Five Empirical Studies on Income Distribution in Sweden

Mårten Palme

AKADEMISK AVHANDLING

som för avläggande av filosofie doktorsexamen
vid Handelshögskolan i Stockholm
framlägges till offentlig granskning
fredagen den 29 oktober 1993
kl. 13.15 i sal Ragnar
Handelshögskolan, Sveavägen 65

STOCKHOLM 1993

Five Empirical Studies on Income Distribution in Sweden

Mårten Palme

A Dissertation for the Doctor's Degree in Philosophy (Economic Statistics)

Stockholm School of Economics 1993

Five Empirical Studies on Income Distribution in Sweden

Papers included:

1. Income Distribution Effects of the Swedish 1991 Tax Reform: An Analysis of a Microsimulation Using Generalized Kakwani Decomposition.

2. The Effect of Economic Incentives on Worker Absenteeism: An Empirical Study Using Swedish Micro Data.

Coauthor Per Johansson.

3. Gender Discrimination and Compensating Differentials in Sweden.

Coauthor Robert Wright.

4. Changes in the Rate of Return to Education in Sweden: 1968-1991.

Coauthor Robert Wright.

5. Earnings Mobility and Distribution: Comparing Statistical Models on Swedish Data.

1 Introduction

In almost all international comparisons of income inequality in the 1980s, Sweden has been found to have the least inegalitarian income distribution (see Barr, 1992, for a recent overview). It is likely that this outcome is very much affected by the policies pursued, primarily in the 1960s and 1970s, by major trade unions and the government. The blue collar workers' Swedish Trade Union Confederation (LO) pursued a "solidaristic wage policy" which was aimed at levelling wage rates within and across occupations. As shown by e.g. Hibbs (1990), this policy led to a compression of the distribution of earnings among blue collar workers. Government policies have affected the post tax and transfer income distribution by several means. The implementation of a compulsory social insurance system (e.g. the sickness insurance in 1955 and the pension scheme gradually in the 1960s) and certain allowances (housing allowances and continuous increases in the amounts of child allowances) are important in this respect. Furthermore, the government pursued a policy of increasingly progressive income taxation throughout this time period. Some of these policies have been reversed in the 1980s and early 1990s, with the decentralisation of wage negotiations and a series of tax reforms. However, recent research indicates that very little has happened to the income distribution (see Gustafsson and Palmer, 1993; Gottschalk, 1993), at least until the end of the 1980s.

For a person with preferences against income inequality, this may be seen as a favourable state of affairs. On the other hand, the policies described above have, apart from ethical considerations, been criticised for two kinds of disadvantages that have been extensively investigated in empirical economic studies on the welfare state. First, it is likely to affect economic incentives and efficiency. Progressive income taxes and social security (e.g. the sickness insurance) have effects on incentives to labour supply and work effort. In addition, a compressed wage structure is likely to decrease the economic incentives for individuals to develop skills that may increase the economic returns from their work. Second, the policies described above may level out earnings differences that have arisen on the labour market as compensations for more broadly defined inequalities, such as undesirable job characteristics. Decreasing monetary compensation for these characteristics may thus actually accentuate inequality. Furthermore, some income differences are levelled out when the income concept is extended from annual incomes to incomes earned over a longer time period – levelled differences in annual incomes may thus also entail increased inequality in total incomes over longer periods.

The discussion thus far suggests that studies on income distribution in the welfare state have a broad substantive and conceptual scope. In this sense all five papers included in this thesis deal with income distribution analysis. Apart from this similarity, the five studies also share the common property of being empirical studies analysing micro-level data. The first study uses the *HINK* sample provided by Statistics Sweden, while the other four use the *Swedish Level of Living Survey* (SLLS) conducted by the Swedish Institute for Social Research, Stockholm University. The first study employs income distribution measures, while the other four use statistical models as empirical method. The papers can be classified into different areas of empirical economics: the first two can be classified as studies in public economics, since they analyse public sector interventions, the tax system and the compulsory sickness insurance. The last three are best described as studies in labour economics, since they analyse wage and earnings distributions.

2. Summary of the Thesis

The *first* study looks at the income distribution effects of the 1991 Swedish tax reform. The 1989 income distribution is compared with a hypothetical distribution obtained with the microsimulation model designed at the Swedish Ministry of Finance. This model only considers the direct effects of changes in the tax/transfer systems, in other words it does not model any behavioural responses to such changes. The income distribution effects of the changes in the tax and transfer systems are analysed separately using a decomposition technique first proposed by Kakwani (1984). This decomposition breaks down the difference between the Gini coefficients for the pre- and post-tax/transfer income distributions into three components:
(1) A component measuring the degree of tax progressivity (or transfer regressivity). For taxes, progressivity is measured as the fraction of the total tax burden shifted from relatively low income earners to relatively high, compared to a proportional tax system. (For transfers, regressivity is measured as the faction of total transfers shifted to the relative low income earners, compared to a proportional transfer system.)
(2) A component measuring the overall mean tax rate (the overall mean share of incomes received from transfers).
(3) A component measuring the horizontal equity of the tax/transfer systems, i.e. rerankings induced by the respective system in the transition from pre- to post-tax or transfer income.
In further analyses of these decompositions, we apply the generalization of the Gini coefficient proposed by Yitzhaki (1983), where different value judgements on aversion to inequality in different parts of the income distribution could be considered in the aggregation. Five main conclusions emerged from this first study. (1) The distribution of pre tax and transfer household equivalent income is unambiguously more equal under the post-reform regime. (2)

The increased amounts of child and housing allowances (a part of the 1991 reform) have a substantial equalising effect on the household income distribution. (3) The pre-reform tax system has a larger vertical equalising component (the progressivity and rate components taken together). The decreased progressivity is the most important explanation to this change in the lower end of the income distribution, while the decreased tax rate explains more of the difference in the upper end. (4) The reform increased the horizontal equity of the tax system. (5) The observer must make a value judgement on how to aggregate different parts of the income distributions, to decide which distribution of household disposable incomes is to be considered most equal.

The *second* study investigates whether, and to what extent, economic incentives affect work absence among blue collar workers in Sweden. In setting up the economic model we assume that the choice of whether to attend work or not is made as day-to-day decision. In other words, we assume that the worker makes this decision every day contingent on his perception of his health, contracted work hours, income and the price of being absent. The individual cost of being absent is the share of the individual's daily earnings not covered by the sickness insurance. We also consider other costs of attending work (e.g. risk exposure and undesirable job characteristics) and not attending work (cost of increased probability of loosing the job, assumed to increasing with the unemployment rate). A linear demand function for work absence is derived. As demand for work absence could not be fully observed, i.e. it could only be observed for full days, the demand for work absence had to be modelled as a threshold value function, with a logit specification for the probability of attending work each day.

Our data-set only permits us to observe the number of days each individual has been absent from work during a year, thus, we have a time-aggregation problem. Under the assumption of no serial correlation and no unobserved heterogeneity, the model could be estimated using binomial maximum likelihood. However, there are reasons to believe that both serial correlation and unobserved heterogeneity can be present. To relax the assumption of no serial correlation, a Markov chain model is estimated using non-linear least squares. To accommodate the problem of unobserved heterogeneity, a mixture distribution model is formulated and estimation is carried out using a semiparametric estimator.

We use two data-sets in this second study – a sample of around 2,000 blue collar workers from the SLLS interview year 1981, as well as an extension of this sample to an annual panel 1981-1984. In the results from the second data-set, Slutsky conditions were not unambiguously fulfilled for either of the models. For the 1981 sample, however, they were fulfilled in the results from the model shown to be best by both the Akaike and the Bayes information criterion: the mixing distribution model with two distinct groups. The results also show that the local unemployment rate has a significant inverse effect on work absence and risk exposure a significant positive effect. The results obtained from the 1981 sample using the

mixing distribution model were used to simulate changes in work absence between 1990 and 1991. It was found that the model could explain 72 per cent of the actual changes. About 79 percent of the predicted changes is attributed to alteration of compensation levels in the sickness insurance and the rest to changes in the unemployment rate.

The *third* study examines the hypothesis that compensating wage differentials can explain male–female wage differentials. The data are obtained from the 1981 Swedish Level of Living Survey. Separate hedonic wage equations are estimated for males and females and since wage offers are more adequate than observed wages when studying discrimination, corrections sample selection bias are made using Heckman's (1979) method. To correct for possible endogeneity of the "undesirable job characteristic" variable, we use an instrumental variable method. Very small compensating wage differentials were found, and when using the instrumental variable technique, the coefficient estimates obtained were negative, i.e. not consistent with theory. The overall conclusion of the study is that gender differentials in working conditions are not important in explaining of the wage differentials between men and women.

The *fourth* study examines changes in mean wage differentials between individuals with different educational attainments, one of the main determinants of the rate of return to education, between 1968, 1981 and 1991 in Sweden. We concentrate on three issues concerning these differentials: (1) Male-female differentials; (2) Wage differentials due to education "quantity" (i.e. years of schooling completed); and (3) Differences due to education "quality" (i.e. education level obtained). Furthermore, two methodological issues are investigated: (1) To what extent does correcting for sample selection bias (using Heckman's 1979 method) affect the estimates? When comparing wages of individuals with different educational attainments, the most relevant wage concept is (again) wage offers rather than observed wages – if one has a further interest in returns to education. However, an additional question is raised as to whether this distinction is practically relevant any longer, given the increase in female labour force participation. (2) To correct for small sample errors, a very common method is to use the Mincerian quadratic earnings function. However, recent research (see Murphy and Welch, 1990) has shown that this earnings function is overly restrictive and can therefore lead to biased estimates of for example wage differentials due to educational differences. Therefore, we will compare the results obtained with the restrictive Mincerian quadratic earnings function with those of a more flexible functional form (a cubic spline function). Five main conclusions emerge from the analysis. (1) The rate of return to education declined considerably between 1968 and 1981. However, since 1981 there has been little change. (2) The magnitude of the decline was not the same across all levels of education. The decline was particularly pronounced for university education. (3) The decline was not equally shared between men and women. The decline was larger for women. (4) The correction for

sample selection bias did not change the results considerably. The only possible exception is indicated by the results for females 1968. (5) The quadratic wage function and the cubic splines give approximately the same results, except when there was an interaction effect of formal schooling and work experience on wages. This was true for the results from the 1968 sample for both males and females.

The *fifth* study compares the human capital and purely stochastic models for predicting earnings mobility and distribution. There are several motives for extending income distribution analysis to include over time. Most income distribution data-sets cover only a few years. However, statistical models for predicting earnings mobility can be used to extend the data-set thus creating a "pseudo-panel". The statistical models which could be used can be classified into human capital models and stochastic models. According to the stochastic theory for earnings mobility a cross-section income distribution could be rationalised by a stochastic process for earnings mobility. The human capital theory asserts that observed individual earnings can be seen as the outcome of accumulation in individual investment. Investments could be made in formal schooling or work experience – on-the-job-training. The purpose of this fifth study is to compare the usefulness of different versions of these models in order to predict future individual earnings and earnings distributions. The data used is a 20-year panel of 651 employed Swedish males aged 23-65 throughout the sample period. The models are estimated on 17 annual observations for each individual and predictions are made for the last three years. The results show that the human capital model gives reasonably good predictions of earnings distributions if individual variances are calculated.

References

Barr, N. (1992). Economic Theory and the Welfare State: A Survey and Interpretation. *Journal of Economic Litterature*, XXX, 741-803.

Gottschalk, P. (1993). Changes in Inequality of Family Income in Seven Industrialized Countries. *American Economic Review, Papers and Proceedings*, 136-142.

Gustafsson, B. and Palmer, E. (1993). The Development of the Distribution of Well-being in Sweden during the 1970s and 1980s. Mimeo Gothenburg University.

Heckman, J. J. (1979). Sample Selection as a Specification Error. *Econometrica*, 47, 153-161.

Hibbs, D. A. (1990). Wage Compression Under Solidarity Bargaining in Sweden. Economic Research Report no. 30. Trade Union Institute for Economic Research. Stockholm.

Kakwani, N. C. (1984) On the Measurement of Tax Progressivity and Redistributive Effect of Taxes with Application to Horizontal and Vertical Equity. *In Advances in Econometrics 3*, JAI Press, Greenwich Conn.

Murphy, K. and Welch F. (1990). Empirical Age-Earnings Profiles, *Journal of Labor Economics*, 8, 202-229.

Yitzhaki, S. (1983). On an Extension of the Gini Index. International Economic Journal, 24, 617-628.

Income Distribution Effects of the Swedish 1991 Tax Reform: An Analysis of a Microsimulation Using Generalized Kakwani Decomposition*

by

Mårten Palme$

Abstract: In 1991 a major tax reform was implemented in Sweden. This study compares the income equalization (income redistribution), vertical and horizontal equity properties of the pre and post reform tax and transfer systems. The method used is a decomposition, first proposed by Kakwani (1984), of the difference between the generalized Gini coefficients of the pre- and post-tax/transfer income distributions. The post-reform income distribution was obtained through a microsimulation carried out by the Ministry of Finance. Several findings emerge from the study. (i) The distribution of original incomes is unambiguously more equal under the post-reform regime. (ii) The increased amounts of child and housing allowances have a considerable equalizing effect. (iii) The pre-reform tax system has a larger vertical redistributive (equalizing) component. The decreased progressivity is the most important explanation in the lower end of the distribution, while the decreased tax rate explains more of the difference in the upper end. (iv) The reform increased horizontal equity. (v) It is a value judgement which of the distributions of disposable income are the most equal. The first four of these main findings apply when the calculations were redone within each separate household group. The overall methodological conclusion is that much more information was gained when different generalizations of the decomposition were considered.

* I am indebted to Ingemar Svensson for collaboration in the project which this paper is a part of; and Ingemar Eriksson at the Ministry of Finance for providing the data. Anders Björklund, Clas Cassel, Björn Gustafsson, Kjell Nyman, Emma Rothschild, Erik Ruist, Brita Schwarz, Amartya Sen, Anders Westlund and Robert Wright have commented on one or several drafts of this paper. Comments from partisipants on seminars at The Institute for Social Research at the University of Stockholm, Univesity of Umeå, FIEF and Stockholm School of Economics are also acknowledged. Of course, all errors and shortcomings are the sole responsibility of the author. This project was partially financed through the research project "Income distribution in Sweden" which is supported by The Bank of Sweden Tercentenary Foundation and the Swedish Council for Research in the Humanities and Social Sciences.

$ Department of Economic Statistics, Stockholm School of Economics, Box 6501, S-113 83 Stockholm, Sweden

1. Introduction

All of the numerous income tax reforms in Western industrialized countries, implemented in the latter part of the 1980's, share two common properties: (1) a reduction in marginal tax rates and (2) a broadening of the tax base[1]. Prior to the tax reform of 1991 Sweden was in the position of having the highest marginal tax rates in the OECD - on average, between 50 and 75 per cent for the majority of the working population. In the 1991 tax reform marginal tax rates were cut to between 30 and 50 per cent, the tax base was extensively broadened, and child and housing allowances were increased. The policy debate preceding the tax reform dealt, apart from the effect of taxes on economic incentives and behavior, very much with equity considerations and taxation's effects on the income distribution. Advocates for the reform stressed the importance of increasing horizontal equity (equal treatment of income units in an equal position)[2]. The skeptics argued that the reform would compromise the old tax system's vertical equity principle of taxation according to ability to pay, and that income units with unequal tax-paying ability, should be taxed unequally.

Since the three main elements of the tax reform have contradictory impacts on the income distribution, it is an empirical question what the overall end result will be. The purpose of this study is to demonstrate how recently developed techniques can be used to, ex ante, estimate the effects of different parts of the tax and benefit reform on income equalization (income redistribution). The tax and benefit systems are analyzed separately, and the effect on the income distribution of three different properties of the respective systems are estimated. These properties are: (1) the overall mean tax or benefit rate, (2) the progressivity of the tax system and the regressivity of the transfer system and (3) the horizontal equity of each system.

The method used is based on decompositions of the equalizing effect measured as the difference between the generalized Gini coefficients for the distribution of income before and after taxes and transfers. This decomposition method, which makes it possible to estimate the effects on the income distribution of the three properties mentioned above, was first proposed by Kakwani (1984).

The aim of the study is to compare, or order, different properties of the pre- and post-reform benefit and tax systems using aggregate measures. As is shown by Sen (1974), ordering of income distributions with the help of aggregate measures could involve value

---

[1] See Whalley (1990) for a description of tax reforms in seven countries following the U.S. 1986 tax reform.

[2] In the proposal to the parliament for a new tax system the Swedish Minister of Finance writes: "An important goal for the tax reform that I have suggested is that the tax paid should to a greater extent be based on the tax liables income, i.e. the horizontal equity should be improved" (Prop. 1989/90:110 page 630).

judgements of where inequality matters most in the income distribution. The same is obviously true for summary measures of tax/benefit progressivity/regressivity and measures of horizontal equity. To make these value judgments explicit, we will use a parametric generalization of the Lorenz based measures, investigated by Yitzhaki (1983), and consider different generalizations of the decomposition.

The data used is from a microsimulation[3] carried out by the *Swedish Ministry of Finance*. The original data base is the 1987 *Household Income Survey* provided by *Statistics Sweden*. The rules that are used in the microsimulation are the outcome of the agreement made in the autumn of 1989 between the Social Democratic government and the Liberals. To obtain the pre-reform income distribution, the original 1987 data has been transformed to correspond to 1991 levels of wages, prices, taxes and transfers. The post-reform distribution has been obtained through a simulation using the "new" tax/transfer rules on the 1991 income distribution data. Household economic behavior, labor supply and savings, were assumed to be fixed. Indirect economic effects of the tax reform, or welfare changes due to changes in distortions of households choices between labor and leisure, could unfortunately not be taken into account. However, a one per cent increase in labor income for all individuals is assumed in order to account for "increases in labor supply or because the functioning of the economy is improved" (Eklind and Johansson (1989)).

Several conclusions emerge from this study. (i) The distribution of original[4] income is unambiguously more equal under the post-reform regime. (ii) The increased amounts of child and housing allowances have a considerable equalizing effect on the income distribution. (iii) The pre-reform tax system has a larger vertical redistributive (equalizing) component than does the post-reform tax system. Decreased progressivity is the most important explanation for this change in the lower end of the income distribution, while the decreased tax rate explains more of the difference in the upper end of the distribution. (v) It is a value judgement which of the distributions of disposable income are the most equal. The first four of these main findings apply when the calculations were redone within each separate household group.

---

[3] The methods of the microsimulation are presented in Eklind and Johansson (1989).

[4] Here defined as gross income less child and housing allowances.

Schwartz and Gustafsson (1991) have previously carried out a microsimulation study of the income distribution implications of the Swedish 1991 tax reform. Their focus is, however, somewhat different - they concentrate on the distribution of disposable income. This study considers the impact of the change of different properties of the tax and benefit systems on the income distribution. The study also has a second potential contribution to make: it is the first time, to the authors knowledge, that the generalized Kakwani decomposition is applied to study the income distribution implications of a tax reform. The study shows that more information is gained when the analysis is extended to several generalizations of the decomposition, and that a single generalization could lead to misleading results.

The paper is organized as follows. Section 2 describes the methods used. Section 3 presents the results of the empirical study. In Section 3.2, the complete population - here defined as households with head of household between 25 and 65 - is analyzed and equivalence scales are used to compensate for differences in needs. In Section 3.3 the same analysis as 3.2 is carried out separately on the different types of households. This is done in order to evaluate if the main conclusions from Section 3.2 hold for the within-group income distribution, or wheather they are due to differences between the pre- and post-reform benefit and tax systems treatment of different types of household (i.e. sensitive to the choice of equivalence scales). Conclusions are given in Section 4.

2. A short description of the method

2.1. Measures of the equalizing effect[5]

It is well known that if the Lorenz curve for one income distribution is above the Lorenz curve of another income distribution at every point, the first income distribution is unambiguously more equal than the second. In addition, it is also well known that the first income distribution is preferred to the second based on *any* concave social welfare function[6]. If, however, the Lorenz curves cross, it is not possible to order the income distributions without further value judgements.

This leads to the problem of how to aggregate the information contained in a Lorenz curve into a single measure. In other words, the problem has to do with how one should weight different segments of the income distribution. The ordinary Gini coefficient is most sensitive to c ıges in the part of the income distribution with highest density. To illustrate this implicit value judgement, consider the following formula for the Gini coefficient[7]:

$$G = 1 + \frac{1}{N} - \frac{2(x_N + 2x_{N-1} + \ldots + Nx_1)}{N^2 \mu}, \qquad (1)$$

where N is the number of income units; μ is mean income; and the income units are ordered in descending order. Assume that a transfer of 1 SEK from the income unit with rank h to the income unit k, where h > k, takes place. The summation within the parenthesis will increase by N + 1 - k and decrease by N + 1 - h. G will decrease by $2(h - k)/N^2\mu$. This decrease will be larger the larger the difference in rank, that is, the difference between h and k. By contrast, the difference in monetary units between the income of the income units does

---

[5] In the descriptions of the methods in this section, we will assume that the distribution of gross income is more equal than the distribution of original income and that the distribution of disposable income is more equal than the distribution of gross income. Tax progressivity and benefit regressivity are necessary condition for this outcome. All empirical studies of the income distribution effects of benefit and tax systems that I have seen support this outcome (see, for example, O'Higgins and Ruggles (1981) for the UK, Fritzell (1991) for Sweden and O'Higgins et. al. for several other countries). This assumption is, however, not necessary for using the proposed methods.

[6] Atkinson (1970).

[7] This definition of the Gini coefficient and the example is from Lambert (1989) chapter 2 p. 44.

not matter. This means that the Gini coefficient gives maximum weight to redistributions where the distribution density is highest.

To make this value judgement explicit, Donaldson and Waymark (1980) and Kakwani (1980) proposed the following generalization of the Gini coefficient:

$$G(\nu) = 1 - \int_{o}^{1} \nu\,(\nu-1) \cdot (1-p)^{\nu-2} \cdot L\,(p)\,dp\,, \qquad (2)$$

where p measures the income unit's order in the distribution, ranging from 0 to 1, and $L(p)$ is the Lorenz curve. By choosing different values of the parameter ν, it is possible to express different views about the importance of equality in different segments of the income distribution. In other words, one can consider the implications of choosing different social welfare functions. The properties of this coefficient have been further investigated by Yitzhaki (1983).

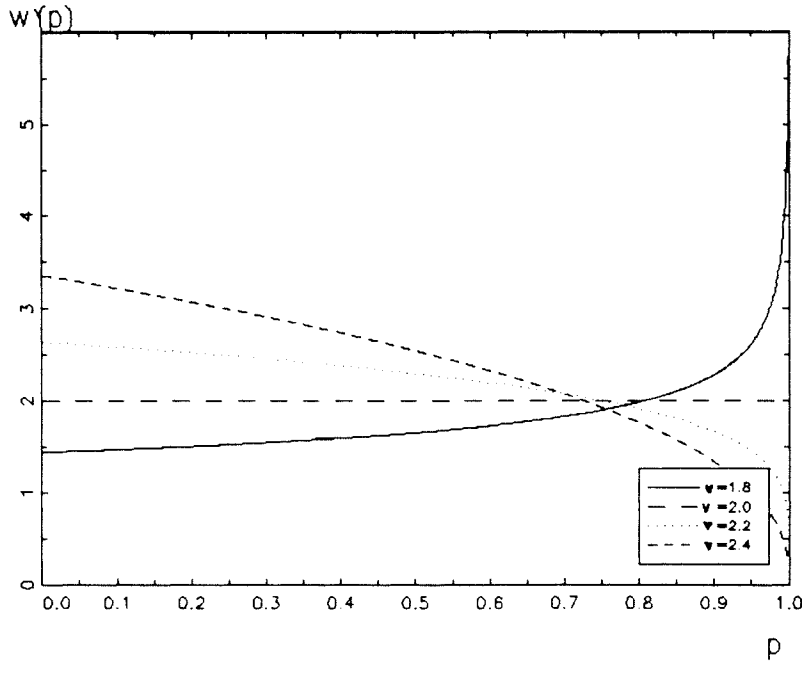
Figure 1 shows the weighting function $w^{\nu}(p)=\nu(\nu-1)(1-p)^{\nu-2}$ for the values 1.8, 2.0, 2.2, 2.4. If $\nu > 2$ the weights will decrease with rank. The opposite is true for $1 < \nu < 2$. G(2) is the ordinary Gini coefficient, when the same weight is given to all income units. As $\nu \rightarrow \infty$, the total weight becomes assigned to the lowest income unit. In this case, the weight structure corresponds to a Rawlsian social welfare function, whereby maximizing the well-being of the worst off is the central criterion for evaluing the desirability of an income distribution.

In this study we will compare the impacts of the pre- and post-reform transfer and tax systems on the income distribution. To do this, we depart from the difference between the Lorenz curves for original and gross income, L_X-L_{X+B}, to analyze the effects of the transfer system, and the difference between the Lorenz curves for gross and disposable income, L_{X+B}-L_{X+B-T}, to analyze the effects of the tax system. We will denote these differences the "equalizing effect" of the transfer and tax systems respectively[8].

If the difference between the Lorenz curves is larger in every point for one system, it is unambiguously more equalizing. If both systems are more equalizing in different parts of the

[8] "Redistribution" is the most common notation for this difference (see, for example Lambert (1989) and Phäler (1987)). I think that notation is misleading for two reasons. First, the two Lorenz curves are not necessarily based on the same ordering of income units. More specifically, the difference for a certain proportion of individuals cannot be interpreted as the share of income redistributed from those on the right side of the point to those on the left, as the two groups do not necessarily consist of the same individuals when comparing the two Lorenz curves. Second, it is not a measure of a balanced budget proposal. What is measured is then shares of *different* sums.

Figure 1. *Weighting functions for the generalized Gini coefficient.*



income distribution, we must make a value judgement as to which system is most equalizing. When aggregating this information into a single measure, we will use the difference between the generalized Gini coefficients to explicitly weight different parts of the income distribution.

2.2. Decomposition of the equalizing effect

In this section we will show how different properties of the transfer and tax systems, and their effect on the income distribution, could be analyzed separately. We will also discuss how these different measures should be interpreted.

In order to study the effect of different properties of the tax system on the income distribution, Kakwani (1984) suggests a method to decompose the difference between Gini coefficients for original and gross income as a measure of the equalizing effect from a tax

system. Lambert (1989) derives a corresponding decomposition for the transfer system. Formula 3 gives the decomposition for the transfer system and formula 4 is for taxes.

$$L_X - L_{X+B} = \frac{b}{1+b}({}_XK_B - L_X) + (L_{X+B} - {}_XK_{X+B}) , \tag{3}$$

$$L_{X+B} - L_{X+B-T} = \frac{t}{1-t}(L_{X+B} - {}_{X+B}K_T) + (L_{X+B-T} - {}_{X+B}K_{X+B-T}) , \tag{4}$$

where b is the overall mean share of transfer income; ${}_XK_B$ is the concentration curve for transfers when the income units are ordered by ascending original income (the first subindex, X, shows the ordering variable, the second, B, the measured variable); ${}_XK_{X+B}$ is the concentration curve for gross income when the income units are ordered by original income; t is the overall mean share of taxes paid; ${}_{X+B}K_T$ is the concentration curve for taxes when the income units are ordered by gross income; and ${}_{X+B}K_{X+B-T}$ is the concentration curve for disposable income when the income units are ordered by gross income.

These decompositions show that the equalizing effects are determined by three properties of the tax or transfer system. They are:

(i) The *departure from proportionality*. A strictly proportional tax or transfer system will not affect the relative income distribution. Two distinct kinds of departure from proportionality may occur. Firstly, the transfer system could be regressive, which means that those households with relatively low income receive a larger share of the sum of transfers relative to their income. Secondly, the tax system could be progressive, meaning that low-income units pay a smaller share of the sum of taxes relative to their to their income than do high-income units. Each of these will have an equalizing effect on the income distribution.

(ii) The *relative volume of total transfers and taxes*. At a given degree of departure from proportionality, the equalizing effect will be larger at a higher average tax/transfer rate. This means that if for example two tax systems are compared, one being more progressive and one having higher overall average tax rates, one could not tell a priori which has the most equalizing effect on the income distribution.

(iii) *Horizontal equity*. If no reranking occurs when moving from the pre- to the post-transfer/tax income distribution, the Lorenz curve for gross/disposable income and the

concentration curve for gross/disposable income will coincide in every point. If at least two income units change place, the curves will depart and the difference between them will be negative in the interval of the reranking. At a given average transfer or tax rate and a given degree of departure from proportionality, the maximum equalization will be in a regime with no rerankings. The more rerankings, the less equalization. As will be discussed below, measures of rerankings could be interpreted as a measures of horizontal equity.

2.2.1. Departure from proportionality

The difference between the concentration curve for transfers and the Lorenz curve for original income, $L_X(p)-{}_{X+B}K_B(p)$, measures the share of total transfer payments shifted to the lowest 100p per cent from the highest 100(1-p) per cent, compared to if all income units received transfers proportional to their original incomes. The degree of departure from proportionality is determined by the distribution of original incomes and the incidence of income units who receive allowances in different parts of the income distribution. However, if the transfer is means tested, the regressivity is also affected by the rules for the transfer and the extent to which the intentions they codify are fulfilled in practice.

Departure from proportionality in the case of taxes (progressivity) can be analyzed in a similar way. If taxes are exactly proportional to the pre-tax income of the income units, the concentration curve ${}_{X+B}K_T$ will be exactly identical to the Lorenz curve for pre-tax incomes. If, however, the taxes paid increase at a rate more than proportional to pre-tax income, the concentration curve for taxes paid will be below the Lorenz curve for pre-tax incomes. The distance $L_X(p)-{}_{X+B}K_T(p)$ is the fraction of the total tax burden shifted from high income units (the top 100(1-p) per cent) to low income units (the bottom 100p per cent) when the tax is progressive rather than proportional[9].

When the departure from proportionality is aggregated into a single new measure (or when ordering two alternative tax/transfer systems), the same problem arises as when measuring the equalizing effect. That is, the outcome is dependent on the value judgement placed on the importance of departure from proportionality in different ranges of the income distribution. The ordinary Kakwani index for departure from proportionality measures the area between the concentration curve and the Lorenz curve. This measure is most sensitive to departures

[9] The definition is given in Lambert (1989) chapter 7 p. 179.

from proportionality in the range of the income distribution where density is highest[10]. In this study we will use the generalized Kakwani measure for departure from proportionality. Formula 5 yields this measure for transfers and formula 6 does so for taxes.

$$\rho^K(\nu) = \nu(1-\nu)\int_0^1 (1-p)^{\nu-2}[{}_XK(p)_B - L_X(p)]dp = G_X(\nu) - {}_XC_B(\nu) , \qquad (5)$$

$$\Pi^K(\nu) = \nu(1-\nu)\int_0^1 (1-p)^{\nu-2}[L_{X+B}(p) - {}_{X+B}K_T(p)]dp = {}_{X+B}C_T(\nu) - G_{X+B}(\nu) , \qquad (6)$$

where ${}_XC_B(\nu)$ is the concentration index for transfers when the income units are ordered by original income and ${}_{X+B}C_T(\nu)$ is the concentration index for taxes when the income units are ordered by gross income.

2.2.2. Horizontal equity

The operation of any tax-and-transfer system involves a transition from an ex ante to an ex post income distribution. For the case of a tax system, the transition from pre-tax to post-tax income distribution, Feldstein (1976) suggests the following formulation of the principle of horizontal equity: "If two individuals would be equally well off (i.e. have the same utility level) in the absence of taxation, they should also be equally well off if there is a tax", i.e. equal treatment of equals. The motives for studying horizontal equity have been extensively debated during the last decades[11]. Musgrave (1991) investigates the requirements of horizontal equity under various formulations of distributive justice and different principles of taxation, ranging from the Lockean *entitlement view* and the *benefit approach* to taxation to *utilitarianism* and the *ability-to-pay approach* to taxation[12]. Musgrave finds that "... the

[10] This could be shown in a similar way as was done above for the ordinary Gini coefficient.

[11] See for example Musgrave (1959, 1991), Feldstein (1976), Atkinson (1980) and Kaplow (1989).

[12] The three different variants of utilitarian tax principles that he considers are J.S. Mills *equal absolute sacrifice rule* and the ability-to-pay approach to taxation, the view of *equal marginal sacrifice rule,* and the "neo-utilitarian" view originally proposed by Vickrey and Harsanyi.

requirement for horizontal equity remains essentially unchanged under the various formulations of distributive justice ..." (Musgrave 1991, p. 116). But is there a case for *measuring* horizontal equity?

According to the entitlement view, the pre-tax and transfer distribution has an intrinsic ethical justification. This has led many authors to claim that a preservation of the initial utility ordering is a central concern of the advocates of this view of justice[13]. On the other hand, the argument that natural law entitles a person to keep what he earns in the market rules out redistribution. Taxes could still be needed to finance social goods, but according to the entitlement logic all should pay an amount equivalent to the benefit received, i.e. the distributions of burdens will be totally dependent on income and substitution elasticities of demand for social goods. Horizontal equity will thereby be achieved but measures of horizontal equity will not have any meaningful interpretation. The same is true for all other theoretical models of distributive justice, based on the ability-to-pay view of the distribution of the tax burden. Musgrave (1991) concludes that horizontal inequity is a second-best problem, horizontal inequity arises because the tax base in reality is imperfectly defined. In other words, the intentions of the tax system are not achieved in reality.

The measure of horizontal equity that we will use measures rerankings. A reranking occurs if an individual does not maintain her rank when comparing the distributions of original and gross income or gross and disposable income. The principle of "equal treatment of equals" has, as King (1983) suggested, been extended to "and unequals treated accordingly".

It is possible to measure the extent of reranking with Lorenz and concentration curves[14]. Let us first consider rerankings resulting from taxation. ${}_{X+B}K_{X+B-T}$ is the concentration curve for disposable income where the income units are ordered in ascending order according to their pre-tax (gross) income. Kakwani (1980) has proved that this curve will be above the conventional Lorenz curve for disposable income,

[13] See for example Atkinson (1980) and Atkinson and Stiglitz (1980).

[14] This measure of horizontal equity was originally used by Atkinson (1980). See Plotnick (1985) for a comparison of different measures of horizontal inequities.

$$_{X+B}K_{X+B-T} \geq L_{X+B-T} \quad \textit{for all} \quad 0 \leq p \leq 1 .$$

If the ranking is the same for the pre- and post-tax income distributions, these two curves will coincide in every point. If the rank is changed, the inequality will be strict for some points, and the more the rank is changed, the larger will be the distance between the curves.

Kaplow (1989) criticizes the reranking measures of horizontal equity for not adequately measuring the degree of horizontal equity: "Minute movements leading to order reversals count as full violation of horizontal equity while substantial disturbances in the initial distribution that result in no order reversal are ignored" (page 141). In the ranges of the income distribution with the highest density, departures from horizontal equity have (a priori) the highest probability of resulting in rerankings. If the area between the concentration curve and the Lorenz curve for disposable income is used as an aggregate measure of horizontal equity[15], too much weight is given to the high density areas of the income distribution. In this case, horizontal inequity should not be seen as being more severe because there are many income units with similar pre-tax incomes (and the probability for rerankings is therefore high). It is a value judgement which part of the income distribution one considers as the most important locus for high horizontal equity.

To make this value judgement explicit, it is possible to use the generalized Gini and concentration indices:

$$H^T(\nu) = {}_{X+B}C_{X+B-T}(\nu) - G_{X+B-T}(\nu) . \tag{8}$$

A corresponding measure is used to measure reranking caused by the transfer system:

$$H^B(\nu) = {}_{X}C_{X+B}(\nu) - G_{X+B}(\nu) . \tag{9}$$

where $G_{X+B}(\nu)$ is the generalized Gini index for gross income and ${}_{X}C_{X+B}(\nu)$ is the generalized concentration index for gross income when the income units are ranked according to their original incomes.

[15] The difference between the ordinary concentration index and the Gini coefficient for disposable income.

To sum up, we will use the following decompositions of the equalizing effect of transfers and taxes when we compare the pre- and post-reform tax and transfer systems:

$$G_X(\nu) - G_{X+B}(\nu) = \frac{b}{1+b}\,\rho^K(\nu) + H^B(\nu)\ , \tag{10}$$

$$G_{X+B}(\nu) - G_{X+B-T}(\nu) = \frac{t}{1-t}\,\pi^K(\nu) + H^T(\nu)\ , \tag{11}$$

i.e. three aggregate measures of three properties of the transfer and tax system respectively, the overall share in income of benefit and taxes respectively, departure from proportionality and horizontal equity. These three properties together determine the equalizing effect of the respective systems.

3. Results: Effects of the Swedish tax reform of 1991

3.1. The data and some additional methodological considerations

In this section we will compare two income distributions: one pre-reform and one simulated post-reform. The simulation is described in more detail in the appendix. The data source of the pre-reform income distribution is the *1987 Household Income Survey* provided by *Statistics Sweden*. It is the principal annual micro-data source on incomes in Sweden. It is a stratified sample of about 10,000 households.

In this study, we select households with the age of head of household between 25 and 65. The reason for excluding the age group between 18 and 25 is that this group has an income distribution that is very different from the rest of the population. This is because there are many observations with very low incomes, sometimes due to studies and sometimes to temporarily very irregular life habits. The individual's income position in this age group tells us very little about his lifetime income[16]. The results of the study, especially when more weight is given to the lower end of the income distribution, would have been very much dependent on this age group. The reason for excluding households headed by people aged 65 or above is primarily that the quality of the data in this age group is not as good as in the rest of the population. Also, the tax reform affects old age pensioners very differently from the other population and would deserve a study of its own. These selection criteria resulted in a sample size of 7,277 households.

The observation unit in this survey, and the unit of analysis in our study, is the household. The households consist of one or two married or cohabiting adults and children aged 17 or younger living in the same family. To generate the pre-reform income distributions from the 1987 data, the income data was transformed to correspond to the 1991 price and cost levels. The post-reform income distribution was obtained through a simulation using the tax-benefit model at the Ministry of Finance. We have used the same results as have been used in the official calculations to evaluate the reform proposal. The only changes in taxes that will be considered in this study are the changes in income and wealth taxes. The increase in consumption taxes was left out of the analysis, as there are no consumption data. The part

[16] This is shown in Björklund (1992).

of the tax reform that we considered is the increase in child and housing allowances and the decrease in income taxes. Therefore, this is not a balanced budget proposal in the simulation results.

The simulation of the effect of the tax reform is purely static in the sense that it does not take the household's economic behavior responses to tax/transfer changes into account. However, a one per cent increase in labor income was imputed due to "increased labor supply or a general improvement in the functioning of the economy"[17]. This is a highly unsatisfactory approximation. Ideally, the kind of data an economist would like to have is data generated by a general equilibrium model that takes all economic effects of the change in the tax/transfer system into account and produces distributions of individual utilities. The data requirements for such a model are, however, unrealistic. In practice, a trade-off exists between detailed information on household income components but very limited information on economic changes or very stylized information on household income components but more realistic computations of economic implications of a tax reform[18]. The focus of this study is to show how recently developed methods could be applied to the analysis of the income distribution implications of a tax reform. These methods could also be used on more realistic data that hopefully will be possible to obtain for ex ante analysis of future tax and benefit reforms.

We define disposable income as the total amount that a household can consume during a year without changing the real value of the wealth. Gross income is defined as disposable income plus taxes paid. Original income is earnings and fringe benefits from work, income from capital and capital gains and income from private and social insurance. Two per cent of the market value of the home is imputed as income for homeowners. Original income is thus, gross income less child and housing allowances.

---

[17] Eklind and Johansson (1989) page 2.

[18] An alternative to using results from a microsimulation is to use income distribution data, collected after the tax reform. This method also has many drawbacks. The main problem is to separate the effects that are caused by the tax reform from those caused by other changes in the economy. With no knowledge about the individual households' labor supply behavior and economic restrictions, it could be misleading to calculate changes in the income distribution. If a relatively well-off person responds to the tax reform by reducing his labor supply, while a poorer person responds by increasing his labor supply, we may observe an income rank reversal but not a utility rank reversal. In our study, however, simulations do not take behavioral responses into account. What we can measure is thus income changes, as if the economic behavior was constant.

Most tax systems, and all transfer systems, take differences in household needs into consideration, either explicitly or implicitly. There are probably several high-income single person households that pay more income tax than a married couple with dependent children and even higher income, due to tax-exempted child allowances and the separate taxation of spouses. Similar examples could be constructed for most tax systems. It would be misleading to measure tax progressivity by comparing different types of households without taking characteristics other than income into account. There are two methods of dealing with this problem: (1) only compare income units from the same household groups, (2) use equivalence scales to transform the incomes of different types of income units. The disadvantage with the first method is that it severely limits the results, since one can only compare intra-household group changes in the income distribution. The disadvantage with the second method is that it always involves value judgements since one must choose the particular equivalence scales used (see, for example, Fisher (1987)).

In this study we will use both methods. First, we will study the entire population and use equivalence scales. The equivalence scales we use are derived from the rules for social welfare allowance, which are determined by the *National Board of Health and Welfare*[19]. To examine whether the conclusions from the first part of the study hold within sub-groups of the population, we will do the same decomposition of the equalizing effect of the benefit and tax systems for different types of households[20].

We have chosen the values 1.8, 2.0, 2.2 and 2.4 for the parameter of the generalized Gini coefficient. The reason for choosing two values above 2 is that most people, including the author of this paper, are more concerned about the distribution in the lower end of the distribution when comparing income distributions.

Table 3.1 reports some summary statistics from the simulation study. A few points should be noted. First, the mean original income is lower under the post-reform regime. This is

---

[19] A single person without children is given the weight 1.15 and a couple without children 1.90. Children in ages 0-3, 4-10 and 11-17 are given the weights 0.55, 0.65 and 0.75, respectively.

[20] A related issue is the choice of unit of analysis. One can choose either the household or the individual. The argument for choosing the individual is that all individuals should be counted equally and that political decisions should be based on the welfare of all individuals. On the other hand, children do not have the right to vote in any democracy. The practical implication of the choice of unit of analysis is that large households, i.e. households with children, are weighted more heavily. We have chosen the household as unit of analysis for practical reasons, but as we do the within-group analysis separately by type of household, the choice is not of great importance.

because the higher costs for employers to provide fringe benefits is assumed to be directly passed through to wages. Second, the types of households that do not gain from increased allowances, households without children, gain more on average on the changes in taxation. Third, all single adult households, as well as households with three or more children, will increase their disposable income (on average) more than other households after the tax reform.

Table 3.1. *Summary statistics.*

Group	Number of			n_j/n	Mean X		Mean X+B	
	Adults	Children	n_j	%	Pre-reform	Post-reform	Pre-reform	Post-reform
1	1	0	1 704	23.4	159 960	159 296	160 074	159 631
2	1	1	171	2.3	173 893	173 199	185 892	193 465
3	1	≥2	97	1.3	173 531	174 249	200 318	214 543
4	2	0	2 345	32.2	336 236	332 791	336 331	332 814
5	2	1	1 205	16.6	344 071	340 052	351 272	350 987
6	2	2	1 245	17.1	349 648	344 855	363 916	366 501
7	2	≥3	510	7.0	326 557	322 459	357 990	367 742
8	All, unadj.		7 277	100	261 520	259 028	267 066	267 457
9	All, equiv.				157 781	156 480	159 909	159 790

Table 3.1. *Continued.*

Group	Mean X+B-T			Mean B			Mean T		
	Pre-reform	Post-reform	Change, %	Pre-reform	Post-reform	Change, %	Pre-reform	Post-reform	Change, %
1	104 574	110 812	6.0	114	335	193.9	55 500	48 819	-12.1
2	134 477	146 499	8.9	12 000	20 266	68.9	51 416	46 966	-8.7
3	158 099	173 033	9.4	26 787	40 294	50.4	42 219	41 510	-1.7
4	218 525	228 590	4.6	5	23	460.0	117 806	104 224	-11.5
5	236 662	248 102	4.8	7 201	10 935	51.9	114 611	102 886	-11.2
6	250 340	262 133	4.7	14 268	21 645	51.7	113 577	104 367	-8.1
7	255 268	271 674	6.4	31 433	45 284	44.1	102 721	96 068	-6.5
8	178 984	188 516	5.3	5 546	8 429	52.0	88 082	78 941	-10.4
9	106 186	111 952	5.4	2 127	3 310	55.6	53 723	47 838	-11.0

3.2. Within- and between-group comparisons using equivalence scales

In this section we will compare both households from the same household groups and households from different groups. As was shown in the previous section, some of the reported income distribution effects of the benefit tax reform are due to the fact that different household groups are affected differently by the reform. These between-group results are highly dependent on the equivalence scales used.

Table 3.2. *Generalized Gini coefficients for original income,* G_X.

ν	Pre-reform	Post-reform	% change
1.8	0.3905	0.3836	-1.78
2.0	0.2561	0.2512	-1.91
2.2	0.1469	0.1429	-2.73
2.4	0.0569	0.0531	-6.63

Table 3.2 depicts the generalized Gini coefficients for the distribution of original incomes before and after the tax reform. As can be seen in the table, the Gini coefficient for all generalizations is lower for the post-reform distribution. This follows from Lorenz dominance: the Lorenz curve for post-reform original income is above the corresponding pre-reform curve at every point of the distribution[21]. The relative change is larger if more weight is given to the lower end of the distribution. This indicates that the largest difference between the distributions is located in that range.

The more equal distribution of the post-reform original incomes can be explained at least in part by reference to two assumptions made in the simulations. First, it is assumed that the higher costs of fringe benefits to employers[22], are passed on directly to employees receiving these fringe benefits, in the form of lower wages. As the individuals receiving these benefits

[21] The Lorenz and concentration curves are calculated on percentile shares (100 points). The comments in the text are not only based on graphical inspection, but also on comparisons between the actual numbers of the percentile shares.

[22] The abolition of the employers' rights to deduct subsidies to employees pensions, car benefits and meals.

are on average relatively well off, this assumption then leads to a more equal observed post-reform income distribution. Second, households in the lower segments of the income distribution tend to receive a larger share of their factor (or original) income from labor. The assumption of a one per cent increase in labor income means that these households increase their original income on average more than proportionally to their pre-reform income shares.

3.2.1. Decomposition of the equalizing effect of child and housing allowances.

The equalizing effect of housing and child allowances can be measured as the difference between the Lorenz curves for gross and original income. The results are unambiguous: the equalizing effect of these transfers is larger for the post-reform system (the difference between the Lorenz curves is larger for that system at every point).

If we use an aggregate measure of the equalizing effect in this case, the difference between the generalized Gini coefficient for original and gross income, we find that the post-reform transfer system has a larger equalizing effect than the pre-reform system. This can be seen in table 3.3. The difference is larger the more the lower parts of the income distribution are weighted.

Table 3.3. *Equalizing effect of child and housing allowances measured as percentage difference between the generalized Gini coefficients for original and gross income.*

ν	Pre-reform	Post-reform	Difference
1.8	2.32	3.71	1.39
2.0	3.53	5.68	2.15
2.2	6.48	10.52	4.04
2.4	18.23	30.86	12.63

What part of the post-reform system's greater equalizing effect depends on differences in (respectively) the average volume of the allowances, the degree of departure from

proportionality and horizontal equity? The answer can be obtained from decomposing the equalizing effect.

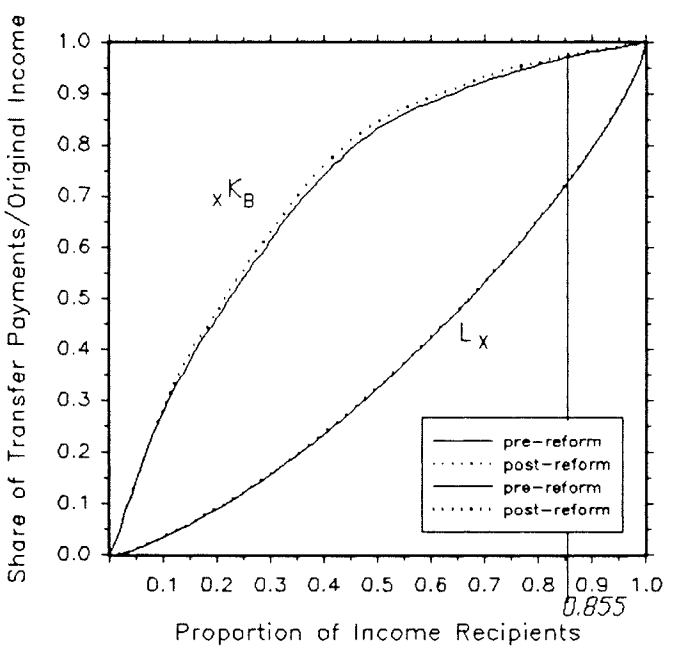
Table 3.4. *Household transfer incomes as shares of gross income, b.*

Pre-reform	Post-reform	Difference
0.0135	0.0212	0.0077

Table 3.4 depicts the overall mean child and housing allowances as a share of total gross household income. There is a considerable increase in these allowances after the reform.

Figure 2 depicts the concentration curves for child and housing allowances under both regimes, along with the Lorenz curves for the original income distributions. As this figure shows, the concentration curve for transfers is above the Lorenz curve for original income. This means that households with relatively low original income receive transfers amounting to a greater proportion of income than the transfers received by higher income households, i.e. both systems studied are regressive. Furthermore, the concentration curve for transfers is above the diagonal in the Lorenz diagram. This means that households with relatively low income receive more child and housing allowances measured in absolute amounts. Households with children, who receive more allowances, are overrepresented in the lower end of the distribution of equivalent incomes, i.e. the value judgements implicit in the equivalence scales are similar to those governing the distribution of the allowances.

Figure 2. *Concentration curves for transfers and Lorenz curves for original income distributions. Pre and post-reform. Equivalent incomes.*



The difference between the concentration curve and the Lorenz curve measures the regressivity of the transfer system. This measure is larger for the post-reform transfer system for all points between 0 and 0.855. Estimates of the Kakwani index for transfer regressivity are given in Table 3.5. It can be seen that the difference between the systems increases as more weight is given to the lower end of the income distribution. The difference between the concentration curves and the Lorenz curves for the post-reform system does not exceed the

Table 3.5. *The Kakwani transfer regressivity index, G_X-C_B. Child and housing allowances.*

ν	Pre-reform	Post-reform	Difference
1.8	0.6943	0.7043	0.0100
2.0	0.6940	0.7111	0.0151
2.2	0.7350	0.7541	0.0191
2.4	0.8039	0.8266	0.0227

corresponding measure for the pre-reform system. This means that there exists a weight coefficient, below 1.8, that will reverse the order in which the two systems are considered as more regressive.

Table 3.6 reports the total vertical equity component of the redistribution imposed by the two transfer systems. Column 3 shows the difference between these two components. It can be seen that this component is considerably larger for the post-reform transfer system. As was shown in Tables 3.3 and 3.4, both the transfer share of total gross income (b) and the transfer regressivity component (G_X-C_B) were larger for the post-reform transfer system. To determine the extent to which each component contributes to the redistribution, the post-reform transfer share component and the pre-reform transfer regressivity component are used together. These results are reported in column 4. As can be seen in Table 3.6, between 92.7 and 96.0 per cent of the differences in redistribution due to vertical equity can be attributed to differences in the transfer share of total income between the two systems.

Table 3.7 gives the estimates of horizontal inequity imposed by the two transfer systems. The measures give higher estimates when the lower end of the income distribution is weighted more heavily. This can be explained by the fact that the transfers studied are concentrated in that end of the distribution. The measures of horizontal equity differ very little when comparing the two transfer systems.

To sum up, the equalizing effect of the transfer system increased considerably. This increase is primarily achieved neither by a higher transfer regressivity nor increased

Table 3.6. *Changes in measures of vertical equity imposed by the transfer system.*

ν	V^B_{pre}	V^B_{post}	V^B_{post}-V^B_{pre}	V^{B*}	V^{B*}-V^B_{pre}	Share of difference
1.8	0.0092	0.0146	0.0054	0.0144	0.0051	96.0
2.0	0.0093	0.0148	0.0055	0.0144	0.0052	94.3
2.2	0.0098	0.0156	0.0058	0.0152	0.0054	93.2
2.4	0.0104	0.0171	0.0064	0.0167	0.0060	92.7

V^B_{pre} = vertical equity before tax reform, $(b/(1+b))*(G_X-C_B)$.

V^B_{post} = vertical equity after tax reform.

V^{B*} = vertical equity index computed using the post-reform transfer regressivity component and the pre-reform transfer regressivity component, i.e. $(b_{post}/(1+b_{post})*(G_X^{pre}-C_X^{pre})$.

Share of difference = the percentage share of the difference between the pre- and post-reform vertical equity component that can be attributed to differences in the amount paid in transfers.

horizontal equity, but by an increase in the total amount of transfer payments.

3.2.2. Decomposition of the equalizing effect of taxes

The equalizing effect of taxes can be measured as the difference between the Lorenz curves for gross and disposable income. The difference is larger for the pre-reform system in all points below 0.955. Above these points it is larger for the post-reform tax system. The aggregate measures of this difference, the percentage change in the generalized Gini coefficient, are shown in Table 3.8. As this table shows, the equalization is bigger for all

Table 3.7. *Measures of horizonal equity of the transfer system.*

ν	Pre-reform	Post-reform	Difference
1.8	-0.0002	-0.0004	-0.0002
2.0	-0.0002	-0.0005	-0.0003
2.2	-0.0003	-0.0006	-0.0003
2.4	-0.0004	-0.0007	-0.0003

generalizations. As there is no dominance for the equalizing effect, we can always find a ν, smaller than 1.8, that reverses this result.

To see which properties of the two tax systems that generate this finding, we decompose the equalizing effect.

Table 3.8. *Equalizing effect of the tax systems measured as percentage difference between the generalized Gini coefficients for gross and disposable income.*

ν	Pre-reform	Post-reform	Difference
1.8	11.25	10.21	-1.04
2.0	15.38	12.25	-3.13
2.2	27.26	19.47	-7.79
2.4	81.38	66.14	-15.24

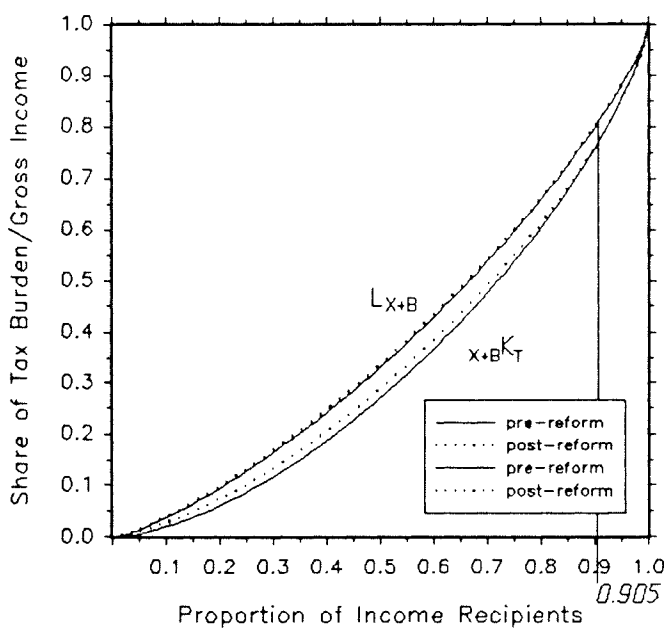
Table 3.9 shows the average share of total gross income collected in income taxes under both regimes. It can be seen that the taxes are on average about 3.7 percentage points larger under the pre-reform tax system.

Figure 3 shows the concentration curve for income taxes paid, along with Lorenz curves for gross income under the two tax regimes. This figure reveals three interesting properties of the two tax systems. First, both tax systems are unambiguously progressive, i.e. the

Table 3.9. *The over all mean income tax share of gross income, t.*

Pre-reform	Post-reform	Difference
0.3360	0.2994	-0.0366

Figure 3. *Concentration curves for taxes and Lorenz curves for gross income distributions. Pre- and post-reform. Equivalent incomes.*



concentration curve for taxes are below the Lorenz curve at every point for both tax systems.

Second, tax progressivity, measured as the vertical distance between the concentration curve for taxes and the Lorenz curve for gross income, gives an ambiguous ordering of the tax systems. The distance is larger for the pre-reform system up to the point 0.905. Third, the relative difference between the systems is larger in the lower end of the income distribution. This is shown in Table 3.10, where generalized Kakwani tax progressivity indices are reported. When more weight is given to the lower end of the income distribution, the difference becomes larger. If some weight sufficiently smaller than 1.8 is used, the progressivity ordering of the tax systems is reversed.

Table 3.10. *The Kakwani tax progressivity index,* G_{X+B}-C_T.

ν	Pre-reform	Post-reform	Difference
1.8	0.1018	0.0946	-0.0072
2.0	0.0876	0.0734	-0.0142
2.2	0.0849	0.0635	-0.0214
2.4	0.0853	0.0624	-0.0229

Table 3.11 depicts the vertical component of the equalization. It is larger for the pre-reform tax system for all chosen values of ν and the difference becomes larger as more weight is given to the lower end of the income distribution. As was shown in Tables 3.9 and 3.10, both the rate component and the progressivity component are larger for the pre-reform system. To determine the extent to which these two components contribute to the total vertical component, we have used the progressivity component of the pre-reform system and the rate component of the post-reform system to construct V^{T*}. As can be seen in Table 3.11, the share of the difference in the vertical component that can be attributed to the rate component is highly dependent on the weight (ν), with the share ranging from 72.2 to 40.6 per cent. This follows from the fact that the difference in progressivity between the two systems varies over the range of the income distribution (see Figure 2). This result shows the importance of value

Table 3.11. *Changes in measures of vertical equity imposed by the tax system.*

ν	V^T_{pre}	V^T_{post}	V^T_{post}- V^T_{pre}	V^{T*}	V^{T*}-V^T_{pre}	Share of difference
1.8	0.0515	0.0404	-0.0111	0.0479	-0.0080	72.17
2.0	0.0443	0.0314	-0.0130	0.0371	-0.0069	53.07
2.2	0.0429	0.0271	-0.0158	0.0321	-0.0067	42.25
2.4	0.0432	0.0266	-0.0166	0.0315	-0.0067	40.60

V^T_{pre} = vertical equity before tax reform, $(t/(1-t))*(G_{X+B}-C_T)$.

V^T_{post} = vertical equity after tax reform.

V^{T*} = vertical equity index computed using the post-reform tax progressivity component and the pre-reform tax progressivity component, i.e. $(t_{post}/(1-t_{post})*(G_{X+B}{}^{pre}-C_T{}^{pre})$.

Share of difference = the percentage share of the difference between the pre- and post-reform vertical equity components that can be attributed to differences in the amount paid in taxes.

judgements when measuring tax progressivity.

Table 3.12 shows measures of the horizontal inequity imposed by the two tax systems.[23]

[23] Ankrom (forthcoming), using the *1981 Household Income Survey*, estimates the horizontal inequities imposed by the whole Swedish tax and transfer systems (transition from market income plus employers contributions to disposable income) to -0.06. An explanation for his very high estimates is that he treats social insurance incomes as transfers.

Table 3.12. *Measures of horizontal equity of the tax system.*

ν	Pre-reform	Post-reform	Difference
1.8	-0.0086	-0.0027	0.0059
2.0	-0.0063	-0.0023	0.0040
2.2	-0.0055	-0.0023	0.0032
2.4	-0.0053	-0.0024	0.0029

It can be seen that horizontal equity has increased considerably after the tax reform, regardless of the measure used. An improvement in horizontal equity was stated as an explicit goal of the 1991 tax reform. The increase has several explanations. Following Gravelle (1992), sources of horizontal inequity can be divided into three categories: (1) family characteristics, (2) sources of income, and (3) disposition of income.

Family characteristics. In this study we use the household as the unit of analysis. If we compare two households with the same gross income, one with a single earner and one with two earners, then the second household will pay less tax due to increasing marginal taxes and separate taxation of spouses. Since marginal income tax rates are lower after the tax reform, this source of horizontal inequity is expected to become less important.

Child and housing allowances targeted to households with dependent children are not taxable. This means that households with children, at a similar level of gross income, pay proportionally relatively less in income taxes than households without children. Since child and housing allowances are higher after the tax reform, this source of horizontal inequity should have increased.

Sources of income. It is well known that people pay different levels of tax depending on the sources of their income. Different forms of compensation for labor are not taxed equally. The same is true for different kinds of capital income. The tax reform act attempted to reduce these differences. For example, employers' subsidies to meals are now taxed at the same rate as wages, and the same is true for the fringe benefit of a free car. These attempts to broaden the tax base will lead to an increase in the degree of horizontal equity.

Disposition of income. The level of the consumption tax varies when it is levied on different kinds of goods and services. The tax reform act also attempted to reduce this source of

horizontal inequity by mandating the same tax rate for all kinds of consumption including, for example, construction services. Since data on household consumption is lacking, we cannot here assess the extent to which this goal was achieved.

In Table 3.12 it can also be seen that measures giving more weight to the upper end of the income distribution gives higher estimates of horizontal inequity, or rerankings. This can be explained by the fact that households in the upper end of the income distribution tend to have incomes from different sources that are not taxed equally.

3.2.3. The distribution of disposable incomes

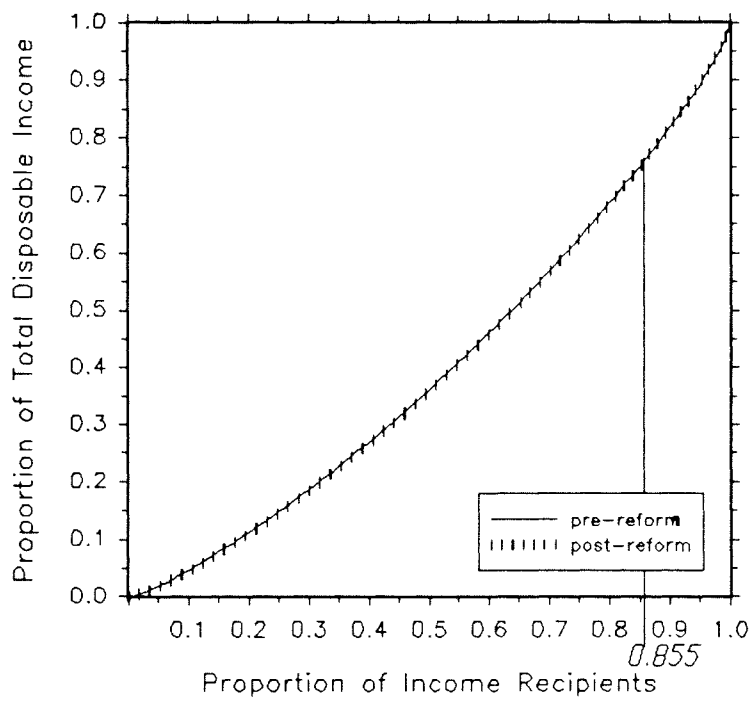
Figure 4 shows the Lorenz curves for the distribution of pre- and post-reform disposable income. The Lorenz curve for the pre-reform distribution is drawn with a thin, solid line and the curve for the post-reform distribution is drawn with a wide, dotted line. As they are very close, it is hard to distinguish the curves. If the thin solid line crosses the short vertical line above the midpoint, the Lorenz curve for pre-reform disposable income distribution is above the post-reform Lorenz curve at that point. As can be seen in the figure, the Lorenz curve for the pre-reform distribution is above the post-reform between 0 and 0.855 where they intersect. Between 0.855 and 1, the Lorenz curve for the post-reform distribution is above the pre-reform Lorenz curve, i.e. the distribution of disposable income is more equal in the lower end of the distribution, but more unequal in the upper end, under the pre-reform tax/transfer system.

Table 3.13. *Generalized Gini coefficients for disposable income.*

ν	Pre-reform	Post-reform	Difference
1.8	0.3385	0.3316	-0.0069
2.0	0.2090	0.2079	-0.0011
2.2	0.0999	0.1029	0.0030
2.4	0.0087	0.0124	0.0037

As can be seen in Table 3.12, the results of the generalized Gini coefficients are ambiguous: if more weight is given to the lower part of the income distribution, the pre-reform income

Figure 4. *Lorenz curves for the distribution of pre- and post-reform disposable incomes. Equivalent incomes.*



distribution is more equal. It is then a value judgement which income distribution a person averse to inequality would prefer.

Figure 4 and Table 3.13 also reveal that the impact of the tax reform on the distribution of disposable incomes is surprisingly small. This result is consistent with the findings of Schwarz and Gustafsson (1991).

3.3. Within-group comparisons

In Table 3.1 we saw that different household groups were not affected equally by the benefit and the tax reforms. Households with children increased on average their income from child and housing allowances, while households without children on average enjoyed a larger income tax reduction. Single parents increased on average their disposable income more than other groups. It is possible that the results obtained in Section 3.2 were dependent on the between group income distribution effects of the tax/transfer reform, i.e. could be very much dependent on the choice of equivalence scales. In this section we investigate if the results from Section 3.2 apply to the income distribution within household groups. To do this we have made the same decompositions as we did in Section 3.2 for each household group separately.

In Table 3.14.a to 3.14.g we report the difference between the pre- and post-reform components of the equalizing effect measured as percentage share of the pre-reform generalized Gini coefficient for original income for different household groups (i.e. $[(Y^{pre}-Y^{post})/G_X^{pre}]*100$, where Y is the different measures). In Table 3.14.h, we present the same measures for the weighted averages of the within group results. Table 3.14.i gives the results for the total population as a reference. In this exercise, the change in the generalized Gini coefficient and the changes in the decomposition component sum up to the total change in the generalized Gini coefficient when moving from original to disposable income. This means that if a decomposition component, vertical or horizontal, is positive, this component contributes less to equalization after the tax reform. If it is negative it contributes more. The share of equalization that was due to the benefit rate and tax rate respectively was obtained using the same method as in the previous section, i.e. the post-reform benefit/tax rate was used together with the pre-reform regressivity/progressivity component. As can be seen in Table 3.14.h and 3.14.i, the main results from Section 3.2 hold for the within group comparison. That is: (i) The distribution of original income was unambiguously more equal under the post-reform regime. (ii) The increased amounts of child and housing allowances have a considerable equalizing effect on the income distribution. (iii) The pre-reform tax

system has a larger vertical redistribution (equalizing) component. The decreased progressivity is the most important explanation in the lower end of the distribution, while the decreased average tax rate explains more of the difference in the upper end. (v) It is a value judgement which of the distributions of disposable income one considers the most equal.

There are, however, differences between the within-group decomposition components. The differences in the increased equalizing effect of the child and housing allowances closely follows the differences in the mean increased amount reported in Table 3.1: the largest equalization is for single parent households. Note that benefit regressivity has decreased in these groups, as well as in the group of households with two adults and three children. The explanation of this phenomenon is that child allowances, which are the same amount for each child, have increased relatively more than the income dependent housing allowances paid to these groups.

The within-group equalizing effect of changes in income taxes are surprisingly homogenous, and are approximately the same as for the total population. It is interesting to note that horizontal equity has increased more in the group of single-member households without children, than it has in the entire population. In this group, the effect of a broader tax base could be isolated from the other two potential sources of horizontal equity mentioned in section 3.2.2, i.e. separate taxation of spouses and (almost) tax exempted allowances.

In no group did the new tax/benefit system generate an unambiguously more unequal income distribution. For the single parent groups and the group with two adults and two children, however, the Gini coefficients for post-reform distribution of disposable income are smaller for all generalizations. Thus, the single-parent households experienced both the greatest increase in their average disposable income and the greatest decline in within-group inequality.

Table 3.14.a. *Percentage change in components of the Kakwani decomposition of equalizing effect due to child and housing allowances and income taxes. Households with one adult and no children.*

γ	1.8	2.0	2.2	2.4
G_X	-2.4	-3.0	-5.7	-63.3
V^B, due to b	-0.3	-0.4	-0.9	-10.7
V^B, due to G_X-C_B	-0.2	-0.3	-0.9	-12.0
H^B (C_{X+B}-G_{X+B})	0.0	0.0	0.0	-0.1
V^T, due to t	2.9	4.2	8.8	100.5
V^T, due to C_T-G_{X+B}	1.1	4.7	13.8	169.6
H^T (C_{X+B-T}-G_{X+B-T})	-2.2	-2.8	-5.2	-54.5
G_{X+B-T}	-1.0	2.3	10.0	128.4

Table 3.14.b. *Percentage change in components of the Kakwani decomposition of equalizing effect due to child and housing allowances and income taxes. Households with one adult and one child.*

γ	1.8	2.0	2.2	2.4
G_X	-2.2	-2.9	-13.2	-3.8
V^B, due to b	-7.2	-12.2	-73.5	-24.4
V^B, due to G_X-C_B	1.5	3.2	23.9	8.8
H^B (C_{X+B}-G_{X+B})	0.1	0.2	1.6	0.6
V^T, due to t	3.7	6.2	31.5	10.0
V^T, due to C_T-G_{X+B}	2.5	5.8	31.1	10.7
H^T (C_{X+B-T}-G_{X+B-T})	-0.2	-5.0	-24.2	-9.0
G_{X+B-T}	-1.8	-4.7	-23.8	-7.1

Table 3.14.c. *Percentage change in components of the Kakwani decomposition of equalizing effect due to child and housing allowances and income taxes. Households with one adult and two children.*

γ	1.8	2.0	2.2	2.4
G_X	-1.0	-1.9	-12.6	-2.3
V^B, due to b	-7.3	-15.6	-130.4	-27.7
V^B, due to G_X-C_B	4.5	6.2	52.7	12.0
H^B (C_{X+B}-G_{X+B})	-0.8	1.7	16.2	3.2
V^T, due to t	1.7	3.4	26.1	5.3
V^T, due to C_T-G_{X+B}	3.3	7.9	62.3	12.6
H^T (C_{X+B-T}-G_{X+B-T})	-1.2	-3.1	-22.4	-3.9
G_{X+B-T}	-0.8	-1.4	-8.1	-0.9

Table 3.14.d. *Percentage change in components of the Kakwani decomposition of equalizing effect due to child and housing allowances and income taxes. Households with two adults and no children.*

γ	1.8	2.0	2.2	2.4
G_X	-2.6	-2.8	-5.6	-22.2
V^B, due to b	0.0	0.0	-0.1	-0.4
V^B, due to G_X-C_B	0.0	0.0	0.0	-0.1
H^B (C_{X+B}-G_{X+B})	0.0	0.0	0.0	0.0
V^T, due to t	2.3	2.5	6.2	25.7
V^T, due to C_T-G_{X+B}	2.7	1.2	10.3	46.7
H^T (C_{X+B-T}-G_{X+B-T})	-1.8	-2.0	-4.3	-16.7
G_{X+B-T}	0.6	-1.2	6.5	32.9

Table 3.14.e. *Percentage change in components of the Kakwani decomposition of equalizing effect due to child and housing allowances and income taxes. Households with two adults and one child.*

γ	1.8	2.0	2.2	2.4
G_X	-2.1	-3.0	-6.4	-11.3
V^B, due to b	-1.0	-1.7	-4.6	-10.0
V^B, due to G_X-C_B	-0.1	-0.2	-0.6	-1.6
H^B (C_{X+B}-G_{X+B})	0.0	0.0	0.0	0.1
V^T, due to t	1.9	2.8	6.6	12.8
V^T, due to C_T-G_{X+B}	0.9	2.7	8.2	17.8
H^T (C_{X+B-T}-G_{X+B-T})	-9.6	-2.4	-4.5	-8.7
G_{X+B-T}	-9.2	-1.7	-1.9	0.9

Table 3.14.f. *Percentage change in components of the Kakwani decomposition of equalizing effect due to child and housing allowances and income taxes. Households with two adults and two children.*

γ	1.8	2.0	2.2	2.4
G_X	-2.9	-3.5	-9.4	-8.7
V^B, due to b	-1.9	-3.1	-10.5	-11.6
V^B, due to G_X-C_B	-0.1	-0.2	-1.1	-1.4
H^B (C_{X+B}-G_{X+B})	0.0	0.0	0.1	0.1
V^T, due to t	1.9	2.7	8.4	8.5
V^T, due to C_T-G_{X+B}	1.8	2.0	12.5	13.7
H^T (C_{X+B-T}-G_{X+B-T})	-1.7	-2.3	-7.0	-7.3
G_{X+B-T}	-2.9	-4.5	-7.0	-6.8

Table 3.14.g. *Percentage change in components of the Kakwani decomposition of equalizing effect due to child and housing allowances and income taxes. Households with two adults and three or more children.*

γ	1.8	2.0	2.2	2.4
G_X	-3.0	-3.5	-8.6	-7.7
V^B, due to b	-4.7	-7.4	-24.1	-27.0
V^B, due to G_X-C_B	0.4	0.5	1.2	1.1
H^B (C_{X+B}-G_{X+B})	0.2	0.3	1.3	1.6
V^T, due to t	3.5	3.4	9.1	9.3
V^T, due to C_T-G_{X+B}	7.6	5.1	13.5	14.5
H^T (C_{X+B-T}-G_{X+B-T})	2.4	-2.1	-7.0	-7.0
G_{X+B-T}	6.3	-3.8	-14.7	-15.4

Table 3.14.h. *Percentage change in components of the Kakwani decomposition of equalizing effect due to child and housing allowances and income taxes. Weighted averages of within-group decomposition.*

γ	1.8	2.0	2.2	2.4
G_X	-2.5	-3.0	-6.8	-32.7
V^B, due to b	-1.2	-2.0	-9.5	-10.0
V^B, due to G_X-C_B	0.1	0.1	1.6	-4.3
H^B (C_{X+B}-G_{X+B})	0.0	0.1	0.5	0.1
V^T, due to t	2.5	3.4	9.2	48.2
V^T, due to C_T-G_{X+B}	2.0	3.4	13.8	81.1
H^T (C_{X+B-T}-G_{X+B-T})	-2.5	-2.6	-6.3	-27.7
G_{X+B-T}	-1.5	-0.6	2.5	54.8

Table 3.14.i. *Percentage change in components of the Kakwani decomposition of equalizing effect due to child and housing allowances and income taxes. All households and equivalence scales.*

γ	1.8	2.0	2.2	2.4
G_X	-2.0	-2.3	-4.0	-43.6
V^B, due to b	-1.5	-2.5	-5.5	-68.8
V^B, due to G_X-C_B	-0.1	-0.2	-0.4	-5.4
H^B (C_{X+B}-G_{X+B})	0.1	0.1	0.3	4.7
V^T, due to t	2.4	3.3	6.7	77.5
V^T, due to C_T-G_{X+B}	0.9	2.9	9.1	113.3
H^T (C_{X+B-T}-G_{X+B-T})	-1.7	-1.9	-3.2	-34.1
G_{X+B-T}	-2.0	-0.5	3.0	43.6

4. Conclusions

Income distribution analysis always entails value judgements, firstly with respect to the choice of an equivalence scale, and secondly regarding how to aggregate different parts of the distribution into one measure. In this study we have shown how different properties of a proposed new tax-and-transfer system could be analyzed ex ante by considering different aggregations of the income distribution and by comparing the results that were obtained using a certain set of equivalence scales for data on each group of households. This enables the reader to consider the impact of different value judgements, and also to evaluate if the results are dependent on these considerations. To consider only one set of equivalence scales and one aggregation could yield a misleading result.

Appendix

This appendix gives a short description of the changes in the tax and transfer system that is taken into account in the microsimulation. The simulations refer to the government's proposal for a new tax system in the autumn of 1989. Some minor changes took place when the reform was implemented in 1991.

1. Changes in the tax base, labor income.

a. The value of a company car is increased to between 35 and 42 per cent of the price of a corresponding new car.

b. The value of subsidized meals at work is added.

c. More stringent taxation on the benefit of low interest rates on loans granted by the employer.

d. The benefit of shares is increased by 1.5 per cent.

e. Increases in the value of insurance benefits provided by the employer.

Workers (AGS+TGL+AGB+TFA) 1.32 per cent of earnings.

Private sector white collar workers (TGL+TFA) proportional to earnings.

f. The deduction for commuting costs is decreased to 12 SEK per 10 kilometers.

g. More stringent rules for tax on allowances for expenses.

h. Tax reduction of 1,800 SEK for couples with only one person working is abolished.

i. More stringent tax on pension funds.

2. Changes in the tax base, capital income.

a. More stringent tax on capital gains from promissory notes and shares.

b. The standard deduction for capital income is abolished.

c. More stringent tax on houses. The taxable value is increased to 2.5 per cent for apartment houses and 1.2 per cent on houses.

d. More stringent tax on private rental income.

e. Reduced subsidization of interest payments on loans to finance apartment construction.

f. Previously tax-exempted savings funds become subject to a 20 per cent tax.

g. More stringent tax on capital gains from houses.

3. Other changes in tax/transfer rules.

a. National income tax is changed from the following scales 1989:

Taxable income	Tax, per cent
0-75,000,	5
75,000-140,000	17
140,000-190,000	31
190,000-	42

To the 1991 income tax scales on earnings:

0-170,000	0
170,000-	20

The tax rate on capital income and capital gains is 30 per cent 1991. The local government tax rate is approximately 30 per cent and is not changed after the reform.

b. The standard allowance is increased from 3,000 to 4,000 SEK.

c. The personal exemption for earnings is 10,000 SEK for taxable income below 58,000. In the interval 58,000-90,000 it is gradually increased to 18,000, and in the interval 95,000-175,000 it is gradually decreased to 10,000.

d. Interest payments are tax deductible up to 100,000 SEK and deductible to 70 percent above 100,000.

f. Deductions for deficits in own trade are decreased.

g. Child allowance are increased from 5,820 SEK in 1989 to 9,000 per child.

h. The total volume of means-tested, housing allowances paid is increased by 1,500 million SEK.

References

Ankrom, J., Two Refinements in the Measurement of Income Redistribution: Sweden, the U.S, and U.K., forthcoming in *Scandinavian Journal of Economics*.

Atkinson, A. B., On the Measurement of Inequality. *Journal of Economic Theory*, 1970, 2, 244-263.

Atkinson, A. B., Horizontal Equity and the Distribution of the Tax Burden. In *The Economics of Taxation* (H J Aaron and M J Boskin (eds.)), pp. 3-18, The Brookings Institution, Washington D C 1980.

Atkinson, A. B. and Stiglitz, J. E., *Lectures on Public Economics*, McGraw Hill, London 1980.

Björklund, A., A Comparison Between Actual Distributions of Annual and Lifetime Income: Sweden 1951-1989. Paper presented at the IARIW 22nd General Conference in Flims, Switzerland. Swedish Institute for Social Research, Stockholm University, 1992.

Cowell, F. A., The Structure of the American Income Inequality, *Review of Income and Wealth*, 1984, 30, 351-75.

Donaldson, D., and Weymark, J., A Single-Parameter Generalization of the Gini Indices of Inequality. *Journal of Economic Theory*, 22, 1980, 67-86.

Eklind, B. and Johansson, R., Skattereformens Fördelningseffekter. SOU 1989:33, Bilaga 8.

Feldstein, M., On the Theory of Tax Reform. *Journal of Public Economics*, 1976, 6, 77-106.

Fisher, F. M., Household Equivalence Scales and Interpersonal Comparisons. *Review of Economic Studies*, 1987, vol LIV, 519-524.

Fritzell, J., *Icke av marknaden allena: Inkomstfördelningen i Sverige*, Almqvist & Wiksell International, Stockholm 1991.

Gravelle, J., Equity Effects of the Tax Reform Act of 1986. *The Journal of Economic Perspectives*, 1992, 6, 27-44.

Jenkins, S. P., Empirical Measurement of Horizontal Inequity. *Journal of Public Economics*, 1988, 37, 305-329.

Kakwani, N. C., *Income Inequality and Poverty: Methods of Estimation and Policy Application*, Oxford University Press, New York 1980a.

Kakwani, N. C., On a Class of Poverty Measures. *Econometrica*, 45, 1980b, 437-446.

Kakwani, N. C., On the Measurement of Tax Progressivity and Redistributive Effect of Taxes with Application to Horizontal and Vertical Equity. In *Advances in Econometrics 3*, JAI Press, Greenwich Conn. 1984.

Kaplow, L., Horizontal Equity: Measures in Search of a Principle. *National Tax Journal*, 1989, XLII, 139-154.

King, M. A., An index of Inequality: With Applications to Horizontal Equity and Social Mobility. *Econometrica*, 51, 1983, 99-115.

Lambert, P. J., *The Distribution and Redistribution of Income: A Mathematical Analysis*, Basil Blackwell, Oxford 1989.

Musgrave, R. A., *The Theory of Public Finance*, McGraw-Hill, New York 1959.

Musgrave, R. A., Horizontal Equity, Once More. *National Tax Journal*, 1991, XLII, 113-122.

O'Higgins, M. and Ruggles, P., The Distribution of Public Expenditures and Taxes Among Households in the United Kingdom. *Review of Income and Wealth*, 1981, 27, 298-326.

O'Higgins, M., Schmaus, G., and Stephenson, G. Income Distribution and Redistribution: A Microdata Analysis for Seven Countries. In *Poverty, Inequality and Income Distribution in Comparative Perspective* (T.M. Smeeding, M. O'Higgins and L. Rainwater (eds.)) 1990, pp. 20-56, New York: Harvester Wheatsheafs.

Phäler, W., Redistributive Effects of Tax Progressivity: Evaluating a General Class of Measures. *Public Finance/Finances Publiques*, 1987, 37, 1-31.

Plotnick, R., The concept and measurement of horizontal inequity. *Journal of Public Economics*, 1982, 17, 373-391.

Plotnick, R., A comparison of measures of horizontal inequity. In *Horizontal Equity, Uncertainty and Economic Well-Being* (M. David and T. Smeeding (eds.)), pp. 239-268, NBER, New York 1985.

Schwarz, B. and Gustafsson, B., Modelling Income Redistribution Effects of Tax Reforms. *Journal of Policy Modeling*, 1991.

Whalley, J., Foreign Responses to U.S. Tax Reform. In *Do Taxes Matter? The Impact of the Tax Reform Act of 1986* (J. Slemrod (eds.), 1990, pp. 286-314, Cambridge, Massachusetts: MIT Press.

Yitzhaki, S., On an extension of the Gini index. *International Economic Journal*, 1983, 24, 617-628.

The Effect of Economic Incentives on Worker Absenteeism: An Empirical Study Using Swedish Micro Data*

by

Per Johansson[£] and Mårten Palme[$]

Abstract: The effect of economic incentives on absenteeism in Sweden is examined. Based on a utility function, frequently used in labour supply studies, absenteeism is modelled as an individual day-to-day decision. Implications of compensating wage differentials and efficiency wage hypotheses are discussed. Two data sets are used: survey data from 1981 and an annual panel 1981-1984. The dependent variable is the number of days absent per year for each individual, i.e. the outcomes of the day-to-day decision are aggregated over the time period of one year. Theory consistent models allowing for state dependence or unobserved heterogeneity are derived. Binomial, overdispersed binomial, mixture distribution and Markov chain models are estimated using maximum likelihood and least squares techniques. The results suggest that changes in the compensation level of the sickness insurance and changes in the unemployment rate could explain about 72 per cent of changes in work absence rate 10 year later.

Key Words: Count Data; Binomial distribution; Serial correlation; Unobserved heterogeneity; Markov chain; Mixture distribution; Maximum likelihood; Non-linear least squares; Semiparametric Estimation; Sickness insurance.

---

*We are indebted primarily to Kurt Brännäs for generous and competent guidance through all stages of the work on this study. Anders Björklund, Erik Ruist, Anders Westlund and Magnus Wikström have commented on one or several versions of the paper and thereby significantly improved the final result. We also want to thank participants in seminars at Ekonometridagarna 1993, ESPE meeting 1993 in Budapest, and the Economics departments at the Universities of Uppsala and Umeå for comments. Finally, we vish to thank the Institute for Social Research at the Stockholm University for making available the Swedish Level of Living Survey. The responsibility for remaining errors and shortcomings is ours.

[£]Department of Economics, University of Umeå, S-901 87 Umeå, Sweden

[$]Department of Economic Statistics, Stockholm School of Economics, Box 6501, S-113 83 Stockholm, Sweden

1. Introduction

Absenteeism has recently attracted increased attention in several of the social sciences (cf. Goodman and Atkin, 1984, for an overview). One reason for this is the long-term increase in the economic importance of this phenomenon, which has been experienced in most industrialised economies (OECD, 1985). However, very few studies have analysed the impact of economic incentives on this issue.

Sweden has a compulsory sickness insurance as a part of the national social insurance system. An insured individual is entitled to benefits from the sickness insurance if his (or her) perception of his state of health is such that he considers that "it does not permit him to do his regular work". The regulations allow the insured person to be absent from work for up to eight days without a certificate from a physician. As a person's state of health can be difficult to monitor even for a qualified doctor, and as the decision to work or to be on sick leave is left primarily to the insured individual, it seems realistic to believe that economic incentives influence the every-day choice of whether to work or to be absent from work. On the other hand, it is fair to say that many individuals who receive sickness insurance are, depending on their state of health, expected to be insensitive to economic incentives.

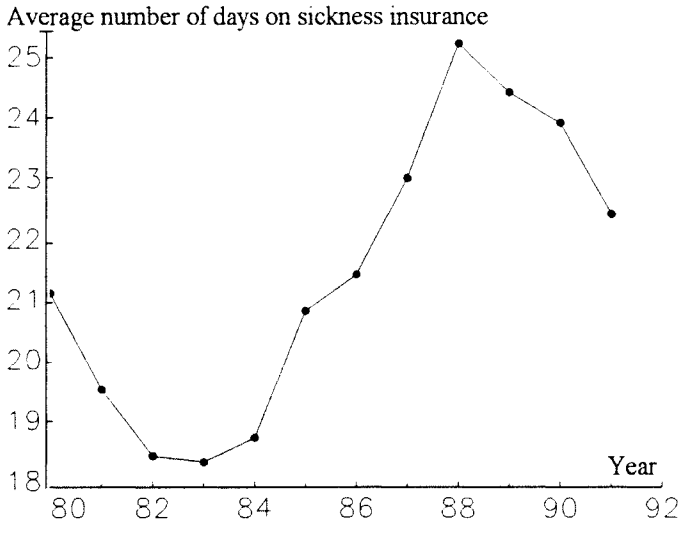


Figure 1. *The average annual number of days on sickness insurance in Sweden,* 1980-1991. *All insured individuals. Source: The Swedish National Social Insurance Board.*

Work absence is defined as time when the employee is absent from work, which can not be referred to as statutory leisure time or absence agreed in advance with the employer. A recent time-use study (SAF, 1986) shows that 97.1 per cent of the work absence for blue collar

workers in Sweden, is covered by the sickness insurance. Work absence and utilisation of the sickness insurance can thus be seen as almost synonymous.

The recent developments regarding the average number of days spent on sickness insurance, shown in Figure 1, reveal some marked fluctuations. It seems unlikely that these fluctuations are caused by changes in the state of health of the Swedish population alone. There have, however, been changes in economic incentives related to absence from work, alterations in the contribution levels and macroeconomic fluctuations, that deserve further investigation.

The aim of this study is to investigate whether, and to what extent, economic incentives affect absence from work. The econometric models are derived from standard economic utility theory. First, a theoretical model for the day-to-day decision is formulated. Since the data we have access to is the number of days absent during each year (1981-1984), models for the number of days absent during a year are devised, under different assumptions on unobserved heterogeneity and serial correlation, that are consistent with the day-to-day decision.

When analysing absenteeism behaviour empirically, it is very important to consider individual heterogeneity. Individuals, primarily due to different state of health, are likely to differ in their work absence behaviour, in this study their sensitivity to economic incentives. Obviously, it would have been advantageous if individuals who are suffering from bad health could be singled out. Unfortunately, this requires more information than can be obtained from our (if any) data set. The data permits, however, that individuals who have been absent less then a certain number of days to be selected and analysed separately. On the other hand, sample selection problems will arise following such strategy. The approach we adopt, is to consider information on the individuals state of health that can be obtained from the data set and to use estimation methods that can handle the possibility of state dependence or unobserved heterogeneity.

Under the assumption of no serial correlation or no unobserved heterogeneity, the model can be estimated using binomial maximum likelihood. However, there are reasons to believe that both serial correlation (see Avery & Hotz, 1984, for a review) and unobserved heterogeneity can be present. Serial correlation can result from the well known fact that individuals state of health tend to persist, or that they persist in their behaviour for other reasons (state dependence). It can also result from omitted variables that change slowly over time (spurious autocorrelation). To relax the assumption of no serial correlation two models are estimated; an overdispersed binomial model using a quasi-likelihood estimator and a Markov chain model using non-linear least squares. To handle the problem of unobserved heterogeneity a mixture distribution model is estimated (cf. Titterington et al., 1987) using a semiparametric estimator.

Few other empirical studies have used economic theory to explain differences in absence from work. Most of these studies share the methodological problem of time aggregation. Allen (1981a) departs from a labour-leisure choice model and obtains, by comparative statics, theoretical predictions of the signs of explanatory variables. He uses a representative sample of U.S. employees and the dependent variable is the individual's absence rate in the two weeks preceding the interview. Negative elasticities for hourly wage rates and household disposable income and positive coefficients for self-evaluated risks at work are found. Dunn & Youngblood (1986) test empirically the hypothesis of absenteeism as a labour supply adjustment by estimating work absence as a function of the difference between a measure of the worker's marginal rate of substitution between consumption and leisure (obtained through interviews) and the actual wage rate. Their empirical results support the hypothesis of absenteeism as a labour supply adjustment. Allen (1981b) estimates an inverse hedonic wage function, where the possibility to be absent from work is viewed as a desirable non-pecuniary element of the compensation package. A linear model with the absence rate as the dependent variable is used and, because the mean absence rate is very close to zero, use is also made of a logit model. Delgado & Kniesner (1992) start from the same economic hypothesis using panel data on individuals working on London buses. The dependent variable is the number of days absent during a year. Maximum likelihood and semiparametric estimators are used. The dynamic specification is interpreted as "sluggish adjustments". Barmby et al. (1991a) stress the importance of what they denote as "the demand side", i.e., the employers option to prevent absenteeism. They use daily absence records from four manufacturing plants belonging to the same firm in the UK, for the time period 1985-1989. Both Heckman's (1981) discrete panel data techniques and duration models are estimated.

The present study extends two aspects of this literature. The everyday economic choice of being absent from work is explicitly modelled contingent on occupation and personal characteristics that may influence the choice. Given the kind of data in this study, i.e., the number of days of work absence aggregated over the time period of one year, theoretically consistent estimation methods are used.

The data are collected from the 1981 *Swedish Level of Living Survey* (SLLS). In order to obtain information on individual work absence, this sample is matched with the *National Social Insurance Board* register on the annual number of days for which sickness insurance for each individual is paid. Individual income data are collected from tax registers. We use two types of data sets. First, data for 1,967 individuals for the year 1981. Second, this sample is extended to cover the period 1982-1984 using register data.

The paper is organised as follows. In Section 2 we derive the empirical specification of the choice between work and work absence and discuss how hypotheses previously used in labour

economics, could be used to explain the relationship between occupation and absenteeism. Section 3 describes the data used and the Swedish sickness insurance and tax systems. Section 4 discusses the econometric model specification, and addresses estimation and testing issues. The empirical results are given in section 5 and the final section contains a discussion of the results.

2. Theoretical Specification

To model the every day choice of whether or not to be absent from work, we adopt an approach that has been widely used for labour supply models. We assume that each individual has a utility function

$$u = U(x, L; s), \tag{1}$$

where x is a composite consumption good, L is leisure and s is a vector of personal characteristics. Leisure consists of contracted leisure time, t^l, and time absent, t^a. The utility function is maximised subject to a budget constraint

$$x = w(h^* - (1-\delta)t^a) + R, \tag{2}$$

where h^* is the contracted number of working hours, R is income from sources other than labour, w is net wage and δ is the share of the income the worker receives when absent. The price of the composite consumption good, x, is normalised to be one. h^* can be divided into desired number of working hours and time absent, hence $h^* = h + t^a$. This gives the identity $T \equiv h + t^a + t^l$, where T is total time available.

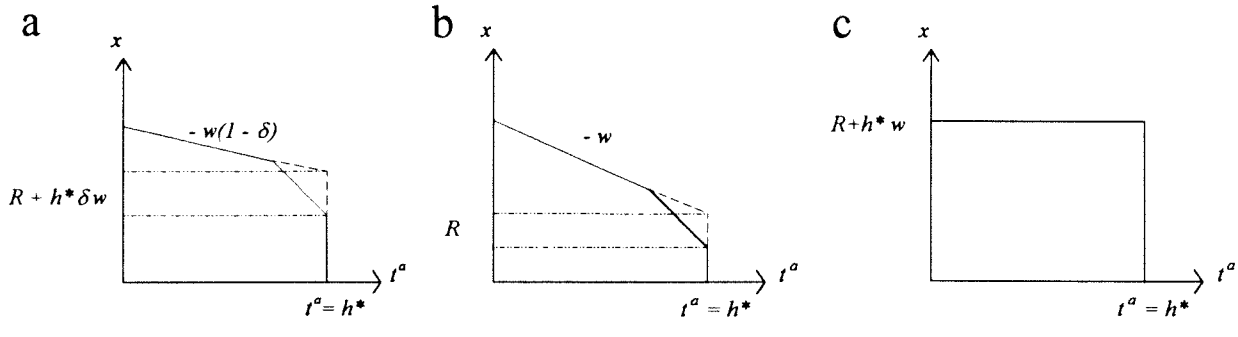


Figure 2: *a) partial compensation,* $\delta \in (0,1)$, *b) no compensation,* $\delta = 0$, *c) full compensation,* $\delta = 1$.

The budget constraint is illustrated in Figure 2. In these figures we have assumed that absence from work is a continuous choice. The income tax has two brackets and the compensation scheme for work absence is linear. In Figure 2a the slope of the budget line is the negative of

the relative price between consumption and time absent, i.e., $-w(1-\delta)$. The virtual income when absent equals the compensation received when absent from work plus non-labour income, i.e., $R+h^{*}\delta w$. If $\delta=0$, i.e. there is no compensation when the individual is absent from work (Figure 2b), the slope of the budget line is $-w$ (the net wage) and income when absent is solely R. If $\delta=1$, i.e., there is full compensation when absent from work (Figure 2c), the relative cost for being absent is zero and the slope of the budget line is vertical.

In order to obtain a linear demand function for time absent, we assume the following functional form for the direct utility function (cf. Hausman, 1980)

$$U(x,t^{a})=\exp\left\{-\left(1+\frac{\beta(x+\bar{s})}{b-T+(t^{l}+t^{a})}\right)\right\}\left(\frac{T-(t^{l}+t^{a})-b}{\beta}\right), \tag{3}$$

where $\bar{s}=s/\beta-\alpha/\beta^{2}$, s is a vector of socio-economic variables, α and β are parameters and $b=\alpha/\beta$. When the utility function (3) is maximized subject to the budget constraint (2), the following demand function for work absence is obtained

$$t^{a}=h^{*}-\alpha w(1-\delta)-\beta(R+h^{*}w\delta)-s=h^{*}+\alpha w^{*}+\beta y-s. \tag{4}$$

The latter expression introduces the notations to be used. This is a linear function of the cost for the individual of being absent from work (net earnings not covered by the health insurance, i.e. the relative cost between absence and consumption), w^{*}, and virtual income when the individual is absent, y. It is easy to show that the demand for absence equation (4) is the dual to the supply function for labour given in Hausman (1980) and Blomquist (1983). If the demand for absence equation (4) is to be consistent with a well behaved utility function, the Slutsky condition

$$\frac{\partial t^{a}}{\partial w^{*}}\Big|_{\bar{U}}=\frac{\partial t^{a}}{\partial w^{*}}-t^{a}\left(\frac{\partial t^{a}}{\partial y}\right)<0 \tag{5}$$

must be satisfied. We have $\partial t^{a}/\partial w^{*}=\alpha$ and $\partial t^{a}/\partial y=\beta$. Thus, it is sufficient that $\alpha<0$ and $\beta\geq 0$, for every $t^{a}\geq 0$.

It is important to emphasise that the contracted leisure time is exogenous in this model. Every day the individuals are assumed to choose between attending work or not, *conditional* on their perception of their health, the contracted number of work hours and the costs of the alternatives. We thus, assume that the contract specifying the hours of work is made between the worker and the employer in advance. An *alternative* specification is to assume that contracted leisure time and time absent are substitutes and that the individual chooses a

combination of these two. An important component of such a model would be to consider the costs of abuse of the sickness insurance, i.e., a moral hazard problem.

In the vector of socio-economic variables, *s*, variables related to health are included. It is also important to consider, and in the empirical analysis control for, individual characteristics that are likely to be correlated with work absence and the cost and income variables. Since the cost variable partly consists of the wage rate, we use the compensating wage differential (e.g., Rosen, 1986) and the efficiency wage (e.g., Shapiro & Stiglitz, 1984) hypothesis to get some idea about which variables should be included in *s*.

Compensating wage differentials. When a person accepts a job offer, he/she not only accepts a wage rate and a contracted number of work hours, but also a number of non-wage characteristics. The worker may accept higher risks at the workplace in exchange for a higher wage rate. If the worker chooses a higher risk level, he is likely to have a higher rate of work absence for two reasons. First, since the worker faces a higher level of risk at the workplace, he will strive to minimise the exposure time to these risks, i.e., he has, ceteris paribus, greater incentives to be absent from work. Second, workers who are exposed to higher risks also have a higher probability of being absent due to work related illnesses or injuries. Ideally, one would like to separate these causes of differences in absenteeism. This requires, however, that work absence can be separated according to cause of absence. Thus, omitting the risk may lead to biased estimates of the relationship between wage rate and absence.

Efficiency wage hypothesis. The employer may influence the level of absence by several means, for example, the controls on the employees could be increased (see Henrekson et al., 1992, p. 79 for a review of an empirical study on the firm level). However, some jobs are by their nature very difficult to monitor. Shapiro & Stiglitz's (1984) shirking model predicts that when the possibilities of monitoring workers' job performance are poor, employers may pay wages above the market clearing level in order to elicit adequate effort from their employees. That is the employer pays a wage higher than the market wage to induce the worker not to be absent from work. Thus, if we do not control for the employer's monitoring level we may get a positive bias of the estimates on the relation between wage rate and absenteeism. Another implication is that when unemployment increases, shirking or absenteeism decreases. The potential cost of being caught shirking and eventually loosing one's job is greater if the unemployment rate is high and it is harder to find a new job. Wages will, because of excess supply on the labour market, ceteris paribus, be lower in times of high unemployment. Thus, if we do not control for unemployment we may get a negative bias on the estimated relation to wages.

3. Insurance System and Data

In this section we present the variables included in the models (tables containing all the variables are given in Appendices 1 and 2). The data are obtained from the 1981 SLLS, see Eriksson & Åberg (1987). The total sample size of the survey is 5,613 individuals. Some selections are made for the present study. First, the sample is restricted to individuals aged 20 to 64 years throughout the sample period 1981-1984. Second, individuals not in the labour force in 1981 are excluded from the sample. We also exclude self-employed, students, military personnel and white collar workers, i.e. the study is restricted to blue collar workers. The reason for excluding these other groups is to limit heterogeneity arising from differences in sickness insurance systems that can not be obtained from the available data. After these exclusions are made, the sample consists of 1,981 individuals.

The dependent variable is the number of days for which each individual receives compensation from the sickness insurance, aggregated over the time period of one year. Data for this variable are obtained from the *National Social Insurance Board* registers, by matching with the SLLS sample. As a measure of work absence, these data contain three deficiencies. First, as mentioned in the introduction, the share of absence from work that is covered by sickness insurance is estimated to 97.1 per cent. Second, during the time period studied the first day in each sickness spell was not compensated by the insurance and consequently was not reported. Third, those days that the individual does not regularly work (e.g., Saturdays and Sundays), are not covered by the sickness insurance for the first ten compensated days in one spell. After that time period, however, they are covered by the insurance as well and are therefore recorded, in our data.[1] The advantage of using data from the register, is that they generally contain very few, if any, measurement errors.

Data to construct the cost and income variables, i.e. the different components of the individual's income, are obtained from tax registers that are matched with the SLLS. As was shown in Section 2, the cost and virtual income variables are influenced by the compensation level in the sickness insurance and by the income tax. In the studied time period, 90 per cent of the insured income below the social insurance ceiling was compensated by the insurance for all days covered by the insurance, i.e. for the days that we have recorded in the data set. Very few individuals in our data set, only about 3 per cent, have labour incomes above the ceiling. The Swedish income tax system consists of two parts; a proportional tax imposed by the local authorities and a progressive tax imposed by central government. The local tax rate varies between Sweden's 286 communes, but as there is a dense clustering around the mean tax rate,

---

[1]In 1987, the rules of the sickness insurance were changed, the first day is compensated and days that individuals do not regularly work are also covered the first ten days in a spell.

we obtain a simplification by assuming that all individuals pay the mean local tax rate. The income tax is based on the individual's gross income from labour, capital and social security. The mean marginal tax rates are exhibited in Appendix 1.

As labour income is not fully compensated by the sickness insurance, the marginal tax rate is dependent on how many days the individual is absent from work, i.e. it is not independent of the individual decision whether or not to attend work. This means that if the marginal tax rate is changed for the individual, the budget set will vary between days within the time period (a year) that we have aggregated, furthermore it means that the marginal tax rate is endogenous for at least one day during that time period. This implies that we have to approximate these separate budget sets with one for each individual during the aggregated time-period. Thus, we have to approximate the marginal tax rate each individual faces during the aggregated time-period. The approximation used, is the marginal tax rate each individual actually pays for the year studied. This marginal tax-rate is then used to calculate a linear budget set for each individual, i.e. to calculate the cost variable and the average daily virtual non-labour income.

The cost variable, w^*, is calculated as 10 per cent (the share of labour income not covered by the sickness insurance) of the net marginal hourly wage rate, i.e. the hourly wage rate multiplied by one minus the individual's marginal income tax rate, for labour incomes below the social security ceiling. For incomes above the ceiling, the cost is counted as 100 per cent of the hourly wage rate. The hourly wage rate is obtained by dividing the potential annual labour income by the number of hours of work stated in the 1981 survey. Thus, when using the 1981-1984 panel we assume that the individuals do not change their number of hours of work from those in 1981. To calculate the potential annual income from labour, we have added 10 per cent for each day recorded as the individual has been compensated by the sickness insurance. For individuals with insured incomes above the social security ceiling, 100 per cent is added for the part of their incomes exceeding the social security ceiling.

The non-labour income, y, consists of two parts. The first one is the average daily income from capital and child and housing allowances. If the individual is married or cohabiting only, 50 per cent of these allowances are incorporated. The second part is net daily income from sickness insurance.

Contracted working hours, h^*, is based on the self reported number of contracted weakly working hours in 1981, obtained from the survey. The self reported contracted working hours are classified into four levels to obtain h^*. h^* takes the values 2, 4, 6 and 8, if the self reported contracted weekly working hours are $\leq$ 10, >10 and $\leq$ 20, >20 and $\leq$ 30 and > 30, respectively.

The specification of the vector of the socio-economic variables was discussed in Section 2. The conclusion from that discussion was that, in addition to variables related to health, variables relating to working conditions, variables measuring the level of monitoring of workers and the unemployment level should also be considered.

The variables relating to the state of health are obtained from the SLLS survey. Questions on some sixty different illnesses are put forward. The questions are of the following kind *"have you during the last twelve months had* headache, *cold, a heart attack, etc*". Since the survey was conducted during the spring and summer 1981, it is likely that some variables are endogenous, depending on the number of days spent on sickness insurance. Since we include working conditions, some types of illnesses that may be symptoms of bad working conditions, e.g., headache, pain in the chest etc., are not included. A subset of eleven variables, for which a physician has most likely made a diagnosis, are chosen (see Appendix 2 for a description of the variables). The illness variables chosen can be considered as exogenous and are hopefully not severely affected by bad working conditions. Since we are only interested in controlling for differences in the state of health we choose to reduce the dimension by performing a principal component (PC) analysis (see Appendix 2). The three first PC:s, *HS1*, *HS2* and *HS3* are included, they account for almost 60 per cent of the variation in the illness variables. A dichotomous variable, *DISAB*, taking the value one if a person is in some way disabled and prevented from doing his work properly, is also used.

We use two types of working condition variables: occupation specific risk indices and self reported information from the 1981 survey.

The occupation specific risk indices consisting of two SIR indices (Standard Incidence Ratio) measure reported accidents at work and work related diseases for the period 1981-1983 per 100 000 worked hours, respectively, and are provided by the *National Board of Occupational Safety and Health.* 195 different occupations are considered and the indices are matched with the SLLS. As these two indices are highly correlated and measure the same expected influence on absence, we perform a PC analysis (see Appendix 2). The first PC, *RISK*, which takes care of 87 per cent of the variation in the risk variables, is included in the model.

The SLLS survey contains questions on several job characteristics some of which we use in this study. The number of such variables is large and they are highly correlated. Furthermore, they have, in most cases, the same expected influence on work absence. Therefore, a PC analysis is performed (see Appendix 2). We choose to include the first three PC:s that take care of about 40 per cent of the variation in the variables. The first PC, *DISS1*, can be interpreted as a physically demanding factor. The second PC, *DISS2*, can be interpreted as a mentally demanding factor. The third PC, *DISS3*, can not be given any firm interpretation.

The risk indices and the self reported job characteristic variables can be interpreted as measuring different aspects of higher costs for actually attending a job; the costs for risk exposure and disutility, respectively. On the other hand, all the job characteristic variables, e.g., contact with poison and monotonous body movements, are likely to cause work related diseases and accidents. The advantage of using the risk indices is that they give a more objective measure of risk. When an individual characterises his job as physically demanding, this could be generated by his perception of his health rather than by the nature of his work. The advantage of using self reported characteristics is that they are more directly tied to the particular individual's workplace.

Two dummy-variables, if it is "important to arrive on time at the work place", *ITIME,* and "if it is required to clock in", *PUNCH,* are used to measure the monitoring level at the individual's workplace.[2] The unemployment variable, *UNEMP,* is the county specific annual average unemployment rates that have been matched with the original data set.

Data for the number of days spent on sickness insurance and the components of income, can be obtained for the individuals in the sample for the years 1982-1984 from the two registers. There are primarily two advantages with extending the analysis to include these years. First, the number of observations is radically increased. Second, as a result primarily to the 1982 marginal tax reform (see the Figure in Appendix 1), some additional variance in the cost variable is introduced. However, there are also some disadvantages. First, the data on the number of days spent on sickness insurance for 1982 differ somewhat compared with the other years. For this year, the variable measures the number of days in spells terminated in 1982, i.e. spells that begun in 1982 but are terminated in 1983 are counted as zero. Second, the measurement of the variables obtained from the 1981 survey (e.g., hours of work and workplace characteristics) will be inferior as changes from 1981 can not be measured.

As it is impossible to judge between these advantages and disadvantages, we carry out the empirical analysis using both of these data sets. Individuals with zero or missing value on labor income in 1981 - 1984 (either not working, working in the "informal sector" or outside Sweden) are excluded from the sample for the year(s) in question. 129 observations are excluded with this criterion (14, 17, 45 and 67 observations in 1981, 1982, 1983 and 1984, respectively). The resulting samples consist of 1,967 individuals for the 1981 sample and of 7,757 observations for the 1981 - 1984 panel. The number of individuals for each year is given in Appendix 1.

---

[2]These two variables have previously been used by Arai (1993) to measure the monotoring level.

4. Econometric Models and Estimation

Demand for time absent, t^a, is a latent variable that is not directly observable. Instead we can observe absence for two days or more.[3] If we partition the socio-economic variables, s, into γz and θ, where z is a vector of exogenous variables, γ is the corresponding parameter vector and θ is a random or fixed variable taking into account nonobservable variables. For individual i at day h, equation (4) may be written

$$\begin{aligned} t_{ih}^a &= h_{ih}^* + \alpha w_{ih}^* + \beta y_{ih} + \gamma z_{ih} + \theta_i + \zeta_{ih} \\ &= h_{it}^* + \alpha w_{it}^* + \beta y_{it} + \gamma z_{it} + \varepsilon_{ih} + \zeta_{ih}, \end{aligned} \tag{6}$$

where subscript t indicates year, ζ_{ih} are identically and independently distributed (iid) random variables and $\varepsilon_{ih} = \theta_i + \varepsilon_h$, with $\varepsilon_h = h_{ih}^* - h_{it}^* + \alpha(w_{ih}^* - w_{it}^*) + \beta(y_{ih} - y_{it}) + \gamma(z_{it} - z_{ih})$. The annual variables are measured as annual averages. If all explanatory variables as well as h^* are constant over the year, we have that $\varepsilon_h = 0$. This is a situation that is unlikely. Thus even if no unobserved heterogeneity is present, it is likely that we have systematic errors due to time aggregation of the explanatory variables.

Let I_{ih} be an indicator variable for absenteeism such that

$$I_{ih} = \begin{cases} 1, \text{ if } t_{ih}^a > k_{ih} \\ 0, \text{ if } t_{ih}^a \le k_{ih} \end{cases},$$

where k_{ih} is a threshold value for day h. If demand for absence exceeds k_{ih}, individual i will be absent. Under the assumption that ζ_{ih} has a logistic distribution and that $k_{it} = k_{ih}$, the probability to be at work on day h is

$$\Pr(I_{ih} = 0) = \Pr(t_{ih}^a - k_{it} \le 0) = 1/\left[1 + \exp(h_{it}^* + \alpha w_{it}^* + \beta y_{it} + \gamma z_{it} + \varepsilon_{ih} - k_{it})\right],$$

with the probability of being absent $\Pr(I_{ih} = 1) = 1 - \Pr(I_{ih} = 0) = 1/\left[1 + \exp(k_{it} - (h_{it}^* + \alpha w_{it}^* + \beta y_{it} + \gamma z_{it} + \varepsilon_{ih})\right]$.

[3]If an individual is absent two days, it will be recorded as one day. Accordingly an individual who is absent for 365 days during a year, enters the year as absent.

4.1 Estimation

The Binomial Model

Let V_{it} be the number of days person i is absent during year t. Under the assumption of no state dependence, i.e., that ε_{ih} is independent of ε_{ig}, for every h and g, V_{it} is binomially (Bin) distributed with parameters $\pi_{it} = \Pr(I_{ih} = 1)$ and N_t, the number of days in year t, i.e.,

$$V_{it} = \sum_{h=1}^{N_t} I_{ih} \sim \mathrm{Bin}(N_t, \pi_{it}).$$

If V_{it} is independent of $V_{jt'}$, for every i, j, t and t' and if there is no unobserved heterogeneity present (i.e., $\theta_i = 0$, for every i), the log-likelihood function can be written

$$\ell(\omega; V_{11}, \ldots V_{N\tau}) = \sum_{t=1}^{\tau} \sum_{i=1}^{n_t} \ln \binom{N_t}{V_{it}} + V_{it} \ln \pi_{it} + (N_t - V_{it}) \ln(1 - \pi_{it}), \tag{7}$$

where τ is the number of years and n_t is the number of individuals in year t. The maximum likelihood (ML) estimates of the parameter vector $\omega = (\beta, \alpha, \gamma')'$ are obtained by maximizing ℓ in (7).

If there is serial correlation (i.e., ε_{ih} is not independent of ε_{ig}, for every h and g) and/or there is individual unobserved heterogeneity (i.e., $\theta_i \neq 0$), the ML estimator of the Bin model yields inconsistent estimates. Unobserved heterogeneity is a main cause for overdispersion (i.e., the predicted variance is greater than the one expected from the theoretical distribution). Depending on the assumption of the cause for overdispersion, different tests for overdispersion and different estimation procedures are applicable.

Overdispersed Binomial Model

When absence is serially correlated over days, V_{it} is not binomially distributed. Under the assumption of a correct mean function, the estimation can, e.g., be performed by regarding ℓ in (7) as a quasi-likelihood function. This estimator is suggested by McCullagh & Nelder (1983) along with the additional assumption $\mathrm{Var}(V_{it}) = \sigma^2 N_t \pi_{it}(1 - \pi_{it})$, where $\sigma^2 > 1$. The σ^2 is a measure of overdispersion. The same parameter estimates as for the original binomial model are obtained, but the covariance matrix is adjusted with an estimate of σ^2.

The Markov Model

Another approach is to allow for state dependence.[4] If we extend equation (6) into

$$t_{ih}^{a} = h_{it}^{*} + \alpha w_{it}^{*} + \beta y_{it} + \boldsymbol{\gamma}\, \mathbf{z}_{it} + \lambda I_{ih-1} + \zeta_{ih},$$

where λ is a parameter to be estimated, we obtain a two-state Markov chain with state one corresponding to absence and state zero to working (cf. Heckman, 1981). There are four possible transitions $(0 \to 1)$, $(0 \to 0)$, $(1 \to 0)$ and $(1 \to 1)$. The four transition probabilities take the forms $p_{00} = 1/\left[1 + \exp(h_{it}^{*} + \alpha w_{it}^{*} + \beta y_{it} + \boldsymbol{\gamma}\, \mathbf{z}_{it} - k_{it})\right]$, $p_{11} = 1/\left[1 + \exp\{-(h_{it}^{*} + \alpha w_{it}^{*} + \beta y_{it} + \boldsymbol{\gamma}\, \mathbf{z}_{it} + \lambda - k_{it})\}\right]$, $p_{10} = 1 - p_{11}$ and $p_{01} = 1 - p_{00}$, where $p_{jk} = \Pr(y_{ih} = k | y_{ih-1} = j)$, for j, k = 0, 1. The probability of being in a state at time period t is written as $\boldsymbol{p}_t = (p_{0t} \quad p_{1t})$. By the Markov property (e.g., Bhat, 1972, Ch. 3), the probability of the present state depends only on the probabilities in the previous period, i.e., we may write $\boldsymbol{p}_t = \boldsymbol{p}_{t-1}\boldsymbol{P}$, where

$$\boldsymbol{P} = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}.$$

The process is time homogenous within a year, since the elements in the transition matrix **P** cannot reflect daily variation in explanatory variables.

Under the Markov assumption, the expected number of days absent during a year, given that the person is present the first day is (cf. Bhat, 1972, Ch. 3 and Andersson & Brännäs, 1992)

$$\mu_{01}^{(N_t)} = \frac{N_t p_{01}}{p_{01} + p_{10}} + \frac{p_{01}(1 - p_{01} - p_{10})\left[1 - (1 - p_{01} - p_{10})^{N_t}\right]}{(p_{01} + p_{10})^2},$$

and the expected number of days absent, given that the person is absent the first day is

$$\mu_{11}^{(N_t)} = \frac{N_t p_{01}}{p_{01} + p_{10}} + \frac{p_{10}(1 - p_{01} - p_{10})\left[1 - (1 - p_{01} - p_{10})^{N_t}\right]}{(p_{01} + p_{10})^2}.$$

The term raised to N_t tends to zero if p_{00} is large (i.e., the probability for a person to work given that he worked the day before) or if p_{11} is large (i.e., the probability to be absent given absenteeism the day before). Thus, we may exclude the term raised to N_t. Therefore, the mean number of days absent is given by

---

[4]We do not allow for duration dependence which is a restrictive assumption. However, Barmby et al. (1991b) could reject the hypothesis of duration dependence, using a model on daily data.

$$\mu^{(N_t)} = \frac{N_t p_{01}}{p_{01}+p_{10}} + \frac{(1-p_{01}-p_{10})}{(p_{01}+p_{10})^2}(\phi p_{01} + (1-\phi)p_{10}), \tag{8}$$

where ϕ = 1 if the individual is present the first day and zero if absent the first day. The following expression is estimated with respect to $\boldsymbol{\rho}' = (\boldsymbol{\omega}', \lambda)$ using non-linear least squares

$$V_{it} = \mu_{it}^{N_t} + e_{it}, \tag{9}$$

where e_{it} is an error term. The only individuals we know with certainty enter as absent are the ones with 365 days of absence during the year, hence for these individuals ϕ = 0. Rather than attempting to use a switching regression approach, we assume known sample separation such that for the remaining individuals have ϕ = 1. It can be noted that if λ = 0 then $\mu_{it}^{(N_t)} = N_t \pi_{it}$.

Asymptotically, it holds that $\sqrt{n}(\tilde{\boldsymbol{\rho}} - \boldsymbol{\rho})$ is normally distributed with mean 0 and covariance matrix $\boldsymbol{Q}$, where $\tilde{\boldsymbol{\rho}}' = (\tilde{\boldsymbol{\omega}}', \tilde{\lambda})$ (cf. Gallant, 1987, Ch. 1 and 2). Under the assumption of homoscedastic errors, the estimator of the covariance matrix is

$$\tilde{\boldsymbol{Q}}_M = \frac{s^2}{\Sigma_{t=1}^{\tau} n_t}(\tilde{\boldsymbol{M}}'\tilde{\boldsymbol{M}})^{-1}, \tag{10}$$

where $s^2 = \Sigma_{i,t}(V_{it} - \tilde{\mu}_{it}^{(N_t)})^2 / \Sigma_{t=1}^{\tau} n_t$ and $\tilde{\boldsymbol{M}} = \partial\boldsymbol{\mu}(\tilde{\boldsymbol{\rho}})/\partial\boldsymbol{\rho}'$, with $\boldsymbol{\mu}(\tilde{\boldsymbol{\rho}})$ a stacked vector of estimated $\mu_{it}^{(N_t)}$. With heteroscedastic errors of an unspecified form the covariance matrix estimator is

$$\tilde{\mathbf{Q}}_F = \tilde{\boldsymbol{J}}^{-1}\tilde{\boldsymbol{F}}\tilde{\boldsymbol{J}}^{-1}, \tag{11}$$

where $\tilde{\boldsymbol{J}} = \tilde{\boldsymbol{M}}'\tilde{\boldsymbol{M}}/\Sigma_{t=1}^{\tau} n_t$ and $\tilde{\boldsymbol{F}} = \Sigma_{i,t}\tilde{e}_{it}^2(\partial\tilde{\mu}_{it}^{(N_t)}/\partial\boldsymbol{\rho})(\partial\tilde{\mu}_{it}^{(N_t)}/\partial\boldsymbol{\rho})'/\Sigma_{t=1}^{\tau} n_t$.

The Mixture Distribution Model

Unobserved heterogeneity can be modelled if we let the stochastic parameter $\varepsilon_{ih} = \theta_j$ in the demand for absence function, hence

$$t_{ihj}^a = h_{ih}^* + \alpha w_{ih}^* + \beta y_{ih} + \gamma z_{ih} + \theta_j + \zeta_{ih}.$$

Heterogeneity is individual if $i = j$ and group specific if the number of groups, K, is less than n. A semiparametric estimator may be used (e.g., Andersson & Brännäs, 1992) to estimate the unknown parameters, $\boldsymbol{\omega}$, and the distribution function $H(\theta)$ of θ. The parameters and the distribution function are estimated jointly. The nonparametric ML estimator of $H(\theta)$ takes the form of a discrete distribution function. By this, the probability for a person to be absent on day h equals

$$\Pr(I_{ih}=1)=\sum_{j=1}^{K} q_j \Pr(I_{ih}=1|\theta_j),$$

where $\Pr(I_{ih}=1|\theta_j)=1/\left[1+\exp(k_{it}-(h_{it}^{*}+\alpha w_{it}^{*}+\beta y_{it}+\gamma z_{it}+\theta_j))\right]$. This yields the log-likelihood function

$$\ell \propto \sum_{i,t} V_{it} \ln \sum_{j=1}^{K} q_j \Pr(I_{ih}|\theta_j)+(N_t-V_{it})\ln\left(1-\sum_{j=1}^{K} q_j \Pr(I_{ih}|\theta_j)\right) \tag{12}$$

that is maximised with respect to $\boldsymbol{\omega}$, the probabilities $q_1,\ldots,q_K$, the mass points $\theta_1,\ldots,\theta_K$ and the finite number K. Leroux (1992) shows that the number of groups may be consistently estimated using the Akaike information criteria (AIC, Akaike, 1973) or the Baysian information criteria (BIC, Schwarz, 1978).

The probability that a given observation is associated with a particular group j (j=1, ..., K) may be obtained by the Bayes rule as

$$d_j=\Pr(j|V_{it})=q_j \Pr(V_{it}|j)\Big/\sum_{j}^{K} q_j \Pr(V_{it}|j). \tag{13}$$

The d_j is a logistic discrimination probability.

4.2 Overdispersion Tests

It is important to test for overdispersion in qualitative dependent variable models because generally, if present, this leads to inconsistent estimates (see Godfrey, 1990, Ch. 6 and Yatchew & Griliches, 1985). On the other, hand if the overdispersion is on the borderline of detectability and the mean function is correctly chosen, the ML estimator without overdispersion is asymptotically efficient (Cox, 1983).

Under the assumption that $\mathrm{Var}(V_{it})=\sigma^2 N_t \pi_{it}(1-\pi_{it})$, tests for overdispersion can be made using, e.g., the Pearson chi-square statistic

$$\chi^2=\sum_{i,t}(V_{it}-N_t\hat{\pi}_{it})^2/\left[N_t\hat{\pi}_{it}(1-\hat{\pi}_{it})\right],$$

that is asymptotically $\chi^2(\Sigma_{t=1}^{\tau} n_t - p)$ distributed and where p is the number of estimated parameters. A satisfactory (when $\Sigma_{t=1}^{\tau} n_t \to \infty$, see McCullagh & Nelder, 1983) estimator for σ^2 is

$$\hat{\sigma}^2=\chi^2/(\Sigma_{t=1}^{\tau} n_t - p).$$

Dean (1992) proposed a score test for extra-binomial variation under the assumption that, $\Pr(I_{ih}=1|\theta_i) = /\left[1+\exp(k_{it}-(h^*_{it}+\alpha w^*_{it}+\beta y_{it}+\gamma z_{it}+\theta_i))\right]$, where θ_i are iid with $E(\theta_i)=0$ and $\mathrm{var}(\theta_i)=\sigma^2_\theta$. The score test for testing $\sigma^2_\theta=0$, i.e., to test if the model (7) is adequate, is based on the log-likelihood function under the alternative hypothesis,

$$\partial\ell/\partial\sigma^2_\theta\big|_{\sigma^2_\theta=0} = T(\hat{\boldsymbol{\omega}}),$$

where $\hat{\boldsymbol{\omega}}$ is the ML estimate when $\sigma^2_\theta=0$ and $T_{it}(\hat{\boldsymbol{\omega}})=(V_{it}-N_t\hat{\pi}_{it})^2-N_t\hat{\pi}_{it}(1-\hat{\pi}_{it})$. The Fisher information matrix, $\boldsymbol{I}(\hat{\boldsymbol{\omega}},\sigma^2_\theta)$, evaluated at $\sigma^2_\theta=0$ is used to obtain an estimator of the variance of $T(\hat{\boldsymbol{\omega}})$. If the information matrix is partitioned as

$$\mathbf{I}(\hat{\boldsymbol{\omega}},\sigma^2_\theta)=\begin{bmatrix} \boldsymbol{I}_{\omega\omega} & \boldsymbol{I}_{\omega\sigma^2_\theta} \\ \boldsymbol{I}'_{\omega\sigma^2_\theta} & I_{\sigma^2_\theta\sigma^2_\theta} \end{bmatrix},$$

the score test takes the form

$$S = T(\hat{\boldsymbol{\omega}})/\hat{W}, \tag{14}$$

where $\hat{W}^2 = I_{\sigma^2_\theta\sigma^2_\theta} - \boldsymbol{I}_{\omega\sigma^2_\theta}\boldsymbol{I}^{-1}_{\omega\omega}\boldsymbol{I}'_{\omega\sigma^2_\theta}$. S is asymptotically normally distributed with expectation zero and variance one.

5. Results

In this section we will discuss the results. First, we give a general description and comparison between the different estimation methods used. We then discuss individual, and groups of, parameter estimates.

The parameter estimates for the 1981 data and the 1981-1984 panel are shown in Table 1 and Table 2, respectively.[5] The first three columns in each table give the parameter estimates and standard errors (*se*) from the binomial ML estimation as well as the QL corrected *se*'s. Using this estimation method, we find that overdispersion is present using both data sets; χ^2 = 195582.02, S = 3174.22 and $\hat{\sigma}^2$ = 100.50, for the 1981 data and χ^2 = 833525.94, S = 6469.48 and $\hat{\sigma}^2$ = 107.41 for the panel. The coefficients of determination, are as expected, quite small, R^2 = 0.11 and 0.08, respectively.

[5]It must be noted that the estimated parameters are standardized by the standard error in the logistic distribution. Thus the estimates can not be interpreted as the structural ones and it would not be correct to restrict the parameter of h^* to be one. Consequently, it is not possible to test if this estimated parameter is different from one.

Table 1. *Parameter estimates and standard errors for the Binomial ML, Markov model NL and mixture distribution semiparametric estimation for the* 1981 sample. *(se(QL) and se(H) indicates standard errors computed from the quasi-likelihood and heteroscedastic covariance matrix, respectively. *indicates a parameter that is insignificant, at the* 0.05 *level).*

Variable/ Parameter	Binomial model			Markov model			Mixture model	
	Estimate	*se*	*se*(QL)	Estimate	*se*	*se*(H)	Estimate	*se*
Const (θ_1)	-5.2054	.0537	.5383	-4.9385	.3708	.5456	-6.2921	.0865
λ/θ_2				2.9736			8.6460	.5830
h^*	.1204	.0051	.0512	.1276	.0346	.0509	.1800	.0074
y	-.0008	.0002	.0024*	-.0020	.0015*	.0025*	-.0019	.0003
w^*	-.0404	.0085	.0850*	.0008	.0634*	.0559*	-.0494	.0125
UNEMP	-.1719	.0058	.0583	-.1964	.0402	.0752	-.2728	.0099
FEMALE	.4688	.0142	.1428	.1672	.0855	.1481*	.5104	.0197
SINGLE	.1659	.0154	.1542*	.0182	.0951*	.1841*	.1346	.0219
DIV	.1589	.0161	.1617*	-.0702	.1008*	.1581*	.1680	.0209
AGE	.0251	.0005	.0053	.0165	.0033	.0054	.0319	.0008
KIDS6	-.1399	.0125	.1251*	.0296	.0874*	.1235*	-.1459	.0202
KIDS16	-.1134	.0076	.0757*	-.2204	.0691	.0751	-.3100	.0167
ITIME	.1705	.0130	.1299*	.2200	.0960	.1484*	.3429	.0212
CLOCK	.1856	.0115	.1152*	.1280	.0706*	.1273*	.1828	.0157
DISS1	.2481	.0073	.0730	.0963	.0459	.0866*	.2539	.0102
DISS2	-.2993	.0098	.0981	-.3751	.0675	.1111	-.4863	.0166
DISS3	.0655	.0095	.0949*	.1227	.0600	.1120*	.1138	.0133
RISK	.0942	.0051	.0510*	.0426	.0320*	.0447*	.1241	.0069
DISAB	.9327	.0108	.1087	.6969	.0686	.1027	1.2174	.0185
HS1	-.1111	.0148	.1484*	-.4070	.1792	.2630*	-.8783	.0426
HS2	-.5750	.0201	.2013	-.9925	.4208	.5581*	-2.4695	.0957
HS3	.3834	.0325	.3254*	1.6064	.8645*	1.1665*	4.1156	.1969
q_1							.9815	.0006

Note: Let $q_1 = 1/(1+e^{-\phi})$ and maximize with respect to the unconstrained ϕ. On convergence we get $\hat{q}_1$. The variance is estimated from the Gauss approximation $V(\hat{q}_1) = \hat{\sigma}^2_{\hat{\phi}} e^{-2\hat{\phi}} / (1+e^{-\hat{\phi}})^4$.

Columns 4, 5 and 6 show the Markov model parameter estimates and *se*'s from the non-linear least squares estimation (grid search over λ is used for the estimation, as there were problems with convergence with unrestrictedly minimizing (9). Thus no standard errors for the estimate of λ, $\tilde{\lambda}$, are estimated). The *se*'s are estimated from the covariance matrix (10) and from the heteroscedasticity consistent covariance matrix (11), respectively. The coefficient of determination, R^2 = 0.15 and 0.09, shows only a small change from the binomial model. The values of λ that minimise the sums of squares are $\tilde{\lambda}$ = 2.97 and 4.70. To test $H_0: \lambda = 0$ against $H_A: \lambda \neq 0$, we use (see, e.g. Gallant, 1987, Ch. 1.5) $F = \tilde{\lambda}\left(\boldsymbol{H}(\tilde{\boldsymbol{M}}'\tilde{\boldsymbol{M}})^{-1}\boldsymbol{H}'\right)^{-1}\tilde{\lambda}/\tilde{\sigma}^2$, which is $F(1, n\tau\text{-}p)$ distributed under H_0. Here, $\boldsymbol{H}$ is a vector that picks out the appropriate variance

element for λ. We obtain F = 0.131 and 0.015, thus the Markov model is not a significant improvement over the binomial model for either of the two data sets.

Table 2. *Parameter estimates and standard errors for the Binomial ML, Markov model NL and mixture distribution semiparametric estimation for the* 1981-1984 panel. (*se*(*QL*) *and se*(*H*) *indicates standard errors computed from the quasi-likelihood and heteroscedastic covariance matrix respectively. *indicates a in-significant parameter, at the* 0.05 *level).*

Variable/ Parameter	Binomial model			Markov model			Mixture model	
	Estimate	*se*	*se*(QL)	Estimate	*se*	*se*(H)	Estimate	*se*
Const (θ_1)	-4.3350	.0249	.2581	-4.6106	.1361	.1986	-5.0183	.0426
λ/θ_2				4.7010			4.8792	1.0969
h^*	.0639	.0023	.0238	.0508	.0123	.0200	.0834	.0031
y	-.0015	.0001	.0010*	-.0015	.0005	.0007	-.0020	.0001
w^*	-.0010	.0010	.0099	-.0008	.0057*	.0018*	-.0016	.0011
UNEMP	-.1342	.0024	.0246	-.0917	.0129	.0193	-.1801	.0036
FEMALE	.3659	.0071	.0738	.1276	.0309	.0453	.4029	.0092
SINGLE	.0873	.0078	.0810*	.0147	.0422*	.0622*	.1011	.0102
DIV	.2896	.0079	.0822	.1308	.0377	.0586	.3224	.0096
AGE	.0191	.0003	.0027	.0115	.0014	.0019	.0244	.0004
KIDS6	-.1661	.0060	.0620	-.0812	.0409	.0400	-.2351	.0091
KIDS16	-.0801	.0036	.0373	-.1016	.0244	.0267	-.1356	.0062
ITIME	.1183	.0064	.0660*	.1462	.0373	.0501	.1763	.0090
CLOCK	.1496	.0058	.0600	.1176	.0280	.0416	.1566	.0074
DISS1	.2315	.0037	.0379	.0397	.0193	.0286*	.2817	.0050
DISS2	-.2518	.0049	.0506	-.0162	.0240*	.0345*	-.3584	.0080
DISS3	.0237	.0047	.0488*	.1026	.0251	.0378	.0504	.0061
RISK	.1048	.0025	.0263	-.0167	.0125*	.0197*	.1303	.0033
DISAB	.7864	.0055	.0572	.5353	.0276	.0380	.9584	.0098
HS1	.0287	.0070	.0722*	.0072	.0396	.0561*	.0049	.0089 *
HS2	.3009	.0136	.1408	.2502	.1282*	.0771	.4788	.0244
HS3	.2064	.0093	.0968	.1861	.0434	.0491	.2199	.0112
q_1							.9858	.0005

Note: Let $q_1 = 1/(1+e^{-\phi})$ and maximize with respect to the unconstrained ϕ. On convergence we get $\hat{q}_1$. The variance is estimated from the Gauss approximation $V(\hat{q}_1) = \hat{\sigma}^2_{\hat{\phi}} e^{-2\hat{\phi}} / (1+e^{-\hat{\phi}})^4$.

The mixture model is semiparametrically estimated (the BFGS algorithm in GAUSS) with K = 1, 2, ..., T treated as fixed. As can be seen from Table 3, the minimum AIC and BIC are obtained for K = 2 using both data sets. Thus, the estimates, shown in columns 7 and 8, are those corresponding to K = 2. The first group is very large for both data sets, 98.7 per cent for the 1981 sample and 98.4 per cent for the panel. The rest of each sample, naturally belongs to the second group. The signs of the intercepts for the two groups reveal that the probability to be absent is larger for the second group.

Table 3. *Information criteria values for different K.*

	the 1981 sample		*the 1981-1984 panel*	
K	AIC	BIC	AIC	BIC
1	152442.23	152995.29	618339.28	618392.33
2	152785.01	152843.65	618055.05	618113.68
3	152787.00	152851.22	618056.80	618121.02

By using Bayes rule or the logistic discrimination probability (13), we classify the individuals into group one if the probability d_1 is larger than 0.5 and into group two if otherwise. Box plots of the distribution of days absent for the two samples are shown in Figure 3. The second group may be interpreted as primarily consisting of the long term sick and thus less sensitive to economic incentives to be absent from work. The classifying rule above (i.e., using the mean prediction, $\hat{V}_{ij} = N_t\hat{p}_{ij}$, for individual i belonging to group j) is used to estimate the coefficient of determination, R^2 = 0.53 and 0.55 for the two samples, respectively. Compared with the other two estimation methods this is a major improvement in goodness of fit.

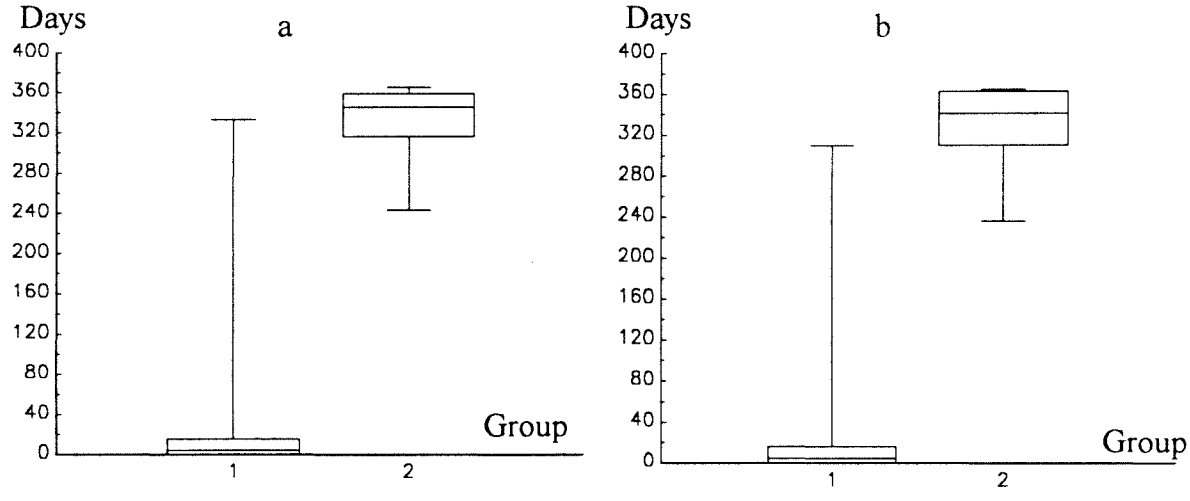


Figure 3. *Box plot of the number of days absent for the two groups in a)* 1981 *and b) the panel* 1981-1984.

(i) Cost and income elasticities. The estimates of α, range from 0.0008 to -0.0494. The only positive estimate, inconsistent with respect to the Slutsky conditions, is that of the Markov model for the 1981 sample. However, it is not significantly different from zero. The largest estimate in absolute value, significantly different from zero, was obtained when using the semiparametric estimator on the 1981 sample. The estimates of β, i.e., the coefficient for virtual income, show very small variation. The estimates are between -0.0020 and -0.0008 and significantly different from zero when the semiparametric estimator is used and when the Markov model is used for the panel 1981-1984. Negative coefficient estimates for both wage rate and disposable income are also obtained by Allen (1981a).

As the income elasticities are negative for all estimation methods, the Slutsky condition (5) is not satisfied for every $t^a > 0$ for any of the estimation techniques and, as mentioned above, not satisfied at all for the Markov model on the 1981 sample. However, if the Slutsky condition is solved for t^a, it can be seen to apply for every $t^a < 50.5$ and 26.0 for the binomial and mixture models, respectively, using the 1981 sample. Hence, it applies for all individuals since t^a is a latent variable measuring hours of work absence per day and it is impossible to work this many hours. Using the panel, the results can not be unambiguously interpreted since the Slutsky conditions are satisfied for every $t^a < 0.67$, 0.53 and 0.80 using the binomial, Markov and mixture models respectively.

Elasticities facilitate the interpretation of the economic content of the parameter estimates. The elasticity of the expected number of days absent, $E(V_{it})$ with respect to the *k*th explanatory variable, z_k, is

$$\eta_{itk} = \frac{\partial E(V_{it})}{\partial z_k}\frac{z_k}{E(V_{ik})} = z_{ik}(1-\pi_{it})\gamma_k,$$

where γ_k is the parameter corresponding to z_k and since $\partial E(V_{it})/\partial z_k = N_t \partial \pi_{it}/\partial z_k = -N_t \pi_{it}(1-\pi_{it})\gamma_k$. The elasticity of the contribution level is

$$\eta_{it\delta} = \frac{\partial E(V_{it})}{\partial \delta}\frac{\delta}{E(V_{ik})} = -\delta(1-\pi_{it})(\alpha w - \beta h^* w),$$

since $\partial E(V_{it})/\partial\delta = N_t \partial\pi_{it}/\partial\delta = -N_t \pi_{it}(1-\pi_{it})(\alpha w - \beta h^* w)$. Using the estimates from the 1981 sample and mixture distribution model (results where the Slutsky condition is satisfied) the elasticity of the mean number of days absent with respect to the contribution level, δ, is calculated for the two groups separately. This yields a mean value of 0.441 and 0.00001 for groups one and two, respectively. Thus a reduction of δ by 10 per cent would decrease the mean level of days absent by about 4.4 per cent in group one. A box plot of the distribution of the elasticities for the 1981 samples are shown in Figure 4.

(ii) State of health. The parameter estimates on *DISAB* are, always positive and strongly significant, whereas the parameter estimates for the PC:s are hard to interpret. The estimates are only significant in one case out of three and at least the first PC, in the 1981 sample, is of a unanticipated sign (see the loading in Appendix 1). Two possible explanations are that the variables relating to health are highly correlated with *DISAB* and that those who are on maternity leave have the right to stay home without being registered as absent.

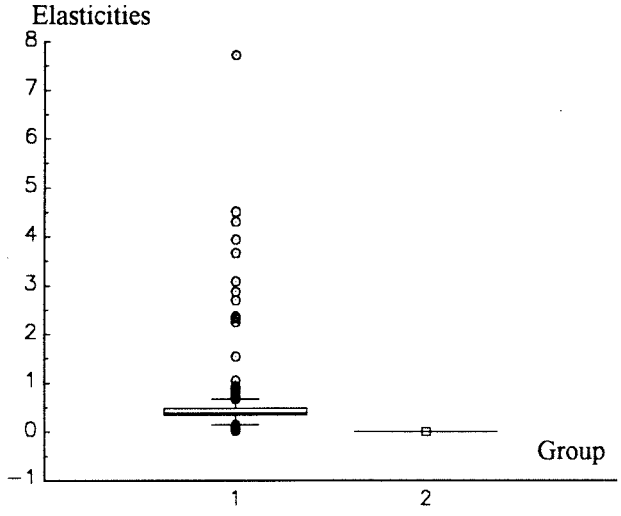


Figure 4. *The distribution of the estimated elasticity of number of days absent with respect to δ for the two groups in* 1981.

(iii) Risk and disutility of work. The risk variable has a positive coefficient for all methods used and, in most cases, also significant. This effect on work absence should be taken into account in, for example, the cost-benefit analysis of increased safety at work. The principal components for the self reported variables measuring the disutility of work have with, most approaches, significant effects on work absence. The effect, however, is not uniform. The second principal components, the contrast between repetitive, monotonous (negative loadings) and outdoor (positive loadings) job characteristics, have a significant (except for the Markov model for the panel) negative coefficient estimate. Thus individuals with the work profile: low stress and outdoor work, have on average a lower absence rate. As the risk and disutility components are highly correlated, it is, unfortunately, not possible to separate these effects.

(iv) Monitoring level. The two dummy variables used to measure the monitoring level, whether or not the individual uses a punch clock and whether or not it is important to be on time, both have positive signs and are significant for almost all techniques. Following the discussion in Section 2, this result was not expected. There are, however, several possible explanations for this. First, a small fraction of work absence is not reported to the sickness insurance (see SAF, 1986), it is likely that the reporting rate is higher in firms with a higher monitoring level. Second, lower work time flexibility will probably induce higher absence as, for example, visits to the dentist or the bank must be reported if the work time flexibility is low.

(v) Unemployment rate. Unemployment has a significant inverse effect on work absence for all estimation techniques. The elasticity with respect to the unemployment rate is, using the 1981 sample and the mixture distribution model, -0.623 for group one and -0.00001 for group two. Thus an increase of one per cent in the unemployment level would decrease the number of days

absent by 0.6 per cent in group one. This is a fairly strong effect and could very well be an important reason for the decline in work absence that has been observed recently.

(vi) Personal characteristics. The results are very similar to those obtained in other empirical studies on this issue (see primarily Björklund, 1991). Age has a significant positive effect. This result was expected since the state of health detoriates on average with age. Women are significantly more absent than men. Unmarried persons are absent more than married.

6. Conclusions and Further Research

To give an illustration of what our results tell us about the interpretation of changes in work absence seen recently, we will predict the change in work absence between 1990 and 1991 and compare the result with the actual outcome on aggregate data. The reason for analysing the change between these two particular years, is because two interesting alterations took place in 1991. First, on March 1 the level of the sickness insurance paid was reduced from 90 to 60 per cent of the wage for the first three days and from 90 to 80 per cent for days 4 - 89. Second, the unemployment level increased from 1.5 to 2.7 per cent. Table 4 shows the change in the average number of days absent from work because of sickness, divided into different length of spell of illness, between these years.

Table 4. *Average number of days on sickness insurance divided into different kinds of spells.* 1990 *and* 1991. *All insured individuals.*

Spell length, days	*1990*	*1991*	*Difference*
1 - 6	4.80	3.82	-0.98
7 - 14	1.44	1.12	-0.31
15 - 89	5.76	4.95	-0.81
90 - 364	4.56	4.28	-0.28
365 -	7.44	8.33	0.89
All	24.00	22.50	-1.50

Under the assumption that the spell length within the first group, 1 to 6 days is uniformly distributed for both 1991 and 1992 and that 10/12 of the work absence occurred after March 1, the cost of being absent from work increased, on average by 8.1 per cent in 1991 compared with 1990.

For this experiment we have chosen the estimates from the mixture distribution model and the 1981 sample. The motive for using the results from this model is, as is shown in Section 5, that it is superior to the binomial model and it could not be shown that the Markov model is superior to the binomial model. Our motive for using the 1981 sample rather than the panel is, in addition to the arguments given in Section 3, that the Slutsky condition is ambiguously

satisfied for the results using this data set and, thus not consistent with our theoretical model. Using these estimates and the presumptions given above, our predictions for the change between 1990 and 1991 is that days of work absence will decrease by 4.50 per cent in the first group and 4.45 in the whole sample. About 79 per cent of this change is attributed to the alteration in the levels of sickness insurance paid and the rest to the changes in the unemployment level.

The actual outcome is, as can be seen in Table 4, that work absence due to sickness decreased from an average of 24.00 to 22.50 days, i.e. a 6.25 per cent decrease. Thus, our predictions underestimate the actual outcome. However, there are several reasons why we failed to predict the actual outcome of the changes between these two years.

The actual outcome is based on all insured individuals while our estimates are for a subsample of blue collar workers. The estimates originate from data ten years before the predictions. One major change in the sickness insurance was realized in 1987 with the result that the first day on sick-leave is compensated and the days that individuals do not regularly work are covered for the first ten days. This would result in an underprediction of days on sickness insurance in our model since the number of absences for just one day is not represented in our model.

In addition, other alterations of potential importance for the changes in the average number of days on sickness insurance may have either not been identified in this study, or have been identified but the change between 1990 and 1991 could not be measured, e.g., changes in working conditions. Finaly, it should again be stressed that the estimates of cost and income elasticities have large variances and vary between different estimation techniques.

Further research

There are several ways to improve the results, both through better data and further methodological research in this area.

The estimates of the cost and income parameters would improve if there was more variation in primarily the cost variable. To obtain this, it is desirable to have a change in the level of the sickness insurance paid in relation to the wage within the sample period, or alternatively a major reform of the marginal tax rates, such as the 1991 income tax reform in Sweden. Several alternative definitions of the income variable are also possible. It can, for instance, be argued that household, rather than individual, income is the most appropriate income concept for the kind of decision modelled in this study.

It would also be desirable to have information on differences in the indirect costs of being absent. According to the dual labour market hypothesis (cf. Piore, 1972), the labour market can be divided into two sectors: one primary sector with relatively high wages and

opportunities for career advancements and one secondary sector with no on-the-job-training and small opportunities for advancement. There are non-economic barriers that prevent the secondary sector workers from entering the primary sector. There are probably higher costs, in terms of forgone career possibilities, for being absent from work in the primary sector. More information, primarily on how to classify different jobs, is, however, needed in order to test this hypothesis.

Differences between different demographic groups deserve further investigation. It could be questioned whether the differences in behaviour between men and women can be properly accounted for by only, as was done in this study, allowing for differences in intercept. Age-groups could also differ more than is allowed for with a linear specification.

In this study we have tried to deal with two methodological problems, time aggregation and individual heterogeneity. Handling individual heterogeneity is, in our opinion, a key issue when studying absenteeism behaviour empirically. In this study a semiparametric estimator is used to empirically separate the individuals in the sample as they respond, primarily because differences in their state of health, very differently to economic incentives. An alternative to this method is to use the dependent variable to sample different groups and use estimation techniques to handle the problem of endogenous sampling (e.g., Cassel, 1989). A subject for further research is to compare these methods.

References

Akaike, H. (1973). "Information theory and an extension of the maximum likelihood principle." In *Second International Symposium on Information Theory* (B. N. Petrov and F. Csaki, eds) 267-281. Budapest: Akademiai Kiado.

Arai, M. (1993). "Compensating wage differentials versus efficiency wages. An empirical Study of Job Autonomy and Wages." Forthcoming in *Industrial Relations.*

Allen, S. G. (1981a). "An empirical model of work attendance." *Review of Economics and Statistics* **63**, 77-87.

Allen, S. G. (1981b). "Compensation, safety, and absenteeism: evidence from the paper industry." *Industrial and Labor Relations Review* **34**, 207-18.

Andersson, T. & Brännäs. K. (1992). "Explaining cross-country variation in nationalization frequencies." *International Review of Economics and Business* **40**, 47-62.

Avery, R. B. & Hotz. V. J. (1984). "Statistical models for analyzing absentee behaviour." In *Absenteeism* (P.S. Goodman, R. S. Atkin, eds.) 158-93. San Francisco: Jossey-Bass Inc. Publisher.

Barmby, T. A., Orme, C.D. & Treble, J.G. (1991a). "Worker absenteeism: an analysis using micro data." *The Economic Journal* **101**, 214-29.

Barmby, T. A., Orme, C.D. & Treble, J.G. (1991b). "Analysis of absenteeism using discrete panel data." Paper presented at the 1991 European meeting of the Econometric Society, Cambridge.

Bhat, U.N. (1972). *Elements of Applied Stochastic Processes.* New York: John Wiley & Sons.

Björklund, A. (1991). "Vem får sjukpenning? En empirisk analys av sjukfrånvarons bestämningsfaktorer." In *Arbetskraft, arbetsmarknad och produktivitet* 289-299. Stockholm: Produktivitetsdelegationen.

Blomquist, N. S. (1983). "The effect of income taxation on the labor supply of married men in Sweden." *Journal of Public Economics* **22**, 169-197.

Cassel, C-M. (1989) "Analysing endogenously selected samples." Mimeo, Stockholm School of Economics.

Chesher, A. (1984). "Testing for neglected heterogeneity." *Econometrica* **52**, 865-872.

Cox, D. R. (1983). "Some remarks on overdispersion." *Biometrica* **70**, 269-274.

Dean, C. P. (1992). "Testing for overdispersion in Poisson and binomial regression models." *Journal of the American Statistical Association* **87**, 451-457.

Delgado, M. A. & Kniesner, T. J. (1992). "Semiparametric versus parametric count-data models - econometric considerations and estimates of a hedonic-equilibrium model of worker absenteeism", Mimeo, Indiana University, Bloomington.

Dunn, L. F. & Youngblood, S. A. (1986). "Absenteeism as a mechanism for approaching an optimal labor market equilibrium: an empirical study." *The Review of Economics and Statistics* **68**, 668-674.

Eriksson, R. & Åberg, R. (1987). *Welfare in Transition - Living Conditions in Sweden 1968 - 1981.* Oxford: Clarendon Press.

Godfrey, L. G. (1990). *Misspecification Tests in Econometrics.* Cambridge: Cambridge University Press.

Goodman, P. S. & Atkin, R. S. (eds.) (1984). *Absenteeism.* San Francisco: Jossey-Bass Inc., Publisher.

Gallant, A.R. (1987). *Nonlinear Statistical Models.* New York: John Wiley & Sons.

Hausman, J. A. (1980). "The effect of wages, taxes, and fixed costs on women's labor force participation." *Journal of Public Economics* **14**, 161-194.

Heckman, J. J. (1981). "Statistical models for discrete panel data." In *Structural Analysis of Discrete Data* (C. Manski, and D. McFadden, eds.) 114-178. Cambridge, Mass.: MIT Press.

Henrekson, M., Lantto, K. & Persson, M. (1992). *Bruk och Missbruk av Sjukförsäkringen.* Stockholm: SNS Förlag.

Leroux, B. G. (1992). "Consistent estimation of a mixing distribution." *The Annals of Statistics* **20**, 1350-1360.

McCullagh, P. & Nelder, J. A. (1983). *Generalized Linear Models.* London: Chapman and Hall.

OECD (1985). *Measuring Health Care.* Paris: OECD.

Piore, M. J. (1975). "Notes for a theory of labor market stratification." In *Labour market segmentation* (R. C. Edwards et al., eds.). Cambridge, Mass.: D. C. Heath.

Rosen, S. (1986). "The theory of equalising differences." In *Handbook of Labor Economics*, **1** (O. Ashenfelter and R. Layard, eds.). Amsterdam: Elsevier.

SAF (1986). *Tidsanvändningsstatistik.* Stockholm: Svenska Arbetsgivareföreningen.

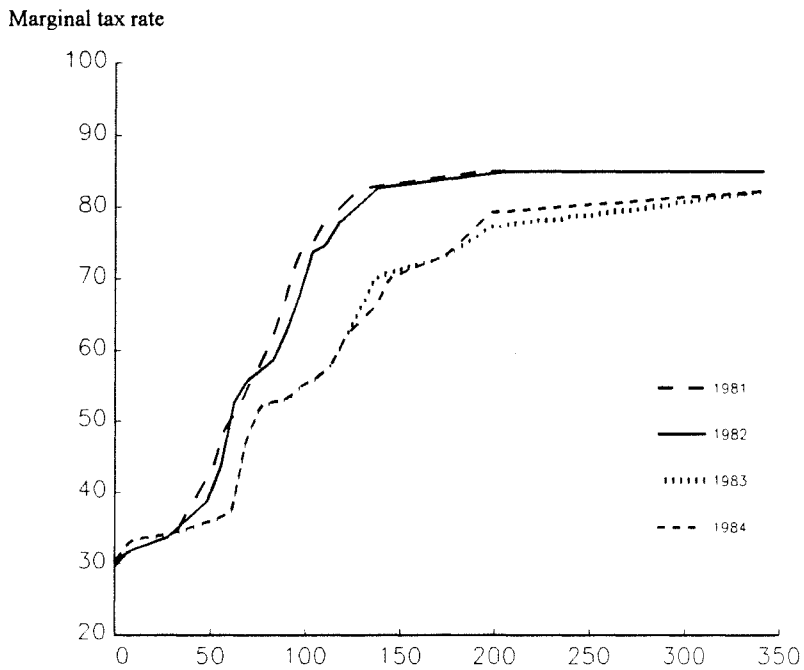
Schwarz, G. (1978). "Estimating the dimension of a model." *The Annals of Statistics*, **6**, 461-464.

Shapiro, C. & Stiglitz, J. E. (1984). "Equilibrium unemployment as a worker discipline device." *American Economic Review* **74**, 433-44.

Titterington, D.M., Smith, A.F.M. & Makov, U.E. (1987). *Finite Mixture Distributions.* Chichester: John Wiley & Sons.

Yatchew, A. & Griliches, Z. (1985). "Specification error in probit models." *The Review of Economics and Statistics* **63**, 134-139.

Appendix 1. Description.



Mean marginal tax rates in 1981, 1982, 1983 *and* 1984.

Descriptive statistics for the number of absences.

Year	Mean	St. dev.	Min	Max	Number of observations
1981	22.31	53.43	0	365	1967
1982	16.89	37.47	0	344	1964
1983	24.89	55.28	0	365	1936
1984	25.13	58.25	0	365	1914
Total	22.27	52.03	0	365	7781

Descriptive statistics for the independent variables, the 1981 years sample.

		Mean	*St.dev.*	*Min*	*Max*
	Economic variables				
h^*	Contracted labor time	7.25	1.38	2.00	8.00
w^*	Cost of being absent	1.40	.94	.14	25.03
y	Virtual income	111.73	29.15	-3.56	199.73
UNEMP	Unemployment rate	2.37	.92	1.00	5.50
	Dangerous or unhealthy work environmens				
DISS1	First PC (see Appendix 2)	1.18	.81	.00	3.36
DISS2	Second PC (see Appendix 2)	-.41	.59	-1.92	1.22
DISS3	Third PC (see Appendix 2)	-.33	.56	-1.84	1.19
RISK	First PC (see Appendix 2)	-.00	1.32	-1.59	6.48
	Control at the workplace				
TIME	Punctuality is important	.74	.44	.00	1.00
PUNCH	Use of punch clock is required	.33	.47	.00	1.00
	Personal characteristics				
SINGLE	Single	.20	.40	.00	1.00
DIVORCED	Divorced	.08	.27	.00	1.00
FEMALE	Female	.53	.50	.00	1.00
AGE	Age of individual	38.37	11.70	20.00	61.00
CHILD6	Number of children below 6 year	.27	.58	.00	3.00
CHILD16	Number of children below 16 year	.47	.79	.00	6.00
	State of health				
DISAB	Disabled	.18	.27	.00	1.00
HS	First PC (see Appendix 2)	.09	.32	-.02	2.28
HS2	Second PC (see Appendix 2)	-2.15	.26	-2.14	.99
HS3	Third PC (see Appendix 2)	-.04	.24	-.18	.50

Appendix 2: Principal component analysis

Loadings for the principal components, HS1, HS2 and HS3.

	1981			1981-1984		
Variable/PC	*HS1*	*HS2*	*HS3*	*HS1*	*HS2*	*HS3*
INJURED	.14	-.88	-.41	.01	.01	-.01
STRUMA	.04	.01	.03	.02	.02	.02
TB	.00	.00	.00	.00	.00	.01
HEARTH	.01	.00	.01	.01	.01	.01
STOMACH	.09	-.19	.02	.08	.17	.95
HEMORR	.96	.07	.21	.98	.16	-.10
PREGNANT	.17	.42	-.88	.17	-.97	.16
BROCK	.01	-.01	.02	.01	.02	.02
MENTAL	-.00	-.06	.00	-.00	.05	.24
CANCER	-.00	.00	.01	-.01	.00	-.01
DIABETIC	.00	.01	-.01	-.01	-.01	.03
NEURO	.00	-.03	-.00	-.00	.02	.05
Variance/total variance in %	26.33	16.56	14.23	30.60	16.79	13.13

Loadings for the principal components (PC), RISK 1981 *and* 1981-1984.

Variable/PC	*RISK*
RISK1	.71
RISK2	.71
Variance/total variance in %	87.15

Loadings for the principal components for (un)healthy conditions, DISS1, DISS2 and *DISS3.*

	1981			1981-1984		
Variable/PC	*DISS1*	*DISS2*	*DISS3*	*DISS1*	*DISS2*	*DISS3*
DIRTY	.19	.19	.04	.19	.18	.04
NOISE1	.13	-.02	.15	.13	-.02	.15
NOISE2	.13	.18	-.14	.13	.18	-.14
OUTSIDE	.17	.45	-.16	.17	.45	-.16
TEMP	.16	-.04	.09	.16	-.04	.09
DRAFT	.27	.09	.02	.27	.08	.02
SMOKE	.24	.02	.21	.24	.02	.20
SHAKE	.10	.02	.03	.10	.02	.03
POISON	.11	.04	.01	.11	.03	.01
LIFT	.26	.22	-.17	.26	.22	-.16
OTHPHY	.41	-.02	-.07	.41	-.02	-.07
SWEAT	.39	.04	.00	.39	.04	.00
PHYEXH	.26	-.24	-.05	.26	-.24	-.05
TIRED	.02	-.21	-.23	.02	-.21	-.23
MENTEXH	.01	-.30	-.54	.01	-.29	-.55
STRESS	.12	-.39	-.49	.11	-.39	-.48
REPET	.10	-.27	.33	.10	-.27	.33
MONBODY	.21	-.50	.38	.21	-.50	.38
UNPBODY	.44	.06	-.04	.44	.05	-.04
Variance/total variance in %	19.21	10.03	9.01	19.17	1.08	8.98

Descriptive statistics for the variables included in the PC analysis, 1981 *years sample.*

		Mean	*St.dev*	*Min.*	*Max*
	Dangerous or unhealthy work environment				
DIRTY	Work is dirty	.15	.36	.00	1.00
NOISE1	Noisy environment	.14	.35	.00	1.00
NOISE2	Noisy environment	.31	.46	.00	1.00
OUTSIDE	Work is outside	.28	.45	.00	1.00
TEMP	Exposed to non-normal temperatures	.17	.38	.00	1.00
DRAFT	Exposed to strong drafts	.23	.42	.00	1.00
SMOKE	Exposed to gas, dust or smoke	.23	.42	.00	1.00
SHAKE	Exposed to strong shakes or vibrations	.06	.24	.00	1.00
POISON	Exposed to poisons, acids or explosives	.11	.32	.00	1.00
	Physically demanding work				
LIFT	Heavy lifting	.25	.43	.00	1.00
OTHPHY	Otherwise physically demanding	.51	.50	.00	1.00
SWEAT	Work causing daily sweating	.32	.47	.00	1.00
PHYEXH	Work is physically exhausting	.26	.44	.00	1.00
TIRED	Feel very tired at end of day	.14	.35	.00	1.00
	Stressful or monotonous work				
MENTEXH	Work is mentally exhausting	.34	.47	.00	1.00
STRESS	Work is stressful	.57	.49	.00	1.00
REPET	Work is repetitive	.23	.42	.00	1.00
MONBODY	Monotonous movements	.49	.50	.00	1.00
UNPBODY	Unpleasant body positions	.47	.50	.00	1.00
	Risk indices				
RISK1	SIR, work accidents	1213.88	984.30	50.00	6000.0
RISK2	SIR, work related diseases	1105.49	999.02	50.00	6000.0
	State of health variables				
INJURED	Persistent injury	.05	.26.	.00	2.00
STRUMA	Struma	.02	.15	.00	2.00
TB	Tuberculosis	.003	.06	.00	2.00
HEARTH	Heart attack	.005	.08	.00	2.00
STOMACH	Gastric ulcer	.03	.21	.00	2.00
HEMORR	Hemorrohoids	.08	.33	.00	2.00
PREGNANT	Pregnant or pregnancy difficulty	.04	.25	.00	2.00
BROCK	Inguinal hernia	.01	.14	.00	2.00
MENTAL	Mentaly sick	.02	.16	.00	2.00
CANCER	Cancer	.01	.12	.00	2.00
DIABETIC	Diabetic	.02	.15	.00	2.00
NEURO	Neurological illnes, e.g., Polio	.01	.13	.00	2.00

Applied Economics, 1992, **24**, 751–759

Gender discrimination and compensating differentials in Sweden

MARTEN O. PALME* and ROBERT E. WRIGHT‡

**Department of Economics, Stockholm School of Economics, Box 6501, S-11383 Stockholm, Sweden* and ‡*Department of Political Economy, Adam Smith Building, University of Glasgow, Glasgow, G12 8RT, UK*

The determinants of the unexplained component of the male–female wage difference in Sweden are examined. The hypothesis that there is a compensating wage differential associated with undesirable job characteristics is tested. Hedonic wage equations are estimated using data collected in the 1981 *Swedish Level of Living Survey*. The estimation takes into consideration possible sample selection bias due to the non-randomness of employment and the potential endogeneity of job choice variables in wage offer equations. Overall, the analysis indicates that compensating wage differentials are not a major determinant of the male–female wage difference in Sweden.

I. INTRODUCTION

Over the past decade or so, there has been a renewed interest among economists in Adam Smith's theory of compensating wage differentials. The main argument is that jobs with 'undesirable' characteristics will command higher wages (other things being equal) because: 'the whole of the advantages and disadvantages of the different employments of labour and stock must, in the same neighbourhood, be either perfectly equal or continually tending toward equality' (Smith, 1937, p. 100). For example, jobs with a greater level of physical risk, in terms of the probability of death, illness or injury, are less desirable, and therefore, *ceteris paribus*, should be characterized by higher wages.[1]

The 'modern' theory of compensating wage differentials is based heavily on the theory of hedonic pricing (see Rosen (1974, 1986)). Within this framework, each job is characterized by a series of non-wage attributes. The value of each of these attributes is the compensating wage differential. These differentials are not directly observed, however, but are implicit in the wage rate of the job. Those firms which have jobs with less desirable attributes will have to offer higher wages in order to attract workers of a given quality.

It follows that one way to test for compensating wage differentials is to estimate human capital wage equations augmented with measures of job attributes. Examination of the rates of return to these attributes provides information on any compensating differentials. For example, industry and/or occupation-specific death and accident rates and/or self-assessed measures of working conditions and duties may be included in cross-sectional wage equations (see for example, Marin and Psacharopoulos (1982); Duncan and Holmlund (1983); Garen (1988); McNabb (1989) and Meng (1989)). However, as Brown (1980), Moore and Viscusi (1990) and Smith (1979) point out, such an approach has not generated unanimous support for the theory of compensating wage differentials.

Nevertheless, only limited empirical attention has been directed towards examining the role of compensating differentials in determining the wages of women. This is surprising given the potential for compensating wage differentials to further our understanding of the determinants of the male–female wage differential. More specifically, there may be differences in undesirable job characteristics between men and women and differences in the rates of return to such characteristics. Clearly such information is important, remembering that a large component of the male–female wage differential is usually not 'explained' by differences in human

[1]An earlier version of this paper was presented at the meeting of the European Society for Population Economics, Marmara University, Istanbul, Turkey, 7–9 June 1990.

capital (such as work experience and education) between men and women, and it is this component that is routinely attributed to gender-based discrimination.

With this in mind, this paper examines whether the male–female wage differential in Sweden can be explained by compensating wage differentials associated with undesirable job characteristics. Sweden provides a valuable context in which to examine this hypothesis for three reasons. The first is that Sweden has probably done more than any other country to ensure the equality of women in the labour market. This is almost certainly reflected in the male–female wage gap, which is the smallest in the industrialized world (OECD, 1988, p. 212). The second reason is that a sizeable component of the male–female wage differential is left unexplained after differences in human capital between men and women are controlled for (see for example, Gustafsson and Jacobsson (1985)). The third reason is there is some support for the claim that for men compensating wage differentials are at work in the Swedish labour market. More specifically, Duncan and Holmlund (1983) find that some undesirable job characteristics receive positive rates of return in the labour market.

II. EMPIRICAL FRAMEWORK

In order to examine the role of compensating differentials in determining male and female wages, cross-sectional wage offer equations are estimated. The methods for distinguishing between observed-wage and wage-offer functions are now well-established in the literature (see for example, Miller (1987) or Reimers (1983)). The first step is to estimate hedonic wage equations separately for men and women, using Heckman's (1979) correction for sample selection bias. That is

$$\ln W_M = \alpha_M X_M + \beta_M C_M + \phi_M \lambda_M + e_M \quad (1)$$

$$\ln W_F = \alpha_F X_F + \beta_F C_F + \phi_F \lambda_F + e_F \quad (2)$$

where the subscripts M and F denote male and female respectively. $\ln W_i$ is the natural logarithm of the hourly wage rate; X_i is a vector of human capital characteristics thought to determine earnings (e.g. education and work experience); C_i is a vector of job attributes (defined below); λ_i is the inverse of the Mills ratio estimated from a reduced-form probit equation of the probability of being observed in the wage sample (i.e. having $W_i > 0$); and e is a well-behaved error term. α, β, and ϕ are parameters to be estimated. In this framework, the α's are the rates of return to human capital and the β's are the rates of return to the job attributes.

Subtracting Equation 2 from Equation 1 and rearranging terms gives a slight variant of the standard Oaxaca (1973) and Blinder (1973) wage decomposition

$$\begin{aligned}\overline{\ln W_M} - \overline{\ln W_F} + (\hat{\phi}_F \bar{\lambda}_F - \hat{\phi}_M \bar{\lambda}_M) &= (\hat{\alpha}_M - \hat{\alpha}_F)\bar{X}_F + (\bar{X}_M - \bar{X}_F)\hat{\alpha}_M + (\hat{\beta}_M - \hat{\beta}_F)\bar{C}_F \\ &\quad + (\bar{C}_M - \bar{C}_F)\hat{\beta}_M \end{aligned} \quad (3)$$

where the hats ^ and–bars denote parameter estimates and mean values, respectively. $\overline{\ln W_M} - \overline{\ln W_F}$ is the difference between the mean wages of males and females – the observed wage gap. The term $(\hat{\phi}_M \bar{\lambda}_M - \hat{\phi}_F \bar{\lambda}_F)$ stands for the part of the observed wage gap due to the difference in average selectivity bias. Therefore, the left-hand side of Equation 3 may be interpreted as an estimate of the average wage-offer gap.

As for the remaining terms in Equation 3, $(\bar{X}_M - \bar{X}_F)\hat{\alpha}_M$ is the component of the wage-offer gap due to differences in human capital characteristics (i.e. differences in the level of human capital between men and women). The term $(\hat{\alpha}_M - \hat{\alpha}_F)\bar{X}_F$ is the component of the wage-offer gap due to differences in the estimated coefficients of the wage-offer equations (i.e. differences in the rates of return to human capital between men and women). It is this component that is usually attributed to gender discrimination. Likewise, $(\bar{C}_M - \bar{C}_F)\hat{\beta}_M$ is the component of the wage-offer gap due to differences in job attributes between men and women. Finally, $(\hat{\beta}_M - \hat{\beta}_F)\bar{C}_F$ is the component due to differences in the rate of return to these attributes between men and women.

Data

The wage offer equations are estimated using data collected in the 1981 *Swedish Level of Living Survey* (see ISF (1984)). The survey was carried out by the Swedish Institute for Social Research at the University of Stockholm. This survey is well suited to the issues raised in this paper because detailed information on wages, work experience, place of employment and socio-economic characteristics was collected for a large sample of Swedish men and women ($N = 5613$). The sample used in our analysis is restricted to individuals between the ages of 20 and 64. Furthermore, farmers, the self-employed, students and the members of the military are excluded from the sample. This resulted in a sample of 3488 individuals, of which 1652 are men and 1836 are women. Of these totals, 1543 men (or 93.4%) and 1487 women (or 81.0%) are employed. The mean hourly wage rates for men and women are 42.3 and 33.2 SK per hour, respectively ($\bar{W}_F/\bar{W}_M = 78.5\%$). This supports a logarithmic wage gap of 0.193 (i.e. $\overline{\ln W_M} - \overline{\ln W_F} = 3.695 - 3.502$). All estimates are weighted in order to reflect population values.

Variables

The human capital vector, X, consists of variables shown to be the important determinants of wages in Sweden (see, e.g. Blomquist (1979)). The list includes: years of schooling, educational qualifications, years of work experience and its square, and place of residence. Further definitions of these variables, along with their means, are given in Table 1.

The empirical framework described above requires the estimation of a probit equation of the probability that the

Table 1. *Human capital variables included in wage-offer equations, Sweden, 1981*

Variable		Women mean	Men mean
$\ln W$	Logarithm of hourly wage rate (Swedish Crowns)	3.502	3.695
Exp	Years of work experience	16.2	21.56
$(Exp)^2$	Work experience squared	382.9	649.91
School	Years of schooling	10.3	10.6
Educational qualifications (%):			
Qual1	Basic compulsory 6 or 7 years (i.e. folkskola).	28.4	28.5
Qual2	Vocational education for at least 1 year in addition to *Qual1*	10.1	15.7
Qual3	Junior high school (e.g. realexamen, grundskola, hogre folkskola, flickskola, folkhogskola)	15.0	8.6
Qual4	Vocational education for at least 1 year in addition to *Qual3*	28.7	22.8
Qual5	High school (i.e. studentexamen)	4.4	7.4
Qual6	Vocational education for at least 1 year in addition to *Qual5*	8.7	7.5
Qual7	University degree	4.7	9.5
Place of residence (%):			
Res1	Stockholm	18.9	17.1
Res2	Gothenburg	6.9	7.0
Res3	Malmo	4.5	5.7
Res4	Other cities > 30,000	22.8	18.9
Res5	Cities < 30,000	21.2	24.3
Res6	Rural areas	25.7	27.0

Source: *1981 Swedish Level of Living Survey.*

individual is employed at the time of the survey (i.e. has $W_i > 0$). The variables included in this reduced form equation are: age and its square, years of schooling, educational qualifications, place of residence, marital status, number and age of children in the household, disability status, father's and mother's social class and educational qualifications, rural background, number of siblings, non-labour income, and the local unemployment rate. From the estimates of this equation, the inverse of the Mills ratio, λ, can be calculated. That is

$$\lambda_i = PDF(-y_i)\ [1 - CDF(-y_i)]$$

where y_i is the deviation from the probit equation and *PDF* and *CDF* are the probability density and cumulative density functions, respectively (see Heckman (1979)).

The wage equation specification is a slightly modified version of Jacob Mincer's influential 'post-schooling investment' model, which has guided much of the empirical research into the determinants of earnings and wages (Mincer, 1974). In this model, earnings are primarily determined by work experience (i.e. post-schooling investment) and education (i.e. pre-labour market investment). In Mincer's model, work experience is usually measured as the number of years of employment after leaving formal education, while education is usually measured as the number of years of formal schooling completed.

We have excluded age from the wage equations for three reasons. First, Mincer (1974) has shown that an individual's cumulative work experience is a more important correlate of wages than age. Furthermore, numerous studies (see, e.g. Heckman and Polachek (1974)) have shown that adding age to a wage equation that already includes education and work experience adds little explanatory power. Second, work experience and age tend to be highly correlated (in cross-sections), so including both variables in wage equations (along with their squares) generates worrying collinearity, leading potentially to biased parameter estimates. In our data, the zero-order correlations between work experience and age for individuals with observed wages are quite high; 0.95 for men and 0.79 for women. Third, from a practical econometric point of view, including work experience in the wage equation and age in the employment equation assists in the identification of what is essentially a two-equation system of equations (i.e. wages and employment).

The simple measure of education, years of schooling completed, is less relevant for the examination of the determinants of earnings in Sweden. It has been shown that educational qualifications are a relatively more important correlate of wages than years of schooling completed (see, e.g. Blomquist (1979)). Therefore, in addition to years of schooling, we have included six binary variables measuring the highest educational qualification held by the individual (the excluded category in the regressions is 'compulsory education only', see Table 1).

The use of both years of schooling and educational qualifications may be problematic because of collinearity. In

Table 2. *Job characteristics included in wage-offer equations, Sweden, 1981*

Characteristic (C_j)	Males (1)	Females (2)	[(2)/(1)] × 100 (3)
Dangerous or unhealthy work environment			
1. *DIRTY* = Work is dirty	20.6	2.5	11.9
2. *NOISE* = Noisy environment (ear-deafening)	17.2	5.1	29.4
3. *OUTSIDE* = Work is outdoors	37.7	11.5	30.4
4. *TEMP* = Exposed to non-normal temperatures	17.0	9.7	57.4
5. *DRAFT* = Exposed to strong drafts	26.0	12.2	47.1
6. *SMOKE* = Exposed to gas, dust or smoke	23.5	11.6	49.5
7. *SHAKE* = Exposed to strong shakes or vibrations	8.7	1.0	11.9
8. *POISON* = Exposed to poisons, acids or explosives	12.8	6.0	47.4
Physically demanding work			
9. *LIFT* = Heavy lifting	28.1	11.3	40.1
10. *OTHPHY* = Otherwise physically demanding	40.5	44.8	110.4
11. *SWEAT* = Work causes daily sweating	28.9	21.4	73.9
12. *PHYEXH* = Work is physically exhausting	16.9	26.2	155.2
13. *TIRED* = Feel very tired at end of day	15.1	20.9	138.1
Stressful or monotonous work			
14. *MENTEXH* = Work is mentally exhausting	43.4	43.3	99.8
15. *STRESS* = Work is stressful	58.9	60.7	102.9
16. *REPET* = Work is repetitive	16.8	20.0	118.6
17. *MONBODY* = Monotonous body hand movements	33.7	46.2	137.2
18. *UNPBODY* = Unpleasant body positions	39.9	38.6	96.7
Time constraints			
19. *TIME* = Punctuality is important	68.9	74.6	108.3
20. *PUNCH* = Use of punch clock is required	40.5	24.3	59.9
21. *FLEX* = Inflexible working hours	59.5	65.8	110.6

Note: Variable is coded 1 if characteristic is present and 0 otherwise. Totals are percentages.

order to evaluate the collinearity generated by this specification, formal collinearity tests were carried out. Following Belsley *et al.* (1980), condition numbers and variance inflation factors (VIFs) were calculated for the variables included in the wage equations. These tests are now widely used for evaluating the 'severity' of collinearity in a multiple regression context. In brief, these indexes are calculated by comparing the relative sizes of characteristic roots and vectors of the correlation matrix of regression variables. In all the regressions reported below, the condition numbers are well below the critical values established by Belsley *et al.* (1980) and the *VIFs* are very small. The results of these tests strongly suggest that the inclusion of both years of schooling and educational qualifications is not a source of degrading collinearity.

Finally, five binary variables denoting the individual's place of residence are included in the wage equations (the excluded category is 'residing in Stockholm'). These variables are intended to proxy any aggregate wage differences across local labour markets.

The vector of job attributes, C, consists of 21 self-assessed measures of working conditions and duties. All are binary variables coded 1 for the 'less desirable' outcome – the direction hypothesized to generate a compensating differential. The variables can be broadly grouped into four categories: (1) Dangerous or unhealthy work environment (*C1–C8*); (2) Physically demanding work (*C9–C13*); (3) Stressful or monotonous work (*C14–C18*); and (4) Time constraints (*C19–C21*). The definition of each of the variables is given in Table 2.

There are numerous problems associated with including self-assessed measures of working conditions and duties in wage equations. It is likely that there is considerable measurement error inherent in such variables, which brings their overall reliability into question. Because they are by definition subjective measures, they are often open to excessive interpretation of the respondent, which raises questions about their lack of comparability across individuals. For example, as Smith (1979) points out, workers who are dissatisfied with their job, due to low rates of pay, may report higher levels of dissatisfaction generally, compared to workers with higher rates of pay. This is not only a source of measurement error, but it also introduces a spurious negative correlation between self-reported job attribute measures and wages (i.e. low wages and high dissatisfaction), and points to the potential endogeneity of such variables in wage-offer equations.

Because of the limiations of self-assessed measures, it is often argued that industry and/or occupational measures of job-related risk (e.g. fatality, injury and illness rates), calculated from independent sources, should be used in empirical studies of compensating wage differentials. In this approach,

the average level of risk is calculated for homogeneous occupation and/or industry groupings. These 'aggregate' estimates are then matched to individuals based on their (usually self-reported) occupation and/or industry. Even though these measures may appear to be more 'objective' compared to self-assessed measures, there is still considerable measurement error. The main source of measurement error is the inaccuracy in assigning the average risk value to all individuals in a given occupation and/or industry. That is, variation in risk across individuals in a given occupation or industry is assumed to be zero. Errors in matching also result when individuals misreport their occupation or industry. Finally, there is still the problem of the potential endogeneity of these risk variables in the wage equations.

It is our belief that there is no basis of choosing one type of measure over the other, since neither is 'ideal' for the empirical investigation of compensating wage differentials. Therefore, we use self-assessed measures of working conditions and duties. However, our findings must be viewed with the limitations (and strengths) of the data used to construct these variables clearly in mind.

Table 2 shows the distribution of the job attribute variables for males and females. What is immediately clear from this table is that there are considerable differences between men and women with respect to levels of these attributes and neither sex has a clear advantage (disadvantage) when they are judged as a group. For example, while 20.6% of men report that their work is 'dirty', only 2.5% of women do so. On the other hand, 46.2% of women report that their work requires monotonous hand or body movements while only 33.7% of men report the same.

As mentioned above, there is a potential problem with the specification of the wage equation. This specification assumes that the job attribute variables are exogenous. This is unlikely to be true since employment and occupational-choice decisions are theoretically endogenous. Therefore, there is no reason to assume that the sorting of individuals into jobs that differ with respect to desirable and undesirable characteristics is exogenous. If this is true, then estimates of equations 1 and 2 that do not take into consideration this potential endogeneity may generate biased estimates of any compensating differentials (see Garen, 1988).

If the variables included in the wage equations were continuous and fewer in number, one way to deal with this problem would be to estimate the equations using simultaneous equation techniques, such as two-stage least-squares (TSLS) or other instrumental variables techniques. However, the variables used here are all binary and there are a large number of them. Therefore, such an approach is not feasible. However, in order to examine this issue, equal weight was assumed for each of the 21 variables and a composite index of job attributes, C^*, was constructed simply by summing the indicators: $C^* = \sum_{j=1}^{21} C_j$. Since the number of attributes is 21, the maximum score of the index is 21 and the minimum score is 0. A higher value of C^* implies that the job the individual holds is characterized by a higher level of 'undesirability'. Wage equations are then estimated with this index in place of the 21 separate job attribute variables.

By using this composite index, we are imposing equality restrictions on the effects of each of the individual job attributes. In order to test the validity of these restrictions, we experimented with weighted indexes constructed after carrying out a factor analysis of the 21 job characteristics. Regressions using these weighted indexes confirmed the findings discussed below and therefore, for brevity, are not reported here. It is interesting to note that the average score of $\bar{C}^*$ is 5.8 for women and 6.9 for men, suggesting a higher degree of self-reported job undesirability amongst men.

In order to examine the potential endogeneity of the job attributes, the wage equations were estimated using an instrumental variable approach. An instrumental variable was created for the composite index of job attributes (denoted by $\hat{C}^*$ in Table 3), by regressing it onto the (exogenous) variables entering the employment (probit) equation (i.e. the variables used to construct the inverse of the Mills ratio). This predicted value was then entered into the wage equations in place of the composite index C^*.

Estimates

The estimated wage-offer equations are summarized in Tables 3 and 4. Table 5 reports the decomposition of the male–female wage-offer gap following Equation 3.

Column 1 of Table 5 shows the decomposition based on the wage equation that includes only the human capital variables (see Table 1). In this specification the estimated wage-offer gap is 0.187. As Table 5 shows, 0.051 or 27.3% of this gap may be attributed to differences in human capital between men and women. On the other hand, 0.136 or 72.7% is due to gender differences in the rates of return to human capital. This decomposition suggests that the wages of Swedish women would be (on average) about 14.6% higher (i.e. exp (0.136) − 1) if their rate of return to human capital were the same as for men.

Column 2 of Table 5 is the decomposition based on the equations in which all the 21 job attribute variables are included. In this specification, the estimated wage-offer gap is slightly lower, i.e. at a value of 0.182. What is clear from this decomposition is that the inclusion of these variables contributes little to the explanation of the male–female wage difference. Only 0.004 of the wage-offer gap is due to differences in job attributes between men and women. Likewise, the component of gap due to differences in the rates of return to these attributes is −0.046. This suggests that the role that gender differences in compensating differentials plays in determining differences in male and female wages is very small compared to differences in the rates of return to human capital.

Table 3. *Male–female wage offer equations, Sweden, 1981*

	(1)		(2)		(3)		(4)	
Specification No.	Males	Females	Males	Females	Males	Females	Males	Females
Variable								
Exp	0.026[a]	0.017	0.026	0.015	0.026	0.018	0.025	0.017
	[11.95]	[8.10]	[10.73]	[7.18]	[11.85]	[8.27]	[10.93]	[7.19]
$(Exp)^2/1000$	−0.367	−0.258	−0.328	−0.220	−0.367	−0.261	−0.363	−0.243
	[8.20]	[5.31]	[7.40]	[4.63]	[8.19]	[5.39]	[8.09]	[4.89]
School	0.020	0.022	0.018	0.019	0.019	0.022	0.018	0.022
	[4.24]	[4.67]	[3.84]	[4.09]	[4.19]	[4.69]	[3.85]	[4.67]
Qual2	0.031	0.039	0.013	0.047	0.031	0.040	0.030	0.037
	[1.27]	[1.57]	[0.54]	[1.93]	[1.26]	[1.62]	[1.22]	[1.47]
Qual3	0.104	0.009	0.077	0.020	0.103	0.010	0.099	0.005
	[3.58]	[0.36]	[2.70]	[0.86]	[3.55]	[0.44]	[3.59]	[0.19]
Qual4	0.127	0.026	0.100	0.035	0.125	0.031	0.113	0.015
	[4.36]	[0.86]	[3.44]	[1.17]	[4.27]	[1.03]	[3.63]	[0.47]
Qual5	0.147	0.035	0.108	0.043	0.140	0.043	0.104	0.016
	[3.79]	[0.81]	[2.78]	[1.01]	[3.56]	[1.00]	[2.21]	[0.36]
Qual6	0.151	0.103	0.110	0.116	0.144	0.112	0.109	0.083
	[3.38]	[2.41]	[2.49]	[2.76]	[3.20]	[2.61]	[2.02]	[1.85]
Qual7	0.280	0.211	0.224	0.221	0.271	0.226	0.227	0.180
	[5.08]	[3.73]	[4.08]	[3.97]	[4.88]	[3.97]	[3.38]	[2.95]
Res2	−0.014	−0.004	−0.014	−0.006	−0.013	−0.003	−0.005	−0.005
	[0.48]	[0.14]	[0.47]	[0.24]	[0.43]	[0.10]	[0.16]	[0.20]
Res3	−0.046	−0.023	−0.044	−0.040	−0.044	−0.023	−0.034	−0.022
	[1.43]	[0.70]	[1.38]	[1.25]	[1.36]	[0.71]	[1.00]	[0.67]
Res4	−0.059	−0.021	−0.051	−0.029	−0.060	−0.019	−0.062	−0.024
	[2.64]	[1.07]	[2.29]	[1.54]	[2.66]	[0.99]	[2.74]	[1.24]
Res5	−0.059	0.003	−0.055	−0.002	−0.058	0.005	−0.052	−0.001
	[2.77]	[0.13]	[2.64]	[0.07]	[2.72]	[0.23]	[2.40]	[0.07]
Res6	−0.081	−0.034	−0.076	−0.039	−0.079	−0.032	−0.070	−0.040
	[3.79]	[1.79]	[3.63]	[2.08]	[3.70]	[1.66]	[3.14]	[2.03]
C_j	*b*	—	See	Table 4	—	—	—	—
C^*	—	—	—	—	−0.003	0.005	—	—
					[1.07]	[2.51]		
$\hat{C}^*$	—	—	—	—	—	—	−0.013	−0.018
							[1.39]	[1.45]
α_0	3.130	3.096	3.226	3.130	3.152	3.056	3.263	3.184
	[63.02]	[70.19]	[59.5]	[64.54]	[58.73]	[65.27]	[30.18]	[42.43]
λ	−0.134	−0.072	−0.120	−0.084	−0.134	−0.078	−0.134	−0.057
	[2.68]	[2.27]	[2.42]	[2.73]	[2.70]	[2.49]	[2.69]	[1.75]
R^2	0.26	0.20	0.30	0.25	0.26	0.20	0.26	0.20

Notes:
[a]Absolute value of t-statistics given in parentheses.
[b]Variable(s) not included in specification.

Column 3 of Table 5 is the decomposition based on the equations where the composite index of job attributes is included in place of the 21 job attribute variables. It is interesting to note that despite the crudeness of this index, the decomposition is very similar to what was found when the variables were entered separately in the equations (column 2). In the wage equations, the parameter on the job attribute index is 0.005 ($t = 2.5$) in the female wage equation and is −0.002 ($t = 1.1$) in the male equation. Taken at face value, this indicates that there are compensating wage differentials associated with undesirable job attributes for women, as measured by the composite index. However, the opposite is the case for men, where such differences have little effect or might even support lower wages. This suggests that compensating wage differentials may be working in a way that reduces the component of the male–female wage differentials usually attributed to gender-based discrimination.

Unfortunately, this finding is not confirmed when the wage equations are estimated using the instrumental variables approach described above. The parameter on the instrumented job attribute index, $\hat{C}^*$, is negative for both

Table 4. *Job characteristics parameters*

Job characteristics (C_i)	Specification No. 2 Cont. Males	Females
Dangerous or unhealthy work environment		
1. *DIRTY*	0.004[a, b]	0.045
	[0.19]	[1.10]
2. *NOISE*	0.081	0.026
	[3.63]	[0.85]
3. *OUTSIDE*	−0.015	−0.142
	[0.92]	[6.90]
4. *TEMP*	0.005	−0.023
	[0.24]	[1.03]
5. *DRAFT*	−0.001	0.013
	[0.03]	[0.65]
6. *SMOKE*	−0.021	0.003
	[1.14]	[0.17]
7. *SHAKE*	−0.027	0.080
	[1.02]	[1.28]
8. *POISON*	0.021	0.020
	[1.00]	[0.79]
Physically demanding work		
9. *LIFT*	−0.052	0.005
	[3.07]	[0.24]
10. *OTHPHY*	−0.016	−0.009
	[0.99]	[0.58]
11. *SWEAT*	0.012	0.046
	[0.67]	[2.63]
12. *PHYEXH*	−0.017	−0.005
	[0.89]	[0.32]
13. *TIRED*	0.005	0.034
	[0.26]	[1.10]
Stressful or monotonous work		
14. *MENTEXH*	0.031	0.029
	[2.04]	[2.06]
15. *STRESS*	0.061	−0.027
	[4.26]	[2.01]
16. *REPET*	−0.024	−0.029
	[1.21]	[1.78]
17. *MONBODY*	−0.022	−0.013
	[1.39]	[0.90]
18. *UNPBODY*	−0.007	−0.010
	[0.41]	[0.68]
Time constraints		
19. *TIME*	−0.022	0.046
	[1.43]	[2.97]
20. *CLOCK*	−0.044	0.021
	[2.96]	[1.38]
21. *INFLEX*	−0.050	−0.012
	[3.32]	[0.82]

Notes:
[a]Absolute value of t-statistics given in parentheses.
[b]Other parameters of the wage-offer equations given in column 2 of Table 3.

men and women and is not statistically significant at the 10% level (see column 4 in Table 3). In the male equation, this parameter is −0.013 ($t = 1.4$) and −0.012 ($t = 1.5$) in the female equation. Furthermore, as column 4 of Table 5 shows, the component of the male–female wage gap that may be attributed to the role of compensating differentials is very small.

In order to further explore the possible endogeneity of the job attribute variables, Hausman–Wu tests were carried out (Wu, 1973; Hausman, 1989). These tests are based on the theory of instrumental variables. In short, an instrumental variable (IV) is created for the potentially endogenous variable. This IV is then entered into the regression equation along with the variable itself. Conventional goodness-of-fit tests are then applied. If the variable is exogenous, then the inclusion of the IV does not significantly improve the overall fit of the equation. In our application, an IV was created for the job attribute index using as instruments the (exogenous) variables entered into the employment equation (i.e. the variables used to construct the inverse of the Mills ratio). This IV, along with the job attribute index, was then entered into the wage equations.

For both males and females, an F-test revealed that the inclusion of IV did not significantly improve the fit of the wage-offer equations. The calculated (test) F-statistic was 2.38 for males and 2.88 for females. Both values are well below the 5% critical (table) value of 3.8. This finding suggests that job attributes, as measured by a composite index of binary job attributes, are exogenous. However, as Nakamura and Nakamura (1985) show, the power of the Hausman–Wu test is directly related to the explanatory power of the auxiliary regressions used to construct the IVs. In our tests, the R^2-values of these regressions were quite low – 22.3% for males and 13.9% for females. These low R^2-values suggest that the results of these exogeneity tests are clearly tentative and should be viewed with much caution.

III. CONCLUSION

On average, female workers in Sweden earn less per hour than male workers. Even though this gap is smaller than what is found in other industralized countries, it is still substantial. Furthermore, much of this gap remains after differences in the level of human capital between men and women are taken into account (e.g. differences in education and work experience). It is this remainder or 'unexplained' difference that is usually attributed to gender-based discrimination. The analysis carried out in this paper indicates that this remaining unexplained difference cannot be 'explained' by compensating wage differentials associated with undesirable job attributes. Therefore, we provide little evidence that is contrary to the view that a large part of the male–female wage gap in Sweden is due to discrimination.

Our analysis also lends little support to the hypothesis that there are compensating wage differentials, as predicted by economic theory, associated with undesirable job attributes in Sweden. That is, male and female wage offers appear not to be affected by the variance across individuals in such factors. There are many possible reasons for this

Table 5. *Male–female wage–offer decomposition, Sweden, 1981*

Component	(1)	(2)	(3)	(4)
Observed wage gap[a]	0.193	0.193	0.193	0.193
Wage offer gap[b]	0.187	0.182	0.185	0.190
Component due to				
Difference in human capital[c]	0.051	0.045	0.050	0.045
Differences in rates of return[d] to human capital	0.136	0.178	0.181	0.166
Differences in job attributes[e]	—	0.004	−0.002	−0.007
Differences in rates of return to job attributes[f]	—	−0.046	−0.044	−0.014

Notes: Expressions used in the calculations are listed below.

[a]$\overline{\ln \bar{W}}_M - \overline{\ln \bar{W}}_F$, [b]$\overline{\ln \bar{W}}_M - \overline{\ln \bar{W}}_F + (\hat{\phi}_F \bar{\lambda}_F - \hat{\phi}_M \bar{\lambda}_M)$, [c]$(\bar{X}_M - \bar{X}_F)\hat{\alpha}_M$, [d]$(\hat{\alpha}_M - \hat{\alpha}_F)\bar{X}_F$, [e]$(\bar{C}_M - \bar{C}_F)\hat{\beta}_M$, and [f]$(\hat{\beta}_M - \hat{\beta}_F)\bar{C}_F$

finding. One relates to the general problem of using self-assessed measures of working conditions and duties in hedonic wage equations. For reasons discussed above, it is often argued that industry and/or occupational measures of job-related risk, calculated from independent sources, are superior variables for the empirical detection of compensating wage differentials. In a companion paper (Palme and Wright, 1991), we matched occupation–sex specific measures of job-related physical risk (i.e. standardized fatality, injury and illness rates), calculated from the *Swedish Information on Occupational Injuries Database*, to the wage and socio-economic data used above (for information about this data base, see Broberg and Lagerlof (1984)). In this analysis, the estimated wage-offer equations generated no support for any compensating wage differentials associated with job-related risk. Furthermore, the analysis reconfirmed that compensating wage differentials are not a major determinant of the male–female wage difference.

Another possible reason for the apparent non-existence of compensating wage differentials is simply the conclusion that such mechanisms do not operate in the Swedish labour market. It is clear that Swedish labour market policy, with its heavy emphasis on equality, has led to reduction in the rates of return on human capital in general and to the compression of the (pre- and post-tax) distribution of earnings. For example, the rates of return to education and work experience tend to be lower in Sweden compared to other industrialized nations. If this is true, then the pursuit of equality may be the factor which has contributed to the negation of compensating wage differentials of the type examined in this paper.

In order to empirically explore this line of reasoning, one would need longitudinal or panel data in order to calculate and compare the size of compensating wage differentials over time. Fortunately, in Sweden, data sets of this type are available. Unfortunately, the required data-matching and analysis is beyond the scope of this paper. However, it is an important topic for future research and will undoubtedly further our understanding of the operation (or non-operation) of compensating wage differentials in the Swedish labour market. Furthermore, if it proves to be the case that compensating wage differentials are not a feature of the Swedish labour market, then the non-confirmatory findings of this paper are not a product of measurement and/or methodological limitations, but represent a snap-shot of an historical process.

ACKNOWLEDGEMENTS

The comments of John Ermisch, Mia Horn, James Heckman, Heather Joshi, the editor and the anonymous referee are gratefully acknowledged. Thanks are also due to the Institute for Social Research at the University of Stockholm for making available the *Swedish Level of Living Survey*. Of course, all errors and shortcomings are the sole responsibility of the authors.

REFERENCES

Belsley, D. A., Kuh, E. and Welsch, R. E. (1980) *Regression Diagnostics*, Wiley, New York.

Blinder, A. S. (1973) Wage discrimination: reduced form and structural estimates, *Journal of Human Resources*, **8**, 436–55.

Blomquist, N. S. (1979) Wage rates and personal characteristics, *Scandinavian Journal of Economics*, **81**, 505–20.

Broberg, E. and Lagerlof, E. (1984) *A Short Description of the Swedish Information System on Occupational Injuries*, National Board of Occupational Safety and Health, Stockholm.

Brown, C. (1980) Equalizing differences in the labor market, *Quarterly Journal of Economics*, **94**, 113–34.

Duncan, G. J. and Holmlund, B. (1983) Was Adam Smith right after all? Another test of the theory of compensating wage differentials, *Journal of Labor Economics*, **1**, 366–79.

Garen, J. (1988) Compensating wage differentials and the endogeneity of job riskiness, *Review of Economics and Statistics*, **70**, 9–16.

Gustafsson, S. and Jacobsson, R. (1985) Trends in female labor force participation in Sweden, *Journal of Labor Economics*, **3**, (suppl.), S256–S74.

Hausman, J. A. (1978) Specification tests in econometrics, *Econometrica*, **46**, 1251–72.

Heckman, J. J. (1979) Sample selection bias as a specification error, *Econometrica*, **47**, 153–61.

Heckman, J. J. and Polachek, S. (1974) Empirical evidence on the functional form of the earnings–schooling relationship, *Journal of the American Statistical Association*, **69**, 350–4.

ISF (1984) *Variables and Codebook for 1981 Level of Living Survey* (in Swedish). Institute for Social Research, University of Stockholm, Stockholm.

Marin, A. and Psacharopoulos, G. (1982) The reward for risk in the labor market: evidence from the United Kingdom and a reconciliation with other studies, *Journal of Political Economy*, **90**, 827–53.

McNabb, R. (1989) Compensating wage differentials: some evidence from Britain, *Oxford Economic Papers*, **41**, 327–38.

Meng, R. (1989) Compensating differences in the Canadian labour market, *Canadian Journal of Economics*, **22**, 413–24.

Miller, P.W. (1987) Gender differences in observed and offered wages in Canada, *Canadian Journal of Economics*, **20**, 225–44.

Mincer, J. (1974) *Schooling, Experience and Earnings*, National Bureau of Economic Research, New York.

Nakamura, A. and Nakamura, M. (1985) On the relationships among several specification error tests presented by Durbin, Wu and Hausman, *Econometrica*, **49**, 322–47.

Moore, M. J. and Viscusi, W. K. (1990) *Compensation Mechanisms for Job Risks*, Princeton University Press, Princeton.

Oaxaca, R. (1973) Male–female wage differentials in urban labor markets, *International Economic Review*, **14**, 693–709.

OECD (1988) *Employment Outlook*, Organisation of Economic Co-operation and Development, Paris.

Palme, M. O. and Wright, R. E. (1991) Risky Jobs and Wages (in Swedish), unpublished manuscript, Department of Economics, School of Economics, Stockholm.

Reimers, C. W. (1983) Labour market discrimination against Hispanic and Black men, *Review of Economics and Statistics*, **65**, 570–9.

Rosen, S. (1974) Hedonic prices and implicit markets: product differentiation in pure competition, *Journal of Political Economy*, **82**, 34–55.

Rosen, S. (1986) The theory of equalizing differences, in *Handbook of Labor Economics* (Vol. 1). Ashenfelter, O and Layard, R. (eds) Elsevier, Amsterdam.

Smith, R. S. (1979) Compensating wage differentials and public policy: a review, *Industrial and Labor Relations Review*, **32**, 339–52.

Smith, A. (1937) *The Wealth of Nations*, Modern Library, New York.

Wu, D-M. (1973) Alternative tests of independence between stochastic regressors and disturbances, *Econometrica*, **41**, 733–50.

Changes in the Rate of Return to Education in Sweden: 1968-1991*

by

Mårten Palme[$] and Robert E. Wright[£]

Abstract: This paper estimates and documents changes in the rate of return to education in Sweden between 1968 and 1991. Both the "quantity" (i.e. years of schooling completed) and "quality" (i.e. education level obtained) dimensions of education are considered. These returns are calculated by estimating human capital wage equations. The data used are from the 1968, 1981 and 1991 Swedish Level of Living Surveys. There are three main conclusions. The first is that the rate of return to education declined considerably between 1968 and 1981. However, since 1981 there has been little change. The second is that the magnitude of the decline was not the same for all levels of education, with the decline being particularly pronounced for university education. The third is that the decline was not equally shared between men and women, with the decline being larger for women.

---

* We wish to thank Anders Björklund, Magnus Henrekson, Bertil Holmlund, Erik Ruist and seminar partisipants at the 1992 ESPE meeting in Gmunden, Austria for comments on earlier drafts of this paper. Of course, the responsibility for remaining errors and shortcomings is ours. Mårten Palme was partially financed through the research project "Income distribution in Sweden" which is supported by The Bank of Sweden Tercentenary Foundation and the Swedish Council for Research in the Humanities and Social Sciences.

$ Department of Economic Statistics, Stockholm School of Economics, Box 6501, S-113 83 Stockholm, Sweden.

£ Department of Political Economy, Adam Smith Building, University of Glasgow, Glasgow, Scotland, G12 8RT and Centre for Economic Policy Research 25-28 Old Burlington Street , London, England, W1X 1LB.

1. Introduction

There seems to be little disagreement that an educated labour force is one of the main prerequisites for sustained economic growth (Phacharopoulos, 1973; Psacharopoulos and Woodhall, 1985). The arguments linking education and economic growth are often based on "human capital" theory (Becker, 1975). "Human capital" is the productive skills and knowledge embodied in individuals, which are enhanced through such factors as work experience, training and education. Based on this perspective, "investing" in human capital is thought to generate high private and social benefits. For example, from a societal point of view, education is assumed to contribute directly to economic growth through increased labour productivity. Likewise, from in individual point of view, education is assumed to yield a positive benefit through, for example, higher earnings. However, if the process of investing in human capital is viewed in the same theoretical light as investing in physical capital, then in countries that have already achieved high levels of income and have highly educated labour forces, the social and private "rates of return" to education should be lower (Psacharopoulos and Woodhall, 1985).

Sweden provides a valuable context in which to examine issues relating to the rate of return to education. Both economic growth and the expansion of the education system have been rapid. However, policies pursued by the trade unions and government have likely had a considerable impact on the education investment decisions of individuals. More specifically, throughout the 1960s and 1970s, the highly centralised trade union movement campaigned vigorously for "solidarity wage" policies, which were aimed at levelling wage rates across occupations among blue collar workers.[1] The success of these policies led (in part) to a significant compression of the pre-tax distribution of earnings (see Hibbs 1990 a, b). In addition, throughout this period, the Swedish government pursued a policy of increasingly progressive income taxation. This, in turn, led to a compression of the post-tax distribution of earnings. The combined policies of the trade unions and the government, with their heavy emphasis on generating earnings equality, likely had a substantial impact on the private rate of return to education.

In the 1980s, however, these trends toward pre- and post tax equality established in the 1960s and 1970s, began to reverse themselves. More specifically, there was a breakdown in solidarity wage policies and central wage negotiating. Since 1983, union negotiations have occurred primarily at the industry level. This has increased the variation in wage settlements reached and has led to a widening of the pre-tax distribution of earnings (see Edin and Holmlund, 1992). In addition, in a series of tax-reforms, the government has reduced the

[1]For a current discussion on Swedish labour market institutions see Nilsson (1992).

progressive nature of the Swedish tax system. This has likely led to a widening of the post-tax distribution of earnings. Therefore, in the 1980s, the combined policies of the trade union and the government, with little explicit emphasis on generating earnings equality, has likely increased (or at least slowed down the decrease) in the private rate of return to education.

The purpose of this paper is to estimate changes in the rate of return to education, measured as mean hourly wage differentials between individuals with different levels of educational attainment. Thus, several aspects likely to affect the individual decision to invest in formal schooling, e.g. income taxes, expectations of future tax-systems, provision of student loans and employment security, is left out of the analysis.[2] The data used are from the 1968, 1981 and 1991 *Swedish Level of Living Surveys* (see Erikson and Åberg, 1987). We will concentrate on three issues concerning changes in the returns to education: (1) Male-female differentials in returns to education; (2) Differences in returns to "quantity" (i.e. years of schooling completed); and (3) Differences in the returns to "quality" (i.e. education level obtained). Furthermore, two methodological issues will be investigated: (1) To what extent does correcting for sample selection bias (using Heckman's 1979 method) affect the estimates?; and (2) To what extent does the adoption of the restrictive Mincerian quadratic earnings function affect the estimates compared to a more flexible functional form (a cubic spline function) ?

Five main conclusions emerge from the analysis. (1) The rate of return to education declined considerably between 1968 and 1981. However, since 1981 there has been little change. (2) The magnitude of the decline was not the same across all levels of education. The decline was particularly pronounced for university education. (3) The decline was not equally shared between men and women. The decline was larger for women. (4) The correction for sample selection bias did not change the results considerably. The only possible exception is the results for females 1968. (5) The quadratic wage function and the cubic splines, give approximately the same results, except when there was an interaction effect of formal schooling and work experience on wages. This was true for the results from the 1968 sample for both males and females.

The paper proceeds as follows. Section 2 describes the data sources. Section 3 summarises changes in the educational composition of the Swedish labour force. Section 4 outlines the methods used to estimate the rate of return to education. The estimated returns are discussed in Section 5. Conclusions follow in Section 6.

---

[2] For recent empirical studies on some of these issues in Sweden, see Edin and Holmlund (1992) and Björklund and Kjellström (1993).

2. Data

The data are drawn from the 1968, 1981 and 1991 *Swedish Level of Living Surveys*. These surveys were carried out at the Swedish Institute for Social Research, Stockholm University. They are well suited to the issues raised in this paper because detailed information about wages, educational attainment and other socio-economic characteristics was collected for large samples of Swedish men and women using the same sampling method for each survey. The samples used in our analysis are restricted to individuals between the ages of 20 and 64. Furthermore, farmers, the self-employed, students and the members of the military are excluded. We shall refer to this population as the "labour force". All estimates are weighted in order to reflect population totals. A breakdown of the samples used is shown in Table 1.

Table 1. *Unweighted sample sizes used in the analysis.*

	Males			*Females*		
Year	N	N:W>0	N:W=0	N	N:W>0	N:W=0
1968	1,675	1,616	59	1,808	1,023	785
1981	1,520	1,543	109	1,836	1,487	349
1991	1,516	1,443	73	1,623	1,499	124

Note: W = wage.

We consider two measures of education. The first is simply the number of years of schooling completed. The second is the education level obtained. In our analysis, years of schooling is intended to capture the "quantity" dimension of education. Education level is intended to capture the "quality" dimension of education. For convenience, we shall refer to the first measure as "schooling" and to the second measure as "education level"

Education level is divided into four categories. These categories represent distinct divisions in the Swedish educational system. More specifically they are:

S1 = Basic compulsory (e.g. folkskola, junior high school, realexamen, grundskola, högre folkskola, flickskola, folkhögskola);

S2 = Vocational education for at least 1 year in addition to basic compulsory and/or vocational education for at least 1 year in addition to education level S1;

S3 = High school (e.g. studentexamen) and/or vocational education for at least 1 year in addition to high school; and

S4 = University education.

3. Changes in the Educational Composition of the Swedish Labour Force

Tables 2 and 3 document the changing profile of educational attainment of the Swedish labour force between 1968 and 1991. As Table 2 shows, the mean number of years of schooling increased for both men and women between 1968 and 1991. For men, the increase was from 8.7 years in 1968 to 11.6 years in 1991. For women, the absolute increase was slightly bigger, from 8.3 years in 1968 to 11.4 years in 1991. Two further points about this table are worth noting. The first is that in 1968, 1981 and 1991, the mean number of years of schooling is higher for men than for women. The second is that the gender gap in schooling did not decrease much between 1968 and 1991. If we measure the "gender gap" as the ratio of female years of schooling to male years of schooling (i.e. S_F / S_M), this ratio increased only slightly from 0.95 in 1968 to 0.98 in 1991. Since a value of "1.0" represents the situation of "gender equality", equality in years of schooling has not yet been achieved in Sweden. This is despite the fact that the growth rate in schooling was more rapid for women compared to men (i.e. 37.3% versus 33.3%, see Table 1).

Table 2. *Sample mean years of schooling completed.*

Year	*Males* (1)	*Females* (2)	*Ratio* (2)/(1)
1968	8.7	8.3	0.95
1981	10.5	9.9	0.94
1991	11.6	11.4	0.98
Δ	2.9	3.1	0.03
Δ%	33.3	37.3	3.20

Table 3 shows the distribution of the labour force by education level in 1968, 1981 and 1991. As a general remark, these estimates show that the educational attainment (in terms of qualifications obtained) of the Swedish labour force increased considerably in this period, despite the modest increase in the mean number of years of schooling completed (see Table 2). A major change was the reduction in the proportion of individuals who have only basic compulsory education. For men, this proportion decreased from 65.7% in 1968 to 27.3% in 1991. For women, the decline was from 72.6% in 1968 to 31.7% in 1991. These sizable declines are reflected in significant increases in the other three education levels. Most notably, the increase in university education was quite marked. In 1968, only 3.9% of men and 1.0% of women had university qualifications. By 1991, these proportions had increased to 11.0%

and 7.6%, respectively.

Table 3. *Highest level of education obtained. Per cent.*

Year	*Males* (1)	*Females* (2)	*Ratio* (2)/(1)
S1: Basic Compulsory			
1968	65.7	72.6	1.11
1981	38.9	49.4	1.27
1991	27.3	31.7	1.16
Δ	-38.4	-40.9	0.05
%Δ	-58.5	-56.3	4.5
S2:Vocational			
1968	25.9	23.1	0.89
1981	37.8	34.9	0.92
1991	42.9	42.0	0.98
Δ	17.0	18.9	0.09
%Δ	65.6	81.8	10.1
S3: Gymnasium/other			
1968	4.5	3.3	0.73
1981	14.3	11.7	0.82
1991	18.8	18.7	0.99
Δ	14.3	15.4	0.26
%Δ	317.8	466.7	35.6
S4: University			
1968	3.9	1.0	0.26
1981	9.0	4.0	0.44
1991	11.0	7.6	0.69
Δ	7.1	6.6	0.43
%Δ	182.1	660.0	165.4

From Table 3 it is clear that the increase in educational attainment was not shared equally between men and women. In both 1968 and 1991, women had lower levels of educational attainment in terms of formal qualifications obtained. However, one may argue that there was a significant decrease in the gender gap between 1968 to 1991. More specifically, the proportion of men with education beyond basic compulsory (i.e. S2+S3+S4) increased from 34.3% in 1968 to 72.7% in 1991. The analogous proportions for women are 27.4% in 1968 and 68.3% in 1991. Therefore, in 1968 the gender gap in "higher education" was 0.80 (i.e. 27.4%/34.3%). In 1991, this ratio had increased to 0.94 (i.e. 68.3%/72.7%). Nevertheless, despite this apparent progress, it important to stress that in 1991 there was still a very large

difference in the gender composition of individuals with university education. The proportion of men in 1991 with university education is still about one and a half times that of women (i.e. 11.0% for men compared to 7.6% for women).

4. Estimation

Our empirical focus is to estimate the wage gain related to education. There is considerable debate surrounding what is the most meaningful way to measure education when estimating the rate of return to education. Should one use years of schooling completed or education level obtained ? In most empirical studies, years of schooling is used (see Lorenz and Wagner, 1990). The popularity of this specification stems directly from Jacob Mincer's influential theoretical work into the determinants of earnings. However, in some countries, it has been shown that qualifications obtained are a more important statistical correlate of earnings than years of schooling completed (see Blinder, 1976).

As Blomquist (1979) points out, a useful way to conceptualise the relationship between schooling and education level is to view years of schooling as the "input" to the education process and education level obtained as the "output". However, as mentioned above, years of schooling and education level can be thought to represent the quantity and quality dimensions of education. Given a certain education qualification, the number of years an individual requires to obtain this qualification, tells us something about his/her productivity. For example, an individual who obtains a basic university degree in a shorter period of time is likely more productive than an individual who takes longer to obtain the same degree. On the other hand, in the four education levels defined above, there is a certain amount of "lumping together" of qualifications that take different amounts of time to obtain. For example, the university education category (S4) consists of all levels of university education, ranging from the basic undergraduate degree to a doctorate. Clearly, years of schooling will differ depending on what type (and perhaps subject) of degree is obtained. In this sense, therefore, years of schooling provides complimentary information about the extent of investment in education.

Since years of schooling and education level potentially measure different aspects of education, we use both in our empirical investigation. Thus, three "concepts" of the wage differential associated with education are considered: (1) The return to an additional year of formal schooling; (2) The return associated with possesing a given level of education; and (3) The return to an additional year of schooling within a given level of education.

Estimating the wage differentials related to educational differences raises two methodological issues. First, like most other studies, we have relatively small sample sizes when the data is broken down by gender and formal schooling attainment. For example, our sample sizes are

between 1,000 and 1,600 individuals. This means that if we consider formal schooling between 10 and 20 years, the average sample size will be between 100 and 160 for each group, which is too small to make inference from since wages are related to work experience and therefore very sensitive to sampling errors of individuals with different work experience. Second, as we can only observe wages for individuals that are employed, we have a potential problem of sample selection bias.

The most common method to handle the first problem (see Willis, 1986), is to control for differences in work experience by fitting a quadratic earnings function, i.e.

$$\ln W = \varphi + E\alpha + X\beta_1 + X^2\beta_2 + \varepsilon, \qquad (1)$$

where $\ln W$ is the natural logarithm of the (observed) hourly wage rate; E is a vector of educational attainment characteristics; X is work experience; ε is a well behaved error term; and φ, α and β are parameters to be estimated. In this framework, the estimated α's provide information that can be used to calculate the wage differentials between individuals with different amounts of education.

This quadratic wage function has been recently criticised for being overly restrictive (see Willis, 1986, footnote 2 or Murphy and Welch (1990). Murphy and Welch (1990) examines the bias of the quadratic approximation and make the following recomendation for empirical research. "While it seems clear that the quadratic must be scrapped for purposes of estimating career earning patterns, it is unclear whether the quadratic can still be used to effectively "control" for life-cycle wage effects when other factors affecting wages are of primary interest. On these matters we can provide no clear answers, only some words of caution." (p. 228).

One way to address this problem is to use a more flexible functional form such as a polynomial spline.[3] A polynomial spline of degree n is defined as a piecewise polynomial function made up of polynomials of degree at most n, such that both the spline and its derivatives up to and including (n-1) are continuous in all points. In statistical terms it may be described as a polynomial of a certain degree, which is estimated separately for different segments under the restriction that it should be continuous in the points defining the segments, denoted "knots" in the spline function literature. Thus, in the univariate case, the researcher has to decide the degree of the polynomial, number and location of the "knots". It is also possible to estimate splines using more than one independent variable. In this case, apart from estimating piecewise polynomial functions for each independent variable, it is

[3]Splines have been used in a number of studies on wages. First probably by Piorier and Watts (1973) and recently by Murphy and Welch (1992). We will only give a brief description of splines. Piorier (1976) gives a broad description. Piorier (1987) gives a short description and suggestions for further readings.

possible to consider interactions between these variables.

In this study we will use polynomials of the third degree, i.e. cubic splines. We also estimate splines with two independent variables, number of years of work experience, years of schooling and the the interaction between the two. We will use the estimation methods outlined by Suits et al. (1978).

The second problem, results from estimating behavioural equations with "self-selected" samples. In this case, the individuals are self-selected in terms of employment. If there is significant sample selection bias then the estimated rate of return to education derived from Eq. (1) will be biased. This may be more of a problem for women compared to men, since the proportion of women employed is generally lower (see Table 1). However, Schultz (1990) finds large differences in the rates of return to schooling for both men and women in the United States depending on whether or not a correction for sample selection bias is carried out.

Various methods have been developed for correcting for sample selection bias of this type. The method used here is Heckman's two-step method. In the first step, a probit regression equation of the probability that the individual is employed is estimated (i.e. on the probability of having an observed wage). The estimates of this equation are used to construct a variable called the "inverse of the Mill's ratio",

$$\lambda_i = \phi(z_i) / [1 - \Phi(-z_i)] \qquad (2)$$

where z_i is the deviate from the probit equation and ϕ and Φ are the probability density and cumulative density functions, respectively. In the second step, this variable is entered into the wage equation (1).

The correction for sample selection bias described above requires the estimation of a probit equation of the probability that the individual is employed at the time of the survey (i.e. has W>0). The variables (assumed exogenous) included in this equation are: age and its square; place of residence; marital status; number and age of children in the household; disability status; father's and mother's social class and educational qualifications; rural background; number of siblings; non-labour income; and the local unemployment rate.

The mean hourly wage rates (in constant 1991 Swedish Crowns) for men and women are given in Table 4, along with the female-male wage ratio (i.e. W_F / W_M). Across all education levels, this ratio increased between 1968 and 1991, from 0.73 to 0.81, indicating an increase in gender equality of earnings. What is interesting to note, however, is that the percentage decline in the gender wage gap was not the same across the four levels of education. More

specifically for basic compulsory education (S1) the decline was 12.5%; for vocational education (S2) it was 9.8%; for the gymnasium/other education (S3) it is 7.5%; and for university education (S4) it was a much lower 6.0% (see Table 4).

Table 4. Average *Male-Female Wage Rates. Sweden 1968-1991. Real 1991 Swedish Crowns.*

Year	*Males (1)*	*Females (2)*	*Ratio (2)/(1)*
S1: Basic Compulsory			
1968	66.2	49.3	0.75
1981	78.6	65.1	0.83
1991	79.9	67.0	0.84
Δ	13.7	17.7	0.09
%Δ	20.7	35.8	12.0
S2: Vocational			
1968	80.6	60.8	0.75
1981	83.7	69.5	0.83
1991	85.1	70.4	0.83
Δ	4.5	9.6	0.08
%Δ	5.6	15.9	10.7
S3: Gymnasium/other			
1968	115.5	84.8	0.73
1981	90.7	74.4	0.82
1991	97.3	76.8	0.79
Δ	-18.2	-8.0	0.06
%Δ	-15.8	-9.5	8.2
S4: University			
1968	163.6	125.7	0.77
1981	111.2	95.6	0.86
1991	115.7	94.2	0.81
Δ	-47.9	-31.5	0.04
%Δ	-29.3	-25.1	5.2
All			
1968	75.8	55.3	0.73
1981	85.5	69.5	0.81
1991	89.5	72.5	0.81
Δ	13.7	17.2	0.08
%Δ	18.1	31.0	11.0

Notes: Adjusted using the *Swedish Consumer Price Index.*

4. Results

Results from Quadratic Wage Functions

The estimates of the average wage differentials due to formal schooling are shown in Table 5. These returns were estimated by fitting wage equations (1) with the only education variable being years of schooling completed. Therefore, the rate of return to schooling is simply the parameter associated with the years of schooling variable, multiplied by 100 in order to express it as a percentage. Turning first to estimates uncorrected for sample selection bias, the rate of return to schooling for both men and women declined considerably between 1968 and 1991. In 1968 for men, each additional year of schooling completed was associated with a 8.2% increase in wages. By 1991, the return had decreased to 4.0%. Likewise, for women in 1968, an additional year of schooling completed was associated with a 7.4% increase in wages. By 1991, this had fallen to 3.9%.

Table 5. *Average percentage wage differentials due to one additional year of formal education. Results from quadratic earnings functions. Sweden 1968-1991. (Standard errors in parentheses).*

Year	*Uncorrected for Selection Bias*			*Corrected for Selection Bias*		
	Males (1)	*Females* (2)	*Ratio* (2)/(1)	*Males* (3)	Females (4)	Ratio (4)/(3)
1968	8.2	7.4	0.90	8.0	7.7	0.96
	(0.29)	(0.42)		(0.33)	(0.43)	
1981	4.2	3.5	0.83	4.0	3.3	0.82
	(0.22)	(0.22)		(0.23)	(0.23)	
1991	4.0	3.5	0.87	3.8	3.4	0.8[illegible]
	(0.23)	(0.19)		(0.24)	(0.20)	
Δ	-4.2	-3.9	-0.03	-4.2	-4.3	-0.07
%Δ	-51.2	-47.3	-3.3	-52.5	-55.8	-7.3

The estimates corrected for sample selection bias also suggest that the rate of return to schooling is slightly lower for women compared to men. In 1968, the rate of return to schooling was 8.0% for men and 7.7% for women - a difference of about half a percentage point. Likewise, in 1991, the male rate of return was 3.8% compared to 3.4% for women - also a difference of half a percentage point. Overall the estimates suggest that there is little difference in the rate of return to schooling between Swedish men and women.

As a general remark, the estimates do not change much after the correction for sample selection bias is carried out. These "corrected" estimates confirm that the rate of return to

schooling for both men and women declined between 1968 and 1991. Furthermore, they suggest that there is little difference between men and women. Nevertheless, in terms of percentage declines, the decline in return to schooling has been "larger" for women. As Table 5 shows, the corrected estimates suggest that the decline was 52.5% for men and 55.8% for women.

The average wage differentials associated with each of the four education levels are presented in Table 6. These were calculated by first fitting wage equations (1) with three binary variables representing the four different education levels. The excluded (reference) category is "compulsory education only". (Years of schooling completed is not included in the equations). The difference associated with each education level is calculated by the following method,

$$\%\Delta W_j = \left[\exp(\alpha_j) - 1\right] * 100 \qquad (3)$$

where α_j is the parameter associated with each of the three education level variables (i.e. S_j where j=2,3,4). The returns are the percentage increase (or decrease) in wages associated with each education level relative to compulsory education only.

Table 6. *Average percentage wage differentials associated with highest level of education obtained. Sweden 1968-1991. (Standard errors in parantheses).*

Year	*Uncorrected for Selection Bias*			*Corrected for Selection Bias*		
	Males (1)	*Females* (2)	*Ratio* (2)/(1)	*Males* (3)	Females (4)	Ratio (4)/(3)
S2: Vocational						
1968	22.6	21.6	0.96	21.4	22.1	1.03
	(2.40)	(3.29)		(2.93)	(3.40)	
1981	12.4	11.4	0.92	11.3	8.9	0.79
	(1.86)	(1.61)		(1.89)	(2.42)	
1991	14.4	10.1	0.70	13.4	9.4	0.70
	(2.01)	(1.45)		(2.06)	(1.50)	
Δ	-8.2	-11.5	-0.26	-8.0	-12.7	-0.33
%Δ	-36.2	-53.2	-27.1	-37.4	-57.5	-32.0
S3: Gymnasium/other						
1968	83.2	81.0	0.97	86.4	81.9	0.95
	(7.62)	(10.64)		(9.84)	(10.73)	
1981	25.7	23.6	0.92	23.7	20.7	0.87
	(2.87)	(2.59)		(2.93)	(3.94)	
1991	33.5	21.7	0.65	31.3	20.9	0.67
	(2.94)	(2.00)		(3.03)	(2.05)	
Δ	-49.7	-59.3	-0.32	-55.1	-61.0	-0.30
%Δ	-59.7	-73.2	-33.0	-63.8	-74.5	-31.6
S4: University						
1968	157.4	160.0	1.02	149.4	161.0	1.07
	(11.05)	(25.63)		(13.02)	(25.70)	
1981	55.9	51.8	0.93	52.8	47.7	0.90
	(4.15)	(4.74)		(7.79)	(4.47)	
1991	49.5	45.6	0.92	47.3	43.8	0.93
	(3.75)	(3.11)		(3.83)	(3.25)	
Δ	-107.9	-114.4	-0.10	-102.1	-117.2	-0.14
%Δ	-68.6	-71.5	-0.10	-68.3	-72.8	-13.1

Note: The standard errors are obtained through the Gauss aproximation $\hat{\sigma}^2_{\%\Delta w_j} = \hat{\sigma}^2_{\alpha_j} 100 \exp(\hat{\alpha}_j)$ as they are a nonlinear function of the results from the regression equation.

As expected, Table 6 shows that in the three years that we consider, the return is higher (for both men and women) the higher the level of education attained. However, more importantly, the returns all declined considerably between 1968 and 1991. Again the estimates do not change much after the correction for sample selection bias is carried out. In percentage terms, the decline is largest for university education. The estimates corrected for sample selection bias suggest that for men in 1968, the rate of return associated with university education was 149.4%. By 1991, the return had declined to 47.3%. This represents a percentage decline of 68.3%. Likewise for women in 1968, the rate of return associated with university education was 161.0%. By 1991, it had decreased to 43.8%. This is a decline of 72.8%.

It is clear from Table 6 that the decline in the rate of return to education (as measured by qualification obtained) was not equally shared between men and women. In short, the decline was much larger for women. More specifically, in 1968 the ratios of the returns for women relative to men (according to the corrected estimates) were very close to 1.0: 1.03 for vocational education; 0.95 for gymnasium/other and; 1.07 for university education. By 1991, these ratios were all well below 1.0 - 0.70 for vocational education; 0.67 for gymnasium/other and; 0.93 for university education. This is confirmed by examining the percentage declines shown in the Table. In all cases, the percentage declines are larger (i.e. more negative) for women compared to men.

Again it must be stressed that like what was found for the return to years of schooling completed, the largest decline occurred between 1968 and 1981. Between 1981 and 1991, there has been little change. In fact, as is shown in Table 6, for males with vocational qualifications or gymnasium, the rate of return actually increased slightly between 1981 and 1991.

Results from Spline Functions

To compare the results from the "years of schooling" specification, we have choosen to estimate a cubic spline function using both years of schooling and years of work experience as independent variables. Knots are placed in 0, 15, 30 and 45 years of work experience and at 11 years of schooling. To model changes in the relation over the life cycle, we introduce an interaction between years of schooling and work experience (*educ*exp*). All calculations are done with sample selection correction, using the method described in the previous section.

In Figure 1, results for men 1981 are presented. This figure reveals some differences between the quadratic wage function and the cubic spline that are common to all six groups, compared in this study. First, the spline suggests a non-linear relationship between years of schooling

and wages. Second, early career wages, for individuals with few years of formal schooling, is seriously underestimated by the quadratic wage function and, thus, early career wages for the highly educated is overestimated, compared to the cubic spline. Third, late career wages for the relatively highly educated are underestimated by the quadratic. This property is modelled by the interaction component in the spline function.

The average wage differentials due to differences in years of schooling are calculated as follows. Eight years of schooling is taken as a reference and the average wage rate is predicted from 12 to 45 years of work experience. The average wage for nine years of schooling is calculated from 11 to 44 years of experience and so on, until twenty years of schooling which is calculated as an average of 0 to 34 years of experience. The percentage difference is then calculated between these mean wage rates. Note that we compare predictions of wages of individual of approximately the same age using this method. However, compared to the quadratic, we will underestimate the average wage differentials due to differences in schooling, as the wages are generally lower in the beginning of the life-cycle than in the end.

The results are reported in Table 7. In Figure 2 these results are plotted along with the predictions using the quadratic wage function. Figure 2 shows that these two methods give, apart from the linear restriction of the quadratic, very similar estimates of the wage differentials for both groups in 1981 and 1991. For 1968, however, the estimates diverge: the quadratic gives a much lower estimate for the highly educated and the difference is largest for females. This difference is also highlighted in Figure 3, where the predictions are plotted. There is a strong interaction effect between wages and work experience that is not modelled using the quadratic specification. The quadratic predicts decreasing late career wages for the whole population, while the spline only predicts decreasing late career wages for women with low education. For highly educated women, the spline predict increasing wages throughout the career.[4] However, these differences should be interpreted with caution, as there are very few observations with high education in this group (only 64 with university education).

The estimates of the wage differentials for different groups are plotted in Figure 4. Two important findings should be noted. First, following the discussion above, the spline function gives a larger estimate of the decrease in the returns to education between 1968 and 1981 for both males and females. Second, the 1981 and 1991 curves diverge after 17 years of schooling. This property could not be observed using the conventional quadratic wage function.

[4]A similar phenomenon is observed for Black American males by Piorier and Watts (1973).

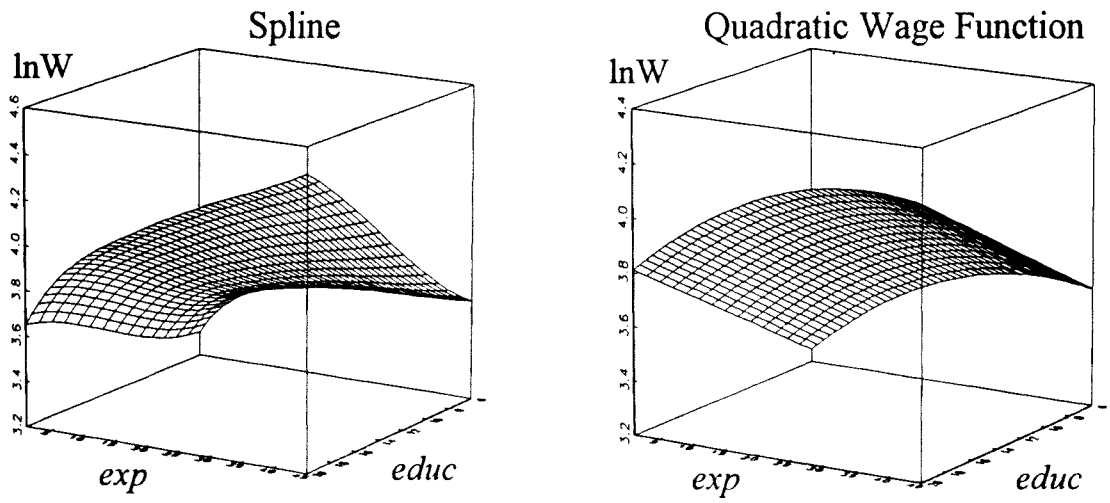


Figure 1. *Plotted predictions from a quadratic wage-equation with years of formal schooling and a cubic spline function. Males 1981*

Table 7. *Average wage differential compared to eight years of formal schooling.*

	Males			Females		
Schooling	1968	1981	1991	1968	1981	1991
9	5.5	3.7	4.0	8.1	1.8	1.5
10	11.4	7.8	8.3	16.7	3.9	3.5
11	19.1	12.4	13.0	25.8	6.7	6.5
12	30.3	17.3	18.0	35.6	10.2	10.7
13	45.0	22.3	23.2	46.0	14.5	16.0
14	62.5	27.5	28.2	56.9	19.6	21.8
15	81.9	32.4	32.7	68.6	25.4	27.5
16	101.4	37.1	36.3	80.9	31.8	32.6
17	118.5	41.1	38.6	94.0	38.7	36.1
18	130.3	44.4	39.2	107.8	46.1	37.3
19	133.5	46.5	37.8	122.3	54.0	35.5
20	125.6	47.5	34.0	137.6	62.0	30.1

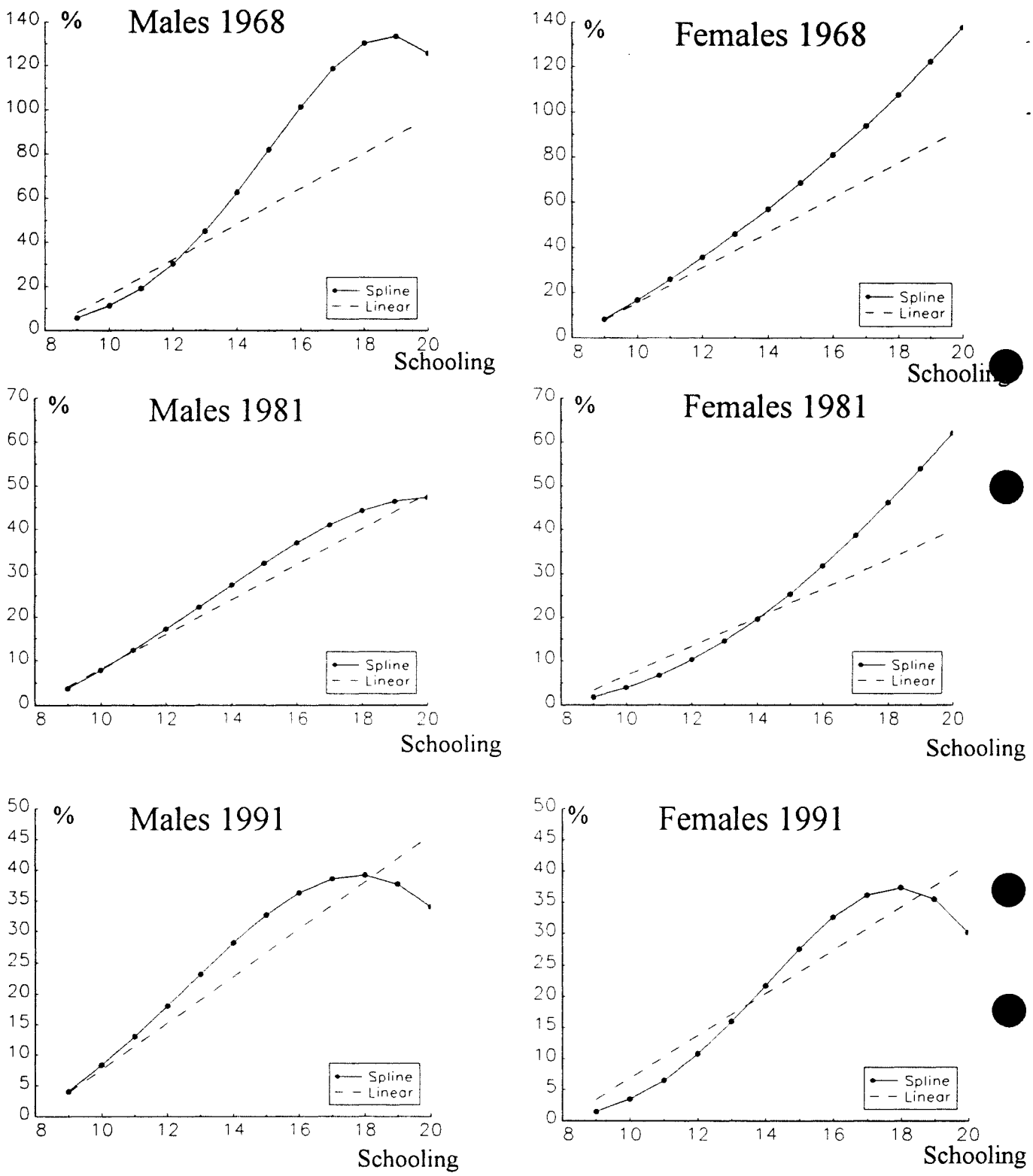


Figure 2. *Wage differentials related to differences in formal schooling. Comparison between linear specification and splines.*

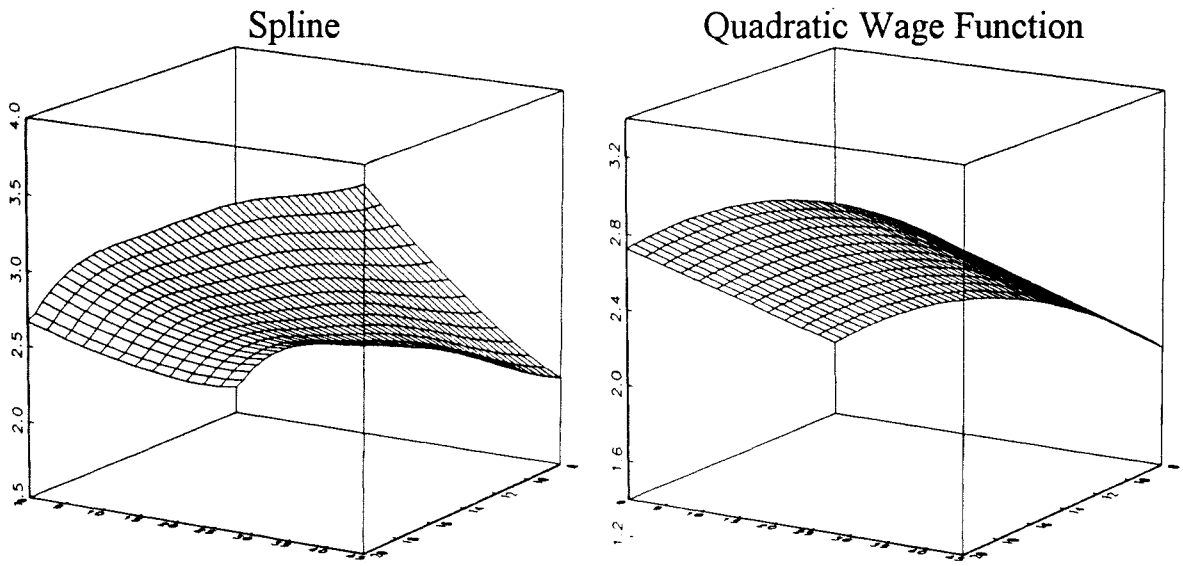


Figure 3. *Plotted predictions from a quadratic wage-equation with years of formal schooling and a cubic spline function. Females 1968.*

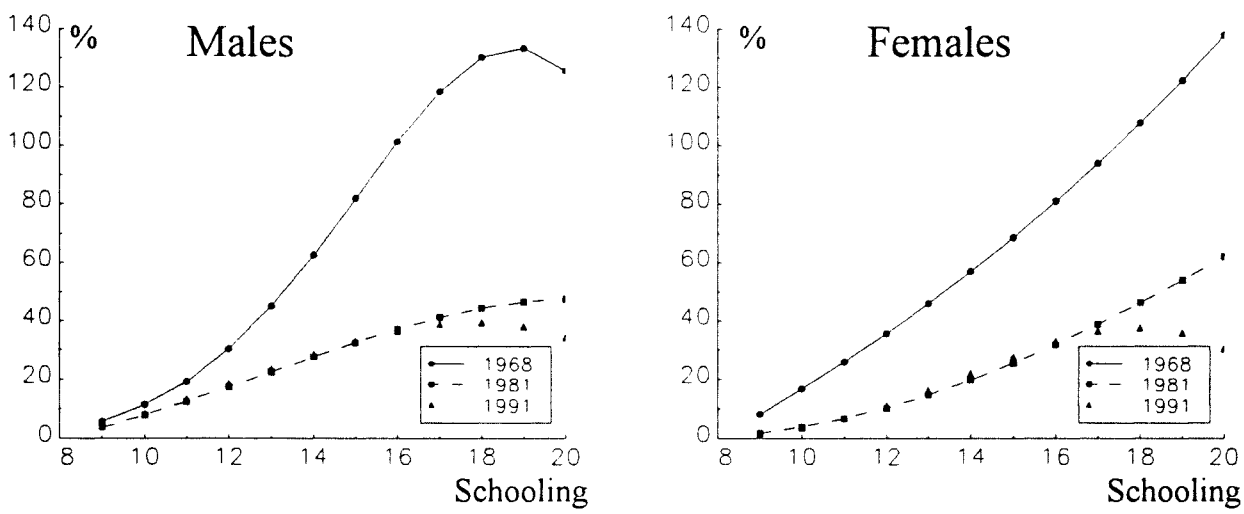


Figure 4. *Comparison of wage differentials related to formal schooling.*

We have also estimated cubic splines for each education level separately, in order to measure the between group wage differentials. For the groups S1 and S2 we have set the "knots" at 0, 10, 20, 30, 40 and 50. For the groups S3 and S4 we have placed the "knots" at 0, 20 and 40 years of work experience, as the sample sizes did not allow us to break down the data further.

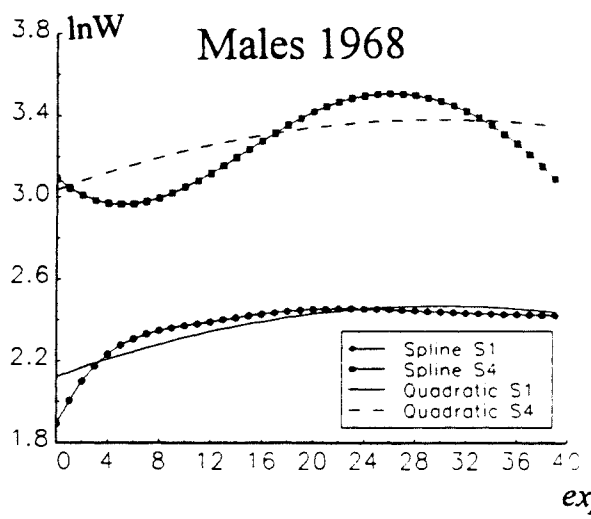
In Figure 5 the estimated splines are plotted along with predictions from the quadratic wage function for the groups S1 and S4. The reason for not showing the results from all groups is that the results from group S1 reveal the properties of the estimates from S2 very well. The same is true for S3 and S4. It can be seen in Figure 5 that the quadratic wage function underestimates early career wages in all cross-sections except for very early career wages for both men and women 1968, compared to the splines. In consequence, late career earnings is overestimated, compared to the splines. Figure 5 also reveals that group S4 is a much "noisier" and smaller sized than S1 as the spline estimates are much more unstable. It can, however, be seen that the quadratic estimates lower late career wages for at least 4 out of the 6 cross-sections.

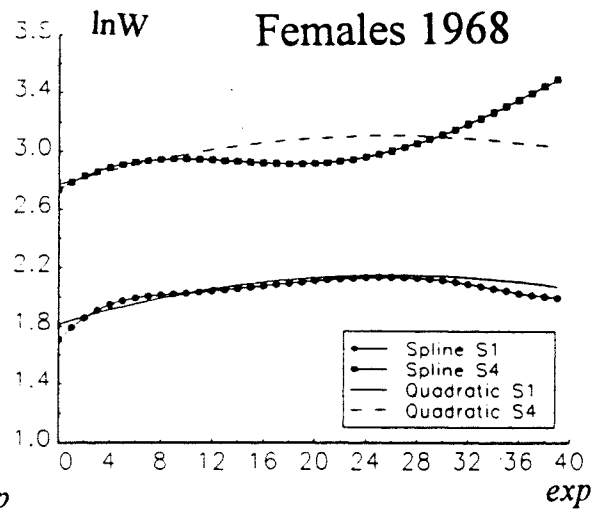
To compute the predicted mean wage differentials we have calculated the mean of the predicted wages from 0 to 39 years of work experience for group S4, 4 to 44 for S3, 8 to 47 for S2 and 11 to for group S1. The reason for choosing these years is that the mean number of years of schooling for men in the 1981 sample is 17.65 13.55 10.80 and 7.52 for groups S4, S3, S2 and S1 respectively. The years of comparison chosen is thus an average between 25 and 64. For convenience these years are chosen for the other cross-sections as well.

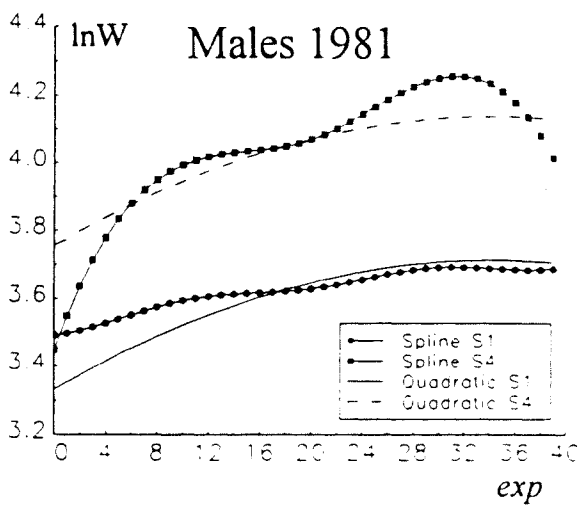
The results are shown in Table 8. It can be seen that the estimates are very similar to those obtained using the quadratic wage function and reported in Table 6. Thus, the specification with the education level dummies and the corresponding spline function gives a more similar prediction of mean wage differentials compared to the specification with years of schooling. The explanation to this is that this specification is not restricted to be linear in years of schooling and thus makes better predictions of the *mean* wages when there is an interaction between education and work experience.

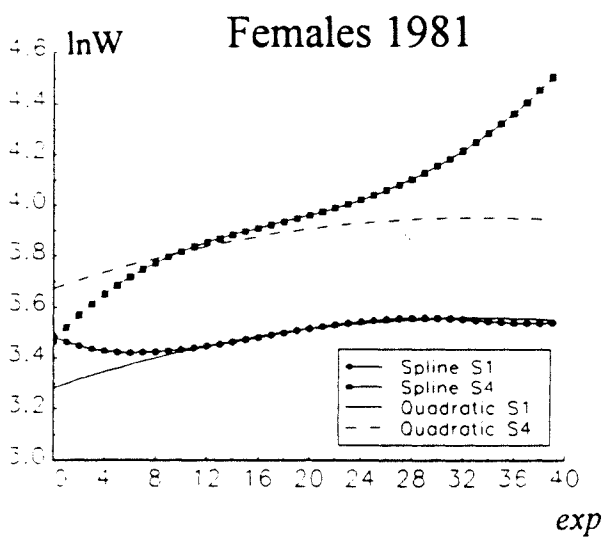
Table 8. *Wage differentials between education groups estimated with splines.*

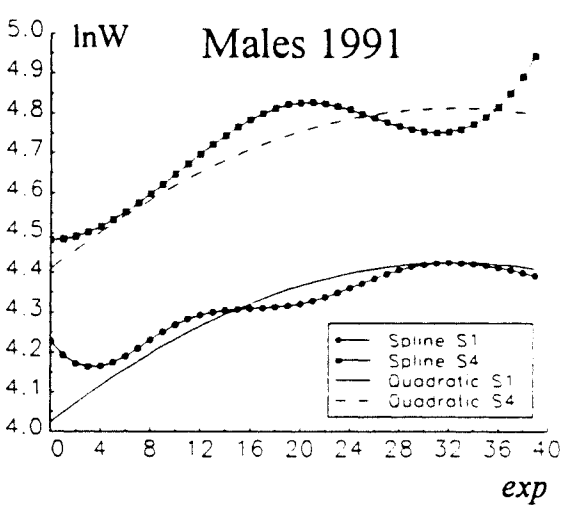
Year	*Males* (1)	*Females* (2)	*Ratio* (2)/(1)
S2: Vocational			
1968	17.7	35.6	2.01
1981	6.4	5.54	1.03
1991	12.2	6.4	0.53
Δ	-5.5	-29.2	5.31
%Δ	-31.4	-82.0	2.61
S3:Gymnasium/other			
1968	96.6	109.3	1.13
1981	28.6	8.5	0.30
1991	39.2	22.2	0.57
Δ	-57.4	-87.1	1.52
%Δ	-59.4	-79.7	1.34
S4: University			
1968	128.4	162.2	1.26
1981	43.8	55.7	1.27
1991	43.9	37.5	0.85
Δ	-84.5	-124.7	1.48
%Δ	-65.8	-76.9	1.17

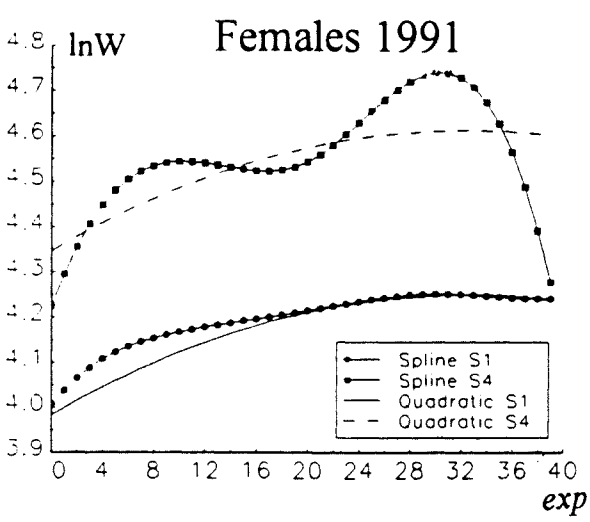

Males 1968
lnW
exp
Spline S1
Spline S4
Quadratic S1
Quadratic S4


Females 1968
lnW
exp
Spline S1
Spline S4
Quadratic S1
Quadratic S4


Males 1981
lnW
exp
Spline S1
Spline S4
Quadratic S1
Quadratic S4


Females 1981
lnW
exp
Spline S1
Spline S4
Quadratic S1
Quadratic S4


Males 1991
lnW
exp
Spline S1
Spline S4
Quadratic S1
Quadratic S4


Females 1991
lnW
exp
Spline S1
Spline S4
Quadratic S1
Quadratic S4

Returns to Schooling Within Education Categories

Finally, the rates of return to an additional year of schooling within a given education level are shown in Table 9. These returns were estimated by fitting Eq. (1) separately for each education level and including years of schooling completed in each of these equations. Therefore, the rate of return to schooling within each education level is the parameter of the years of schooling variable multiplied by 100.

Four points about this table are worth making. First, the estimates are qualitatively unchanged after the correction for sample selection bias is carried out. Second, with the exception of university education in 1991, the rates of return to schooling are all positive. That is, except for university education, each additional year of schooling completed is associated with an increase in wages. This is the case for men and women. Third, with the exception of "gymnasium/other" education amongst men, the returns to schooling have all declined between 1968 and 1991, with the decline being concentrated in the 1968 to 1981 period. And finally, again with the exception of university education, the percentage declines are larger for women compared to men.

Table 9. *Wage differentials due to additional schooling within each education level.*

Year	*Uncorrected for Selection Bias*			*Corrected for Selection Bias*		
Year	Males (1)	Females (2)	Ratio (2)/(1)	Males (3)	Females (4)	Ratio (4)/(3)
S1: Basic Compulsory						
1968	8.3	5.4	0.65	7.3	5.9	0.81
	(0.85)	(0.85)		(0.85)	(0.87)	
1981	5.2	1.8	0.34	4.7	0.9	0.19
	(1.01)	(0.66)		(1.00)	(0.70)	
1991	3.5	1.5	0.43	3.5	1.2	0.34
	(0.80)	(0.55)		(0.85)	(0.58)	
Δ	-4.8	-3.9	-0.22	-3.8	-4.7	-0.47
%Δ	-57.8	-72.2	-33.8	-52.1	-79.7	-58.0
S2:Vocational						
1968	5.0	7.7	1.54	5.0	7.7	1.54
	(0.79)	(1.14)		(0.80)	(1.15)	
1981	3.2	2.3	0.72	3.0	2.3	0.76
	(0.59)	(0.46)		(0.59)	(0.47)	
1991	2.6	2.7	1.04	2.5	2.7	1.08
	(0.55)	(0.40)		(0.55)	(0.41)	
Δ	-2.4	-5.0	-0.50	-2.5	-5.0	-0.46
%Δ	8.0	-64.9	-32.5	-50.0	-64.9	-29.9
S3: Gymnasium/other						
1968	1.6	10.5	6.56	1.4	10.7	7.64
	(2.14)	(3.11)		(2.11)	(3.22)	
1981	1.3	2.3	1.77	1.2	2.3	1.92
	(1.02)	(0.85)		(1.04)	(0.87)	
1991	2.0	1.3	0.65	2.0	1.3	0.65
	(0.88)	(0.73)		(0.87)	(0.73)	
Δ	0.4	-9.2	-5.91	0.6	-9.4	-6.99
%Δ	25.0	-87.6	-90.1	42.9	-87.9	-91.5
S4: University						
1968	0.6	9.0	15.00	0.8	8.7	10.88
	(2.21)	(5.02)		(2.15)	(5.66)	
1981	1.1	5.3	4.82	1.1	5.4	4.91
	(0.95)	(1.76)		(0.96)	(1.77)	
1991	-1.8	-0.5	0.28	-1.8	-0.6	0.33
	(1.24)	(1.30)		(1.21)	(1.30)	
Δ	-2.4	-9.5	-14.7	-2.6	-9.3	-10.55
%Δ	-400.0	-105.6	-98.1	-325.0	106.9	-96.97

5. Conclusions

In this paper, changes in the rate of return to education in Sweden were examined. The period considered was 1968 to 1991. The returns associated with years of schooling completed as well as education level reached were considered. The returns were estimated using wage equation analysis and are therefore measures of the differences in wage rates associated with differences in education. The data used were drawn from the 1968, 1981 and 1991 Swedish Level of Living Surveys. Three main conclusions emerge from the analysis. The first is that the rate of return to education declined considerably between 1968 and 1981. However, since 1981 there has been little change. The second is that the magnitude of the decline was not the same across all levels of education: the decline was particularly pronounced for university education. The third is that the decline was not equally shared between men and women: the decline was larger for women.

The quadratic wage function with a linear specification of the years of schooling variable and the cubic spline gave very different results for the 1968 sample. Therefore, the decrease in returns to education between 1968 and 1991 was underestimated using the quadratic wage function compared to the spline. The within group splines and the quadratic wage function with dummy variables on schooling levels yielded approximately the same estimates of mean wage differentials. However, the estimates of the relation between wages and work experience differed. This is of importance if one believes that the individual decision of investment in education is influenced by the cross-section wage distribution and the individual discounts future expected earnings. Results from the quadratic specification could then be misleading.

References

Björklund, A. and Kjellström, C. (1993). *Avkastningen på utbildning i Sverige 1968-1991.* Mimeo, Swedish Institute for Social Research, Stockholm University.

Blinder, A. S. (1976). On Dogmatism in Human Capital Theory. *Journal of Human Resources*, 11, 8-22.

Blomquist, N. S. (1979). Wage Rates and Personal Characteristics. *Scandinavian Journal of Economics*, 81, 505-520.

Edin, P-A. and Holmlund, B (1992). *The Swedish Wage Structure: The Rise and Fall of Solidarity Wage Policy?*. Working Paper 1992:13, Department of Economics. University of

Uppsala.

Eriksson, R. and Åberg, R. (1987). *Welfare in Transition - Living Conditions in Sweden 1968-1981.* Oxford: Clarendon Press.

Heckman, J. J. (1979). Sample Selection as a Specification Error. *Econometrica*, 47, 153-161.

Hibbs, D. A. (1990). Wage Compression Under Solidarity Bargaining in Sweden. Economic Research Report no. 30. Trade Union Institute for Economic Research. Stockholm.

Hibbs, D. A. (1990). Wage Dispertion and Trade Union Action. In *Generating Equality in the Welfare State: The Swedish Experience,* (ed.) I Persson. Norwegian University Press: Oslo.

Lorenz, W. and Wagner, J. (1990). *A Note on the Returns to Human Capital in the Egthties: Evidence from Twelve Countries.* Luxembourg Income Study Working Paper no. 54 (Walferdange, Luxembourg: Luxembourg Income Study).

Mincer, J. (1974). *Schooling, Experience and Earnings.* New York: Columbia University Press.

Murphy, K. and Welch F. (1990). Empirical Age-Earnings Profiles, *Journal of Labor Economics*, 8, 202-229.

Murphy, K. and Welch F. (1992). The Structure of Wages, *Qurterly Journal of Economics*, 107, 285-326.

Nilsson, C. (1992). *The Swedish Model: Labour Market Institutions and Contracts.* Working Paper 1992:4. Department of Economics. University of Uppsala.

Poirier, D. J. (1976). *The Econometrics of Structural Change*. North-Holland Publishing Company. Amsterdam.

Piorier, D. J. (1987). Spline Functions. In *The New Palgrave: A Dictionary of Economics,* (eds.) John Eatwell et al. The Macmillan Press: London.

Poirier, D. J. and Watts, H. W. (1973). The Measurement of Wage Rates. In New Jersey Graduate Work Incentive Experiment Final Report. Institut for Reasearch on Poverty. University of Wisconsin, Madison.

Psacharopoulos, G. (1973). *Returns to Education: An International Comparison.* Elsevier Scientific Publishing Company: Amsterdam.

Psacharopoulos, G. and Woodhall, M. (1985). *Education for Development: An Analysis of Investment Choices.* Oxford University Press: Oxford.

Schultz, T. P. (1990). *Economic and Demographic Consequences of Educating Women: Social Returns to Women Schooling.* Mimeo, Department of Economics. Yale University.

Suits, D. A., Mason, A. and Chan, L. (1978). Spline Functions Fitted by Standard Regression Methods. *Review of Economics and Statistics,* 60, 132-9.

Willis, R. (1986). Wage Determinants: A Survey and Interpretation of Human Capital Earnings Functions. In Ashenfelter, O. and Layard, R. (eds.) *Handbook of Labor Economics,* vol 1, North-Holland. Amsterdam.

Earnings Mobility and Distribution: Comparing Statistical Models on Swedish Data*

by

Mårten Palme$

Abstract: There are at least two instrumental motives for studying earnings mobility. First, to extend the analysis of income distribution to more than one time period. Second, to predict future individual earnings. For both these motives, adequate models of earnings mobility are needed. This study compares the usefulness of different statistical models (human capital and stochastic models), previously used to estimate earnings mobility, in predicting future individual earnings and earnings distributions. Special attention is given to the effect of considering individual heterogeneity. A 20-year panel, collected from the *Swedish Level of Living Survey*, of 651 employed men is used. The models are estimated on 17 periods and predictions are made for 3. It is found that a dynamic human capital model gives the best predictions of individual earnings.

---

* Thanks to Anders Björklund for showing me the data used in this study and for comments on several drafts of the paper. To Claes Cassel for extensive , patient and inspiering help with analysing the data. Thanks also to Per Johansson, Sune Karlsson, Anders Klevmarken, Joakim Palme, Ed Palmer, Erik Ruist and Anders Westlund for comments on earlier drafts of this paper. The paper was prepared under financial grants 89/209 from Bank of Sweden Tercentenary Foundation and F503/89 from The Swedish Research Council for the Humanities and Social Sciences.

$ Department of Economic Statistics, Stockholm School of Economics, Box 6501, S-113 83 Stockholm, Sweden.

1. Introduction

Apart from any intrinsic value of measuring earnings mobility, i.e. efficiency (labour market dynamics) or inefficiency (uncertainty in individual planning), there are at least two instrumental motives for such studies. First, if an empirical analysis of income inequality is extended to a longer time period than is covered by the data, income mobility has to be measured and predicted. A broad range of authors, from John Rawls to Milton Friedman, have argued that lifetime income would be the most appropriate unit of analysis for studying income distribution and the distribution of economic welfare. In a world of no uncertainty and perfect capital markets this position would not be controversial, and lifetime inequality would be the *only* relevant unit of analysis. However, as is well known, there are no societies where either of these conditions are satisfied. Therefore, transfers from one point of an individual's life to another are not costless and will sometimes not take place, especially if it is a long time period between the two points of time. This means that there is no, for all purposes, best time period for analysing income inequality. To the authors' knowledge there are no samples, of any reasonably broad population, covering such a long time period as a lifetime. Nevertheless, there are lot's of situations where the empirical analysis would be more appropriate if the data set could have been extended with some time periods, even with the help of predictions, if these are resonably accurate. It is, thus, relevant to evaluate *both* long term and short term predictions of earnings mobility.

The second motive is to predict individual incomes. This could be useful in public policy analysis, e.g. analysing the income distribution implications of, or future expenditures for, a pension scheme where benefits are based on individual earnings during a certain time period.

The purpose of this study is to compare the usefulness of different statistical models in serving the abovedescribed motives, i.e. in predicting individual earnings, earnings distribution and relative earnings mobility. The data we have used is 20 years of individual earning histories (annual income from labour), from 1970 to 1989, collected from the *Swedish Level of Living Survey*. The models are estimated on 17 periods and predictions are made for the last three.

Most statistical models used to analyse earnings mobility can be classified according to their basis in one of two schools of thought (1) "stochastic theory" (following Galton (1889) and Champernowne (1953)) or (2) "human capital theory" (following Mincer (1958) and Becker (1962)).[1] Human capital theory is derived from economic behaviour of individual choice. The

[1]See Sahota (1978) for a review of different theories of determination of the income distribution and Atkinson et al. (1991) for a recent survey on different statistical models for earnings mobility.

individual chooses between formal education and work, in order to maximise lifetime utility. Stochastic theory has no such connections to economic theory – earnings mobility is approximated using stochastic processes. To quote the seminal paper by Champernowne (1953) "The forces that determine the distribution of incomes in any community are so varied and complex, and interact and fluctuate so continuously, that any theoretical model must either be unrealistically simplified or hopelessly complicated." It is evidently true that it is impossible to take all possible determinants of future individual earnings into account in a statistical model. On the other hand, it is possible to improve the estimates using an adequate model and it is an empirical question to what extent this is true for models based on the human capital theory.

Models derived from both these theories have been used in empirical studies to construct hypothetical distributions of lifetime income or earnings. Creedy (1977) can be seen as representing the former school of thought, and Blomquist (1981) and Lillard (1977) can be identified with the latter school.

The first criterion for judging the usefulness of these two aproaches is a practical one: that is, whether or not they generate good predictions of future individual earnings and earnings distributions. But there is also a more fundamental criterion. Friedman (1953) argues that economic theories should be judged by their predictive power. Adopting this criterion implies testing whether human capital earnings functions could predict future earnings and earnings distribution better than "theory free" stochastic models.[2]

There are, however, several difficulties with answering this question empirically. (i) *Data*. It would require a nearly infinite data set. Firstly, the whole life-cycle of the individuals in the data-set have to be considered. Secondly, since human capital theory aims to predict the distribution of hourly wage rates, detailed information on each individuals work hours would be required. It would then be necessary to model the labour supply decision to obtain an earnings distribution. This could not be done in this study, since there are not sufficient information on work hours in our data-set. To minimise the impact of differences in labour supply, we have selected a relatively homogenous population, employed males between ages of 23 and 65 throughout the sample period. Third, information on the human capital investment components in different jobs and the quality of different educations would be required. (ii) *Functional form*. Human capital theory only predicts that a relationship emerges between the accumulation of human capital (through work experience and education) and wages. Choosing a functional form is then an additional problem. The quadratic functional form has recently been subject to some criticism. Murphy and Welch (1990) compare different functional forms on cross-sectional data and conclude: ".. it seems clear that the

[2]See Jochumzen (1993) for an application of a similar aproach to the CAPM model.

quadratic must be scrapped for purposes of estimating career earnings patterns.." (p. 228). However, Murphy and Welch (1990) also find some simple alternatives that fit the data better. These will be compared on longitudinal data in this study. (iii) *Dynamic structure*. As will be discussed in Section 3, there are several reasons why the dynamic error or system structure, might have different specifications. This problem, however, has very little to do with human capital theory and from that point of view the choice of specification of the error structure is arbitrary.

Thus, the test performed in this study is limited to a sub-set of specifications previously used in empirical studies and severely limited by the data available. Therefore, this study can not be seen as a decisive test of human capital theory, but as a piece of evidence from a limited experiment.

We give special attention to the issue of aggregation level, i.e. what is the loss or gain in assuming a common rather than an individual trend or error structure. Assuming a common structure facilitates both inferences on and predictions of earnings mobility.

The paper is organised along the following lines. Section 2 describes different specifications of the statistical models used. Section 3 describes the data set and the selections made. Section 4 presents results obtained by the models from all 20 observed years. There are two goals for this section. The first is to evaluate different properties of the statistical models. The second is to describe properties of the sample used, and compare these properties with results from other studies. The latter goal is partly intrinsic in value, since it is interesting to compare the Swedish earnings structure with earnings structures prevailing in other economies. However, it can also help in evaluating whether our findings in Section 5 can be held to apply only to our sample, or whether they are likely to have wider applicability.

Section 5 presents the results from two empirical experiments in which several versions of longitudinal human capital earnings functions and stochastic models are estimated on 17 years and predictions are made for the last 3 years. Two criterion for comparing the models are then computed. First, the sum of squared prediction error. Second, income distributions for each of the last three years, and for the sum these three years, are simulated using each of the models considered. The results are compared between the models and with the real world outcome. We can thus compare four aspects of the predictions of the different models: individual earnings, the earnings distribution for each year separately, the distribution of the sum of three years and the predictions of earnings mobility.

2. Models for Earnings Mobility

2.1 The Stochastic model

The basic idea of the stochastic theory of earnings distribution is that a cross-sectional distribution could be rationalised by a stochastic process. It was discovered already by Gibrat (1931) that departing from any distribution which undergoes a series of random independent *proportionate* changes will ultimately generate a log-normal distribution ("the law of proportionate effects"). Aitchison and Brown (1957) applied this law of proportionate effects to earnings distributions. Thus, the change in an individuals relative earnings position is stochastic, i.e. $\frac{d(y_{it} - \mu_t)}{dt} = u_{it}$, where i and t are subindices for individual and time period respectively, y is the log of earnings, μ is the mean of logged earnings and u is an independent, identically distributed stochastic component. This means that individual earnings could be predicted by a simple Markov process

$$y_{it} - \mu_t = y_{it-1} - \mu_{t-1} + u_{it}, \tag{2}$$

The "law of proportionate effects" then asserts that the distribution of log earnings is $N(\mu_t, \sigma_0^2 + \sigma^2 t)$. A problem with this formulation, as was pointed out by Kalecki (1945), is that the variance in logged earnings is linearly increasing – a property not consistent with empirical observations on many variates, and certainly not for empirical observations on cohort earnings distribution. Kalecki suggests a "regression towards the mean" parameter β, modelled as follows:

$$y_{it} - \mu_t = \beta(y_{it-1} - \mu_{t-1}) + u_{it}. \tag{3}$$

This model will generate the difference equation for the variance of normalised logged earnings, $\sigma^2(t+1) = \beta^2\sigma^2(t) + \sigma_u^2$, where σ_u^2 is the variance of the error component. Assuming that this process started off from a zero variance distribution, the variance will converge monotonically to $\sigma_u^2 / (1-\beta^2)$.[3] These two models rest on five assumptions – pointed out, among others, by Atkinson et al (1991):

(i) The same process applies to all individuals ("population homogeneity").

(ii) Constancy over time ("time homogeneity").

(iii) Earnings at date t are independent of earnings before t-1 (first order Markov assumption).

[3]Atkinson et al. (1991).

(iv) The stochastic components are uncorrelated over time (no serial correlation).

(v) The stochastic components are uncorrelated across individuals.

Several authors have extended these two basic models by relaxing at least one of the assumptions. In the study by Creedy (1977) quoted above, a first order autoregressive process in the residuals, u_{it}, is assumed, i.e., $u_{it} = \rho u_{it-1} + \varepsilon_{it}$, where ε_{it} is assumed to be i.i.d. A positive value of ρ may be interpreted as "success breeds success", growth in earnings is followed by further growth in the next period. A negative ρ indicates that earnings growth in one period tends to be followed by a decline in the next. The important property of this extension, in analysing income mobility, is that it allows for lowered earnings mobility. Substituting this into equation (3) yields

$$z_{it} = (\rho + \beta) z_{it-1} - \rho\beta z_{it-2} + \varepsilon_{it}, \tag{4}$$

where z_{it} is the deviation from mean logged earnings.

The population homogeneity assumption is often imposed if the number of time periods included in the data set is not sufficient for estimating separate processes for each individual. The population homogeneity assumption does not hold if there for example is a segmented labour market, with a primary sector where workers receive relatively stable earnings and a secondary sector where workers are exposed to relatively high earnings fluctuation.

To use these models to make forecasts of future individual earnings and earnings distributions, it is necessary to estimate the trend in mean earnings, μ_t. Creedy (1977) and Creedy and Hart (1979) assume that median earnings follow a quadratic trend in age, $\mu_t = \mu_0 + \theta t - \delta t^2$, i.e. that the average earnings profile has a biological origin. However, the stochastic theory gives no guidance as to why average earnings may change over the life cycle.

2.2 The Human Capital Model

According to human capital theory, observed individual earnings could be seen as the outcome of accumulation of personal investment. Personal investment could be made either through formal schooling or through work experience (on-the-job training). Rational allocation requires that much of the investments are made in early years. However, as it is realistic to assume that there are diminishing returns to production of human capital in each time period, the investments will not be made all at once. In this framework there are three factors that can explain the observed decelerating growth of earnings over the life-cycle. First, the payoff period (remaining working life) becomes shorter. Second, the alternative costs of investment become larger. Third, the value of the human capital accumulated

depreciates.

Thus, in an ideal econometric specification based on human capital theory, earnings mobility should be predicted by investment in human capital, measured in monetary units. In addition, the functional form must allow for the staggered growth pattern predicted by the theory. However, as it is very difficult to obtain such information, especially for on-the-job training, most studies use years of schooling and work experience as measures of human capital investments, and a semi-log specification.

The quadratic specification in years of work experience, originally used by Mincer (1974) on cross-sectional data, is by far the most popular earnings function based on human capital theory, i.e. satisfying the requirements given above. This specification has also been estimated on longitudinal data in several studies (cf. Lillard and Willis, 1978 or Lillard and Wiess, 1979). The specification is as follows:

$$y_{it} = \overline{\beta}_0 + \beta_1 educ_{it} + \beta_2 exp_{it} + \beta_3 exp_{it}^2 + \upsilon_{it}, \qquad (5)$$

where the sub-indices i and t represent individuals and year respectively; y is logged earnings, *educ* years of schooling, and *exp* years of work experience.

There are several possible specifications of the error component, υ. Lillard and Willis (1978) assumed the specification $\upsilon_{it} = \mu_i + \eta_{it}$ and $\eta_{it} = \rho\eta_{it-1} + \varepsilon_{it}$, i.e. υ could be decomposed into an individual effect, μ, and a serially correlated component, η; ε is assumed to be i.i.d. The presence of autocorrelation in longitudinal earnings models could be interpreted in several ways. First, it could be caused by individual random shocks that persist over more than one period but deteriorate over time. Examples of such shocks include the effect of firm closure or wage increases following individual innovations. Second, it may be measuring the influence of unobserved, or omitted, variables which are serially correlated (change slowly) over time (spurious autocorrelation). Third, autocorrelation could result from a too restrictively specified functional form, or be caused by individual earnings growth rate heterogeneity. Although its use has been tremendously widespread, the quadratic earnings function has been questioned on grounds of excessive restrictivity in several recent studies (see Willis, 1986 p 538 footnote 2, and Murphy and Welch 1990,1992). Murphy and Welch (1990) study the bias of the quadratic approximation on cross-sectional CPS data for different experience levels. Their main conclusion is that early career earnings are over estimated and mid-career earnings underestimated by the quadratic human capital earnings function.

As was first discussed by Mincer (1962), the on-the-job training hypothesis in the human capital framework allows for growth rate heterogeneity. Consider two jobs, a and b, that require the same formal schooling and same abilities while b requires more post schooling

OJT investments. Optimizing behaviour of the individuals would generate a labour market in equilibrium in which the earning profiles of the two jobs have the same present value. This is illustrated in Figure 1. If we estimate the model with common work experience coefficients on earnings data of the two individuals in possession of jobs a and b, we will generate a predicted earnings profile similar to profile c in figure 1 for both individuals and the misspecification will generate an autocorrelated error component. Note that growth rate heterogeneity is not consistent with the stochastic model.

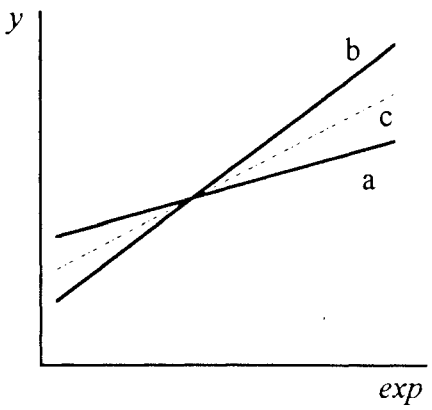


Figure 1. *Heterogeneous on-the-job training.*

There are several ways to measure and accomodate this phenomenon econometrically. First, it is possible to estimate separate equations for each individual. Second, in the random effects framework, it is possible to extend the error component with an individual growth rate component, i.e. $\upsilon_{it} = \mu_i + b_i \exp_{it} + \eta_{it}$ (c.f. Lillard and Wiess, 1979, or Hause, 1980). Third, one might difference each individual earnings series (see MaCurdy, 1982).

Autocorrelation could arise from state dependence, i.e. the individual's earnings in the current period are, at least partly, directly determined by his earnings in the previous period. Thus, models with serial correlation are not the only possible dynamic specification of a human capital earnings function. A specification with lagged dependent variables is also possible, i.e. $y_{it} = \overline{\beta}_0 + \gamma y_{it-1} + \beta_1 educ_{it} + \beta_2 exp_{it} + \beta_3 exp_{it}^2 + \upsilon_{it}$. This could be seen as a combined model between the stochastic model, where all information on current individual earnings is contained in previous earnings, and the human capital model, where earnings are entirely determined by investments in individual traits. As is shown by for example Maddala (1987 p. 313) models with first-order autoregressive errors could be seen as a special case of models with lagged dependent variables, i.e. $y_{it} = \beta x_{it} + \upsilon_{it}$, where $\upsilon_{it} = \rho\upsilon_{it-1} + \varepsilon_{it}$ could be written $y_{it} = \rho y_{it-1} + \beta x_{it} - \rho\beta x_{it-1} + \varepsilon_{it}$. This could be seen as a dynamic model $y_{it} = \gamma y_{it-1} + \beta_0 x_{it} + \beta_1 x_{it-1} + \varepsilon_{it}$, with the non-linear restriction $\gamma\beta_0 + \beta_1 = 0$. If the restriction

holds, then the apparent effect of y_{t-1} is due to serial correlation in the errors. If the restriction does not hold, then y_{t-1} has an effect on y_t, i.e. "state dependence" exists.

There are several reasons why one would expect state dependence in longitudinal earnings functions. It is frequently observed that people tend to persist in their behavior. This could be an element of their labour supply behavior – for example, the probability that a person works full time is probably greater if he worked full time in the preceding period. Furthermore, an element of wage setting could be that the previous period's wage constitutes a baseline for the current period's wage.

3. Data

The sample used is obtained from the *Swedish Level of Living Survey (SLLS). SLLS* is a panel survey and the interviews were made in 1968,1974, 1981 and 1991. The sample size each year is about 6,000 individuals and it is a representative sample of the Swedish population in each survey year.[4] To obtain data on the dependent variable, individual annual earnings, the selected sample is matched with tax registers. To measure earnings, taxable labour income ("inkomst av tjänst") is used, i.e. taxable income received from the employer plus income from sickness insurance minus deductions made for expenses made to earn labour income, for example expenses for commuting. This definition of labour income is chosen as it is the only labour income concept for which data is available for such a long time period as was desired for this study.[5] A deficiency using this income concept is that earnings from social security, e.g. compulsory sickness insurance, is included after 1974 but not before. This will evidently create some additional variation in the earnings distribution before 1974.

The population is defined as employed men between the ages 23 and 64 throughout the sample period, 1970-1989. When the selections were made, the aim was to obtain a representati sample of this population. The reason for selecting out women is that female labour supply varies a lot more than male. As we will not attempt to model the labour supply decision, and minimise earnings mobility due to variations in labour supply, we will therefore restrict the study to men. As earnings data is highly unreliable for young people, due primarily to studies and irregular life habits, individuals younger than 23 years of age were excluded. The reason for excluding the self-employed, is that their earnings and earnings histories are not comparable to employees for a number of reasons not dealt with here.[6] To exclude the

[4]The Level of Living Survey is described in detail in Eriksson and Åberg (1984).
[5]This definition is, however, not adequate for all purposes. If the purpose is to study the distribution of that part of households incomes which could be used for consumption, it could be argued that to live a relatively long distance from one's workplace could be a form of consumption chosen by the worker. On the other hand, the situation could also result from failure in urban planning or a sluggish housing market.
[6]See Johansson and Hedström (1979) for a comparison of earnings data and living conditions for self

self-employed, information on occupations from the surveys in 1974 and 1981 were used. If the individual was self-employed in either 1974 or 1981, he was deleted from the sample. When the selections described above had been made, the sample contained 691 individuals.

There is only information on deaths of individual respondents in the *SLLS* up until the 1st of January 1985. 9 individuals in our sample fell into that category and were excluded from the sample. 19 individuals were not registered as Swedish citizens in 1989, i.e. they have either died between 1985 and 1989 or emigrated permanently. These were also excluded from the sample.

Some of the observations on the dependent variable were zero or reported as missing. If income from earnings is zero, the individual has filed an income tax return, but has reported no income from labour. The reason could be that the individual is working in the "informal sector" of the economy, or that his labour supply is zero, for example due to studies. Missing values could result either because of administrative errors or because the individual in question has not filed an income tax return. The reason for the latter possibility could be, for example, that the individual received all income during residence outside of Sweden in a given year, or because the individual has died. The zero observations are, for our purposes, equivalent to the missing values. Since labour supply decisions are not modelled in this study, the zero observations give us no information on the amount of earnings the individual would have received if the labour supply would have been "as usual" in the year of the zero observation. 14 individuals have five or more missing values or zeros on the dependent variable, and were selected out. Thus, in total, 40 individuals (5.8 per cent of the original sample) constitutes the attrition of the panel. The judgement of the author of this study is that such limited attrition does not have a great effect on the results reported in Sections 3 and 4 and no attempts is therefore made to correct for a possible attrition bias.

When these additional selections were made, 651 individuals remained in our sample. Of these, 584 have complete records on earnings, 36 have one missing or zero observation, 14 have two, 8 have three and 9 have four missing or zero observations. For the missing or zero observations remaining in the sample, values have been imputed. The motive for imputing is that it enables us to use a rectangular data-set. There are two alternatives to imputation. First, it is possible to exclude the individuals containing missing or zero observations. Since the number of individuals containing missing observations are relatively large (67 individuals, i.e. more than 10 per cent of the total sample), this would possibly cause selectivity bias. Second, it is possible to use estimation methods for unbalanced panels. However, these methods are computationally very tedious to use at least for some of the methods in this study. Since the number of observations to be imputed is relatively small (124 of 13,020, less than 1 per cent

---

employed and employees.

of the potential sample size), the error resulting from the imputation should be relatively small.

The imputations were made under the assumption that the missing observation[7] is missing at random (denoted MAR in the missing value literature). This means that the probability to observe a value or not, is not dependent on the "true" value of the dependent variable in question. We have used the following imputation method. If observations immediately surrounding the missing value (for the particular individual) were not missing, the missing value was simply set equal to the average of these two surrounding observations. If two, three or four observations were missing in a sequence, weighted averages of the surrounding observations were used. If the missing or zero observation was located on the end points, the last, or first, non-missing or non-zero observation was multiplied by the sample average change between the observed and the imputed year to yield an estimate. Intuitively, it seems that this method will yield a good approximation of the "true" value of the missing observation. Satisfactory theory to analyse the effects of this imputation method is lacking, however. The reason for choosing this particular imputation method, rather than a model-based method yielding estimable errors under certain conditions , is that a goal of this study is to compare different models. To first use every model to impute missing values and then estimate them on the completed sample would be very tedious, since the error caused by the imputation is very limited, due to the small share of missing observations.

The dependent variable, earnings, is normalised by the average hourly wage rate for the entire Swedish economy. This normalisation will remove changes in individual earnings that depend on macroeconomic fluctuations (price changes and changes in real growth) and changes in the overall functional income distribution. Another method to control for these macroeconomic changes is to add a dummy variable for each year to the equation (see for example Lillard and Willis, 1978). Since one of the goals of this study is to compare the predictive usefulness of different statistical models, and since the macroeconomic changes are exogenous to these models, this method could not be used. Table 1 shows the annual sample means of earning normalised by the 1970 average wage rate.

[7]Not necessarily the individuals containing the missing observations.

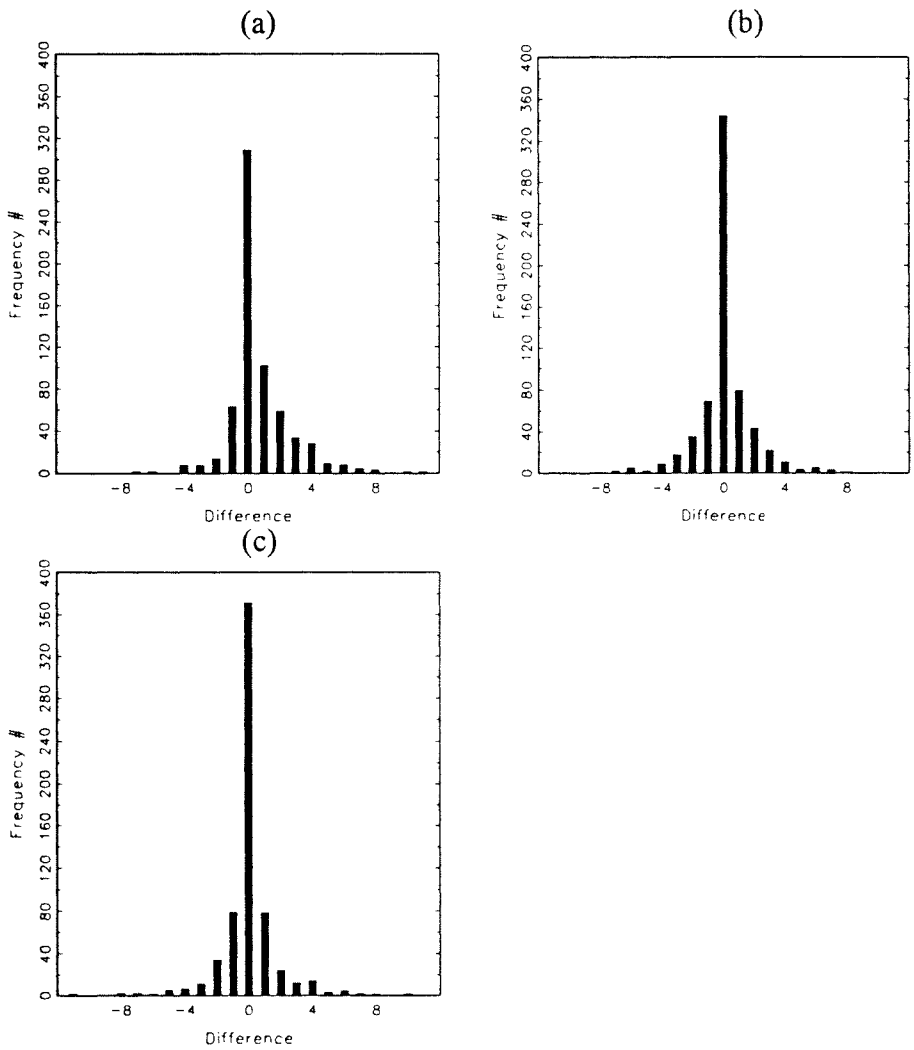
Table 1. *Earnings normalised by the average hourly wage rate, number of years of work experience and years of schooling. Sample averages.*

Year	Earnings	Standard deviation of earnings	Work Experience	Schooling
1970	30,013	12,264	14.74	9.42
1971	30,575	12,301	15.87	
1972	29,917	11,748	16.82	
1973	29,984	11,804	17.74	
1974	31,663	11,460	18.72	10.10
1975	32,074	11,456	19.70	
1976	32,358	10,988	20.68	
1977	32,438	11,071	21.67	
1978	31907	10,805	22.65	
1979	31981	10,658	23.64	
1980	32,035	10,833	24.63	
1981	32,325	10,872	25.56	10.21
1982	33,132	11,484	26.50	
1983	32,770	11,780	27.48	
1984	33,013	12,450	28.46	
1985	33,267	12,851	29.44	
1986	32,938	13,509	30.42	
1987	32,850	13,344	31.41	10.22
1988	33,645	14,229	32.39	
1989	33,501	14,134	33.37	

Information on the independent variables used in the human capital framework, years of education and years of work experience, are collected from the 1968, 1974, 1981 and 1991 surveys. More specifically, for each individual the years of schooling variable for 1970 and 1971 is coded equal to the answer from the 1968 survey; for 1972-77 this variable is obtained from the survey response in 1974, for 1978-85 from that of 1981, and 1986-89 from that of 1991. Twenty missing values for this variable were reported in 1968, and one was reported in 1981. In these cases, years of education are set equal to the amount reported in 1974. In the 1970s the number of adults in education increased considerably in Sweden. The longitudinal data used in this study could be used to estimate the effect on earnings of this education. Figures 1a through 1c depict the frequencies of the difference in years of schooling stated in 1974 and 1968, 1981 and 1974, and 1991 and 1981 respectively. It is possible to increase the number of years of schooling during the time period studied, but impossible to decrease it. As could be seen in Figure 1, a considerable number of individuals have actually stated that they

have decreased the number of years of schooling, which could only be interpreted as measurement error.

Figure 1. *(a) Differences in years of schooling stated in 1968 and 1974. (b) Same for 1974 and 1981. (c) Same for 1981 and 1991.*



Interview responses were also used construct the years of work experience variable. This information obviously contains measurement errors as well. However, as is shown in Table 2, the difference between the mean years of work experience is fairly consistent. To impute values for the work experience variable for the years between the surveys, the following strategy is used. If the individual has a labour income above 10,000 SEK[8] one year of work

[8]This amount of money corresponds to the average wage rate in 1970, thus approximately 51,000 SEK in

experience is added. The notations for two surveys are considered to be consistent if the difference between the observed and the imputed value for the year of the survey is plus or minus one or zero when the abovedescribed imputation method is used. If the notations from two surveys are consistent, years of work experience are imputed for the observations between the surveys using the abovedescribed principles, departing from one of the two surveys. If the notations from two surveys closest in time are not consistent, the imputations are made departing from two consistent notations, regardless of whether the surveys are closest in time or whether the years to be imputed are between the surveys.

This method could be used to impute the years of work experience variable for 601 individuals. The information on 50 individuals did not meet the consistency requirements. However, work experience for 21 more could be imputed when the consistency requirement was loosened to plus or minus two years. 9 of the remaining 29 individuals have a plausible work experience value for 1968, i.e. their reported years of work experience was less than or equal to their potential work experience.[9] For these individuals, work experience was imputed on the basis of the 1968 notation. For the remaining 20 individuals, with either an implausible reported work experience for 1968 or a missing value for this year, work experience was imputed on the basis of the potential work experience in 1968.

Table 2. *Differences in mean years of work experience and mean years of schooling between surveys.*

Survey	Difference	Mean no. of Years of Work Experience	Difference	Mean no. of Years of Schooling	Difference
1968		13.37		9.42	
1974	6	18.31	4.94	10.10	0.68
1981	7	24.52	6.21	10.21	0.11
1991	10	34.49	9.97	10.22	0.01

1989 current prices (approximately 5,000 £ or 7,500 US$).

[9]Potential years of work experience is defined as the individuals age in 1968 minus reported years of schooling in 1968 minus 8 years (AGE68-EDU68-8).

4. Estimation Methods and Descriptive Results

In this section we describe the estimation methods and present some results from the entire data-set, i.e. all 20 time periods.

4.1 Stochastic Models

To estimate the cohort mean income we use a quadratic function in age, i.e. $y_{it} = \beta_0 + \beta_1 age_{it} + \beta_2 age_{it}^2 + \upsilon_{it}$, where y is individual annual logged earnings and *age* is individual age minus 20. The reason for using this method rather than calculating a geometric mean for each age cohort, is that we do not have enough data for each birth-year cohort to get a stable estimate for each year. Our method is based on the assumption that there is no "cohort-effect" on earnings, i.e. that the birth cohort does not affect expected individual earnings apart from overall economic growth or macroeconomic fluctuations. The estimation was carried out using the fixed effects method. Since the Durbin-Watson statistics indicates positive autocorrelation, the method to correct for this using panel data proposed by Anderson and Hsiao (1982) was used. The following results were obtained,

$$y_{it} = 10.21 + \underset{(0.0026)}{0.0332} age_{it} - \underset{(0.0001)}{0.0006} age_{it}^2, \qquad \rho=0.6189, \quad R^2 = 0.7607.$$

with standard errors within parantheses. To estimate model (2), we used the OLS procedure recommended in Cassel and Lundquist (1990), where it is shown to be unbiased when survey sampling as well as model aspects are taken into account. The following estimates were obtained:

$$z_{it} = \underset{(0.0043)}{0.8661} z_{it-1}, \qquad \hat{\sigma}_u^2 = 0.0347, \qquad R^2 = 0.7863.$$

The autoregressive coefficient is significantly smaller than 1 (t-test). Thus, the "law of proportional effects" could be rejected in favour of the "regression towards the mean" hypothesis.[10]

Figure 2 shows the distribution of AR-coefficients when model (3) was estimated for each individual separately. Very few estimates, only 68 of 651, are above 1. The R^2 was improved slightly to 0.8151 and the overall variance of the disturbance decreased to $\hat{\sigma}_u^2 = 0.0300$.

[10]This result has been obtained in most empirical studies on this subject. See e.g Thatcher (1976) table 8.5.

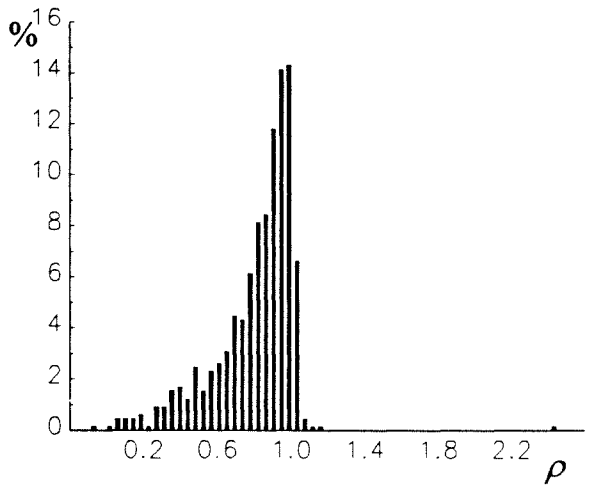


Figure 2. *Distribution of individual AR-coefficients.*

The assumption of no autocorrelation could also be relaxed. Model (4), from Section 2, is estimated using OLS and the following results were obained

$$z_{it} = \underset{(0.0090)}{0.7666} z_{it-1} + \underset{(0.0089)}{0.1100} z_{it-2}, \qquad \hat{\sigma}_{\varepsilon}^{2} = 0.0335, \qquad R^2 = 0.8098.$$

The estimated second-order coefficient is significantly different from zero by a simple t-test. By comparing the Akaike criterion, AIC(k)=$\ln \hat{\sigma}_k^2 + \frac{2k}{NT}$, where k is the order of the AR-process, N is the number of individuals and T is the number of time periods, model (4) is preferred to (3), i.e. AIC(2)=-3.397 and AIC(1)=-3.362. As was shown in Section 2, this model could be interpreted as allowing for autocorrelation in a first order model. Thus, from the results we obtain estimates of β and ρ to 0.89 and –0.12. The negative coefficient of the autocorrelation coefficient could, following Atkinson et al. (1991), be interpreted as "success in a given year would breed failure in the following one". However, the results should be interpreted with some caution since there is likely to be colinearity between z_{it-1} and z_{it-2}.

4.2 The Human Capital Model

Table 3 shows the results from estimates of the human capital model (5). The first column shows the results from the fixed effects (or "within-groups") and the second the random effects (or "between groups") estimation. As the estimates clearly indicate autocorrelation, with ρ = 0.6158 and Durbin-Watson, d = 0.7623, the method to correct for first order autocorrelation proposed by Andersson and Hsiao (1982) and reviewed in Hsiao (1986) are used for the respective estimation method. As pointed out by Hausman and Taylor (1981), these methods generate somewhat different interpretations of the coefficient estimates. In contrast to the fixed effects estimator, the random effects estimator uses information on

differences between observation units. The difference between the estimates of the years of schooling coefficient could thus be interpreted to mean that returns to schooling obtained before the sample period is on average larger than returns to schooling during the sample period. The reason for this is that the fixed effects estimator only measures effects of changes in the independent variables, but the random effects estimate is a weighted average between individual differences and these changes.

Table 3. *Coefficient Estimates of the Quadratic Human Capital Equations. Years of Work Experience and Schooling. Standard Errors in Parentheses.*[11]

	(1)	(2)
exp	0.0322	0.0354
	(0.0025)	(0.0035)
$exp^2/100$	-0.0551	-0.0606
	(0.0053)	(0.0053)
educ	0.0111	0.0140
	(0.0053)	(0.0036)
ρ	0.6158	0.6158
R^2	0.7623	0.5243
R^2, adjusted	0.7242	0.5243
Durbin-Watson	0.7623	0.7623

(1) Fixed effects model
(2) Random effects model

The Hausman test is appropriate for testing for correlation between the independent variables and υ. The test is defined by $m = (\hat{\beta}_F - \hat{\beta}_R)'\left[\mathrm{Var}(\hat{\beta}_F) - \mathrm{Var}(\hat{\beta}_R)\right](\hat{\beta}_F - \hat{\beta}_R)$, which is χ^2 distributed with degrees of freedom here equal to 3 (the number of independent variables), $\hat{\beta}_F$ and $\hat{\beta}_R$, are coefficient estimates for models with fixed (consistent) and random (efficient) effects respectively. The test statistics is estimated to $m = -3.77$, i.e. the hypothesis of no

[11]The results could be compared to those of several other studies on longitudinal income data. Klevmarken (1992), using 1984, 1986 and 1988 waves of the Swedish HUS-panel and the hourly wage-rate as dependent variable, obtained much lower estimates of the coefficient for work experience, 0.00105 *exp* and -0.0007 exp^2. Jäntti (1993) got estimates ranging between 0.036 and 0.051 (*age*), and -0.653 and -0.417 ($age^2/1000$) on data from a Finish four wave panel (1970, 1975, 1980 and 1985), using annual earnings as dependent variable. Lillard and Willis (1978) obtained estimates very similar to ours, 0.038 *exp* and 0.0007 exp^2, using PSID-data (1967-1973) and annual earnings as dependent variable. Their estimate of the coefficient for years of schooling is considerably higher than ours, 0.084.

correlation between the independent variables and the error term could not be rejected.

In Section 5, when making predictions, we will use the fixed effects estimator as we will use the estimated dummy variables in the predictions.

4.2.1 Functional Form

As mentioned in Section 2, the quadratic functional form is not the only possible specification of an earnings function based on human capital theory. Following Murphy's and Welch's (1990) analysis of cross-sectional data, we will also evaluate higher order polynomials in work experience. Figure 3a through 3c shows predictions from quadratic, cubic and quartic earnings functions respectively, along with means for each years of work experience group. The estimation of these equations was carried out using a fixed effects model with correction for first-order autocorrelation.[12] The properties of the quadratic earnings function revealed in Figure 3 are similar to the findings of Murphy and Welch. The quadratic severely overestimates earnings over the first two years, but underestimates early career earnings (up to about 20 years of work experience). The same is true for the cubic, though to a lesser extent. By far the best prediction of the earnings development in that part of the career is given by the quartic. Mid-career earnings (between 20 and 35 years of work experience) are overestimated by the quadratic, but fairly well predicted by both the cubic and the quartic. Late career earnings are overestimated by all specifications, most severely by the cubic that predicts increasing late-career earnings. The quartic predicts falling earnings too late in the career path.

If the functional form gives an unbiased estimate of the earnings path, the difference between mean earnings and predicted earnings for the respective groups with the same number of years of work experience should be randomly distributed. Thus, serial correlation for this error may indicate that this is not the case and, consequently, that the specification yields biased results. For the quadratic, the first order autocorrelation coefficient is estimated at 0.253. For the cubic and quartic, it is estimated at 0.625 and 0.284 respectively. The

---

[12]The following coefficient estimates are obtained for the cubic function:

$$y_{it} = 9.801 + \underset{(0.0025)}{0.0103}\,educ + \underset{(0.0059)}{0.0648}\,exp - \underset{(0.0003)}{0.0022}\,exp^2 + \underset{(0.000004)}{0.00002}\,exp^3,$$

ρ=0.615, R^2=0.762.

The following estimates is obtained for the quartic:

$$y_{it} = 9.653 + \underset{(0.0025)}{0.0097}\,educ + \underset{(0.0116)}{0.1097}\,exp - \underset{(0.0009)}{0.0061}\,exp^2 + \underset{(0.00003)}{0.00015}\,exp^3 - \underset{(0.0000003)}{0.000001}\,exp^4,$$

ρ=0.614, R^2=0.764

hypothesis of no serial correlation could be rejected using a Durbin-Watson test for the cubic specifications but not for the other two functional forms.

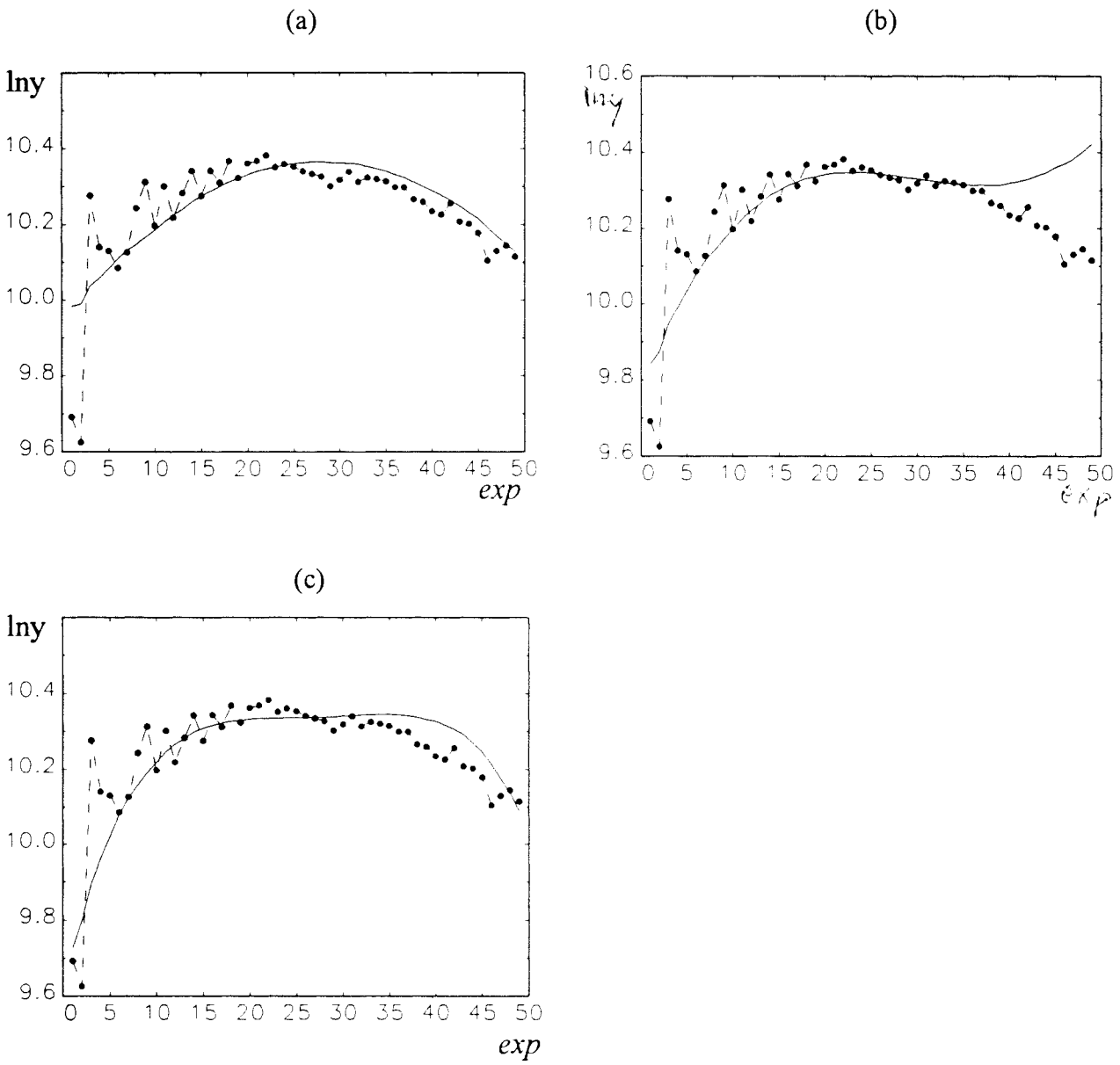


Figure 3. *Predictions of human capital functions compared to work experience group means. (a) Quadratic. (b) Cubic. (c) Quartic. (–) Human capital function. (•) Group means.*

4.2.2 Error Structure

As the reader will recall from Section 2, autocorrelation could occur for several reasons. Furthermore, it is not evident that the residuals follow an AR(1)-process as it could very well follow another, more complicated process. To find out whether this is happening we will estimate the autocorrelation and partial autocorrelation functions of the residuals from the fixed effects earnings function. We will follow a procedure proposed by MaCurdy (1982).

To estimate the *k*th order autocorrelation, the following set of equations is estimated:

$$\begin{bmatrix} \upsilon_{i,T}\upsilon_{i,T-k} \\ \cdot \\ \cdot \\ \upsilon_{i,k+1}\upsilon_{i,1} \end{bmatrix} = \begin{bmatrix} 1 \\ \cdot \\ \cdot \\ 1 \end{bmatrix} \gamma_k + \begin{bmatrix} \varepsilon_{iT} \\ \cdot \\ \cdot \\ \varepsilon_{i(k+1)} \end{bmatrix},$$

where υ is the residual of the earnings function, γ_k is the parameter to be estimated and ε_{it} is an error term independently distributed between individuals. Thus, for each individual there is a set of seemingly unrelated equations with $E(\varepsilon_i\ \varepsilon'_j) = \sigma^2_\varepsilon$ if i = j and 0 otherwise. This equation system could be estimated under the restriction that γ_k is constant throughout the sample period.

The sample partial correlation function could be estimated in a similar way. Consider a following set of equations for each individual

$$\begin{bmatrix} \upsilon_{it} \\ \cdot \\ \cdot \\ \cdot \\ \upsilon_{i,k+1} \end{bmatrix} = \sum_{j=1}^{k} \begin{bmatrix} \upsilon_{iT-j} & & 0 \\ \cdot & \cdot & \cdot \\ 0 & \cdot & 0 \\ \cdot & \cdot & \cdot \\ 0 & & \upsilon_{i,(k+1-j)} \end{bmatrix} \begin{bmatrix} 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix} \rho_j + \begin{bmatrix} e_{iT} \\ \cdot \\ \cdot \\ \cdot \\ e_{i(k+1)} \end{bmatrix}$$

with $E(e_i e'_j) = \sigma^2_e$ if i = j and 0 otherwise, where ρ_k is a parameters representing the *k*th order partial correlation coefficient. This seemingly unrelated regression model could be estimated under the restriction that ρ_k is equal for all $T-k$ time periods.

The results for the residuals of the fixed effects model for the first 7 periods are shown in Table 4a and for all 19 periods in Figure 4. The autocorrelation function is positive for the

first periods and then become strongly negative. An explanation for these results is that they are generated by growth rate heterogeneity. If individual earnings follow a steeper growth curve than average, the residuals will on average be negative in the beginning of the sample period and positive in the end, thus generating a positive correlation for time periods close in time and negative for distant time periods. For individuals following a flatter growth rate than the average, the residuals will on average be positive in the beginning and negative in the end and the same correlations will result.

If quadratic OLS equations are estimated separately for each individual, we obtain the result revealed in Table 4b and Figure 5. It can be seen that the pattern of these functions changes radically from the model with common β.

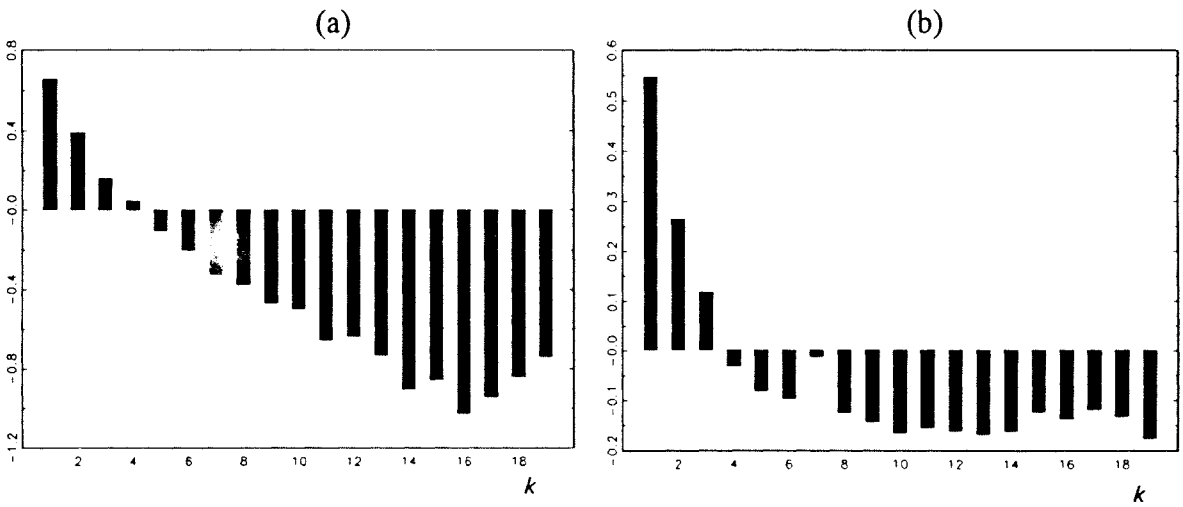


Figure 4. *Autocorrelation function (a) and partial autocorrelation function (b) for OLS residuals from equation with common β.*

Table 4a. *Autocovariance, autocorrelation and partial autocorrelation. OLS residuals from equations with common β (standard errors in parantheses).*

k	0	1	2	3	4	5	6	7
Autocovariance	0.025	0.017	0.010	0.004	0.001	-0.003	-0.005	-0.008
	(0.003)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Autocorrelation	1	0.656	0.389	0.158	0.046	-0.104	-0.200	-0.324
Partial autocorrelation		0.547	0.264	0.118	-0.030	-0.081	-0.096	-0.113
		(0.006)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)

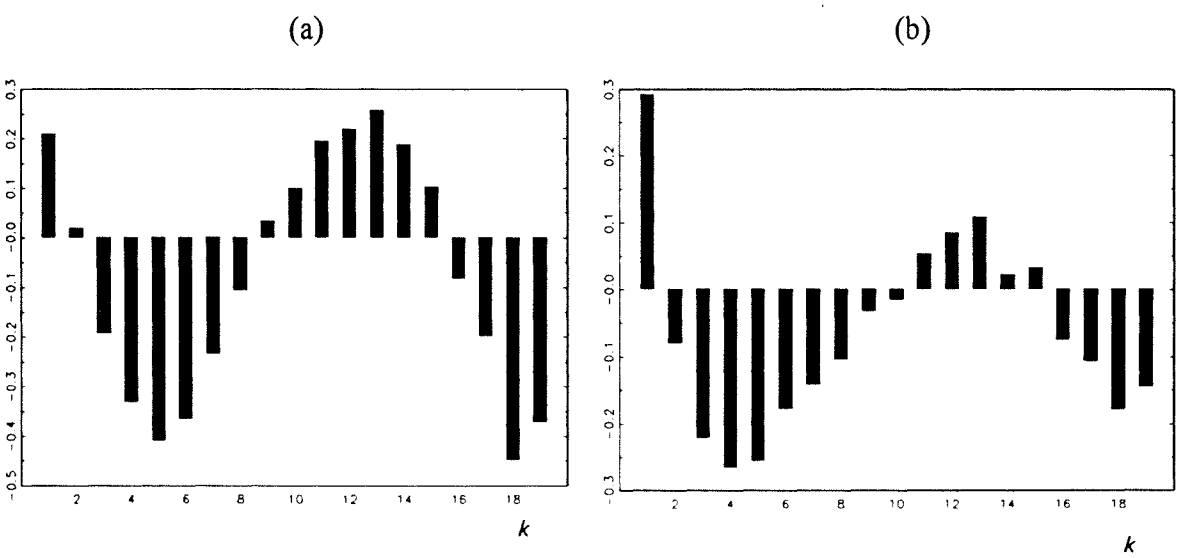


Figure 5. *Autocorrelation function (a) and partial autocorrelation function (b) for OLS residuals from separate equations.*

Table 4b. *Autocovariance, autocorrelation and partial autocorrelation. OLS residuals from separate equations (standard errors in parantheses).*

k	0	1	2	3	4	5	6	7
Autocovariance	0.015	0.003	0.0003	-0.003	-0.005	-0.006	-0.005	-0.003
	(0.002)	(0.001)	(0.0005)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Autocorrelation	1	0.209	0.018	-0.191	-0.330	-0.408	-0.363	-0.232
Partial autocorrelation		0.293	-0.079	-0.220	-0.265	-0.256	-0.178	-0.140
		(0.006)	(0.006)	(0.006)	(0.006)	(0.006)	(0.007)	(0.007)

It is very hard to interpret a common dynamic error structure, especially as there are no theoretical reasons to suspect any particular form. However, a few observations can be made from Figure 5. The oscillating pattern of the autocorrelation function suggest a higher order autoregressive process with a negative second-order coefficient. This means that a divergence from an individual trend will only lanst one year and turn into an opposite effect the next period. Thus, the same phenomenon as was found for the stochastic model with serially correlated error term. When analysing the residuals from the the model with common earnings growth rates this effect was dominated by earnings growth rate heterogeneity.

4.2.3 Earnings Growth Rate Heterogeneity

Several empirical studies have used different methods to measure the presence and extent of growth rate heterogeneity (see Baker, 1990 for a review). To measure growth rate heterogeneity in this sample, we will use the random coefficient model originally used by Lillard and Weiss (1979). In this model the error component of the human capital earningsfunction is modified to $\upsilon_{it} = \mu_i + \eta_{it} + \xi_i exp_{it}$, where $\eta_{it} = \rho\eta_{it-1} + \varepsilon_{it}$ and ξ represents the effect of omitted variables which affect the earnings growth rate. This model gives the following covariance matrix:

$$\Sigma = \Sigma^* + (\mathbf{iexp}_t')\Sigma_{\mu\xi} = \Sigma^* + \sigma_\mu^2 \mathbf{ii}' + \sigma_\xi^2 \mathbf{exp}_t \mathbf{exp}_t' + \sigma_{\mu\xi}(\mathbf{exp}_t \mathbf{i}' + \mathbf{iexp}_t'),$$

where $\Sigma^* = \sigma_\eta^2 \begin{bmatrix} 1 & \rho & \dots & \rho^{T-1} \\ \rho & 1 & & \rho^{T-2} \\ \dots & . & \dots & \\ \rho^{T-1} & \rho^{T-2} & .. & 1 \end{bmatrix}$ and **i** is a unit column vector, σ_μ^2 is the variance of individual intercepts, σ_ξ^2 is the variance of individual earnings growth rates, $\sigma_{\mu\xi}$ is covariance between individual intercepts and growth rates and σ_ε^2 is the variance of the residual.

The components of this covariance matrix are estimated using the likelihood function given in MaCurdy (1982):

$$\ell_N(\hat{\beta}; \Sigma(\omega)) = -\ln|\Sigma(\omega)| - \mathrm{tr}\left\{\Sigma^{-1}(\omega) S(\hat{\beta})\right\}$$

where $S(\hat{\beta}) = (1/N)\sum_{i=1}^{N} \hat{\upsilon}_i \hat{\upsilon}'_i$ and $\hat{\upsilon}$ is a vector of estimated residuals from the human capital earnings function. The estimates are given in Table 4.

Table 4. *Maximum Likelihood Estimates of the Parameters of the Covariance Structure.*

	(1)	(2)	(3)	(4)
ρ	–	0.7472	–	0.7423
		(0.1186)		(0.1180)
σ^2_ε	0.0408	0.0328	0.0407	0.0327
	(0.0020)	(0.0029)	(0.0020)	(0.0022)
σ^2_μ	0.3057	0.1330	0.3037	0.1399
	(0.0370)	(0.0237)	(0.0209)	(0.0122)
σ^2_ξ	0.00048	0.000025	0.00046	0.000031
	(0.00004)	(0.000016)	(0.00003)	(0.000017)
$\sigma_{\mu\xi}$	–0.0097	–0.0016	–0.0095	–0.0018
	(0.0014)	(0.0008)	(0.0015)	(0.0008)

Table 4 reveals four different sets of results. The first two columns is results obtained on residuals from the quadratic earnings function. In model (1) first order autocorrelation is ignored. Columns 3 and 4 show the corresponding results from residuals of the quartic earnings function.

Two interesting results from Table 4 should be noted. First, the estimate of the variance in individual growth rates (σ^2_ξ) is, although small, significantly differently from zero in the estimates from the model without autocorrelation and the ordinary p-values for model (2) and (4) are 0.056 and 0.027 respectively. Second, the covariance between the individual intercept and earnings growth rate, $\sigma_{\mu\xi}$, is negative and significantly different from zero.[13] As follows from the discussion above, this result is suggested by the OJT hypothesis and human capital theory.

In this context the importance of including sufficient number of time periods should be stressed. For instance, this model could be estimated on data from only two time periods. However, a large amount of random fluctuations would then be estimated as heterogenous growth rates. As the number of time periods is increased, this problem will diminish. However, it will remain to some extent unless the whole life-cycle is covered.

[13] These results could be compared to those from several other studies. Using an almost identical population for his 20-year earnings history PSID-sample, Baker (1990, 1991) obtained estimates between 0.00020 and 0.00072 for σ^2_ξ and between -0.013 and 0.0008 (insignificantly different from zero). Jäntti (1992), using the same data as Jäntti (1993), received very similar result 0.0006. However, some methodological differences should be noted. In neither of these studies has e.g. the effect of increased formal schooling been filtered out before these estimates was done.

4.2.4 Dynamic Model

Table 4 shows the estimates of the dynamic model $y_{it} = \overline{\beta}_0 + \mu_i + \gamma y_{it-1} + \beta_1 educ_{it} + \beta_2 exp_{it} + \beta_3 exp_{it}^2 + u_{it}$. The estimation methods for fixed effects dynamic models given in Andersson and Hsiao (1982, p. 70) are used. If the number of individuals and time-series observations tend to infinity, and under the assumptions that u_{it} is normally distributed and the initial observations on the dependent variable are constant, this method will yield consitent parameter estimates. However, if T is fixed they will be incosistent. An expression for the assymptotic bias is given in Andersson and Hsiao (1982). As T is fairly large in this study we will ignor this bias.

Table 4. *Coefficient estimates from a dynamic model.*

y_{t-1}	0.6533
	(0.1749)
exp	0.0130
	(0.0012)
$exp^2/100$	-0.0210
	(0.0023)
educ	0.0063
	(0.0018)
R^2	71.89

5. Predictions of Individual Earnings, Earnings Distribution and Mobility

In this section, results from two experiments are presented using the models described in the previous section. In the first, the different statistical models are estimated on 17 periods (1970-1986) and two predictions are made for each of the last three periods (1987-1989). In the second, a stochastic component is added using random numbers from a normal distribution and earnings distributions for the last three years are generated. The same random numbers are used for all models in order to eliminate differences caused by chance. To generate the autoregressive series a recursive procedure is used.

There are three properties of the predictions that we are interested in – these can be listed as follows: (1) *Accuracy of predicted individual earnings.* This property is measured using squared prediction errors. (2) *Earnings inequality for the three years taken separately and together.* These two earnings inequality concepts are measured using an ordinary Gini coefficient. (3) *Earnings mobility.* A special measure is needed for estimating mobility which is explained below. It was developed by Shorrocks (1978).

As is pointed out by Shorrocks, earnings inequality in the included years and earnings mobility between these years determine the earnings inequality of the sum of the annual earnings. Thus, the income inequality measures of these two concepts could be used to measure earnings mobility. Shorrocks income mobility measure is expressed as follows:

$$M = 1 - \frac{I[Y(t_0, t_m)]}{\sum_{k=1}^{m} w_k I[Y(t_{k-1}, t_k)]},$$

where I is an income inequality measure (in this study we will use the Gini coefficient), Y is income received during the time period defined by the subindices of t, w_k is the share of total income received in period k and m is the number of time periods (years) considered.

There are four kinds of individual heterogeneity that could be considered in the simulations which are listed below together with the ways they will be accounted for:

(1) *Individual levels*, i.e. individual intercepts in the human capital models and differences from mean earnings for the stochastic models. This kind of heterogeneity will be taken into account in all models;

(2) *Individual autoregressive coefficients.* This will be considered for the human capital models, for the serially correlated error term, as well as the stochastic models, the autoregression coefficients;

(3) *Individual variances.* Individuals are likely to differ in their year to year deviations from a common or individual trend. Therfore, we will use both individually calculated and common model variances in the simulations;

(4) *Individual earnings growth rates.* There are at least two ways to accomodate this problem in the simulations. Firstly, to estimate separate models for each individual, and secondly, to generate coefficients from the random coefficient model. We will make use of both these methods.

If one wishes to extend a data-set which originally only contains individual observations from one or very few time-periods, it is not possible to consider all these kinds of individual heterogeneity. It is thus of vital importance to know the losses incurred by not considering them.

We will compare the three different functional forms of the human capital earnings function that are considered in Section 4 and three different specifications of the error structure:

1. $y_{it} = \overline{\beta}_0 + \beta_1 educ_{it} + \beta_2 exp_{it} + \beta_3 exp_{it}^2 + \upsilon_{it}$,

2. $y_{it} = \overline{\beta}_0 + \beta_1 educ_{it} + \beta_2 exp_{it} + \beta_3 exp_{it}^2 + \beta_4 exp_{it}^3 + \upsilon_{it}$,

3. $y_{it} = \overline{\beta}_0 + \beta_1 educ_{it} + \beta_2 exp_{it} + \beta_3 exp_{it}^2 + \beta_4 exp_{it}^3 + \beta_5 exp_{it}^4 + \upsilon_{it}$,

a. $\upsilon_{it} \sim \text{iid}(0, \sigma_\upsilon)$

b. $\upsilon_{it} = \rho\upsilon_{it-1} + \varepsilon_{it}$,

c. $\upsilon_{it} = \rho_i \upsilon_{it-1} + \varepsilon_{it}$,

e. $\upsilon_{it} = \mu_i + \eta_{it} + \xi_i exp_{it}$, where $\eta_{it} = \rho\eta_{it-1} + \varepsilon_{it}$.

The 1,2 and 3 d models are human capital models estimated separately for each individual.We will consider two types of stochastic models. Firstly, we use ordinary AR(1) models with a constant:

4 a. $y_{it} = \alpha + \beta y_{it-1} + \varepsilon_{it}$,

4 b. $y_{it} = \alpha_i + \beta y_{it-1} + \varepsilon_{it}$,

4 c. $y_{it} = \alpha_i + \beta_i y_{it-1} + \varepsilon_{it}$.

Model 4a has a common constant and autoregressive component while 4b includes an individual constant and 4c includes both individual constant and autoregressive component. These models are included as a reference, since they are the simplest stochastic model possible.

Second, we consider the stochastic models presented in Section 4:

5 a. $y_{it} - \mu_t = y_{it-1} - \mu_{t-1} + u_{it}$,

5 b. $y_{it} - \mu_t = \beta(y_{it-1} - \mu_{t-1}) + u_{it}$,

5 c. $y_{it} - \mu_t = \beta_i (y_{it-1} - \mu_{t-1}) + u_{it}$,

5 d. $y_{it} - \mu_t = \beta_1 (y_{it-1} - \mu_{t-1}) - \beta_2 (y_{it-2} - \mu_{t-2}) + \varepsilon_{it}$,

All these autoregressive models are estimated using procedures recommended in Cassel and Lundquist (1990), where they are shown to be unbiased when survey sampling as well as model aspects are taken into account. 4c are estimated using OLS for each individual.

We will also consider two versions of the combined model, i.e. a human capital model with a lagged dependent variable,

6 a. $y_{it} = \overline{\beta}_0 + \gamma y_{it-1} + \beta_1 educ_{it} + \beta_2 exp_{it} + \beta_3 exp_{it}^2 + \upsilon_{it}$.

Table 5 shows the squared prediction errors from the first experiment described above. Table 6 and 7 show Gini coefficients from the predicted earnings distributions, along with values of the Shorrocks earnings mobility measure from the second experiment. The results shown in Table 6 are obtained using the same variance for all individuals, i.e. the variance of the model error term. The results shown in Table 7 are obtained using variances calculated separately for each individual.

5.1 Human Capital Models

As can be seen from Tables 5, the quadratic and quartic functional forms give approximately the same sum of squared prediction errors, while the cubic, as could be suspected from Figure 3, give inferior results.

The OLS models (1a, 2a and 3a) give very good estimates of the annual earnings distribution taken separately, but very poor predictions of individual earnings (Table 5) and earnings mobility. Earnings mobility is seriously overestimated. Thus, the OLS model will

underestimate the earnings distribution for the sum of several years. Both these aspects of the predictions are very much improved when the first order autocorrelation is considered. Nevertheless earnings mobility is still overestimated and the predictions of 1989 individual earnings are not much better than the OLS estimates.

The effect of the three kinds of individual heterogeneity mentioned above, can also be discerned in the results shown in the tables. First, differences in the variances of the error term can be studied by comparing Table 6 and 7. The predictions of earnings inequality within each year is somewhat overestimated – the difference is about 0.015 compared to the former results. However, earnings mobility is also overestimated, and by almost the same amount. Thus, the combined effect of these two errors is that inequality in total over the 3 year period is very similar to the results obtained with individual variances.

Second, the effect of considering individual autocorrelation coefficients could be studied by comparing the b and c models. The results are very similar, but earnings mobility is more severly overestimated when using individual autocorrelation coefficients. Thus, these results indicate a loss from considering these individual differences.

Third, the two methods of modelling earnings growth rate heterogeneity could be studied in the d models, where individual earnings functions have been estimated, and the e models where the estimates from the random coefficient model have been used to generate individual intercepts and coefficients for work experience. This was carried out using random numbers from a normal distribution and 1,000 replications were used to eliminate random errors. The individual earnings equations give, as can be seen in Table 5, very poor estimates of individual future earnings. It is evident that the average growth rate is a much better predictor for the individual growth rate, even if the R^2 in the sample period is higher (about 83 percent) compared to the models with common growth rates. The earnings distribution within each year and earnings mobility are overestimated using these models. In all aspects considered in this experiment they give inferior estimates compared to the model with common growth rates and individual intercepts. The random coefficient model gives increasing predictions of within-year earnings inequality. This is consistent with the on-the-job-training hypothesis. The increase is, however, overestimated by this model. The random coefficient model overestimates earnings mobility more than the model with common coefficients. This is not surprising since there is a negative covariance between the intercept and the earnings growth rate. Thus, the results indicate that earnings growth rate heterogeneity is overestimated using this sample. However, it should be stressed that the estimates, as can be seen in Table 3[14], are obtained with considerable uncertainty. The lower bound of the 95 percent confidence interval is zero variance in individual growth rates.

[14]The estimates for the entire 17 year period give almost identical results.

5.2 Stochastic Models

Two general properties of the predictions of the stochastic models should be noted. First, as can be seen in Table 5, the stochastic models generally give better predictions of individual future earnings, at least for the first predicted year. Second, the predictions using common variances generally give better estimates of earnings distributions than individual variances, i.e. the opposite relation to that seen for the human capital models.

The ordinary AR(1)-models, 4a, 4b and 4c, give comparatively good predictions of individual future earnings when the model allows for individual constants, as 4b and 4c do. However, the models' predictive power for earnings distributions decreases both when individual and common variances are used. This is a bad sign for the practical usefulness of this model. By comparing models 4b and 4c it can be seen that the predictions using a common autoregressive coefficient are superior both for predicting future individual earnings and earnings distributions, to these obtained by using individual estimates.The Gibrat model (5a) gives, as predicted by the theory, increasing predictions of within-year earnings inequality – consequently, predictions of the earnings inequality are overestimated for each year. The Kalecki model (5b) predicts decreasing within-year earnings inequality when using individual variances. These results are very much improved when using common variances. However, these predictions overestimate earnings mobility more than the human capital models – this estimation is not improved when the second order autocorrelation is taken into account.

The dynamic model (6) gives the best predictions of individual earnings. Annual earnings inequality is, however, predicted to increase and, thus, overestimated when common variances are used.

To sum up the results, the combined model was shown to be most useful in predicting future individual earnings and, if individual variances are calculated, also best in predicting future earnings distributions and mobility. An explanation to this advantage is that this model most adequately considers persistence that is likely to influence both wage setting and individual labour supply decisions. It also follows from the discussion that the main models yielded very similar and ambiguous results. If common variances is used, the stochastic model 5b gives, compared to the human capital models and the combined model, fairly stable and accurate predictions of future earnings distributions.

Table 5. *Sums of squared prediction errors of different models for annual earnings, 1987-1989.*

Human Capital Models

	1a	1b	1c	1d	1e
1987	45.91	26.31	24.24	68.35	–
1988	54.54	44.74	47.46	159.72	–
1989	48.20	41.83	42.98	272.14	–

	2a	2b	2c	2d	2e
1987	45.85	27.39	25.29	101.31	–
1988	54.92	47.25	48.79	160.42	–
1989	50.10	48.30	47.22	340.02	–

	3a	3b	3c	3d	3e
1987	45.42	26.41	24.55	71.02	–
1988	54.60	45.57	47.52	147.98	–
1989	48.51	42.97	43.19	282.45	–

Stochastic Models

	4a	4b	4c
1987	26.35	17.77	22.58
1988	61.06	41.07	42.87
1989	72.22	42.59	43.20

	5a	5b	5c	5d
1987	17.10	17.18	18.58	16.55
1988	42.43	40.36	46.36	41.78
1989	42.49	40.43	47.82	50.36

Combined Models

	6a
1987	17.67
1988	39.31
1989	38.86

Table 6. *Predictions of Gini coefficients for annual earnings distribution 1987-1989, the sum of the annual incomes and Shorrocks index for earnings mobility between th years. Common variances.*

	Year 1987	1988	1989	*Sum*	*M*
Actual earnings	0.2013	0.2070	0.2069	0.2007	0.0216
Human Capital Models					
1a	0.2163	0.2249	0.2247	0.1964	0.1150
1b	0.2091	0.2195	0.2270	0.2064	0.0559
1c	0.2089	0.2198	0.2294	0.2062	0.0598
1d	0.2670	0.2959	0.3452	0.2917	0.0401
1e	0.1826	0.1970	0.2047	0.0606	0.1831
2a	0.2301	0.2404	0.2359	0.1986	0.1570
2b	0.2180	0.2296	0.2368	0.2107	0.0771
2c	0.2039	0.2125	0.2214	0.1990	0.0642
2d	0.2703	0.3001	0.3550	0.3001	0.0430
3a	0.2146	0.2228	0.2236	0.1949	0.1152
3b	0.2079	0.2196	0.2277	0.2062	0.0561
3c	0.2063	0.2175	0.2259	0.2034	0.0610
3d	0.2501	0.3031	0.3408	0.2804	0.0410
3e	0.1740	0.1900	0.1982	0.1750	0.0663
Stochastic Models					
4a	0.1500	0.1199	0.1047	0.1126	0.0996
4b	0.2001	0.1983	0.1981	0.1920	0.0346
4c	0.1980	0.1974	0.1984	0.1907	0.0366
5a	0.2312	0.2562	0.2785	0.2455	0.0400
5b	0.2057	0.2079	0.2101	0.1946	0.0637
5c	0.2095	0.2136	0.2222	0.2026	0.0579
5d	0.1987	0.2031	0.2057	0.1903	0.0602
Combined Model					
6a	0.2084	0.2179	0.2252	0.2059	0.0520

Table 7. *Predictions of Gini coefficients for annual earnings distribution 1987-1989, the sum of the annual incomes and Shorrocks index for earnings mobility between th years. Individual variances.*

	Year 1987	1988	1989	*Sum*	*M*
Actual earnings	0.2013	0.2070	0.2069	0.2007	0.0216
Human Capital Models					
1a	0.2018	0.2170	0.2106	0.1928	0.0808
1b	0.2035	0.2094	0.2070	0.1985	0.0392
1c	0.2012	0.2066	0.2112	0.1972	0.0443
1d	0.2580	0.2902	0.3347	0.2895	0.0195
1e	–	–	–	–	–
2a	0.1978	0.2114	0.2038	0.1872	0.0838
2b	0.1970	0.2010	0.1963	0.1900	0.0410
2c	0.1956	0.1993	0.2010	0.1893	0.0464
2d	0.2702	0.3001	0.3507	0.2880	0.0202
3a	0.2004	0.2144	0.2098	0.1913	0.0810
3b	0.2025	0.2099	0.2080	0.1987	0.0391
3c	0.1990	0.2052	0.2084	0.1951	0.0447
3d	0.2588	0.3040	0.3602	0.2990	0.0273
3e	–	–	–	–	–
Stochastic Models					
4a	0.1591	0.1281	0.1094	0.1182	0.1082
4b	0.2080	0.2035	0.1936	0.1941	0.0378
4c	0.2027	0.2023	0.1970	0.1915	0.0454
5a	0.2269	0.2478	0.2479	0.2348	0.0248
5b	0.2026	0.1979	0.1815	0.1846	0.0483
5c	0.2039	0.1962	0.1943	0.1878	0.0520
5d	0.1918	0.1771	0.1777	0.1718	0.0575
Combined Model					
6a	0.2049	0.2097	0.2050	0.1993	0.0353

6. Conclusions

What conclusions can be drawn from the comparison between the human capital and stochastic models ?

The human capital models with common earnings growth rate,and in particular the specification with lagged dependent variable, yielded reasonably good estimates if heterogeneity in individual variances were allowed for. However, if a common model variance was used all human capital models overestimated future earnings inequality and a simple stochastic model made the best predictions.

The results for the random coefficient model, which indicated earnings growth rate heterogeneity negatively correlated with the individual intercept, can be seen as a support for the on-the-job-training hypothesis and, thus, the human capital theory. However, as these results yielded poor predictions, they should be interpreted with some caution since it could be an indication that this kind of heterogeneity is overestimated when short sample periods are used and therefore give misleading results when applied to an extended time period. Segmented labour market theory prescribes that individual earnings growth rates could be divided into at least two groups (see e.g. Piore, 1975). An alternative approach to model growth rate heterogeneity is to empirically divide the sample into groups with different earnings growth rates.

Furthermore,the earnings mobility is overestimated in the predictions from all models. This could possibly result from failure to consider both earnings growth rate heterogeneity and a dynamic error structure beyond first order autocorrelation.

References

Anderson, T. W. and Hsiao, C. (1982). Formulation and Estimation of Dynamic Models Using Panel Data. *Journal of Econometrics*, 18, 67-82.

Atkinson, A. B., Bourguignon, F. and Morrisson, C. (1991). Empirical Studies of Earnings Mobility. Working Paper, DELTA.

Baker, M. (1990). Essays on Unemployment Spell Duration and Life Cycle Earnings, doctoral dissertation, Department of Economics, University of Michigan.

Baker, M. (1991). Growth Rate Heterogeneity and the Covariance Structure of Life Cycle Earnings, mimeo Department of Economics, University of Toronto.

Becker, G. S. (1964) *Human Capital*. New York: National Bureau of Economic Research.

Blomquist, N. S. (1981). The Distribution of Liftime Income: A Case Study of Sweden. In *The Statics and Dynamics of Income* (N.A. Klevmarken and J.A. Lybeck (eds.)), pp. 105-133, Clevedon: Tieto.

Blomquist, N. S. (1976) *The Distribution of Lifetime Income: A Case Study of Sweden.* Ph D dissertation Princeton University.

Cassel, C. M. and Lundquist, P. (1990). Microbased Time Series Analysis: Estimating the Autocorrelation Function Using Survey Samples. Working Paper, HHS Stockholm.

Cassel, C. M. and Lundquist, P. (1991). Microbased Time Series Analysis: Optimal Prediction of Aggregated AR(1) Survey Samples. Working Paper, HHS Stockholm.

Champernowne, D. G. (1953). A Model of Income Distribution. *Economic Journal*, 63, 318-351.

Creedy, J. (1980). The New Government Pension Scheme: A Simulation Analysis. *Oxford Bulletin of Economics and Statistics*, 42, 51-64.

Creedy, J. (1982). The British State Pension: Contributions, Benefits and Indexation. *Oxford Bulletin of Economics and Statistics*, 44, 97-112.

Creedy, J. (1985). *Dynamics of Income Distribution.* Basil Blackwell, Oxford.

Erikson, R. and Åberg, R. (1984). Välfärd i Förändring: Levnadsvillkor i Sverige 1968 - 1981. Prisma, Stockholm.

Friedman, M. (1953). *Essays in Positive Economics*, Chicago, University of Chicago Press.

Hart, P. E. (1976) The Dynamics of Earnings: 1963-1973. *Economic Journal,* 86, 551-565.

Hart, P. E. (1983) The Size Mobility of Earnings. *Oxford Bulletin of Economics and Statistics*, 45, 181-194.

Hause, J. C. (1977) The Covariance Structure of Earnings and the On-the-job Training Hypothesis. *Annales of Economic and Social Measurement,* 6, 335-365.

Hause, J. C. (1980). The Fine Structure of Earnings and the On-the-job Training Hypothesis. *Econometrica,* 48, 1013-1029.

Hsiao, C. (1986). *Analysis of Panel Data.* Cambridge, Cambridge University Press.

Jochumzen, P., (1993). Evaluation of Economic Theory. Mimeo Department of Economics, University of California at Los Angeles.

Johansson, S. and Hedström, P. (1979). Jordbrukares och Företagares Inkomster och Levnadsförhållanden. Mimeo SOFI, Stockholm University.

Jäntti, M., (1992). Earnings Growth Rate Heterogeneity: Minimum Distance Estimation, paper presented at Congress on Longitudinal Perspectives on Income Inequality and Poverty, August 21-22, Åbo Akademi University, Finland.

Jäntti, M., (1993). Earnings Dynamics and Industrial Structure in Finland. Mimeo Department of Economics, Åbo Akademi University.

Kalecki (1945). On the Gibrat Distribution. *Econometrica*, 3, p. 161.

Klevmarken, A. (1992). On Ageing and Earnings: A Review. Mimeo, Department of Economics, Gothenburg University.

Lillard, L. A. (1977). Inequality: Earnings vs Human Wealth. *American Economic Review*, 62, 43-53.

Lillard, L. A. and Willis, R. J. (1978). Dynamic aspects of earning mobility. *Econometrica*, 46, 985-1012.

Lillard, L. A. and Weiss, Y. (1979). Components of Variation in Panel Earnings Data: American Scientists 1960-70. *Econometrica*, 47, 437-454.

MaCurdy, T. E. (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.

Maddala, G. S. (1987). Limited Dependent Variables Models Using Panel Data. *Journal of Human Resources*, 22, 308-338.

Mincer, J. (1962). On the Job Training: Cost, Returns and Some Implications. *Journal*

of Political Economy, 70, 50-79.

Murphy, K. M. and Welch, F. (1990). Empirical Age-Earnings Profiles, *Journal of Labour Economics*, 8, 202-29.

Piore, M. J. (1975): "Notes for a theory of labor market stratification. " In Labour market segmentation (R. C. Edwards el al., eds.). Cambridge Mass.: D. C. Heath.

Sahota, G. S. (1978) Theories of Personal Income Distribution: A Survey. *Journal of Economic Literature,* XVI, 1-55.

Shorrocks, A. (1978). Income inequality and income mobility. *Journal of Economic Theory*, 19, 376-393.