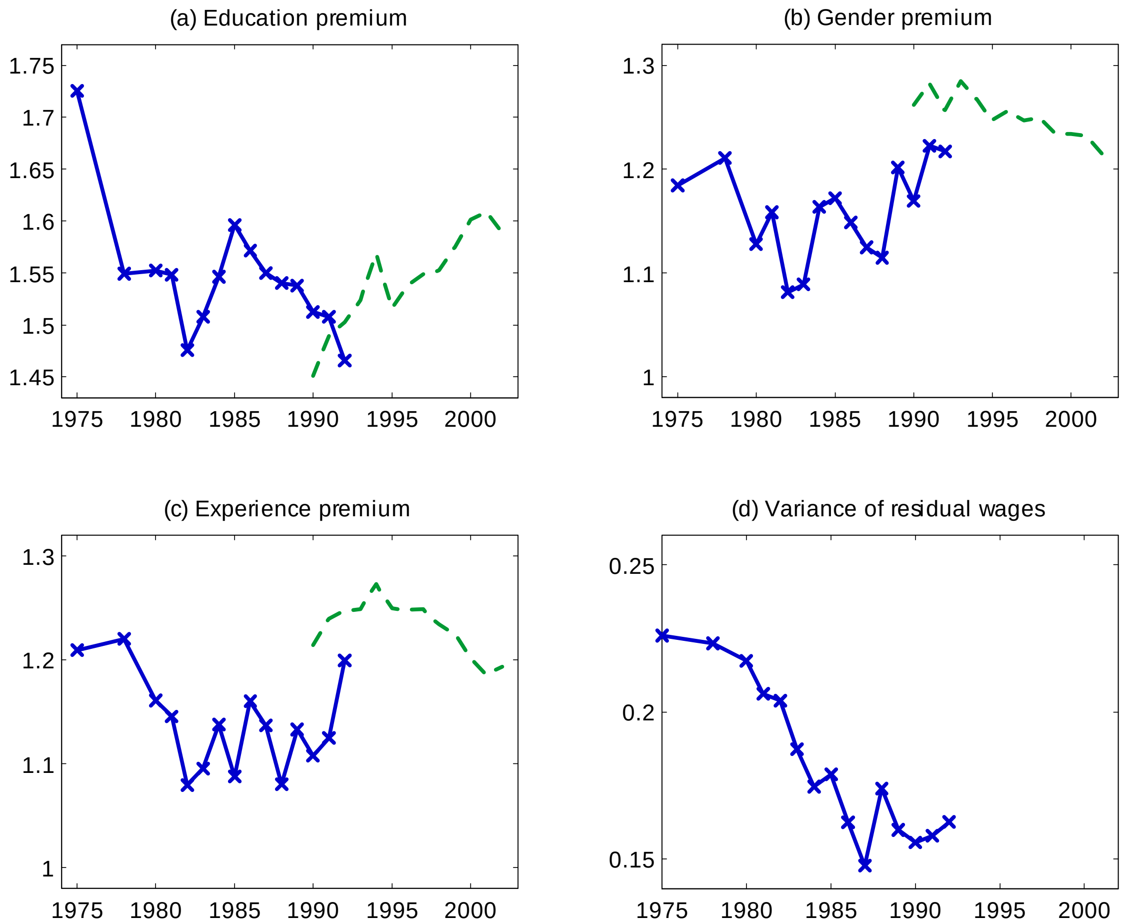
Note: The figure shows total hours relative to the population aged 25-59. "NIPA" data are from Statistics Sweden's Labor Force Survey (AKU).

Figure 4: Basic inequality in wages



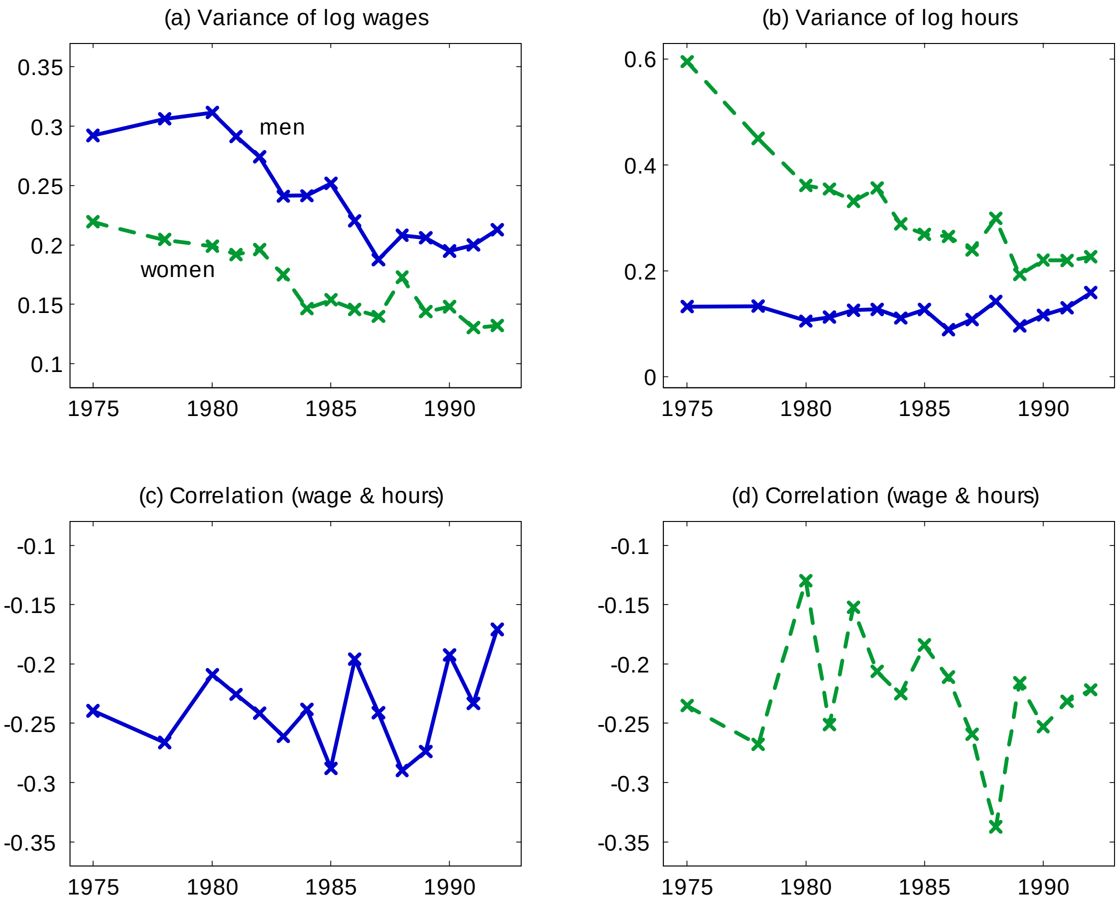
Note: The figure shows different measures of inequality in hourly wages, using data from HINK.

Figure 5: Wage premia



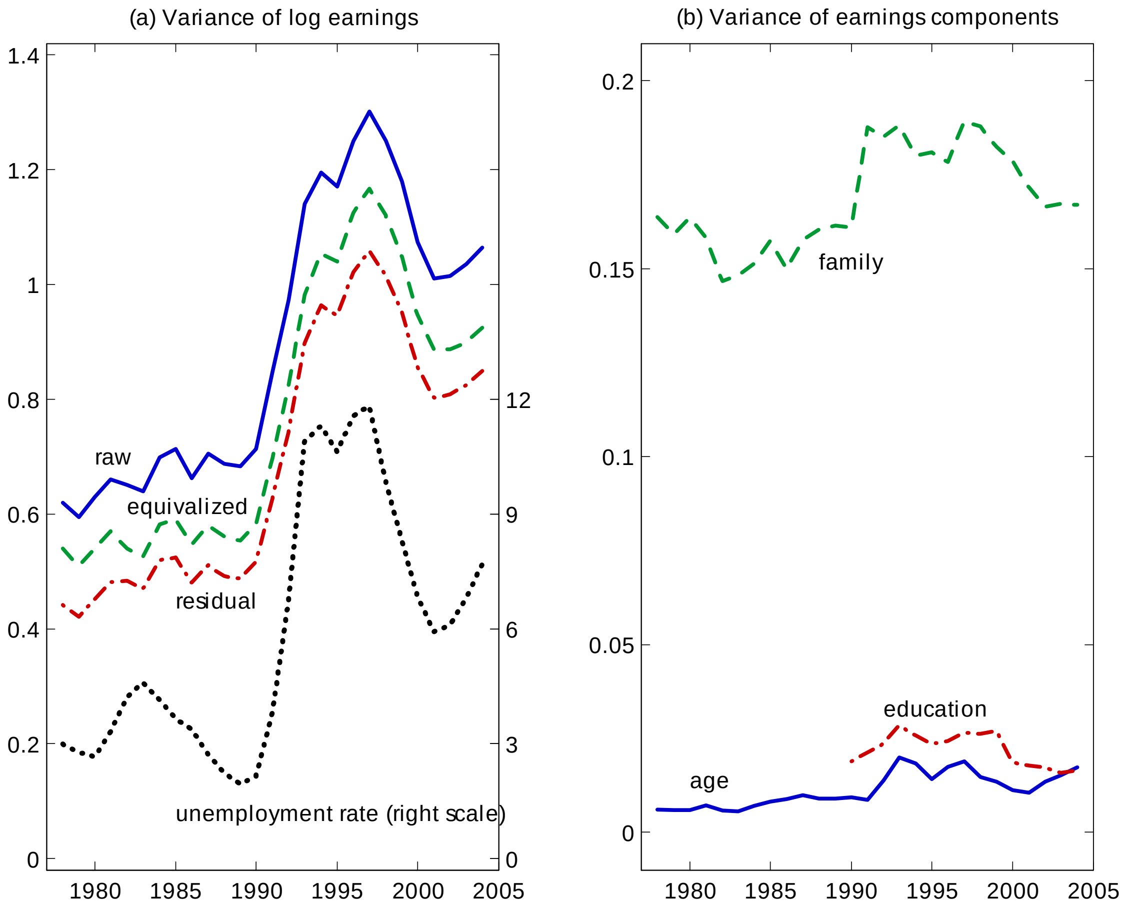
Note: The figure shows wage premia along different dimensions. The solid lines are based on hourly wages from HINK. The dashed lines are based on data from LOUISE. The education premium relates wages for college educated men to non-college educated men, where by college we mean at least three years of university. In HINK we classify individuals as college educated if they belong to socioeconomic groups with typical university education of at least three years, and as low-educated if the typical education is shorter. Socioeconomic groups with no typical education are not included. The experience premium relates the hourly wage of men aged 45-54 to men aged 25-34.

Figure 6: Inequality in labor supply



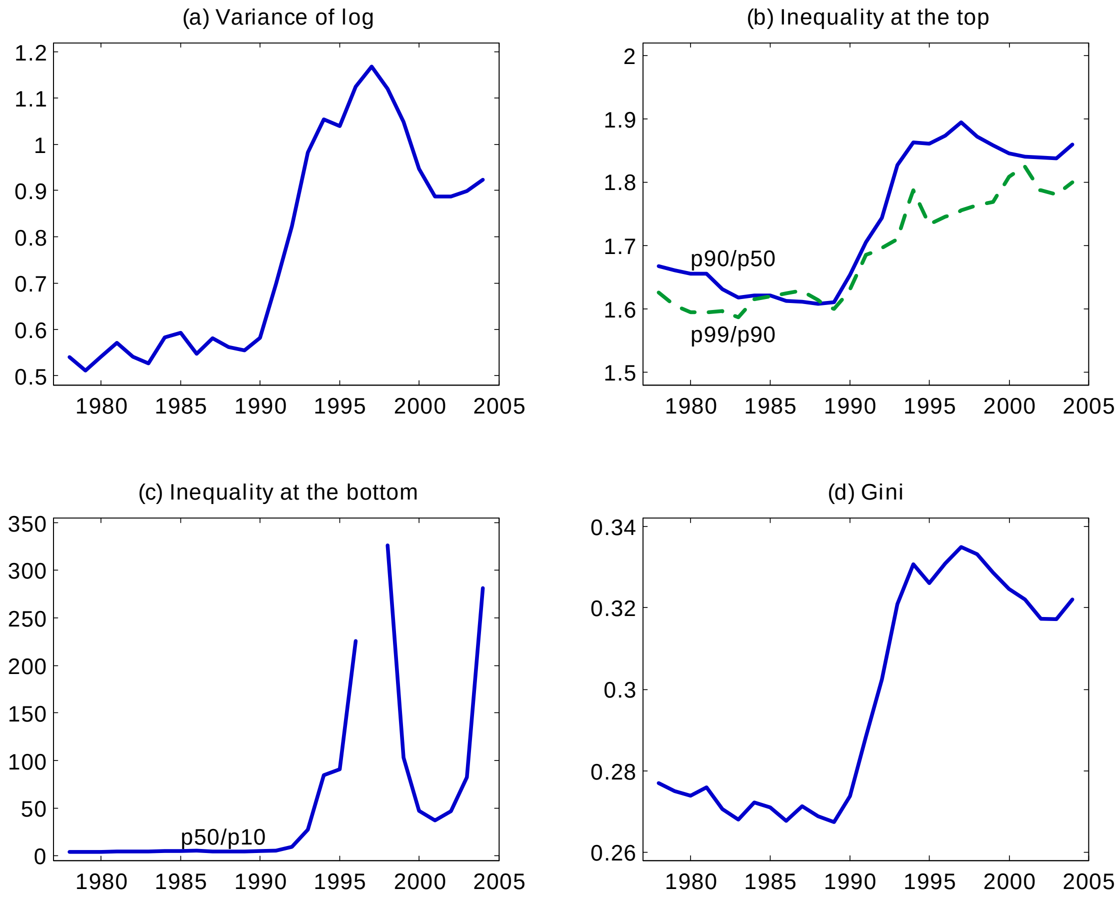
Note: The figure uses hours and hourly wages from HINK. Solid lines refer to men and dashed lines to women.

Figure 7: Earnings inequality and its decomposition



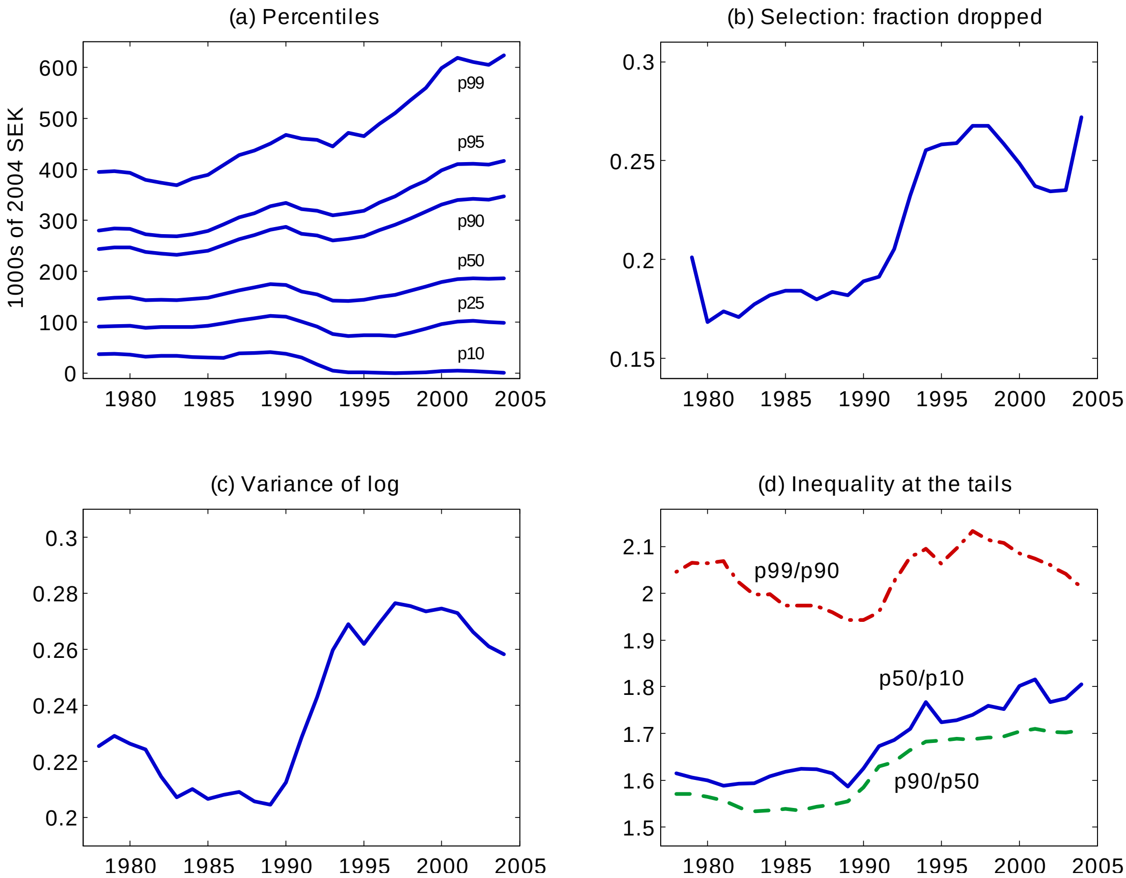
Note: The earnings residual is ε^y as specified in equation (1). The age and family components in panel (b) are the variances of $\beta_{1,t}D^a_{i,t}$ and $\beta_{2,t}D^f_{i,t} + \beta_{3,t}E_{i,t}$ estimated from equation (1), while the education component is the variance of $\beta_{4,t}D^e_{i,t}$ estimated from equation (2).

Figure 8: Inequality in equivalized earnings



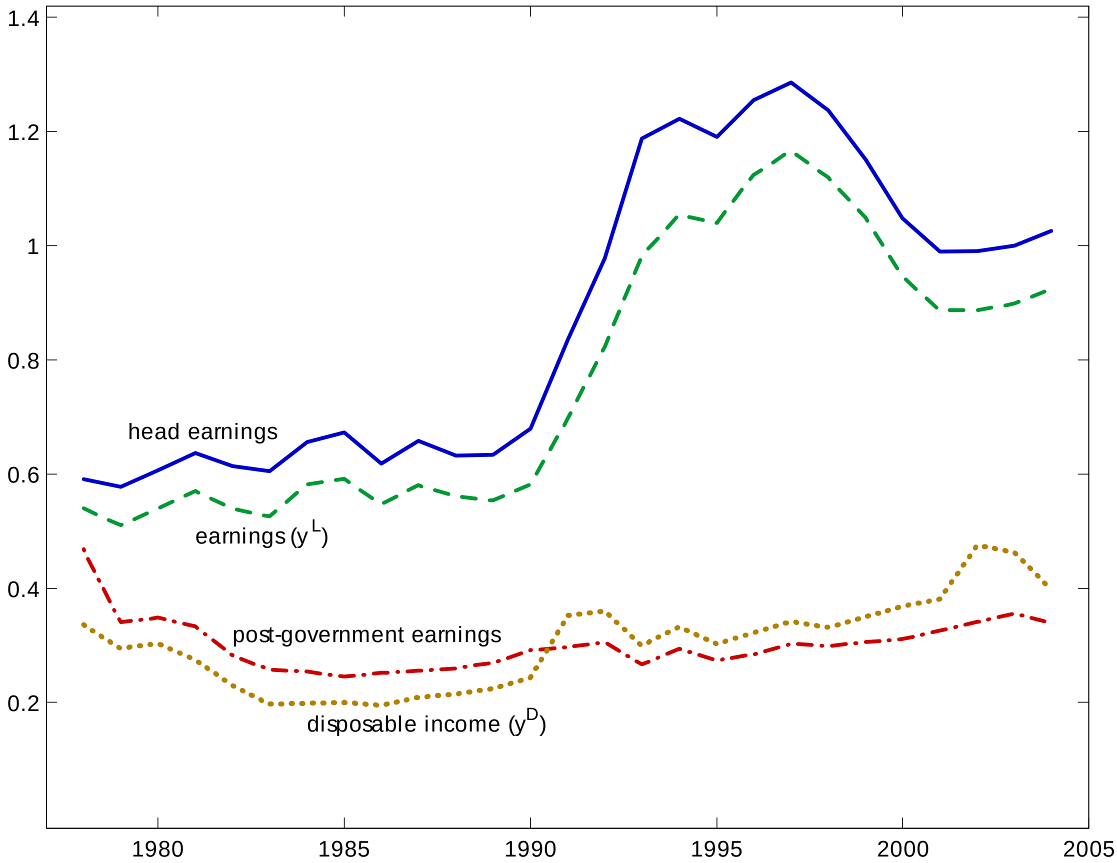
Note: 10.1 percent of the households report zero earnings in 1997, which explains the missing value in panel (c).

Figure 9: The importance of sample selection for inequality in equivalized earnings



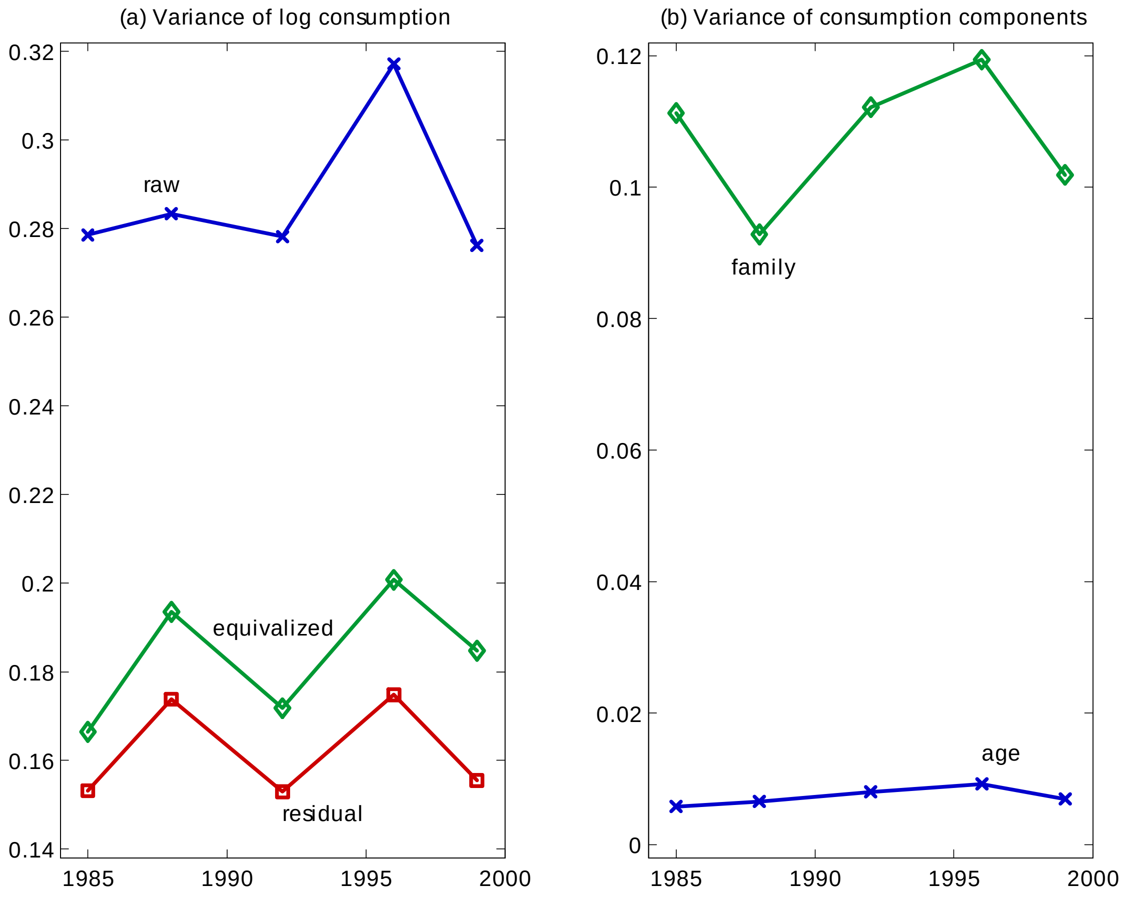
Note: Panel (a) reports earnings at different percentiles in the full sample. Panel (b) shows the fraction of households in the full sample that do not satisfy our benchmark criteria for inclusion in the sample used to estimate earnings processes. Panels (c) and (d) show measures of earnings inequality for that sample.

Figure 10: From earnings to disposable income



Note: The figure shows the variance of log equivalized income measures.

Figure 11: Consumption inequality and its decomposition



Note: Consumption refers to non-durable consumption. See also Figure 7.

Figure 12: Inequality in equivalized non-durable consumption

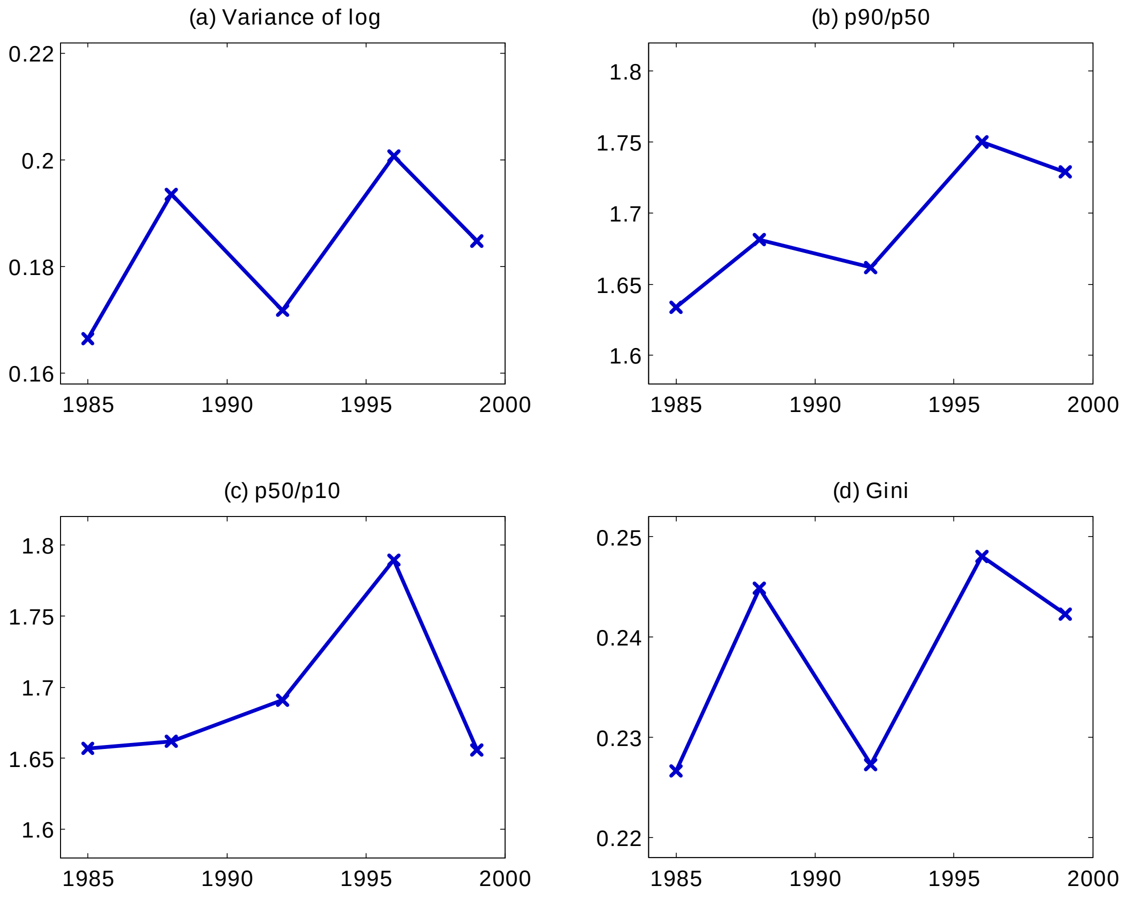
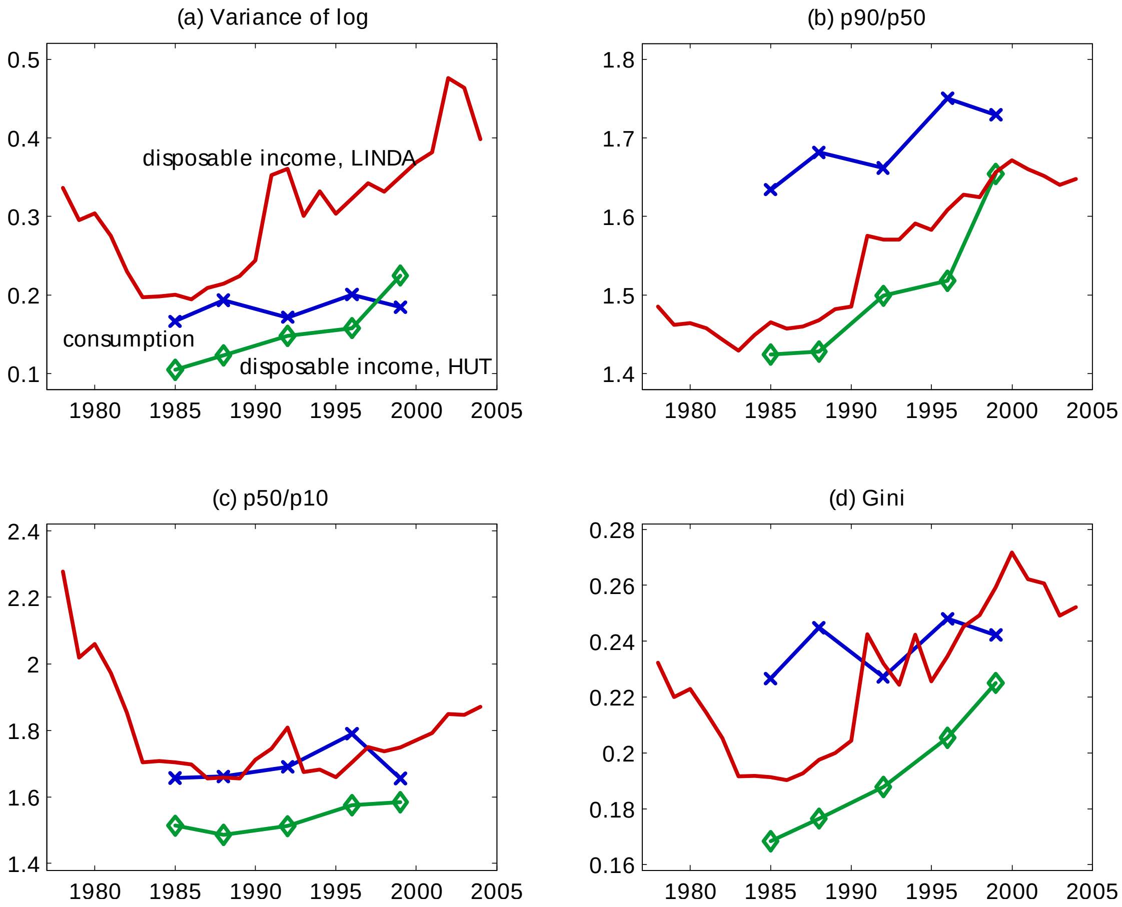
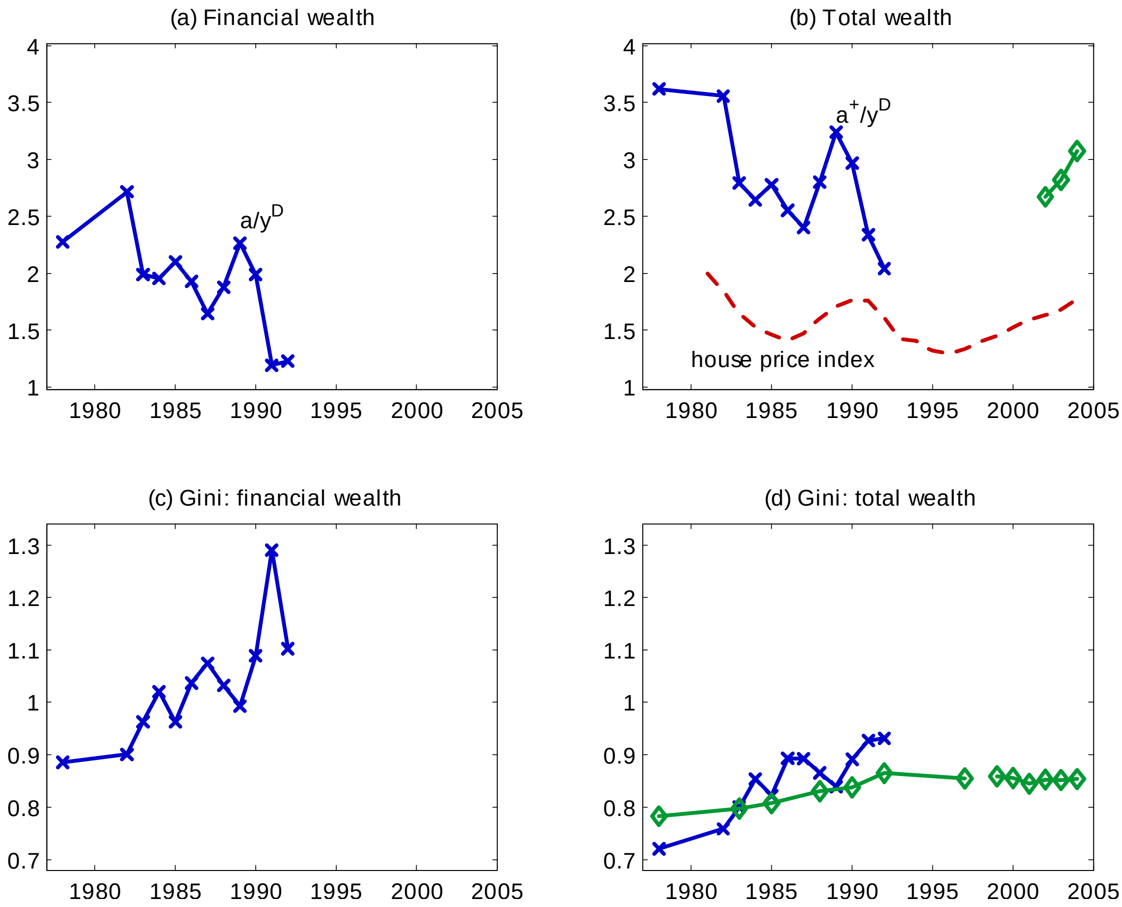


Figure 13: From disposable income to consumption



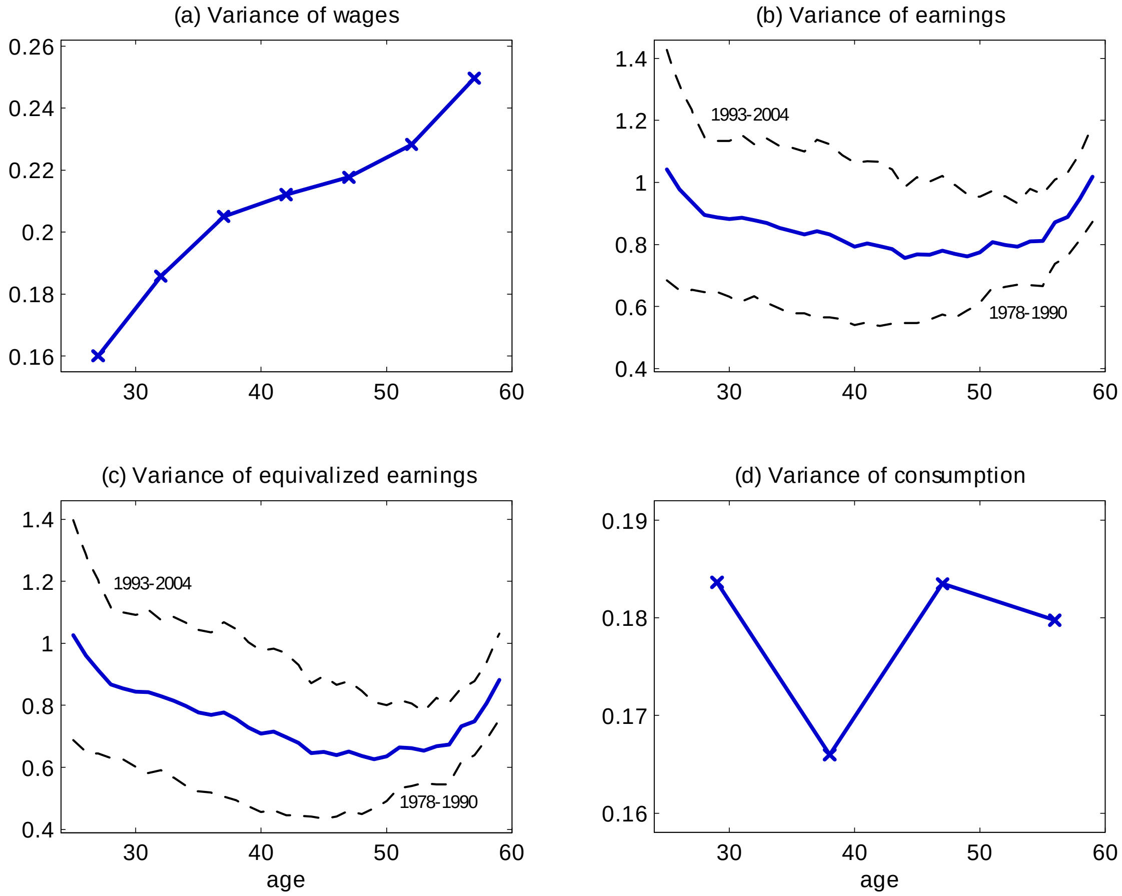
Note: The unmarked lines and lines marked '$\diamondsuit$' are equivalized disposable household income from LINDA and HUT, respectively. The lines marked '$\times$' are equivalized household non-durable consumption from HUT.

Figure 14: Wealth inequality



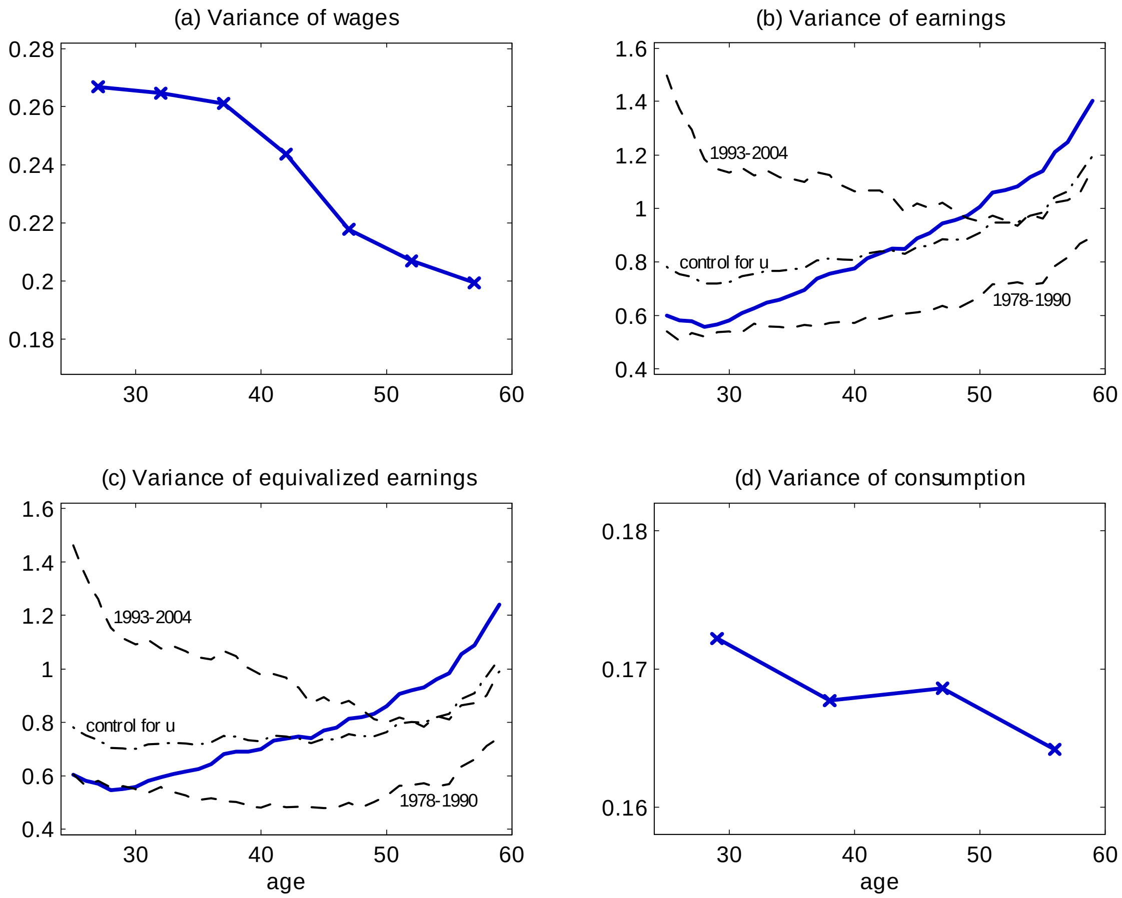
Note: Panels (a) and (b) show net financial and net total wealth relative to disposable income. Lines marked '◇' are based on Statistics Sweden (2000, 2006b). The dashed line in panel (b) shows an index of residential house prices relative to GDP per capita (normalized to 2 in 1981) based on data from Statistics Sweden and our own calculations.

Figure 15: Inequality over the life-cycle (time effects)



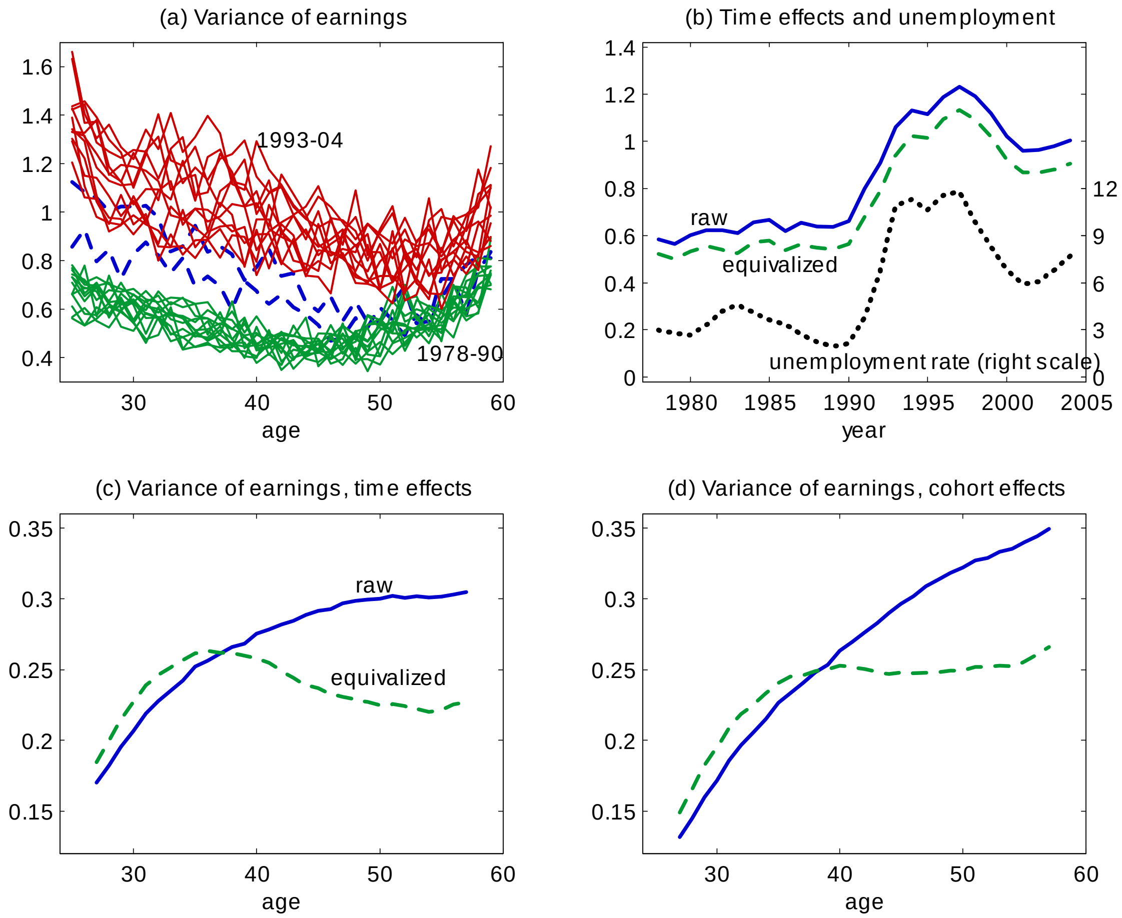
Note: The figure shows various measures of inequality over the life-cycle when controlling for time effects. In panels (b) and (c), the LINDA sample has also been split in the subperiods 1978-1990, and 1993-2004.

Figure 16: Inequality over the life-cycle (cohort effects)



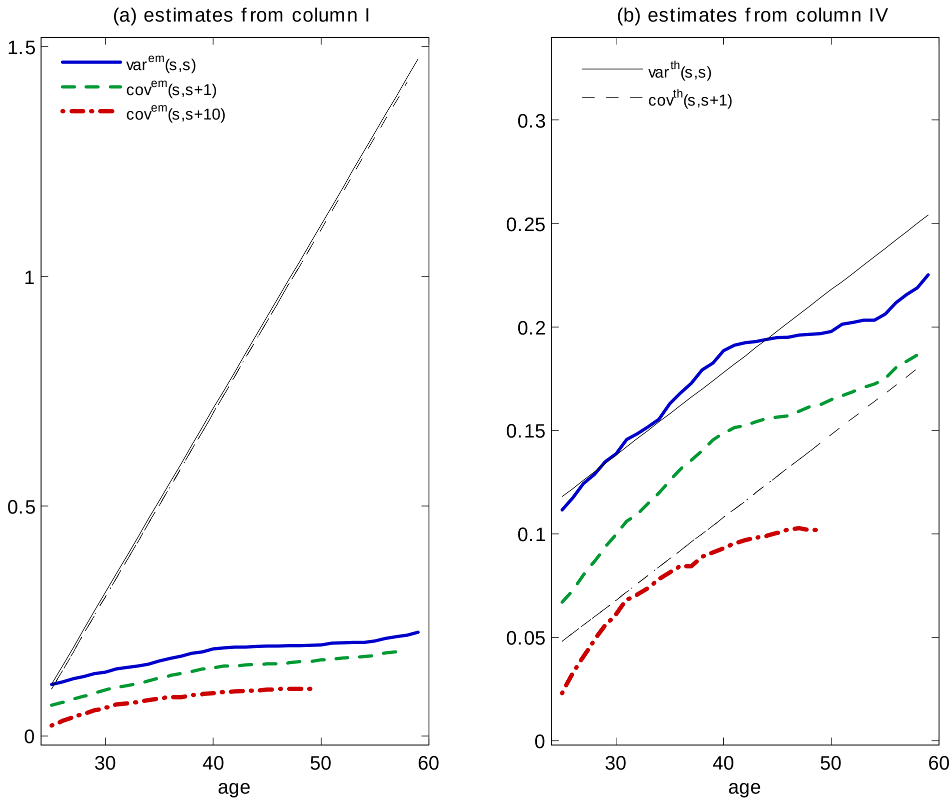
Note: The figure shows various measures of inequality over the life-cycle when controlling for cohort effects. In panels (b) and (c), the LINDA sample has also been split in the subperiods 1978-1990, and 1993-2004. The line marked 'control for u' controls for the unemployment rate in addition to cohort effects.

Figure 17: Time and cohort effects



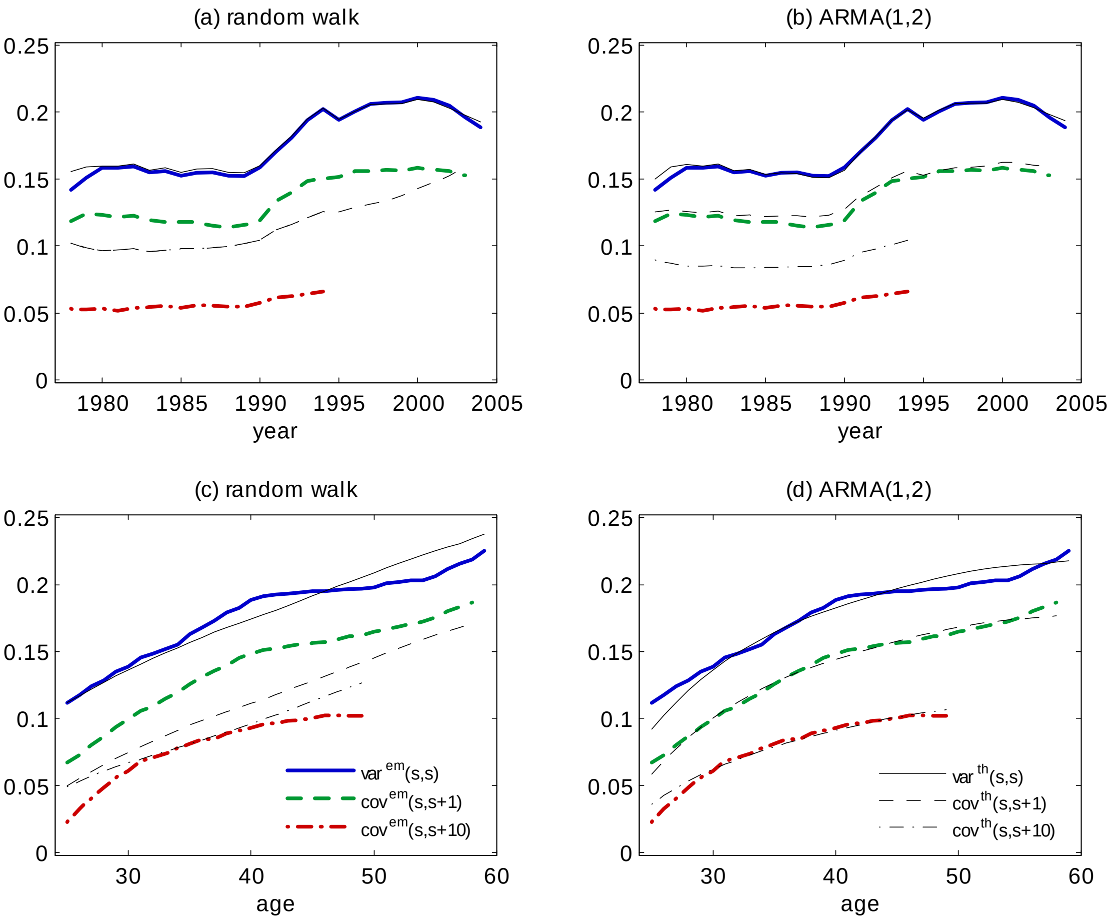
Note: Panel (a) shows the evolution of inequality in pre-government earnings by age over time. Each line shows the cross sectional variance for a particular year; the two dashed lines are for 1991 and 1992, the lower solid lines are for 1978 to 1990, and the upper solid lines for 1993 to 2004. Panel (b) shows the time effects estimated from equation (3) for raw and equivalized pre-government earnings (left scale) and the unemployment rate (right scale). Panels (c) and (d) show the life-cycle profiles of inequality for the benchmark estimation sample used in Section 5, i.e. a sample that excludes households that are not strongly attached to the labor market.

Figure 18: Empirical life-cycle and time-series moments and implications from estimates in Table 5



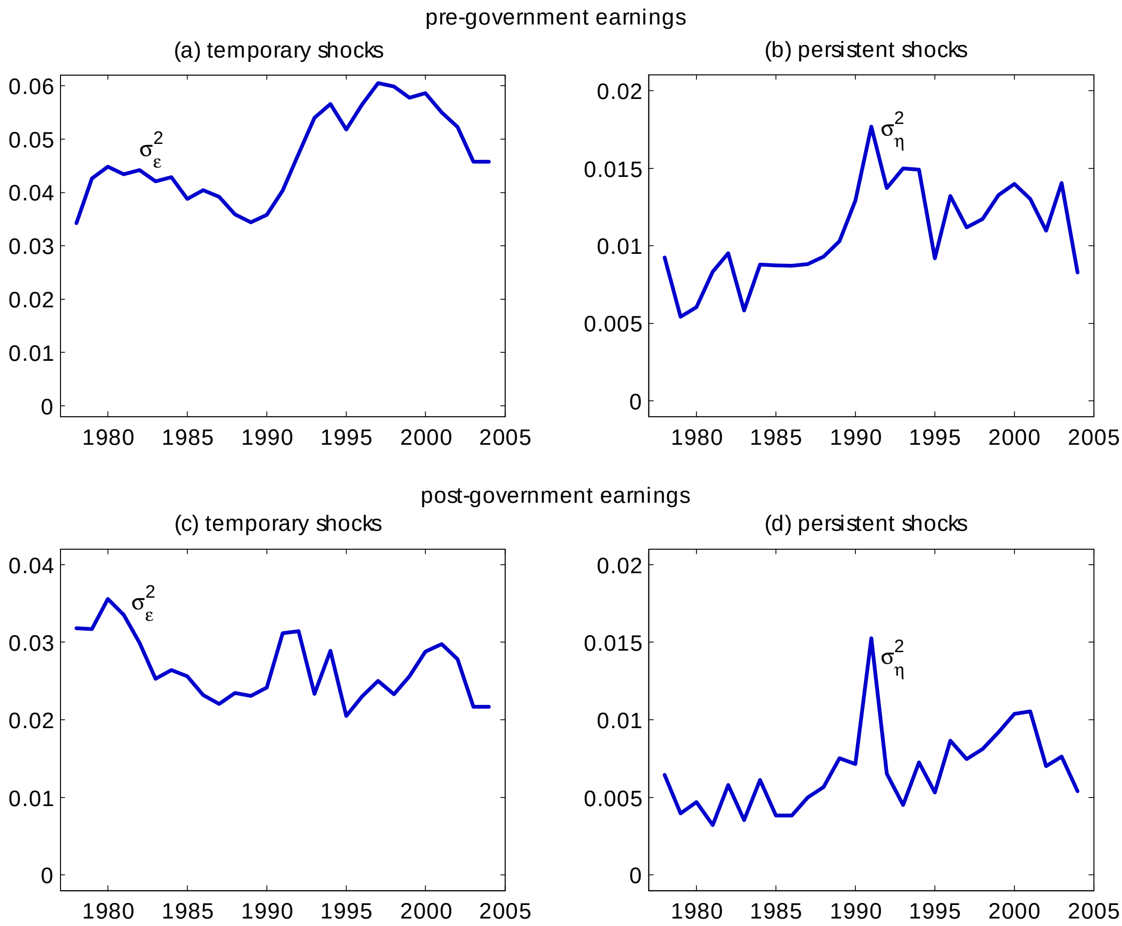
Note: Empirical moments (with superscript ^{em}) and theoretical moments implied by the estimated random-walk process (4)-(6) (with superscript th) for pre-government earnings using different moment conditions. Panel (a) is based on difference moments as in (7)-(8) and column I in Table 5, whereas panel (b) uses the level moments as in (11)-(12) and as in column IV in Table 5. Both panels report the same empirical moments, which were aggregated over years (as in Appendix A.2).

Figure 19: Implications of estimated processes for pre-government earnings



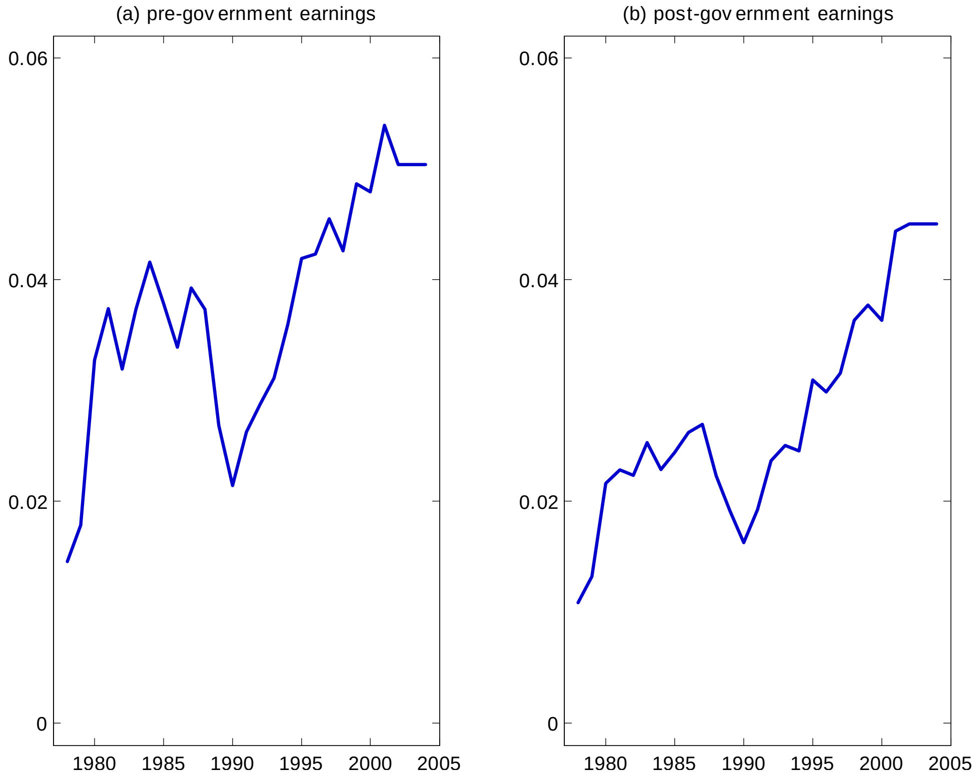
Note: The figure report implications of the process (13)-(15) estimated using both time-series and life-cycle moments, and allowing for year–specific temporary and persistent shocks. Panels (a) and (c) are based on column I in Table 6 whereas panels (b) and (d) are based on column IV in Table 7. See also Figure 18.

Figure 20: Estimated year-specific variances



Note: The figure shows estimates of time-varying variances of temporary shocks $\sigma^2_{\varepsilon,t}$ in panels (a) and (c), and persistent shocks $\sigma^2_{\eta,t}$ in panels (b) and (d). Estimates are as in columns IV and VIII in Table 7 for pre- and post-government earnings, respectively.

Figure 21: Estimated cohort-specific fixed effects



Note: The figure shows estimates of cohort-specific fixed effects, $\sigma^2_{\beta,t-h}$, for the variance of earnings assuming an ARMA(1,2) process.