ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 20/21 — 1991

Wages and Measurement Errors

Joop HARTOG, Hans van OPHEM *

ABSTRACT. – In this paper two wage observations which relate to the same point in time will be compared and the differences analyzed. The OSA-labor market survey contains information about wages received in April 1985 as reported in April 1985 and as reported in October 1986 by the same respondents. Large differences appear to exist. This paper intends to investigate whether these differences are random, by developing and estimating a maximum likelihood model. The estimation results indicate that the differences are not random. In particular, changes in the individual's labor market position seem to be important in the explanation of the observed differences. This conclusion casts strong doubt on the quality of wage variables as reported from the memories of individuals.

Erreurs de mesure sur les salaires

RÉSUMÉ. – Dans cette article, on considère deux mesures d'un même taux de salaire, l'une contemporaine, l'autre effectuée par appel à la mémoire du salarié dix-huit mois après. On estime un modèle de la différence entre ces mesures par le maximum de vraisemblance. La partie systématique est significativement liée au niveau de qualification du salarié et à des changements de postes et de salaires. Il convient donc de se méfier de la qualité des données obtenues par appel à la mémoire des intéressés.

---

* J. HARTOG : Department of Micro-economics, Faculty of Economics, University of Amsterdam; H. van OPHEM : Department of Econometrics, Faculty of Economics, University of Amsterdam. We are grateful for comments received from the conference participants (in particular Zvi Griliches, whose comments led to equation (9)) and from two anonymous referees.

1 Introduction

The quality of the measurement of wages determines for an important part the quality of the estimations of wage equations. Systematic over- or underestimation of individually reported wages will lead to biased estimates. Measurement errors need not cause any problems if they have zero expectations.. In this paper we will investigate whether this is the case. We can do so because of the availability of a unique Dutch dataset: the OSA-labor market survey. Two wage observations, relating to the same date, are available. By comparing these reported wages and by estimating wage equations with ordinary least squares and a maximum likelihood model some insight can be gained with respect to measurement errors.

The analysis in this paper differs from the research of e.g. MELLOW and SIDER [1983], GREENBERG and HALSEY [1983] and DUNCAN and HILL [1985], because our information is deduced from repeated reports of employees. These other authors used matched employee-employer reports. Such matched reports may be considered as two indicators of the "true" measure, each with unknown errors, presumably random and independent. In our case, however, a large correlation of the errors of the observations will be expected to exist. This means that the expectation of one of the errors of the observations given the other need not equal 0 (see further).

The structure of the paper is as follows. In section 2 the two available wages will be compared and two ordinary least squares wage equations will be estimated. In the third section a simultaneous model will be developed to get a better insight in the characteristics of the measurement errors in the reported wages. The maximum likelihood estimation results are presented in section 4. Finally, section 5 will contain some conclusions.

2 The Reported Wages Compared

The total OSA-labor marked survey consists of two parts, [1] relating to 1985 and 1986. The respondents of the OSA-labor market survey 1985 all

1. At this point only two waves are available. The third wave will become available in due course. It is the intention of the OSA to obtain a new wave every one-and-a-half years. OSA is the abbreviation (in Dutch) for Organization for Strategic Labor Market Research, an active government funding agency.

come from the group of individuals of age between 16 up to 60, as far as they were not engaged in full time schooling or in the military service. This group was meant to represent the potential working force of the Netherlands. The interviews took place in April and May 1985. 2387 households were visited and 4020 interviews resulted.

For the OSA-labor market survey 1986 all households who participated in the 1985-survey were again approached. The interviews took place in August, September and October 1986. From the 4020 participants in the first survey 2775 individuals were reinterviewed. The panel character of the surveys thus applies to 68.5% of the 1985-respondents (the extension with new households to make up for attrition is not relevant for the present research).

Three categories of information can be distinguished in the OSA-labor market-1985 and -1986 surveys: about the present labor market situation of the respondents, about their past labor market experience and about their individual characteristics. The present labor market situation is documented extensively. Information is available about wages, working conditions, promotion possibilities, etc. Information about the past labor market experience consists of information about the labor market history of the respondents from 1980 up to 1985 (or 1986). For each of the occupied jobs only a limited amount of information is available: e.g. the wages received in each subsequent job are not available. This information has a retrospective nature. Information about the individual characteristics of respondents for the larger part deals with the educational background, but the usual variables marital status, age, gender, etc. are also available.

In the comparison of the reported wages it is necessary to restrict the sample to the respondents who were employees and reported their wage on both observational points of time: 1289 observations remained. The reported 1985 wage in April 1985 is the answer to the question: "What is your net wage with your present employer? Please, exclude financial compensation received because of working in shifts, financial compensation for working overtime, tips, financial compensation to cover travel expenses, financial compensation for representation costs, etc. Of course, all taxes and social premiums have to be excluded also." Furthermore, the respondents were asked whether they received this wage every week, every four weeks or every month. Combining this information with the response to the question "How many hours do you have to work according to the labor contract you have with your present employer?" a net hourly wage rate, in Dutch guilders, was calculated. This net hourly wage rate in April 1985 as calculated from the 1985-OSA labor market survey will be denoted by w855. Combining the corresponding questions concerning the wage in April 1985, of the OSA-labor market survey 1986, yields w856: the net hourly wage rate in April 1985 as reported in October 1986. The variables w855 and w856 will be the key variables of the present paper. It will be investigated whether there exist differences between these variables and if so, whether these differences are random. One would expect that the variable w855 is of a higher quality than w856 simply because individuals' memories are not perfect. However, this need not be the case, as e.g.

individuals might be more likely to cheat about their present wage than about their former wage.

Note that there may exist an endogeneity problem. [2] The number of working hours is likely to influence the net wage of individuals. On the other hand economic theory predicts that the amount of working time in general will be determined by the wage. The correct procedure would be to estimate wage equations and an hours equation simultaneously (see e. g. MOFFITT [1984]). For now this problem will be solved by considering only the subpopulation of males working between 36 and 41 hours a week of the sample. The endogeneity problem is further reduced by considering only men: social conventions "force" men to work full-time. The remaining sample has 811 observations, so 478 observations were lost.

Another problem might be that the number of working hours is reported with an error. If this error is correlated with the errors of the wage equations, better estimation results can be obtained by estimating the wage equations and hours equation simultaneously. We probably do not have to worry too much about this point because we utilize the number of working hours according to the labor contract, which is less likely to be incorrectly observed.

The simplest indicator for the quality of the reported wages can be derived by directly comparing the net wage of April 1985 reported in 1985 and in 1986. Sample characteristics for the two measurements and some derived characteristics are given in Table 1. The results suggest an upward bias in the later measurement as mean, minimum and maximum all increase, and the mean difference is positive.

TABLE 1

Some Sample Statistics of the Reported Wages (guilders per hour)

	Mean	Variance	Maximum	Minimum
w855	12.61	15.59	44.44	2.17
w856	12.97	17.30	47.22	2.94
Δw(86/85)	0.36	3.31	13.45	−12.28
log(w855)	2.49	0.08	3.79	0.77
log(w856)	2.52	0.08	3.86	1.08
Δlog(w856, 855)*	0.03	0.02	1.10	−0.63

* Δlog(w856, 855) = log(w856) − log(w855).

The differences (Δw(86/85) = w856-w855) in the reported wages are tabulated in Table 2. It indicates that only 42.7% of the employees in the sample give answers which deviate less than Dfl. 0.50. The remaining individuals give answers which are farther apart. Over 15% of all employees in the sample give answers which deviate more than 16% of the

2. As is well known neglecting endogeneity in ordinary least squares estimation will result in biased estimates. Of course, in the reduced form estimations this bias is not present.

mean wage (*i.e.* more than Dfl. 2,-). Thus, it has to be concluded that the quality of the wage variables is suspect.

TABLE 2

The Difference in the Net Hourly Wages of 1985 as Reported in 1986 and in 1985

$\Delta w(86/85)$	Percentage
Dfl. $3.00 \leqslant \Delta w(86/85)$	4.9
Dfl. $3.00 \leqslant \Delta w(86/85) <$ Dfl. 2.00	5.5
Dfl. $2.00 \leqslant \Delta w(86/85) <$ Dfl. 1.00	13.6
Dfl. $1.00 \leqslant \Delta w(86/85) <$ Dfl. 0.50	12.4
Dfl. $0.50 \leqslant \Delta w(86/85) <$ Dfl. 0.00	17.0
Dfl. $0.00 \leqslant \Delta w(86/85) <$ $-$Dfl. 0.50	25.7
$-$Dfl. $0.50 \leqslant \Delta w(86/85) <$ $-$Dfl. 1.00	9.3
$-$Dfl. $1.00 \leqslant \Delta w(86/85) <$ $-$Dfl. 2.00	6.4
$-$Dfl. $2.00 \leqslant \Delta w(86/85) <$ $-$Dfl. 3.00	2.9
$\Delta w(86/85) <$ $-$Dfl. 3.00	2.3

In the next section, a model will be developed that explicitly acknowledges the error structure of the two observations of the wage rate. Here, as a simple benchmark two OLS estimations are presented, one for each wage observation. They indicate what results would be obtained if one would only have either one of these observations available. The explanatory variables are defined in an appendix. Education is measured with dummies for the consecutive levels (which on average differ in duration by 3 years). The variable "certificate" indicates whether the individual graduated from the highest level of education attended. The estimated coefficient of "graduation" will give an average of the graduation effect over and above the years of education effects (a more detailed analysis of graduation effects is given in HARTOG [1983]).

Table 3 contains the results of the OLS-estimation on the logarithm of the net hourly wage rates in April 1985.

From Table 3 it can be concluded that the wage equations conform closely to the results reported in the empirical literature. [3] Education and age determine for a large part the wage of the employees. The coefficients of the dummy variables relating to educational level have the correct order. The age of the individual has a positive influence on the wage but at a diminishing rate. The significance of the variable "certificate" is out of line with the human capital theory literature, but finds an explanation in the signalling theory. Employees from a non-Dutch origin do not seem to earn significantly less than Dutch employees. This conforms closely to the human capital theory and is at odds with the segmented labor market

3. Given the focus on the error structure, our attention to the specific results will be quite modest.

TABLE 3

OLS-Estimates of the Wages Equations

Dependent variable Variable name	log(w855)		log(w856)	
	coeff.	(t-value)	coeff.	(t-value)
Constant	0.636	(5.063)*	0.816	(6.519)*
Educational level 3	0.056	(1.695)	0.086	(2.607)*
Educational level 4	0.156	(4.433)*	0.192	(5.463)*
Educational level 5	0.287	(7.490)*	0.321	(8.382)*
Educational level 6	0.342	(6.190)*	0.400	(7.270)*
Age	0.076	(11.712)*	0.067	(10.363)*
Age^2 (*10e-3)	−0.801	(−10.528)*	−0.709	(−9.352)*
Tenure (*10e-2)	0.188	(1.585)	0.142	(1.200)
Labor experience (*10e-2)	−0.163	(−0.725)	−0.003	(0.013)
Certificate	0.109	(5.524)*	0.096	(4.887)*
Foreigner	−0.071	(−1.668)	−0.051	(−1.205)
Civil servant	−0.006	(−0.396)	−0.003	(−0.210)
Residual sum of squares	37.094		36.880	
Adjusted R-square	0.418		0.400	
Standard deviation error	0.215	(40.274)*	0.215	(40.274)*

*=significant at 2.5%, 10e-2=coefficient has to be multiplied by 10e-2, 10e-3=coefficient has to be multiplied by 10e-3.

theory. Note the insignificance of the variables tenure, labor market experience and civil servant. The former two variables are apparently redundant if age is included.

The comparison of the coefficients estimated in the case of w855 and w856 reveals that for the education and age variables the coefficients in the last case are absolutely larger than in the 1985-case. All the other variables exhibit smaller absolute values. So, there seems to exist both systematic overestimation and underestimation. The quality of the ordinary least squares estimations is only slightly better for the w855-case.

3 An Empirical Model for Analyzing the Differences between w855 and w856

Suppose, the logarithm of the net hourly wage of April 1985 rate as reported in April 1985 is determined by the following relation:

$$\log(w855) = X'\beta + \varepsilon_1. \tag{1}$$

Furthermore, assume that ε_1 is normally distributed with expectation 0 and variance σ_1. ε_1 is simply an error. Assume that the log of the net hourly wage rate of April 1985 as reported in October 1986 is determined by the following relation:

$$\log(w856) = X'\beta + \tau + \varepsilon_2. \tag{2}$$

Again, ε_2 is simply an error with a normal distribution and expectation 0 and variance σ_2. τ will be named the "recall error": this error in the wage equation results from the time lag between the two interviews. Assume that τ is normally distributed with variance σ_3. The aim of the analysis is to investigate whether the expectation of τ is 0. The results of the previous section suggest that it is not. A simple specification for the "recall error" is:

$$\tau = X'\gamma + Z'\alpha + \nu \tag{3}$$

(3) assumes that the recall error is determined by the factors influencing the net hourly wage and some variables which do not have an impact on wages. Furthermore, a non-observed stochastic error term is added. It is assumed that ν is normally distributed with variance σ_3. Furthermore, if the vector X contains a constant, the expectation of ν can be assumed 0.

Substituting (3) in (2) yields:

$$\log(w856) = X'\beta + X'\gamma + Z'\alpha + \varepsilon_2^*, \tag{4}$$

where $\varepsilon_2^* = \varepsilon_2 + \nu$ has a normal distribution with expectation 0 and variance $\sigma_2^* = \sigma_2 + \sigma_3 + 2\sqrt{\sigma_2}\sqrt{\sigma_3\rho_{23}}$ and ρ_{23} is the correlation of ε_2 and ν. Estimating equations (1) and (4) simultaneously with maximum likelihood methods yields estimates of: β, γ, α, σ_1, σ_2^* and ρ_{12}^*, where ρ_{12}^* is the correlation of ε_1 and ε_2^*. This last correlation can be written as:

$$\rho_{12}^* = \operatorname{cov}(\varepsilon_1, \varepsilon_2^*)/\sqrt{\sigma_1}\sqrt{\sigma_2^*} = [\operatorname{cov}(\varepsilon_1, \varepsilon_2) + \operatorname{cov}(\varepsilon_1, \nu)]/\sqrt{\sigma_1}\sqrt{\sigma_2^*}. \tag{5}$$

The main interest of the present study is whether γ and α differ significantly from 0 and to uncover the stochastic error structure. Under the above stated assumptions the error structure $(\varepsilon_1, \varepsilon_2, \nu)$ is completely characterized by $E\varepsilon_1 = 0$, $E\varepsilon_2 = 0$, $E\nu = 0$, σ_1, σ_2, σ_3, ρ_{12}, ρ_{23} and ρ_{13}.[4] Only three estimates are available (the errors from (1) and (4) and the correlation between them), so three restrictions on the last six parameters have to be imposed to identify the structural error structure. Given the available data, it is not possible to estimate all six freely. The following three assumptions are proposed:

$$\begin{cases} \sigma_1 = \sigma_2 \\ \rho_{13} = 0 \\ \rho_{23} = 0. \end{cases} \tag{6}$$

4. ρ_{12} is the correlation of ε_1 and ε_2, ρ_{23} is the correlation of ε_2 and ν and ρ_{13} is the correlation of ε_1 and ν.

The first assumption states that the errors ε_1 and ε_2 of the wage equations are drawn from normal distributions with the same expectation and variance. The second assumption states that there exists no correlation between the error of the wage equation with respect to w855 and the recall error. The third assumption comes to assuming that the recall error is independent of the error ε_2. An alternative to this assumption is assuming $\rho_{12}=1$. This assumption states that the recall error is purely additional in the sense that the errors ε_1 and ε_2 are exactly equal. We opted for the assumption $\rho_{23}=0$, because it appears to be the most plausible.

Due to the assumption of normally distributed error terms of (1) and (4) the following relation holds:

$$(7)\quad E(\varepsilon_1 \mid \varepsilon_2^*)=\rho_{12}^*\left(\sqrt{\sigma_1}/\sqrt{\sigma_2^*}\right)\varepsilon_2^*=\rho_{12}^*\left(\sqrt{\sigma_1}/\sqrt{\sigma_2^*}\right)\times(\log(w856)-X'\beta-X'\gamma-Z'\alpha).$$

This means that given a positive error for the wage equation (4), the sign of the error of wage equation (1) depends upon the sign of the correlation coefficient ρ_{12}^*. We expect a positive correlation coefficient as in that case a positive error in the one wage equation is likely to be accompanied by a positive error term in the other wage equation.

The estimation results of the simultaneous maximum likelihood model constituted by equations (1) and (4) will be presented and discussed in the next section. In the vector Z, with explanatory variables for recall errors only, we use some variables measuring changes in the employee's situation occurring after April 1985:

– Job change: a dummy variable equal to 1 if the employee has changed employer between April 1985 and October 1986. One would expect that a job change, which is usually accompanied by a wage change, will disturb the recall of the wage received in April 1985, but, of course, will not affect the wage received in April 1985.

– Function change: a dummy variable equal to 1 if the employee has changed function although remaining with the same employer between April 1985 and October 1986. Again, such a change will usually be accompanied by a wage change and this may reduce the reliability of the wage observation in 1986 (w856).

– Wage change: the difference of the logarithm of the net hourly wage received in October 1986 and the logarithm of the net hourly wage rate received (and reported) in April 1985. A large difference can be expected to result in a poorer recall of the wage of April 1985 in October 1986.

The inclusion of the variable "wage change" in the vector Z is apt to result in significantly better estimation results. This is due to the following observaton. Denote the coefficient of the variable "wage change" by α_1. (4) is equal to:

$$(8)\quad \log(w856)=X'\beta+X'\gamma+Z_0'\alpha_0+\alpha_1[\log(w86)-\log(w855)]+\varepsilon_2^*$$

where Z_0 is a vector equal to Z excluding the variable "wage change" and α_0 the corresponding vector of coefficients. From (1),

$$\log(w855) - \varepsilon_1 = X'\beta,$$

so we get:

$$(9) \quad \log(w856) = (1-\alpha_1)\log(w855) + \alpha_1 \log(w86) + X'\gamma + Z_0'\alpha_0 + \varepsilon_2^{**}.$$

The dependent wage variable is simply some weighted average of the logarithm of the wages w855 and w856 plus some other variables and an error term $\varepsilon_2^{**} = \varepsilon_2^* - \varepsilon_1$. Note from (3), that the case that the wage observations relating to 1985 only differ in a random fashion is represented by: $\alpha_1 = 0$, $\gamma = 0$ and $\alpha_0 = 0$. α_1 can be interpreted as a measure of (the mean) quality of the memory of the respondents over a period of 1.5 years. If $\alpha_1 \approx 0$, respondents have good memories, if $\alpha_1 \approx 1$ memories are seriously distorted in the direction of the present wage. [5]

4 Estimation Results [6]

Table 4 contains the simultaneous maximum likelihood estimates of equations (1) and (4). The first column of Table 4 gives the maximum likelihood estimates of the model discussed in the previous section including the three variables expected to have influence on recall errors but not on wages introduced in the previous section. The results indicate that a large number of the wage-determining variables are strongly significant. No surprises are encountered. The results for β are almost completely equivalent to the results presented in Table 3. In the vector γ, repeating the variables from the wage equation for the recall error a large number of variables is significant. Only the constant and the variables tenure, labor experience and civil servant are insignificant. The positive signs of the coefficients of the dummies for education level indicate that the employees with higher education levels than level 2 tend to overestimate their wages of April 1985 in October 1986. [7] This overestimation is more significant

5. Note that in order to get consistent estimates, it has to be assumed that the variables suggested for the vector Z are independent of the error term in equation (4). In particular, this requires that the observed wage growth is independent of ε_2^*. This is a strong assumption. Further modelling (of w86) and additional data (the next panel wave) could improve on this.

6. The maximum likelihood estimates are obtained by making use of GRMAX (RIDDER [1984]).

7. Recall that it is not certain which of the wages w855 and w856 is more reliable. The overestimation of w856 compared to w855 might well be an underestimation of w855 compared to w856.

the higher the educational level of the individual. The overestimation is reduced if the employees have a certificate at the highest level of the education they attended or work a lot of hours. Older individuals tend to give a smaller overestimation than younger individuals. The positive sign (and the significance) of the square of the age of the individual indicates that the overestimation decreases strongly with age. From the large number of significant coefficients it can be concluded that recall errors are not purely random. This conclusion is strengthened by the estimation results with respect to the nonwage determining variables listed in the vector Z. Two out of three variables are significant. Employees who changed jobs between April 1985 and October 1986 appear to underestimate their wages of April 1985 in October 1986, whereas employees who change function (without changing employer) appear to neither under- or overestimate their wage of April 1985 in October 1986. A very strong positive impact is found for the variable wage change. Employees who (claim to) have experienced a large wage increase during the period April 1985 to October 1986 tend to overestimate their wages of April 1985 in October 1986. The estimates with respect to the parameters of the model suggest that there exists a strong correlation between the error term of the equations used in the simultaneous maximum likelihood estimation model. The variance of the error term of the wage equation with respect to w855 (σ_1) is slightly larger than the variance of the error term of the wage equation with respect to w856 (σ_2^*). Of course this is a consequence of the fact that more explanatory variables were utilized in the estimation of the second wage equation (w856). The fact that the correlation of both error terms (ρ_{12}^*) differs significantly from 1 (as well as from 0) indicates that the errors of the wage equations (1) and (4) are not totally equivalent.

The second column of Table 4 gives the estimation results under the hypothesis that $\alpha = 0$. Judging from the likelihood-ratio statistic[8] ($LR = 380.60$) this hypothesis is strongly rejected. Note that a smaller number of variables now significantly influence the recall error. Furthermore note that the estimated coefficients of the wage equation do not change at all by the imposition of $\alpha = 0$. Finally note that the estimated correlation between the error terms of both wage equations is larger in the case of the estimation under the restriction $\alpha = 0$ than in the estimation without that restriction. The estimated coefficient of the variable "wage change" is both significantly different from 0 and from 1. A coefficient close to 0.5 indicates that w856 is reported as the mean of w855 and w856. Therefore, the memories of the respondents are far from perfect.

All in all it can be concluded, because of the fact that the hypothesis $\alpha = 0$ was rejected and the significance of some elements of the vector γ, that recall errors are not purely random. Clearly, in particular the changes in the individual's situation affect his recall error systematically.

In order to uncover the error structure we now impose assumptions (6). Doing so by making use of the results presented in the

8. The null hypothesis is a model with only two constants as explanatory variables.

TABLE 4

ML-Estimates of the Wages Equations

Dependent variable Variable name	log(w855) and log(w856)	
	coeff. (t-value)	coeff. (t-value)
β:		
Constant	0.635 (5.095)*	0.635 (5.095)*
Educational level 3	0.056 (1.714)	0.056 (1.714)
Educational level 4	0.156 (4.468)*	0.156 (4.468)*
Educational level 5	0.287 (7.539)*	0.287 (7.539)*
Educational level 6	0.342 (6.234)*	0.342 (6.234)*
Age	0.076 (11.791)*	0.076 (11.791)*
Age^2 (*10e-3)	−0.801 (−10.560)*	−0.801 (−10.560)*
Tenure (*10e-2)	0.187 (1.587)	0.187 (1.587)
Labor experience (*10e-2)	−0.163 (−0.729)	−0.163 (−0.729)
Certificate	0.109 (5.580)*	0.109 (5.580)*
Foreigner	−0.072 (−1.684)	−0.072 (−1.684)
Civil servant	−0.006 (−0.373)	−0.006 (−0.373)
γ:		
Constant	0.051 (0.560)	0.180 (2.403)*
Educational level 3	0.181 (7.300)*	0.030 (1.519)
Educational level 4	0.259 (9.475)*	0.036 (1.704)
Educational level 5	0.310 (10.204)*	0.024 (1.489)
Educational level 6	0.369 (8.746)*	0.059 (1.792)
Age	−0.038 (−7.862)*	−0.009 (−2.302)*
Age^2 (*10e−3)	0.396 (7.012)*	0.090 (1.991)*
Tenure (*10e-2)	−0.120 (−1.398)	−0.043 (−0.616)
Labor experience (*10e-2)	0.225 (1.390)	0.170 (1.266)
Certificate	−0.062 (−4.333)*	−0.013 (−1.093)*
Foreigner	0.051 (1.650)	0.019 (0.757)
Civil servant	−0.011 (−0.956)	0.003 (0.272)
α:		
Job change	−0.039 (−3.353)*	–
Function change	0.001 (0.105)	–
Wage change	0.472 (21.853)*	–
Parameters model:		
σ_1*	0.046 (20.137)*	0.046 (20.137)*
σ_2	0.018 (15.916)*	0.046 (20.137)*
ρ_{12}*	0.692 (29.680)*	0.819 (79.759)*
Loglikelihood	842.191	651.893
LR-ratio statistic	859.886	479.290

*=significant at 2.5%, 10e-2=coefficient has to be multiplied by 10e-2, 10e-3=coefficient has to be multiplied by 10e-3.

first column of Table 4 yields the following estimates relevant for the underlying error structure of the model: $\sigma_2 = 0.046$ and $\sigma_3 = -0.027$. The impossibility of a negative variance forces us to conclude that assumptions (6) can not be correct. In any case ρ_{23} has to be smaller than 0.

Finally, from (7), some information can be obtained about whether there exists a relationship between the signs of the error terms of wage equations (1) and (4). Our expectation of a positive relationship is strongly supported: the estimated correlation coefficient is strongly significant and closer to 1 than 0.

An additional indication about the importance of introducing the variables of the vector Z in the explanatory variables of wage equation (4) can be obtained by subtracting (1) from (4) and carrying out two ordinary least squares estimations. Because of the assumption of normally distributed error terms this is a legitimate procedure. The following equation is estimated:

$$(10) \quad \Delta\log(w856, w855) = \log(w856) - \log(w855) = X'\gamma + Z'\alpha + \nu + \varepsilon_2 - \varepsilon_1.$$

Table 5 contains the ordinary least squares estimates of this equation. For the first column Z is part of the explanatory variables and for the second column it is not (assumption $\alpha = 0$). The most important conclusion with respect to the comparison of these two columns of Table 5 is the large drop in the adjusted R^2 from 0.407 to 0.00019 by introducing the assumption $\alpha = 0$.[9] This means that the variables of the vector Z, which represent changes in the individual's labor market position are very important, and in any case more important than the individual characteristics listed in the vector X, for the explanation of the difference between w856 and

TABLE 5

OLS-Estimates of the Wage Difference Equation

Dependent variable Variable name	Δlog (w856, w855) coeff. (t-value)	 coeff. (t-value)
γ:		
Constant	0.013 (0.221)	0.180 (2.394)*
Educational level 3	0.015 (0.974)	0.030 (1.506)
Educational level 4	0.004 (0.229)	0.036 (1.689)
Educational level 5	0.011 (0.622)	0.033 (1.446)
Educational level 6	0.008 (0.317)	0.058 (1.763)
Age	−0.002 (−0.623)	−0.009 (−2.295)*
Age² (*10e-3)	0.014 (0.407)	0.009 (2.004)*
Tenure (*10e-2)	−0.031 (−0.558)	−0.460 (−0.646)
Labor experience (*10e-2)	0.144 (1.383)	−0.166 (−1.229)
Certificate	−0.011 (−1.196)	−0.013 (−1.101)
Foreigner	0.018 (0.909)	0.020 (0.776)
Civil servant	−0.015 (1.956)*	0.003 (0.311)
α:		
Job change	−0.038 (−3.226)*	–
Function change	−0.001 (−0.090)	–
Wage change	0.502 (23.359)*	–
Residual sum of squares	7.897	13.373
Adjusted R-square	0.407	0.00019
Standard deviation error	0.100 (40.274)*	0.129 (40.274)*

*=significant at 2.5%, 10e-2=coefficient has to be multiplied by 10e-2, 10e-3=coefficient has to be multiplied by 10e-3.

9. The inclusion of the variable wage change is likely to be the most important factor in the better estimation results.

w855. Another conclusion is that by taking account of the simultaneity in the estimations the number of significant variables strongly increases.

6 Conclusion

The most important conclusion with respect to the analysis carried out in the previous sections is that the differences between the two observed wages, which relate to the same period of time, do differ systematically. In particular, changes in the labor market position of employees, listed in the vector Z, seem to be important, but other factors are important as well. The estimated coefficient of the variable wage change (close to 0.5) indicates that the memories of individuals are far from perfect even for a relatively short period of time (1.5 years). This result indicates that estimations on wages from the recollection of individuals are likely to give rise to biased estimates. Furthermore, it is unlikely that only the quality of the wage variable is suspect, other variables may be reported with serious errors as well.

By choosing the subsample of male employees working between 36 and 41 hours a week the problem of the endogeneity of the working hours was avoided. An obvious step for future research is to investigate whether it is possible to include all observations in the estimations. In that case a working hours equations has to be estimated as well. MOFFITT [1984] offers a model which may be valuable in this respect. Another direction in which the analysis can be taken is estimating an adjusted model on the third wave of the OSA-labor market survey. This wave will become available in due course.

APPENDIX

The following explanatory variables were used in estimations of the wage equations:

– a constant;

– educational level. The standard educational classification (Dutch Central Bureau of Statistics) distinguishes: 1 = elementary schooling, 2 = extended elementary schooling, 3 = extended general or vocational education (e.g. MAVO), 4 = intermediate general or vocational education (e.g. Atheneum, MBO), 5 = higher general or vocational education (e.g. HBO), 6 = university. The levels are indicated with four dummy variables, with level 1 or 2 as reference category;

– age: the age of the employee in years;

– age^2: the square of age;

– tenure: the number of years the employee is employed by his present employer;

– labor experience: the number of years the employee reports to have been employed;

– certificate: a dummy variable equal to 1 if the employee has a certificate of the last educational institution he visited;

– foreigner: a dummy variable equal to 1 if the employer is from a non-Dutch origin;

– civil service: a dummy variable equal to 1 if the employee is employed as a civil servant.

• References

DUNCAN, G. J. and HILL, D. H. (1985). – "An Investigation of the Extent and Consequences of Measurement Error in Labor-Economic Survey Data", *Journal of Labor Economics*, vol. 3, pp. 508-532.

GREENBERG, D. and HALSEY, H. (1983). – "Systematic Misreporting and Effects of Income Maintenance Experiments on Work Effort: Evidence from the Seattle Denver Experiment", *Journal of Labor Economics*, vol. 1, pp. 380-407.

HARTOG, J. (1983). – "To Graduate or Not: Does it Matter?", *Economics Letters*, vol. 12, pp. 193-199.

MELLOW, W. and SIDER, H. (1983). – "Accuracy of Response in Labor Market Surveys: Evidence and Implications", *Journal of Labor Economics*, vol. 1, pp. 331-344.

MOFFITT, R. (1984). – "The Estimation of a Joint Wage-Hours Labor Supply Model", *Journal of Labor Economics*, vol. 2, pp. 331-344.

RIDDER, G. (1984). – *Handleiding bij GRMAX*, SEO, Amsterdam.