Discussion Paper No. 11-013

Give Missings a Chance

Combined Stochastic and Rule-based Approach to Improve Regression Models with Mismeasured Monotonic Covariates Without Side Information

Stephan Dlugosz

ZEW

Zentrum für Europäische Wirtschaftsforschung GmbH

Centre for European Economic Research

Discussion Paper No. 11-013

Give Missings a Chance

Combined Stochastic and Rule-based Approach to Improve Regression Models with Mismeasured Monotonic Covariates Without Side Information

Stephan Dlugosz

Non Technical Summary

The education variable in the IAB datasets suffers from problems like missing and misclassified values. The data problems do not occur randomly, but are highly associated with other variables from the dataset. This issue has become more and more important over the last years and it severely influences empirical findings.

The education variable should represent a person's highest formal degree. People can only attain degrees over time but not lose them. This property of the education variable imposes restrictions leading to inconsistencies. Recently, this problem has been addressed in the literature by using correction rules.

The newly developed procedure utilises these rules to identify misclassified values and replace them with missing values. We derive a new estimator for this kind of data correction based on a EM-based estimator for incomplete data. The estimation results of this new procedure are unbiased and consistent under the classical MAR assumption.

We apply this new estimator to a set of Mincer-type wage regression for the years 1993–2003 separately in order to observe changes in the impact of the educational degrees on the wage. These coefficient estimates clearly show that the quality of education is more important than the number of years of education. We also find a rising wage differential between the different educational degrees over time. This indicates that the educational expansion of this decade does not exceed the request for high-skilled workers. Thus, we did not find any evidence that would suggest an inflation of formal education or a devaluation of degrees.

Das Wichtigste in Kürze

Die Bildungsvariable in den IAB Datensätzen enthält viele fehlende und fehlerhafte Angaben. Diese Datenfehler stehen in Abhängigkeit zu anderen Variablen in den Datensätzen. Diese Datenproblematik hat über die Zeit an Quantität gewonnen. Dadurch können Analyseergebnisse, insbesondere auf Basis der letzten Jahre, stark verzerrt sein.

Die Bildungsvariable stellt den höchsten bisher jemals erreichten formalen Bildungsabschluss einer Person dar. Unter Zuhilfenahme dieser Rahmenbedingungen wurden in der Literatur verschiedene Korrekturmöglichkeiten inkonsistenter Zeitreihen vorgeschlagen.

Unter Verwendung dieser Korrekturregeln wird eine neue Methode entwickelt, um fehlklassifizierte Daten zu identifizieren und aus dem Datensatz zu entfernen. Ausgehend von einem erwartungsmaximierenden (EM) Schätzer für unvollständige Daten wird ein neuer Schätzer für diese Art der Datenkorrektur hergeleitet. Schätzer, welche mit dieser neuen Technik gewonnen werden, sind unverzerrt und konsistent unter der klassischen Annahme des bedingt-zufälligen Fehlens von Werten (MAR).

Dieser neue Schätzer wird genutzt, um mit Hilfe getrennter Lohnregressionen nach Mincer für die Jahre 1993 bis 2003 Änderungen in den Lohnwirkungen der Bildungsabschlüsse zu messen. Die Ergebnisse zeigen, dass die Zahl der Bildungsjahre ein schlechter Proxy für die Messung von Bildungsrenditen ist, da es mehr auf die Qualität als auf die Quantität der Bildung ankommt. Zudem wird ein steigendes Lohndifferential zwischen den verschiedenen Bildungsabschlüssen gefunden. Damit wird die These unterstützt, dass die Bildungsexpansion dieser Dekade den Bedarf an hochqualifizierten Arbeitnehmern nicht überschritt und somit per se nicht von einer Bildungsinflation mit einer Abwertung der Bildungsabschlüsse gesprochen werden kann.

Give Missings a Chance

Combined stochastic and rule-based approach to improve regression models with mismeasured monotonic covariates without side information*

Stephan Dlugosz†

Register data are known for their large sample size and good data quality. The measurement accuracy of variables highly depends on their high importance for administrative processes. The education variable in the IAB employment sub-sample is an example for information that is gathered without a clear purpose. It therefore severely suffers from missing values and misclassifications. In this paper, a classical approach to deal with incomplete data is used in combination with rule-based plausibility checks for misclassification to improve the quality of the variable. The developed correction procedure is applied to simple Mincer-type wage regressions. The procedure reveals that the quality of years in education is very important: The German labour market rewards general education less than vocational training. Furthermore, using this method, no indication of an inflation in formal education degrees could be found.

Keywords: measurement error, EM by the method of weights, wage regression, expansion of educational degrees, misclassification, imputation rules

JEL: C13, J24, J31

* I thank Bernd Fitzenberger and Verena Niepel for helpful discussions and comments. This work was supported by the German Research Foundation through the *Statistical Modelling of Errors in Administrative Labour Market Data* grant

† ZEW Centre for European Economic Research, L7 1, 68161 Mannheim, Germany, E-mail: dlugosz@zew.de

1. Introduction

Administrative individual data have gained growing importance for empirical economics and policy evaluation because of their high reliability. In contrast to survey-based data, nonresponse, missing and mismeasured values are not a big issue. At first glance, sample selection in register data is determined by clearly defined rules (i.e. job seekers). Any mismeasurement is due to errors during coding and occur totally at random. While this is true for certain information within the register data (i.e. wages), this does not hold for information that has been collected additionally and bears no (more) relevance for any administrative tasks. One example for such a piece of information is the education variable in the IAB (Institute of Employment Research) datasets (e.g. IABS (Bender et al., 2000), IEBS (Jacobebbinghaus and Seth, 2007)). The very divergent quality of register data is been known for a quite long time (cf. e.g. Cramer, 1985; Schmähl and Fachinger, 1994).

While research on data problems in survey data has a long history (cf. e.g. Rubin, 1987), the specific issues in administrative data have only recently gained attention (see Johansson and Skedinger (2009) for Swedish data and Fitzenberger et al. (2006), as well as Wichert and Wilke (2010) for the case of the IAB data from Germany).

The education variable is not completely ordered and is increasing monotonically over time. These and other rules of the system of formal education in Germany imply restrictions on the education variable. The proposed method explicitly uses the constraints deduced from these restrictions. In this paper, these constraints are combined with an incomplete data algorithm (Ibrahim, 1990) to improve the quality of the education variable in a maximum likelihood estimation framework. The introduced idea may also be used for other time-monotone discrete variables.

This paper focuses on the education variable as the most important factor for measuring returns on education and skill biases in labour market outcomes. These are two very important issues (cf. e.g. Card, 1999; Katz and Autor, 1999), which require a reliable measure of formal education.

The new methodology is compared with other (simple) methods for missing variables. The simplest way to deal with missing values is case-deletion (Little and Rubin, 2002). Unfortunately, this reduces the number of observations. Furthermore, if the data are not missing totally at random, the parameter estimates will be biased. Introducing an additional dummy variable indicating that the value for the current observation is missing keeps the number of observations. In this approach, "missing" represents a category of its own. Unfortunately, the estimates for correlated covariates are biased if the data are

not missing totally at random (Jones, 1996).

A two stage procedure is another classical approach (Little and Rubin, 2002), which is closely related to propensity score methods. First, a prediction model for the non-missing data generation process is estimated ('propensity score'). The resulting probabilities are used as weights for the target analysis at the second stage. Unfortunately, this requires specification and estimation of an auxiliary model, reducing the efficiency of the resulting estimator.

More sophisticated methods make use of either external validation data or—in the case of multiple-imputation—additional (model) assumptions regarding the missing data process (cf. e.g. Little and Rubin, 2002). Therefore, these methods are not considered as contenders here.

Wage estimations, which take misclassification explicitly into account, require conditional (mean) independence of wages and measurement error given the (unobserved) true education (Kane et al., 1999). Fitzenberger et al. (2006) have shown that this is not the case for the IABS data. This and other potential alternatives suggested in the literature are not applicable here.

This paper intends to find indicators for changes in the return on education in the years 1993 to 2003. During these years, substantial changes in the wage structure and the technological background took place (Antonczyk et al., 2010). Additionally, a larger part of the work force gained access to higher education. The natural question is, whether the changes in technology—favouring high-skilled labour—dominate or the greater availability of formal education.

In the first case, we would observe a higher wage differential for higher skilled workers. In the second case, we would see an inflation of educational degrees resulting in lower wage differentials.

We will make use of a set of separately estimated simple Mincer wage regressions (Mincer, 1974). To be precise, we regress the log-wage on education, age, age squared, work experience and the additional control variables: foreign, industries and task groups. Other authors use the panel dimension explicitly (Gebel and Pfeiffer, 2010). Thanks to the large dataset, we can estimate separate models for the different years, which is a less restrictive modelling. We have to admit that we do not want to contribute to the discussion on returns to education or skill premia (cf. e.g. Boudarbat et al., 2010).

Our goal is to show the differences that arise from the correction of the formal education variable, rather than measure the return on education itsef.

2. The Education Variable in the IAB Employment Subsample (IABS)

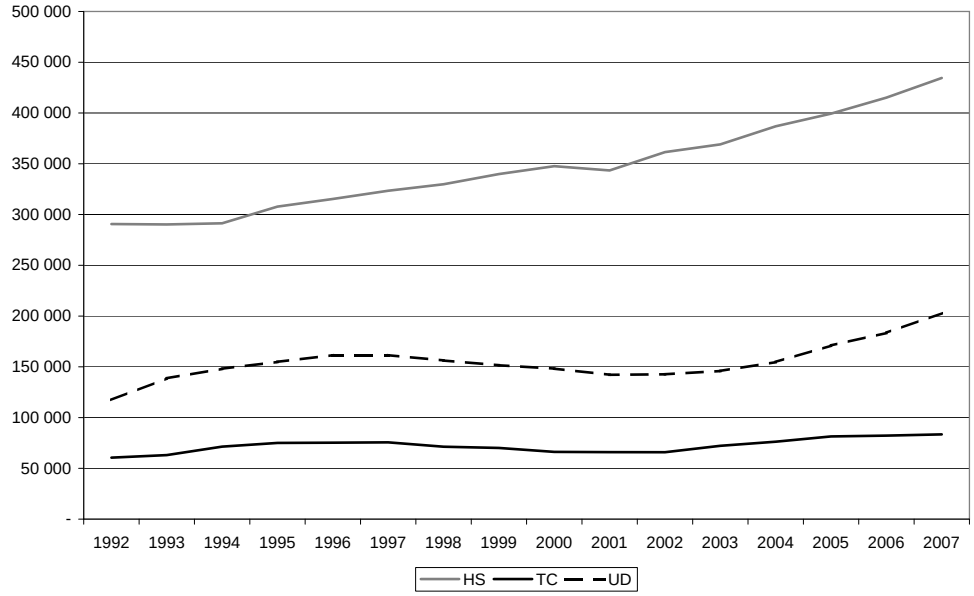
We use the IABS version for the period 1975–2004 released with detailed regional information (see Drews, 2008, for a detailed description). Our method is relevant for all versions of the IABS. The data contain daily register data for 1 360 949 individuals in Germany on employment spells and on spells with transfer payments from the Federal Labour Office. The IABS is a representative 2% sample of employment. After the end of the year and with the termination of an employment, employers have to report earnings and other socio-demographic information about their employees, such as educational degree. The earnings information and the length of the employment spells are used to calculate contributions to and benefits from the social insurance system and, hence, are very reliable. Periods of self-employment and employment as life-time civil servants, which are not subject to (mandatory) social insurance, are not included in the data.

The education information has to be reported with every employment spell but it bears no relevance for the social security system. Reporting the employee's education incorrectly has no consequences (Fitzenberger et al., 2006). This explains why the education variable in the IABS is less reliable compared to information on earnings or the beginning and ending of spells. Other spells than employer reported information—like transfer payments or other technical spells reporting gaps in the employment history—do not provide additional information on the education variable. Thus, we restrict our analysis to the information given in employment spells. There are about 24 936 176 spells in the sample, 20 644 256—i.e. 82.8%—of these are are employment spells. The reporting system did not change since it had been introduced in 1973 for West Germany. The same system was implemented for the eastern parts of Germany in 1991 after the reunification. Thus, inconsistencies in the education variable over time can only be attributed to employer's unreliability (Fitzenberger et al., 2006).

The education variable in the IABS is a partially ordered categorical variable describing the current formal educational status of a worker as it is reported to the employer (usually at the beginning of the working contract). There are 6 possible values: "no degree" (ND), "high school degree" (HS), "vocational training" (VT), "high school degree and vocational training" (HSVT), "technical college degree" (TC), "university degree" (U). The two categories VT and HS are not ordered.

During the last years, there has been a steady expansion of educational degrees (cf. figure 1). At the labour market, however, we can only observe a slight increase in university degrees, but none in technical college degrees. In contrast to figure 1, we

Figure 1: Graduates over time



Source: Federal Statistical Office

even do not observe a significant increase in high school degrees (cf. figure 2). This is not surprising, because workers in Germany are supposed to have additional education beyond a general schooling degree; either a vocational training or a higher educational degree. High school graduates (without additional education) are restricted to certain auxiliary tasks or short-time employment.

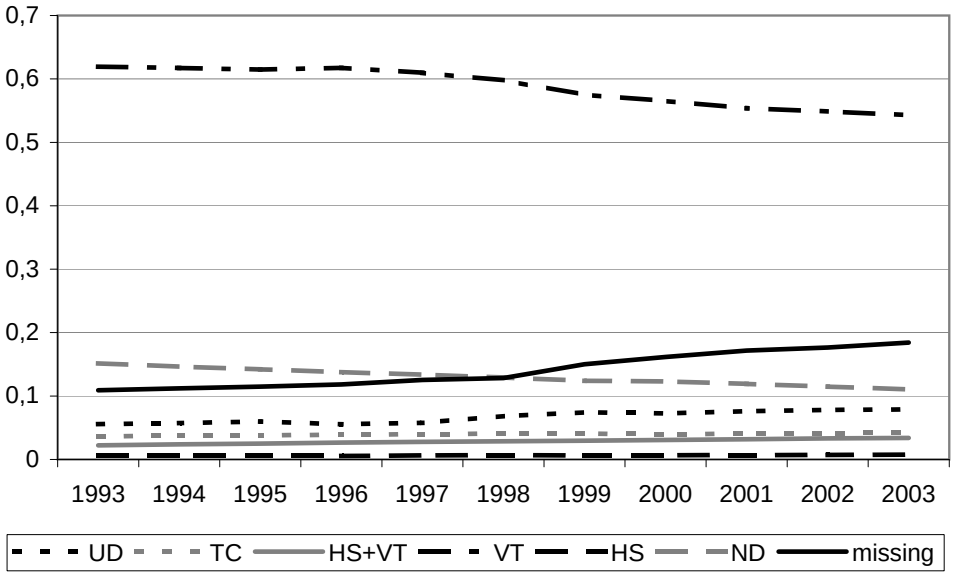
The education variable suffers seriously from missing (Fitzenberger et al., 2006) and misclassified values (Wichert and Wilke, 2010). Moreover, the fraction of missing values has increased over the last years (cf. figure 2) and we have to expect that the employer's reports are losing on reliability. Thus, it is quite reasonable to expect an increasing fraction of misclassified values.

2.1. Rules of Consistency

The education variable represents formal educational degrees. This implies, that there are some rules that restrict the number of plausible values for this variable.

MR The most prominent rule is the monotonicity rule: In Germany, it is impossible to lose your formal educational degree. The variable is partially ordered, i.e. UD $>$ TC $>$ HSVT $>$ HS or VT $>$ ND. HS and VT are not ordered. These formal

Figure 2: Formal educational degrees over time



Source: IABS, own calculations
(jobs with social security protection, weighted by employment duration within year)

educational degrees are stable, i.e. once you have reached a certain level, you will never fall back to a lower level.

Other rules can be obtained from age and time constraints:

TR Nobody is able to obtain any degree without investment of time. For example, studying at an university requires a high school degree. It is impossible to get your high school degree and obtain a university degree one year later. Therefore, we can establish some time rules:

- at least 2 years for vocational training degrees
- at least 2 years for high school degrees
- at least 2 years for university degrees
- at least 1 year for technical college degrees

Usually, it takes much longer to obtain one of the above mentioned degrees. There are combined education paths, where you can get a VT degree and a bachelor (usually TC) after three years of education, for example.

AR Age rules are partly related to the time rules. They describe the minimum age of a person for having a certain degree:

- no vocational training degree under 18
- no high school degree under 16
- no vocational training + high school degree under 18
- no university or technical college degree under 19

These rules are not very stringent, but they are suited to correct (absolutely) implausible values without introducing any overcorrection bias.

Except for the age rules, all rules rely on the panel dimension of the data.

2.2. Rules of plausibility

In addition to these rules for the consistency of the education variable, Fitzenberger et al. (2006) introduced plausibility rules that rely on knowledge of the data generating process. They combined different assumptions for establishing sets of rules with increasing complexity.

IP1 Information from all employment spells are valid. Higher degrees are forward extrapolated over time without restrictions. This should result in an overestimation of the true educational degree.

IP2a Information from all employment spells is considered. Higher degrees are forward extrapolated over time only if the degree has been reported at least three times. This is slightly more conservative than IP1.

IP2b Information from all employment spells is considered. If a sequence is inconsistent with MR, the inconsistency has to be reported three times before this educational information is backward extrapolated.

IP3 Inconsistencies are corrected like in IP2b, but the decision of the correct imputation is related to the reliability of the employer. An employer is reliable, if he changes the educational variable only once for a certain employee. If he reports a too high degree once and corrects himself, this over-reported value is replaced.

These rule sets are used to decide, whether a specific entry of the education variable is plausible. If the imputation strategy decides to replace the value, it is regarded as missing for our analysis. Thus a high number of replacements produces a high proportion of missing values.

3. (Generalised) Linear Models for Misclassified and Missing Education Information

We are in need of a methodology that can deal with both data problems at the same time: missing values and misclassifications. Furthermore, we are interested in a methodology that does not rely on any kind of validation data.

In a series of papers, Ibrahim and co-authors introduced parametric models to deal with the missing data problem (Ibrahim, 1990; Ibrahim et al., 1999). The basic model is:

$$\sum_{i=1}^{n} \sum_{\boldsymbol{x}_{mis,i}} w_{i,o} l(\boldsymbol{\theta}; \boldsymbol{x}_i, y_i) \ , \tag{1}$$

with complete data likelihood $l(\cdot) = l_{y_i|x_i}(\boldsymbol{\beta}, \phi) + l_{x_i}(\boldsymbol{\gamma})$, covariates $\boldsymbol{x}_i = (\boldsymbol{x}_{mis,i}, \boldsymbol{x}_{obs,i})$, parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \phi, \boldsymbol{\gamma})$, and weight updates

$$w_{i,o} = p(\mathbf{x}_{mis,i}|\boldsymbol{x}_{obs,i}, y_i, \boldsymbol{\theta}^{(o)}) = \frac{p(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}^{(o)}) p(\boldsymbol{x}_i|\boldsymbol{\theta}^{(o)})}{\sum_{\boldsymbol{x}_{mis,i}} p(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}^{(o)}) p(\boldsymbol{x}_i|\boldsymbol{\theta}^{(o)})} \ . \tag{2}$$

Model (1) is designed for a single cross-section dataset. People are changing their occupations not only at the end of a year and therefore, we need to adjust the Ibrahim-Model for repeated observations.

Let T_i denote the number of observations for person i. Note that T_i might vary over the persons, i.e. we have an unbalanced panel. Now, we have to look at trajectories, thus we observe not only a set of variations on one observation, but a set of variations on a group of observations over time $t \in \{1, \ldots, T\}$:

$$\sum_{i=1}^{n} \sum_{\boldsymbol{x}_{mis,i}} w_{i,o} l(\boldsymbol{\theta}; \boldsymbol{x}_{it}, y_{it}) \ , \tag{3}$$

with complete data likelihood $l(\cdot) = l_{y_{it}|x_{it}}(\boldsymbol{\beta}, \phi) + l_{x_{it}}(\boldsymbol{\gamma})$, covariates $\boldsymbol{x}_{it} = (\boldsymbol{x}_{mis,it}, \boldsymbol{x}_{obs,it})$, parameters $\boldsymbol{\theta} = (\boldsymbol{\beta}, \phi, \boldsymbol{\gamma})$. Calculation of weights is slightly more complicated, as we have to incorporate some additional assumptions:

i) independence of errors: $p\big((y_{it})_t|(\boldsymbol{x}_{it})_t\big) = \prod_t p\big(y_{it}|(\boldsymbol{x}_{it})_t\big)$

ii) irrelevance of covariates from other points in time 1: $p\big(y_{it}|(\boldsymbol{x}_{it})_t\big) = p\big(y_{it}|\boldsymbol{x}_{it}\big)$

iii) missing probabilities are (conditionally) independent from history (covariates carry full information): $p\big((\boldsymbol{x}_{mis,it})_t|(\boldsymbol{x}_{obs,it})_t; \boldsymbol{\theta}^{(o)}\big) = \prod_t p\big(\boldsymbol{x}_{mis,it}|(\boldsymbol{x}_{obs,it})_t; \boldsymbol{\theta}^{(o)}\big)$

iv) irrelevance of covariates from other points in time 2: $p\big(\boldsymbol{x}_{mis,it}|(\boldsymbol{x}_{obs,it})_t;\boldsymbol{\theta}^{(o)}\big) = p(\boldsymbol{x}_{mis,it}|\boldsymbol{x}_{obs,it};\boldsymbol{\theta}^{(o)})$

We have to discuss assumption iii): The probability for a certain value of $\boldsymbol{x}_{mis,it}$ depends naturally on $\boldsymbol{x}_{mis,i(t-1)}$ for each $t \in \{2,\dots,T\}$ through the hierarchy of formal educational degrees. This, as well as time-constraints are taken into account by excluding impossible trajectories. Other possible sources of variation like timing of events and duration of having a certain degree are ignored. Theoretically, one could include these pieces of information, but we think that the additional gain in precision is too low to justify the effort. With the assumptions i) to iv), we get the weights:

$$w_{i,o} = p\big((\mathbf{x}_{mis,it})_t|(\boldsymbol{x}_{obs,it},y_{it})_t,\boldsymbol{\theta}^{(o)}\big) = \frac{\prod_t p(y_{it}|\boldsymbol{x}_{it},\boldsymbol{\theta}^{(o)})p(\boldsymbol{x}_{it}|\boldsymbol{\theta}^{(o)})}{\sum_{\boldsymbol{x}_{mis,it}}\prod_t p(y_{it}|\boldsymbol{x}_{it},\boldsymbol{\theta}^{(o)})p(\boldsymbol{x}_{it}|\boldsymbol{\theta}^{(o)})} \tag{4}$$

Note, we still assume that the errors are independent for all observations (assumption i)). Consider using a panel data model like fixed effects or random effects otherwise. There is now just a single weight for the whole trajectory.

In many applications the MAR assumption holds for a set of variables but only a certain subset of these variables is used as covariates in the empirical model. Let $\mathcal{V} = (\mathcal{X},\mathcal{Z},\mathcal{Y})$ be the set of variables necessary for the MAR assumption to hold. $\mathcal{X}$ is the subset of $\mathcal{V}$ that is used in our empirical model and the remaining variables $\mathcal{Z}$ are independent from the regressant $\mathcal{Y}$. Again, we note $\boldsymbol{z} = (\boldsymbol{z}_{mis,i},\boldsymbol{z}_{obs,i})$ for the observed and the missing part of the data. Then, model (1) (analog for model (3)) is:

$$\sum_{i=1}^{n}\sum_{(\boldsymbol{x}_{mis,i},\boldsymbol{z}_{mis,i})_i} w_{i,o}\,l(\boldsymbol{\theta};\boldsymbol{x}_i,y_i)\ , \tag{5}$$

with complete data likelihood $l(\cdot) = l_{y_i|x_i}(\boldsymbol{\beta},\phi) + l_{x_i}(\boldsymbol{\gamma})$ and weights

$$\begin{aligned} w_{i,o} &= p(\mathbf{x}_{mis,i}|(\boldsymbol{x}_{obs,i},\boldsymbol{z}_{obs,i})_i,y_i,\boldsymbol{\theta}^{(o)}) \\ &= \frac{p(y_i|\boldsymbol{x}_i,\boldsymbol{\theta}^{(o)})p\big((\boldsymbol{x}_i,\boldsymbol{z}_i)|\boldsymbol{\theta}^{(o)}\big)}{\sum_{(\boldsymbol{x}_{mis,i},\boldsymbol{z}_{mis,i})_i} p(y_i|\boldsymbol{x}_i,\boldsymbol{\theta}^{(o)})p\big((\boldsymbol{x}_i,\boldsymbol{z}_i)|\boldsymbol{\theta}^{(o)}\big)}\ . \end{aligned} \tag{6}$$

We have to assume that y is independent from $\boldsymbol{z}$, i.e. $p(y_i|\boldsymbol{x}_i,\boldsymbol{\theta}^{(o)}) = p(y_i|\boldsymbol{x}_i,\boldsymbol{z}_i,\boldsymbol{\theta}^{(o)})$ to obtain result (6).

This model can deal with missing values in a repeated observations setting. To include a solution to the misclassifications problem, we have to transfer the problem of misclassification to the problem of missing values. A missing value is a special kind of misclassi-

fication because we know that a value is misclassified. The basic idea of this paper is to find a way to detect misclassifications and convert them into missing values. Then, we are able to apply model (3) to the revised data.[1] Therefore, we need a concept on misclassification analogue to the concept of missing values as presented in Rubin (1976). We have to assume, that misclassification occurs at random, i.e. $X_m \neq X_m* \perp\!\!\!\perp X_m|X_c$. To be exact, we will replace possibly misclassified data with missing values. This procedure does not harm the MAR assumption for all ('old' and 'new') missing values.

3.1. Estimation via Expectation Maximisation

Ibrahim (1990) has introduced a method he called "EM by the method of weights", which he elaborated for the class of generalised linear models (McCullagh and Nelder, 1989). This method provides consistent point estimates for the model parameters under the missing-at-random (MAR, (Rubin, 1976)) assumption. Ibrahim et al. (1999) have generalised the methodology to situations, where the missing value generating process is not random. The developed "EM by the method of weights" algorithm can also be formulated for general ML estimation problems (Ibrahim et al., 2005). In this paper the Tobit model (Tobin, 1958) is used.

This algorithm works in three steps that can be performed by standard software quite easily:

init Generate an extended dataset, where all observations with missing values are replicated and all possible values for the missing values are imputed. The replicates are equally weighted such that the weights for each set of replicates belonging to a certain missing value sum to one. The sum of the weights over all observations remains n.

M estimate the model (1), (3) or (5), i.e. optimise the likelihood, with the (re")""weighted extended data set. Usually you can use the weighted version of the ML estimation procedure for your preferred parametric model (e.g. GLM, Tobit).

E estimate the probabilities for the specific patterns using equation (2), (4) or (6) and use it as the new weight.

iter iterate E and M until convergence

During the generation of the missing-patterns, we can incorporate restrictions on the education variable to restrict the set of possible patterns to consistent patterns. For

[1] The major shortcoming of this approach is the loss of information which takes place by removing possibly misclassified values from the data. This problem is discussed in section 6 in more detail.

example, if the preceding value to a missing value is 3 and the subsequent value is 4, only two different values would generate a consistent series with the monotonicity rule: 3 and 4. In other words, the weights for other alternatives are set to zero in advance.

Asymptotic (Co-)Variances Ibrahim (1990) derives a closed form for the standard errors of the EM-estimates in the case of a GLM with canonical link function:

$$\mathcal{I}(\hat{\theta}) = \frac{1}{\phi}\boldsymbol{X}'\boldsymbol{W}\boldsymbol{M}\boldsymbol{V}\boldsymbol{X} - \frac{1}{\phi^2}\boldsymbol{X}'\boldsymbol{W}\boldsymbol{M}^2\boldsymbol{H}^2(I-\boldsymbol{W})\boldsymbol{X} \ ,$$

where $\boldsymbol{X}$ denotes the augmented design matrix, ϕ is the variance parameter of the chosen link function for the GLM, $\boldsymbol{W}$ are the estimated weights, and $\boldsymbol{M}$ is the diagonal matrix of weights. $\boldsymbol{H} = diag(y_i - \mu_i)$ is the estimated linear error from that last step and

$$\boldsymbol{V} = diag\left(\frac{d^2b(\delta_i)}{d\delta_i^2}\right)$$

is the Fisher information from the last iteration of the scoring algorithm (see McCullagh and Nelder, 1989, for more detailed descriptions).

Unfortunately, the Tobit model (Tobin, 1958) does not belong to this class of regression models. We have to calculate the standard errors with the following, slightly more complex formula (Louis, 1982):

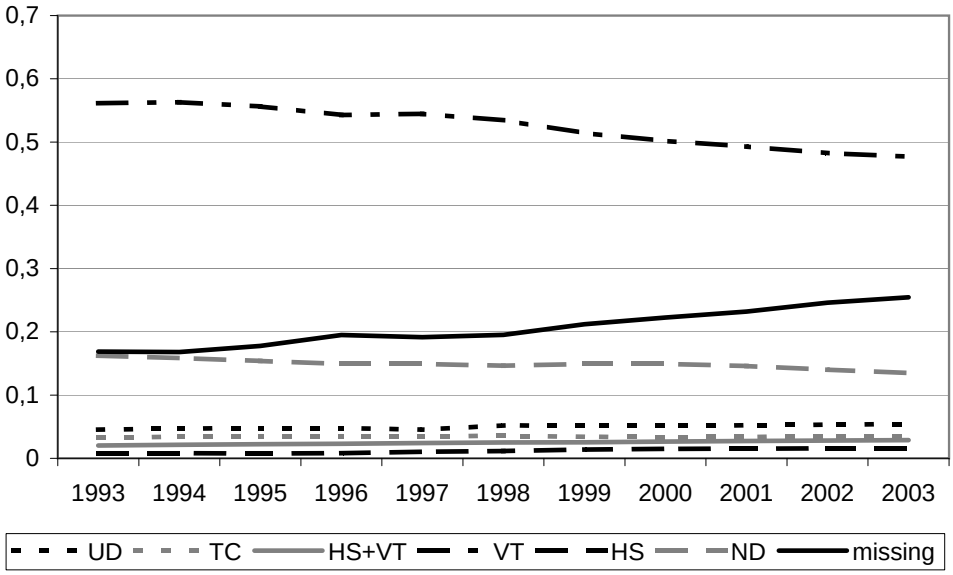
$$\begin{aligned}\mathcal{I}(\hat{\theta}) &= CoVar^{-1}(\hat{\theta}) \\ &\quad - \sum_{i=1}^{n}\left\{\sum_{x_{mis,i}(j)} \hat{w}_{ij}S_i(x_i,y_i,\hat{\theta})S_i(x_i,y_i,\hat{\theta})'\right\} \\ &\quad + \sum_{i=1}^{n}\left\{\sum_{x_{mis,i}(j)} \hat{w}_{ij}S_i(x_i,y_i,\hat{\theta})\right\}\left\{\sum_{x_{mis,i}(j)} \hat{w}_{ij}S_i(x_i,y_i,\hat{\theta})\right\}'\end{aligned}$$

where S denotes the score function. The estimates for the asymptotic covariance matrix of $\boldsymbol{\beta}$ is the upper $p \times p$ block of $(\mathcal{I})^{-1}$.

4. Data Analysis

We use version 4 of the IABS employment subsample data with detailed regional information for the years 1975–2004 (IABS-R04 Drews, 2008). This is a representative 2% sample of all employees that are subject to the social security system in Germany. As already mentioned, we restrict the information in the IABS to employer reported employ-

Figure 3: Values of the Education Variable



Source: IABS, own calculations

ment spells. Males and females are hard to compare due to their different labour market participation patterns and certain wage differences. We therefore restrict our analysis to males. This sample consists of 11 749 334 spells for 738 417 individuals.

The distribution of the education variable differs significantly between the sample and this sub-sample (compare Figures 2 and 3). The main difference is the higher proportion of missing values and a lower proportion of vocational training. The proportions of the other values remain roughly the same.

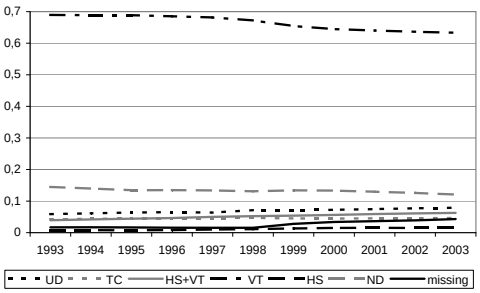
Figure 4 shows the distribution of the education variable for the different imputation procedures.[2] The different imputation procedures IP1 to IP3 reduce the number of missing values significantly in favour of vocational training (cf. Figure 3). Deleting values that are possible misclassified—indicated by diverging values in the original variable and its corrected version—generates a quite high proportion of missing values.

For the wage regression, we have to further select our sample. The labour market of the Eastern part of Germany has been quite different from its western counterpart, especially during the phase after the reunification. To obtain valid results for the whole period of time, only spells from the western part are considered. We further restrict our sample to spells that represent fulltime work and have positive wages[3]. The sample

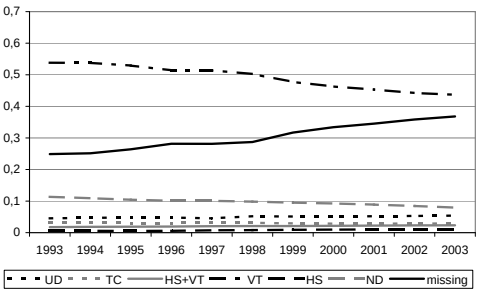
[2] We did not obey the [TR] rules to simplify calculations.

[3] Employment spells show a wage of zero if employment is suspended (e.g. maternity leaves, more

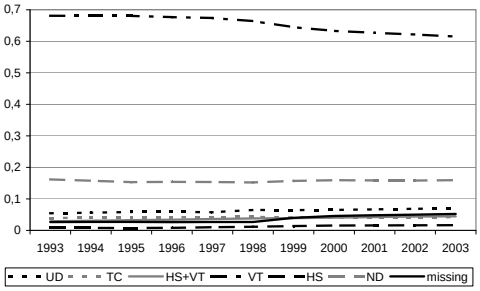
Figure 4: Values of the Education Variable after Correction



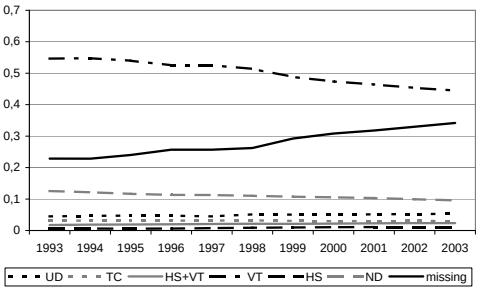
(a) IP1



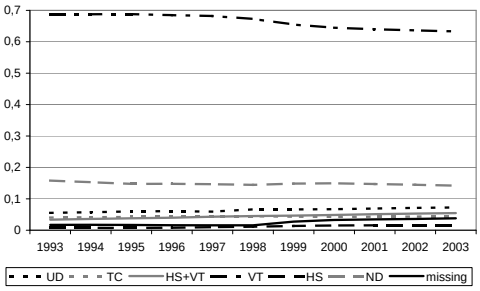
(b) IP1 EM



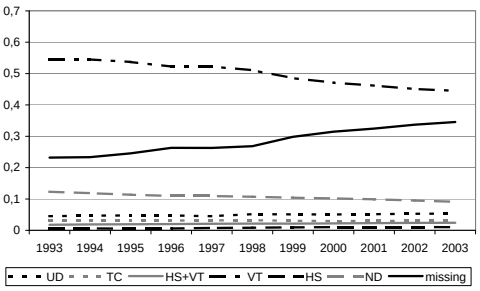
(c) IP2



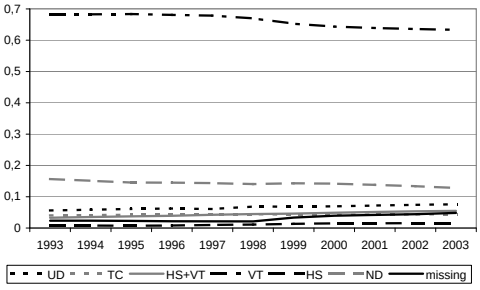
(d) IP2 EM



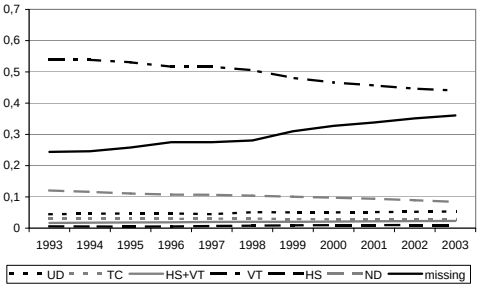
(e) IP2B



(f) IP2B EM



(g) IP3



(h) IP3 EM

Source: IABS, own calculations

consists of 3 485 195 employment spells for 415 284 persons; nearly equally distributed over the whole time period of 11 years.

We calculate Mincer-type wage regressions for log daily wages $\ln W$ for the years 1993 to 2003. For each year y, we estimate the model (Mincer, 1974):

$$\ln W_{iy} = \boldsymbol{X}_{iy}\boldsymbol{\alpha}_y + \boldsymbol{E}_{iy}\boldsymbol{\beta}_y + \epsilon_{iy} \ , \tag{7}$$

where $\boldsymbol{X}_{iy}$ denotes a design matrix for covariates like age, work experience, industry, occupation and nationality; $\boldsymbol{E}_{iy}$ is a matrix of dummy variables for the education variable and (in some cases) further dummies for missing value indication, over- or underreporting (cf. Table 6). We are interested in the estimates for $\boldsymbol{\beta}_y$. The other variables serve as controls.

The used observations are employer reported employment spells, i.e. there could be more than one observation per person for a certain year. Thus, some of the error terms ϵ_{iy} are possibly correlated and we have to take care of that. Wages in the IABS are left and right censored due to regulations for the social security system. Therefore OLS estimates would be biased and we have to use interval regression, also known as Tobit model (Tobin, 1958) which is asymptotically consistent (Amemiya, 1973). This also implies to assume Gaussian error distributions for ϵ_{iy}.

When estimating (7), there are two conceptual sources of bias. First, if individuals with high absolute earnings capacity acquire more education and earn higher wages, the education $\boldsymbol{E}_{iy}$ will be positively correlated with ϵ_{iy} Griliches (1977). This ability bias adds an upward bias to the estimates of the education coefficients (Behrman and Rosenzweig, 1999). Second, a bias can exist if individuals differ in their relative earnings capacity and act upon their comparative advantage when choosing their level of education (Willis and Rosen, 1979). If return on education is homogenous, the latter bias would be absent. It is well known for Germany—and we will provide additional evidence—that return on education is not homogenous, if they are measured in years in education (Gebel and Pfeiffer, 2010).

There are a lot of different approaches in the literature to reduce these biases. You could use an instrumental variable approach to estimate the local average treatment effect (see Ichino and Winter-Ebmer, 1999; Pischke and Wachter, 2008, for examples on return to years in education). Wooldridge's (2004) conditional mean independence approach and Garen's (1984) control function approach have been used to estimate the average partial effect (Gebel and Pfeiffer, 2010).

extended sick leaves).

However, a bias-free estimate of the return on education is not our main concern. We want to show, how the calculations are affected by missing and misclassified values. For a valid interpretation of the results of our application, we have to assume a constant bias over time. This is quite plausible as the bias stems from such unobservable personal characteristics as intrinsic motivation, which are generally stable over time.

The propensity-score like weights for the alternative modelling of missing values are estimated with help of a binary generalised linear model (McCullagh and Nelder, 1989):

$$\mathbb{E}(Y = 1|\boldsymbol{X}) = F\{\boldsymbol{X}\boldsymbol{\gamma}\} \ , \tag{8}$$

where $\boldsymbol{X}$ denotes the design matrix for the covariates log of wage, foreign, occupation, industry, age, age squared, work experience and spell length in years (see Appendix for coefficient estimates).

The panel design implies correlated error terms for spells of a certain person. Usually, the number of spells for a person is quite low; we mostly observe a single spell per year and person. Thus, special panel data models like Fixed Effects or Random Effects models are not applicable (Wooldridge, 2002). Therefore, we use a pooled estimation with the robust Huber-White estimator of the variance-covariance matrix

$$V(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}'\boldsymbol{X})^{-1} \left(\sum_i \hat{u}_i^2 \boldsymbol{x}_i' \boldsymbol{x}_i \right) (\boldsymbol{X}'\boldsymbol{X})^{-1}$$

to calculate the standard errors (Huber, 1967; White, 1980).

Fitzenberger et al. (2006) have shown that the missing value procedure highly depends on the length of the reported spells. Therefore, we use the model (5) for our estimations, where we included a discretised version of the spell length (0–30 days, 31–181 days, 182–366 days).

5. Results

5.1. Detailed Analysis of the Wage Regression Estimates for 1999

We first applied the EM algorithm on the wage regression for 1999. The estimated weights $w_{i,o}$, which represent the a-posteriori probability or a missing to have a certain real value, have a special distribution. The value zero dominates, but also one and a half. This shows that there is a unique optimal imputation for many cases. In other cases, there are obviously two good imputation candidates, which produce more or less the same results.

Table 1: Contingency tables of reported and imputed values for the education variable for the different imputation strategies

IP1		reported education					
imputed values	missing	ND	VT	HS	HS+VT	TC	UD
ND	13.36	56.92	0.05	0.41	0.03	0.00	0.01
VT	48.91	24.63	93.83	3.03	0.08	0.02	0.01
HS	8.32	2.31	0.12	63.83	0.08	0.00	0.01
HS+VT	7.40	5.92	1.23	9.05	83.58	0.00	0.00
TC	7.93	5.89	1.52	6.49	2.76	86.87	0.01
UD	14.07	4.32	3.25	17.19	13.47	13.11	99.95

(a) IP1

IP2		reported education					
imputed values	missing	ND	VT	HS	HS+VT	TC	UD
ND	16.39	68.01	0.36	4.43	0.47	0.14	0.09
VT	49.28	22.11	95.85	4.28	1.88	0.70	0.37
HS	8.29	1.85	0.16	70.14	0.63	0.08	0.12
HS+VT	5.91	3.20	0.83	6.06	84.64	0.28	0.16
TC	7.46	2.98	1.01	4.68	2.94	90.26	0.25
UD	12.67	1.86	1.78	10.41	9.44	8.54	98.99

(b) IP2

IP2b		reported education					
imputed values	missing	ND	VT	HS	HS+VT	TC	UD
ND	14.66	64.41	0.19	2.28	0.24	0.09	0.05
VT	49.90	23.36	95.40	3.63	1.33	0.43	0.29
HS	8.36	2.00	0.14	67.84	0.43	0.07	0.08
HS+VT	6.63	4.09	0.98	7.73	85.24	0.21	0.14
TC	7.51	3.80	1.17	5.68	2.82	89.98	0.20
UD	12.93	2.34	2.11	12.84	9.94	9.22	99.24

(c) IP2B

IP3		reported education					
imputed values	missing	ND	VT	HS	HS+VT	TC	UD
ND	13.83	61.97	0.21	1.54	0.34	0.14	0.09
VT	49.32	22.76	94.97	3.30	1.43	0.47	0.27
HS	8.35	2.20	0.14	68.10	0.29	0.07	0.07
HS+VT	7.19	4.90	0.92	7.52	82.46	0.10	0.13
TC	7.76	4.75	1.18	5.49	3.03	87.39	0.15
UD	13.55	3.43	2.59	14.04	12.45	11.84	99.28

(d) IP3

Source: IABS, own calculations

Table 2: Mincer-type wage regressions (rule based correction (RB), combined correction (EM)) for 1999

	IP1		IP2		IP2B		IP3	
	RB	EM	RB	EM	RB	EM	RB	EM
ND	-0.106***	-0.135***	-0.106***	-0.132***	-0.104***	-0.133***	-0.106***	-0.132***
VT	-0.012***	-0.585***	0.019***	-0.505***	-0.008***	-0.540***	0.005***	-0.566***
HS	0.161***	0.137***	0.179***	0.154***	0.156***	0.148***	0.166***	0.146***
TC	0.317***	0.290***	0.324***	0.308***	0.319***	0.302***	0.318***	0.297***
UD	0.489***	0.515***	0.504***	0.538***	0.499***	0.531***	0.494***	0.527***

Source: IABS, own calculations

reference: vocational training, *** p<0.01, ** p<0.05, * p<0.1

Next, we looked at the contingency tables showing the frequencies of the reported and the imputed values of the education variable (see Table 1). The values are not weighted for the spell length, but are a result of the EM procedure and thus weighted by the corresponding $w_{i,o}$.

The IP1-based version corrects the education variable to higher values and therefore prefers over-reported values. All imputation procedures retain most of the observations. IP2 is the most conservative strategy. Here the least replacements of all strategies took place. IP2b has nearly the same rate of value replacements, but acts more like IP1, i.e. it avoids downgrading. IP3 has nearly the same replacement rate as IP1, but with a downgrading rate which lies between IP2 and IP2b.

Table 2 shows the results for a Mincer-type wage regression for 1999.

The major difference between the two sets of estimates arises from the coefficients for "high school degree". In Germany, both HS and VT (the reference) represent 13 years of education. These 13 years are not comparable, as HS represents general education whilst VT represents 10 years in general education and 3 years in special training for a specific job. This close attachment to a specific job is honoured by the German labour market through higher payments. This difference is absent without proper treatment of the missing values and misclassifications.

By this line of reasoning, the difference between VT and HS is not surprising, but rather the fact that having no degree at all seems to be more valuable than having a high school degree. Most of the employments of people with HS degrees are usually either short termed, auxiliary jobs or vocational training agreements with relatively low wages. In other words, people with a HS degree are still studying for a university or technical college degree or an additional vocational training. Employees with no formal degree—if not unemployed—often have other qualifications that the few categories of formal education do not encompass.

Table 3: Blinder-Oaxaca decomposition of high school graduates versus the (vocational) trained workers

version	wage differential	education	characteristics	coefficients
IP1	-0.893	-1.118	-0.302	0.526
			-0.791	1.015
IP2	-0.776	-1.439	-0.263	0.925
			-0.657	1.320
IP2B	-0.826	-1.281	-0.275	0.730
			-0.731	1.185
IP3	-0.859	-1.289	-0.281	0.711
			-0.741	1.172

Source: IABS, own calculations

Nevertheless, it is surprising that workers without any degree perform much better. We computed a Blinder-Oaxaca-type decomposition (Blinder, 1973; Oaxaca, 1973):

$$\begin{aligned} \ln \bar{W}_{HS} - \ln \bar{W}_{VT} &= (\beta_{0,HS} - \beta_{0,VT}) + (\bar{\mathbf{x}}_{HS} - \bar{\mathbf{x}}_{VT})\boldsymbol{\beta}_{VT} + (\boldsymbol{\beta}_{HS} - \boldsymbol{\beta}_{VT})\bar{\mathbf{x}}_{HS} \\ &= (\beta_{0,HS} - \beta_{0,VT}) + (\bar{\mathbf{x}}_{HS} - \bar{\mathbf{x}}_{VT})\boldsymbol{\beta}_{HS} + (\boldsymbol{\beta}_{HS} - \boldsymbol{\beta}_{VT})\bar{\mathbf{x}}_{VT} \end{aligned} \tag{9}$$

The decomposition is not unique. Both versions are shown in equation (9). As the group of VT is much larger than the HS group, we think that version 2 is more representative than version 1.

Table 3 shows the results of the decomposition between high school graduates and (vocational) trained workers. The raw log-daily-wage differential is about -0.6 log-points. It is predominantly driven by the personal and job characteristics as well as by the degree discrimination. The coefficient effects reduce the difference to about the same degree as the education discrimination. This indicates that there are major differences in the personal or job characteristics between the groups.

About 10% of the workers with high school degree (HS) in 1999 are working students. Usually, this is not a full time employment and thus the daily wages become extremely low for this group if the job is treated as being full-time by the data research centre. We excluded working students as well es interns and workers in partial retirement, reducing the sample to 2 857 799 observations. The results are shown in Table 4.

The strong negative effect of a high school degree vanished under IP2 correction and declines under the other correction rules. This is just the start of a process leading to a significant positive effect of a high school degree (see Subsection 5.2).

The results of the Blinder-Oaxaca-type decomposition for the smaller sample are shown

Table 4: Mincer-type wage regressions (rule based correction (RB), combined correction (EM)) for 1999 (excluding interns, working students and partial retirement)

	IP1		IP2		IP2B		IP3	
	RB	EM	RB	EM	RB	EM	RB	EM
ND	-0.106***	-0.135***	-0.106***	-0.135***	-0.104***	-0.133***	-0.106***	-0.132***
VT	0.077***	-0.552***	0.099***	-0.009***	0.079***	-0.483***	0.092***	-0.517***
HS	0.166***	0.140***	0.181***	0.169***	0.160***	0.150***	0.172***	0.149***
TC	0.316***	0.290***	0.322***	0.321***	0.317***	0.303***	0.317***	0.299***
UD	0.486***	0.564***	0.501***	0.558***	0.496***	0.531***	0.491***	0.527***

Source: IABS, own calculations

reference: vocational training, *** p<0.01, ** p<0.05, * p<0.1

Table 5: Blinder-Oaxaca decomposition of high school graduates versus the (vocational) trained workers

version	wage differential	education	characteristics	coefficients
IP1	-0.836	-1.024	-0.281	0.469
			-0.776	0.964
IP2	-0.205	-1.047	-0.193	1.035
			-0.290	1.133
IP2B	-0.739	-1.155	-0.250	0.665
			-0.688	1.104
IP3	-0.780	-1.160	-0.255	0.635
			-0.712	1.092

Source: IABS, own calculations

in Table 5. Especially the raw distance under the IP2 rule is now smaller. To some surprise, the discriminating effect of the educational degrees stays at the high level of 1 log point under any correction rule. The reduction of the wage differential is driven by a lower personal/job characteristics differential. In both versions of the decomposition, the effect of the characteristic's differential and the effect of the coefficient's differentials are quite similar.

In a conclusion, the results show that a worker with a general high school degree earns about 1 log-point less than a comparable worker with a vocational training degree. Additionally, the personal/job characteristics are also in favour of the vocational trained worker. This is mainly driven by differences in age and work experience (HS graduates are younger and have less experience). However, there is also some selection on unobserved abilities into formal degrees. This is reflected by the high coefficient effect. It shows that high school graduates display stronger aptitude and thus have a steeper increase in earnings after improving their personal characteristics and/or enhancing their working

conditions.

5.2. Changes of the Returns to Education over Time

We ran sets of Tobit regressions, one for each year of the period 1993 to 2003 (see Appendix for detailed results). The estimates for case-deletion, weighted case-deletion, missing dummy regression, and all rule-based IP corrections do not differ substantially (please refer to the Appendix). The new method, however, produces significantly different results from all other methods.

Figure 5 represents the coefficients for the dummies of the education variable under the different correction rule sets. The left figures show the estimated coefficients neglecting the missing value problem (and additional misclassification). The figures in the right column show the estimates using our new approach, i.e. with deletion of possible misclassified values and estimated by the EM method.

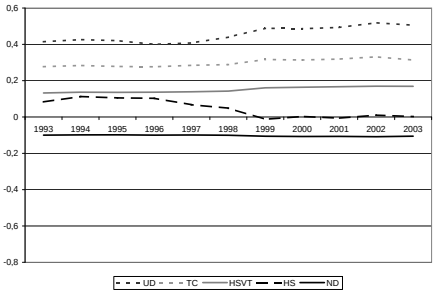
First, the coefficients tend to be spread wider over time, which support literature on rising wage inequality in western industrialised countries. Moreover, this spread is driven by increasing wages for the high-skilled workers (the constant, which represents the average wage for VT workers, does not change significantly over time).

There seems to be a change to the high school graduates. As already mentioned in the former subsection, the estimates for the high school graduates are driven by working students. Thus, we exclude these as well as interns and workers in partial retirement. The new results (starting in 1999) are shown in Figure 6 (see appendix for detailed estimation results).

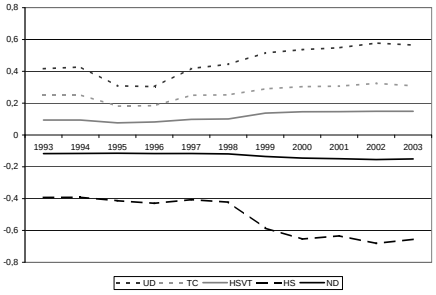
Excluding IP2, the new results show an "upgrading" of the high school degree. In 2003, the coefficient for workers with a high school degree and the coefficient for workers with high school degree and a vocational training degree are equal. There are two possible explanations for this. First, Wichert and Wilke (2010) have shown that misclassification between these two groups are very likely to occur. Additionally, the group of high school graduates is quite small and its members predominantly work in the 'business services' sector (2003: 28.74%). Especially between 1999 and 2000, there was a significant increase of high school graduates working in this sector, which also includes IT-services. During this period of time, the IT sector recruited a lot of young workers with sufficient computer skills but without higher educational degrees. This also partly explains the relative high wages for high school graduates.

Introducing new values to one of the variables changes the reporting system only slightly. Maybe, there was some persistency in reporting the old, less detailed values instead of the new ones. Unfortunately, the new values can also be observed with a quite

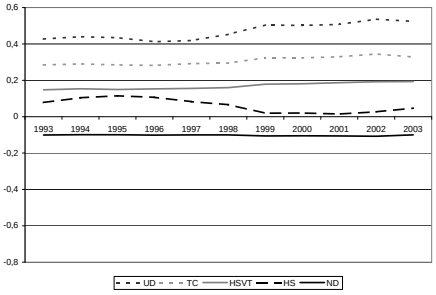
Figure 5: Estimated coefficients for education variable



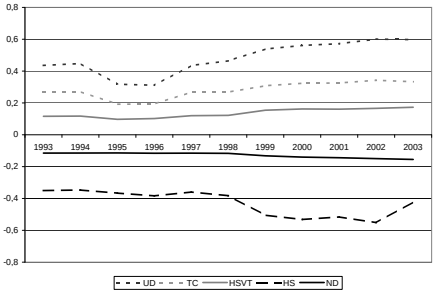
(a) IP1



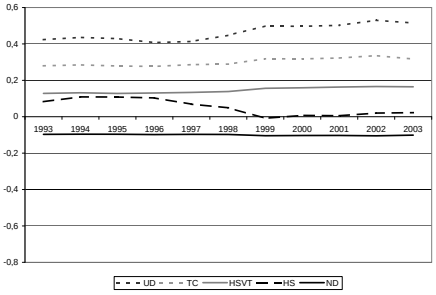
(b) IP1 EM



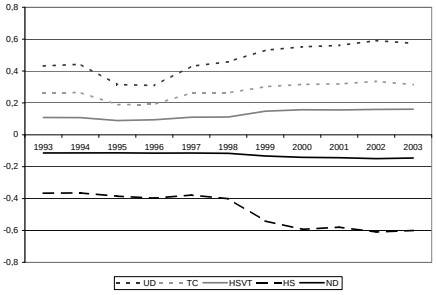
(c) IP2



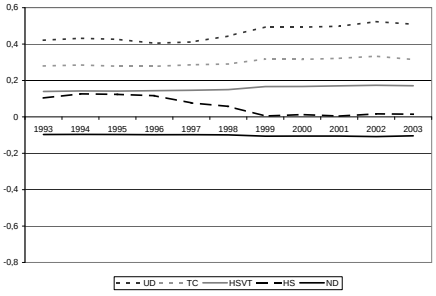
(d) IP2 EM



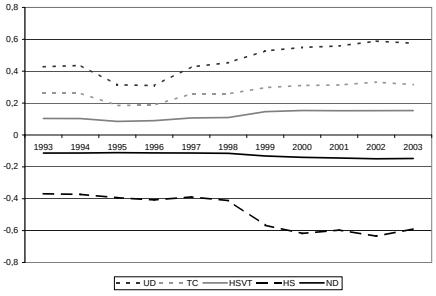
(e) IP2B



(f) IP2B EM



(g) IP3

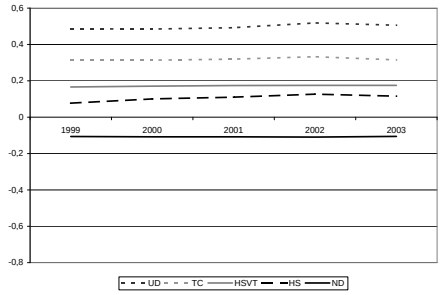


(h) IP3 EM

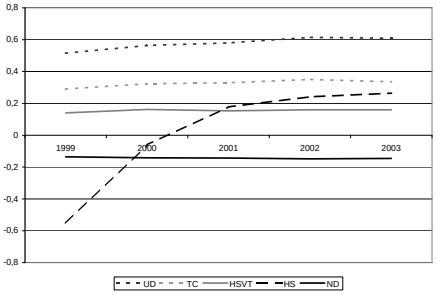
Source: IABS, own calculations

reference: vocational training

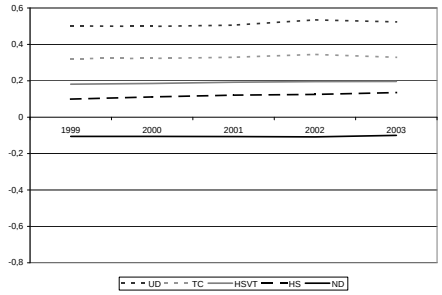
Figure 6: Estimated coefficients for education variable (excluding interns, working students and partial retirement)



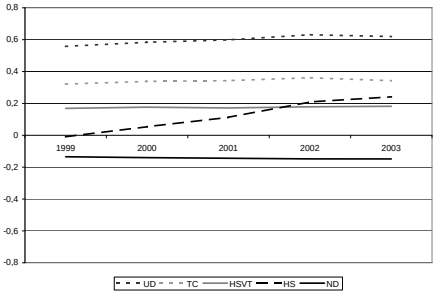
(a) IP1



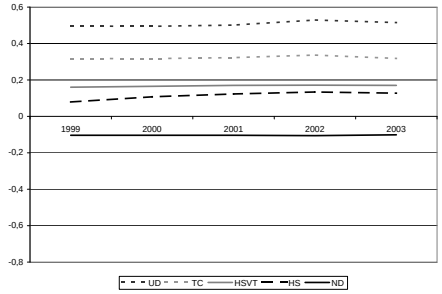
(b) IP1 EM



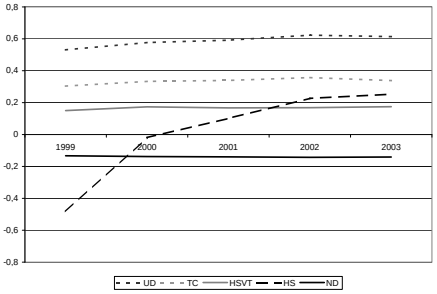
(c) IP2



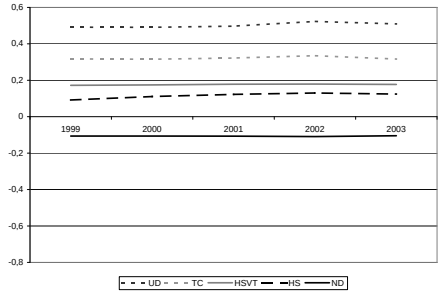
(d) IP2 EM



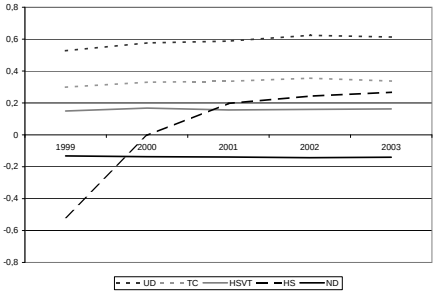
(e) IP2B



(f) IP2B EM



(g) IP3



(h) IP3 EM

Source: IABS, own calculations

reference: vocational training

constant frequency.

The coefficient estimates under the IP2 rule differ significantly from the estimates under the other correction rules (cf. Figure 5). The IP2 is the most conservative correction rule with the lowest number of missing values. The results of the EM procedure are therefore closer to the standard estimates. The high numbers of misclassified values (Wichert and Wilke, 2010) require a less conservative correction procedure. The estimation results under the other correction rules are more plausible in this case.

6. Conclusion

The education variable in the IAB datasets suffers seriously from data problems like missing and misclassified values. The data problems do not occur randomly, but are highly associated with other variables from the dataset. This issue has become more and more important over the last years and it severely influences empirical findings.

Based on the notion that the education variable should represent a person's highest degree, and that people can only attain degrees over time but not lose them, different correction procedures from the literature have been compared. A newly developed procedure utilises these rules to identify possibly misclassified values. These values are removed from the dataset, increasing the proportion of missing values. We use an expectation maximisation algorithm for missing values to take care of that problem. As the procedure removes information from the education variable, the parameter estimates become unbiased, though less efficient under the MAR assumption.

We applied the newly defined estimator to a set of Mincer-type wage regression for the years 1993–2003 separately in order to observe changes in the impact of educational degrees on wages. These coefficient estimates clearly show that the quality of education is more important than the number of years in education. We also find a rising wage differential between the different educational degrees over time. This indicates that the educational expansion of this decade does not exceed the request for high-skilled workers and that the return on education is still rising.

Furthermore, the results of a Blinder-Oaxaca decomposition show that there is significant selection into educational degrees. This partly compensates the a-priori less rewarded high school degree compared to a vocational training degree.

Future research could improve the wage regressions to obtain bias-free estimates for an exact measure of the return on education. Furthermore, the measurement concept should be extended to the total return of education including the probability of unemployment and investments into one's education.

We have seen that data problems in the education variable causes severe problems when used as an explanatory factor in wage regressions. Ignoring these data quality issues might lead to wrong interpretations and misguiding policy recommendations.

References

Takeshi Amemiya. Regression analysis when the dependent variable is truncated normal. *Econometrica*, 41(6):997–1016, 1973. ISSN 00129682. URL `http://www.jstor.org/stable/1914031`.

Dirk Antonczyk, Thomas DeLeire, and Bernd Fitzenberger. Polarization and rising wage inequality: comparing the u.s. and germany. ZEW Discussion Papers 10-015, ZEW - Zentrum für Europäische Wirtschaftsforschung / Center for European Economic Research, 2010. URL `http://ideas.repec.org/p/zbw/zewdip/10015.html`.

Jere R. Behrman and Mark R. Rosenzweig. biases in schooling returns and twins: a test and new estimates. *Economics of Education Review*, 18(2):159–167, 1999. ISSN 0272-7757. doi: 10.1016/S0272-7757(98)00033-8.

S. Bender, A. Haas, and C. Klose. The IAB employment subsample 1975–1995. *Schmollers Jahrbuch, Zeitschrift für Wirtschafts-und Sozialwissenschaften*, 120(4):649–662, 2000.

Alan S. Blinder. Wage discrimination: Reduced form and structural estimates. *The Journal of Human Resources*, 8(4):436–455, 1973. ISSN 0022166X. URL `http://www.jstor.org/stable/144855`.

Brahim Boudarbat, Thomas Lemieux, and W. Craig Riddell. The evolution of the returns to human capital in canada, 1980-2005. *Canadian Public Policy*, 36(1):63–89, March 2010. URL `http://ideas.repec.org/a/cpp/issued/v36y2010i1p63-89.html`.

D. Card. The causal effect of education on earnings. In *Handbook of Labor Economics*, volume 3, pages 1801–1863. Elsevier, 1999.

U. Cramer. Probleme der genauigkeit der beschäftigtenstatistik. *Allgemeines Statistisches Archiv*, 69:56–68, 1985.

Nils Drews. Das regionalfile der iab-beschäftigtenstichprobe 1975–2004. Datenreport 02/2008, FDZ, 2008.

B. Fitzenberger, A. Osikominu, and R. Völter. Imputation rules to improve the education variable in the IAB employment subsample. *Journal of Applied Social Science Studies (Schmollers Jahrbuch)*, 126(3):405–436, 2006. URL `http://schmollersjahrbuch.diw.de/schmollersjahrbuch/index.jsp?n=0010&p=7&c=2006/summary/s_06_3_7`.

John Garen. The returns to schooling: A selectivity bias approach with a continuous choice variable. *Econometrica*, 52(5):pp. 1199–1218, 1984. ISSN 00129682. URL http://www.jstor.org/stable/1910996.

M. Gebel and F. Pfeiffer. Educational expansion and its heterogeneous returns for wage workers. *Schmollers Jahrbuch*, 130(1):19–42, 2010.

Zvi Griliches. Estimating the returns to schooling: Some econometric problems. *Econometrica*, 45(1):1–22, 1977. ISSN 00129682. URL http://www.jstor.org/stable/1913285.

P.J. Huber. The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 221–33, 1967.

J.G. Ibrahim, M.H. Chen, S.R. Lipsitz, and A.H. Herring. Missing-data methods for generalized linear models. *Journal of The American Statistical Association*, 100(469): 332–346, 2005. doi: 10.1198/016214504000001844.

Joseph G. Ibrahim. Incomplete data in generalized linear models. *Journal of The American Statistical Association*, 85(411):765–769, 1990. ISSN 01621459. URL http://www.jstor.org/stable/2290013.

Joseph G. Ibrahim, Stuart R. Lipsitz, and Ming-Hui Chen. Missing covariates in generalized linear models when the missing data mechanism is non-ignorable. *Journal of the Royal Statistical Society. Series B (Methodological)*, 61(1):173–190, 1999. ISSN 13697412. URL http://www.jstor.org/stable/2680744.

Andrea Ichino and Rudolf Winter-Ebmer. Lower and upper bounds of returns to schooling: An exercise in iv estimation with different instruments. *European Economic Review*, 43(4-6):889–901, 1999. ISSN 0014-2921. doi: 10.1016/S0014-2921(98)00102-0.

P. Jacobebbinghaus and S. Seth. The German integrated employment biographies sample IEBS. *Schmollers Jahrbuch*, 127(2):335–342, 2007.

Per Johansson and Per Skedinger. Misreporting in register data on disability status: evidence from the swedish public employment service. *Empirical Economics*, 37:411–434, 2009. ISSN 0377-7332. doi: 10.1007/s00181-008-0238-6. 10.1007/s00181-008-0238-6.

Michael P. Jones. Indicator and stratification methods for missing explanatory variables in multiple linear regression. *Journal of The American Statistical Association*, 91(433): 222–230, 1996. ISSN 01621459. URL http://www.jstor.org/stable/2291399.

Thomas J. Kane, Cecilia Elena Rouse, and Douglas Staiger. Estimating returns to schooling when schooling is misreported. Technical report, National Bureau of Economic

Research, Inc, July 1999. URL http://ideas.repec.org/p/nbr/nberwo/7235.html.

L. Katz and D. Autor. Changes in the wage structure and earnings inequality. In *Handbook of labor economics*, volume 3, pages 1463–1555. North-Holland, 1999.

Roderick J. A. Little and D.B. Rubin. *Statistical analysis with missing data*. Wiley Series in Probability and Statistics. Wiley, Hoboken, 2nd edition, 2002.

Thomas A. Louis. Finding the observed information matrix when using the em algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 44(2):226–233, 1982. ISSN 00359246. URL http://www.jstor.org/stable/2345828.

P. McCullagh and J.A. Nelder. *Generalized linear models*. Chapman & Hall/CRC, 1989.

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

Ronald Oaxaca. Male-female wage differentials in urban labor markets. *International Economic Review*, 14(3):693–709, 1973. ISSN 00206598. URL http://www.jstor.org/stable/2525981.

J.S. Pischke and T. Von Wachter. Zero returns to compulsory schooling in germany: evidence and interpretation. *The Review of Economics and Statistics*, 90(3):592–598, 2008.

Donald B. Rubin. Inference and missing data. *Biometrika*, 63(3):581–592, 1976. ISSN 00063444. URL http://www.jstor.org/stable/2335739.

Donald B. Rubin. *Multiple Imputation for Nonresponse in Surveys*. Wiley, 1987.

W. Schmähl and U. Fachinger. Prozeßproduzierte daten als grundlage für sozial- und verteilungspolitische analysen - erfahrungen mit daten der rentenversicherungsträger für längschnittsanalysen. In R. Hauser, N. Ott, and G. Wagner, editors, *Mikroanalytische Grundlagen der Gesellschaftspolitik*, volume 2. Akademie Verlag, Berlin, 1994.

James Tobin. Estimation of relationships for limited dependent variables. *Econometrica*, 26(1):24–36, 1958. ISSN 00129682. URL http://www.jstor.org/stable/1907382.

Halbert White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48(4):817–838, 1980. ISSN 00129682. URL http://www.jstor.org/stable/1912934.

Laura Wichert and Ralf A. Wilke. Which factors safeguard employment? an analysis with misclassified german register data. Methodenreport, FDZ, 2010.

Robert J. Willis and Sherwin Rosen. Education and self-selection. *The Journal of Political Economy*, 87(5):7–36, 1979. ISSN 00223808. URL http://www.jstor.org/stable/1829907.

J.M. Wooldridge. *Econometric analysis of cross section and panel data*. MIT press, 2002.

J.M. Wooldridge. Estimating average partial effects under conditional moment independence assumptions. CEMMAP Working Papers CWP03/04, The Institute for Fiscal Studies, Department of economics, UCL, London, 2004.

A. Additional Tables

Table 6: Variable Descriptions

Variable	Description
ND, HS, HSVT, TC, UD (Reference: VT)	formal educational degrees: “no degree”, “high school”, “high school + vocational training”, “technical college”, “university degree”, “vocational training”
age, age_sq	age divided by ten and age divided by ten to the square
foreign	dummy variable indicating if the person has no German passport
occupation (reference: salaried employee)	set of dummy variables indicating farmers, service workers, sales workers, clerical workers, admin workers
industry (reference: investment goods)	set of dummy variables indicating agriculture, basic industry, clothes, papers & print, food industry, construction, trade, transport & communication, business services, consumer services, education, and public administration
work experience	3–6 months, 6–12 months, 12–24 months, 2–5 years, 5–10 years, >10 years
reportmiss	dummy indicating that the education information is missing
underreport	dummy indicating that a higher educational degree has been imputed than reported
overreport	dummy indicating that a lower educational degree has been imputed than reported
ln wage	natural logarithm of the daily wage (monthly wages divided by number of working days)
spell length	length of spell in years (at least one report per year, i.e. less than 1)

Table 7: Estimates for Case Deletion

	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
foreign	-0.033***	-0.027***	-0.023***	-0.028***	-0.027***	-0.022***	-0.018***	-0.014***	-0.015***	-0.012***	-0.010***
farmer	-0.103***	-0.099***	-0.109***	-0.100***	-0.098***	-0.104***	-0.149***	-0.158***	-0.173***	-0.169***	-0.096***
service worker	0.020***	0.025***	0.027***	0.026***	0.027***	0.033***	0.023***	0.014***	0.010***	0.015***	0.025***
sales worker	0.261***	0.269***	0.272***	0.281***	0.294***	0.296***	0.313***	0.315***	0.316***	0.324***	0.331***
clerical worker	0.205***	0.213***	0.209***	0.220***	0.228***	0.236***	0.248***	0.248***	0.249***	0.263***	0.264***
admin worker	0.299***	0.301***	0.296***	0.303***	0.310***	0.308***	0.324***	0.321***	0.317***	0.332***	0.323***
agriculture	0.015***	0.013***	0.005***	0.014***	0.015***	0.007***	-0.017***	-0.039***	-0.030***	-0.025***	-0.206***
basic industry	0.000***	0.014***	0.004***	0.004***	0.002***	-0.000***	-0.009***	-0.020***	-0.017***	-0.016***	-0.092***
clothes, paper & print	-0.069***	-0.061***	-0.079***	-0.078***	-0.084***	-0.090***	-0.101***	-0.111***	-0.108***	-0.115***	-0.116***
food industry	-0.096***	-0.087***	-0.113***	-0.108***	-0.108***	-0.122***	-0.128***	-0.142***	-0.138***	-0.138***	-0.178***
construction	-0.042***	-0.031***	-0.060***	-0.064***	-0.081***	-0.099***	-0.112***	-0.135***	-0.141***	-0.150***	-0.195***
trade	-0.149***	-0.145***	-0.163***	-0.161***	-0.169***	-0.178***	-0.192***	-0.199***	-0.198***	-0.211***	-0.249***
transport & communication	-0.093***	-0.103***	-0.127***	-0.124***	-0.131***	-0.140***	-0.155***	-0.173***	-0.174***	-0.186***	-0.226***
business services	-0.053***	-0.054***	-0.078***	-0.078***	-0.089***	-0.103***	-0.129***	-0.147***	-0.145***	-0.145***	-0.201***
consumer services	-0.216***	-0.208***	-0.230***	-0.218***	-0.218***	-0.231***	-0.257***	-0.269***	-0.271***	-0.279***	-0.376***
education	-0.054***	-0.059***	-0.067***	-0.085***	-0.086***	-0.082***	-0.085***	-0.099***	-0.093***	-0.105***	-0.146***
public administration	-0.150***	-0.160***	-0.166***	-0.162***	-0.168***	-0.177***	-0.169***	-0.174***	-0.168***	-0.173***	-0.200***
ND	-0.119***	-0.119***	-0.119***	-0.121***	-0.118***	-0.120***	-0.129***	-0.129***	-0.130***	-0.131***	-0.130***
HS	0.114***	0.130***	0.119***	0.117***	0.117***	0.118***	0.108***	0.125***	0.123***	0.149***	0.147***
HSVT	0.165***	0.168***	0.162***	0.163***	0.164***	0.169***	0.189***	0.187***	0.190***	0.194***	0.190***
TC	0.276***	0.280***	0.276***	0.271***	0.276***	0.278***	0.310***	0.308***	0.314***	0.333***	0.310***
UD	0.428***	0.438***	0.431***	0.400***	0.404***	0.446***	0.496***	0.489***	0.494***	0.520***	0.499***
age	0.254***	0.234***	0.217***	0.197***	0.180***	0.176***	0.187***	0.168***	0.189***	0.216***	0.207***
age_sq	-0.027***	-0.024***	-0.022***	-0.020***	-0.018***	-0.017***	-0.019***	-0.017***	-0.020***	-0.023***	-0.021***
3–6 months	0.029***	0.064***	0.056***	0.078***	0.031***	0.070***	0.058***	0.101***	0.081***	0.047***	0.050***
6–12 months	0.049***	0.111***	0.109***	0.136***	0.096***	0.160***	0.167***	0.120***	0.121***	0.110***	0.167***
12–24 months	0.116***	0.147***	0.178***	0.207***	0.165***	0.192***	0.224***	0.210***	0.205***	0.217***	0.221***
2–5 years	0.267***	0.273***	0.281***	0.322***	0.289***	0.317***	0.332***	0.313***	0.316***	0.338***	0.358***
5–10 years	0.344***	0.382***	0.407***	0.449***	0.430***	0.457***	0.477***	0.453***	0.442***	0.464***	0.500***
>10 years	0.471***	0.505***	0.531***	0.578***	0.569***	0.601***	0.634***	0.617***	0.609***	0.633***	0.663***
intercept	3.429***	3.414***	3.451***	3.437***	3.473***	3.457***	3.418***	3.491***	3.454***	3.371***	3.400***
σ	0.268***	0.271***	0.272***	0.271***	0.274***	0.286***	0.321***	0.326***	0.330***	0.334***	0.318***

Table 8: Estimates with Missing Dummy

	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
foreign	-0.041***	-0.035***	-0.032***	-0.037***	-0.036***	-0.031***	-0.028***	-0.023***	-0.027***	-0.024***	-0.025***
farmer	-0.113***	-0.106***	-0.117***	-0.110***	-0.106***	-0.111***	-0.154***	-0.162***	-0.181***	-0.171***	-0.094***
service worker	0.015***	0.021***	0.022***	0.021***	0.023***	0.029***	0.015***	0.005***	-0.000***	0.002***	0.013***
sales worker	0.261***	0.269***	0.272***	0.281***	0.293***	0.294***	0.313***	0.316***	0.319***	0.325***	0.338***
clerical worker	0.213***	0.222***	0.220***	0.230***	0.240***	0.251***	0.268***	0.271***	0.276***	0.293***	0.296***
admin worker	0.308***	0.311***	0.308***	0.315***	0.322***	0.323***	0.344***	0.345***	0.345***	0.362***	0.355***
agriculture	0.012***	0.010***	0.001***	0.010***	0.011***	0.004***	-0.021***	-0.045***	-0.038***	-0.032***	-0.206***
basic industry	0.000***	0.014***	0.004***	0.004***	0.002***	-0.000***	-0.008***	-0.018***	-0.015***	-0.015***	-0.091***
clothes, paper & print	-0.071***	-0.062***	-0.081***	-0.080***	-0.084***	-0.090***	-0.099***	-0.109***	-0.106***	-0.114***	-0.112***
food industry	-0.098***	-0.087***	-0.116***	-0.111***	-0.111***	-0.122***	-0.128***	-0.142***	-0.139***	-0.140***	-0.177***
construction	-0.035***	-0.024***	-0.054***	-0.058***	-0.076***	-0.093***	-0.104***	-0.126***	-0.133***	-0.142***	-0.185***
trade	-0.149***	-0.145***	-0.163***	-0.160***	-0.166***	-0.175***	-0.186***	-0.192***	-0.190***	-0.203***	-0.243***
transport & communication	-0.099***	-0.112***	-0.136***	-0.133***	-0.142***	-0.156***	-0.174***	-0.191***	-0.194***	-0.209***	-0.246***
business services	-0.056***	-0.057***	-0.081***	-0.079***	-0.087***	-0.098***	-0.124***	-0.142***	-0.140***	-0.138***	-0.209***
consumer services	-0.240***	-0.235***	-0.258***	-0.251***	-0.250***	-0.263***	-0.289***	-0.302***	-0.310***	-0.323***	-0.378***
education	-0.052***	-0.058***	-0.066***	-0.084***	-0.086***	-0.080***	-0.084***	-0.096***	-0.090***	-0.100***	-0.136***
public administration	-0.150***	-0.159***	-0.166***	-0.160***	-0.167***	-0.176***	-0.168***	-0.174***	-0.168***	-0.174***	-0.199***
ND	-0.115***	-0.115***	-0.115***	-0.116***	-0.114***	-0.115***	-0.123***	-0.123***	-0.122***	-0.122***	-0.119***
HS	0.116***	0.132***	0.120***	0.118***	0.117***	0.117***	0.106***	0.121***	0.118***	0.144***	0.143***
HSVT	0.166***	0.169***	0.162***	0.162***	0.162***	0.165***	0.185***	0.180***	0.182***	0.186***	0.183***
TC	0.276***	0.278***	0.274***	0.268***	0.273***	0.272***	0.304***	0.300***	0.305***	0.325***	0.300***
UD	0.432***	0.441***	0.434***	0.401***	0.404***	0.445***	0.496***	0.488***	0.493***	0.521***	0.499***
age	0.250***	0.232***	0.218***	0.197***	0.181***	0.174***	0.186***	0.167***	0.185***	0.213***	0.199***
age_sq	-0.026***	-0.024***	-0.022***	-0.020***	-0.018***	-0.017***	-0.019***	-0.017***	-0.020***	-0.023***	-0.021***
3–6 months	0.034***	0.063***	0.049***	0.055***	0.040***	0.071***	0.077***	0.075***	0.064***	0.040***	0.022***
6–12 months	0.057***	0.118***	0.118***	0.133***	0.102***	0.136***	0.177***	0.079***	0.144***	0.096***	0.133***
12–24 months	0.130***	0.160***	0.173***	0.192***	0.171***	0.166***	0.225***	0.193***	0.215***	0.179***	0.202***
2–5 years	0.270***	0.287***	0.286***	0.311***	0.293***	0.295***	0.327***	0.274***	0.324***	0.301***	0.341***
5–10 years	0.344***	0.391***	0.408***	0.436***	0.432***	0.432***	0.467***	0.410***	0.444***	0.427***	0.484***
>10 years	0.475***	0.517***	0.532***	0.565***	0.569***	0.576***	0.625***	0.570***	0.607***	0.589***	0.648***
reportmiss	-0.093***	-0.096***	-0.098***	-0.104***	-0.106***	-0.113***	-0.141***	-0.148***	-0.152***	-0.162***	-0.155***
intercept	3.439***	3.412***	3.451***	3.455***	3.473***	3.488***	3.432***	3.543***	3.470***	3.430***	3.442***
σ	0.278***	0.282***	0.284***	0.284***	0.287***	0.299***	0.341***	0.349***	0.356***	0.362***	0.343***

Table 9: Logit on Education Information Missing

	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
logwage	-0.608***	-0.637***	-0.708***	-0.743***	-0.711***	-0.699***	-0.679***	-0.640***	-0.722***	-0.724***	-0.825***
foreign	0.838***	0.785***	0.762***	0.760***	0.710***	0.622***	0.564***	0.504***	0.541***	0.473***	0.526***
farmer	0.704***	0.816***	0.648***	0.591***	0.671***	0.696***	0.746***	0.750***	0.801***	0.809***	0.464***
service worker	0.418***	0.395***	0.396***	0.425***	0.443***	0.417***	0.410***	0.428***	0.396***	0.429***	0.462***
sales worker	-0.291***	-0.360***	-0.337***	-0.269***	-0.219***	-0.226***	-0.242***	-0.193***	-0.200***	-0.222***	-0.144***
clerical worker	-0.109***	-0.136***	-0.128***	-0.025***	0.037***	0.083***	0.069***	0.157***	0.185***	0.195***	0.232***
admin worker	-0.192***	-0.173***	-0.124***	-0.080***	-0.056***	-0.060***	-0.090***	0.005***	0.018***	0.051***	0.121***
agriculture	-0.150***	-0.192***	-0.087***	-0.100***	-0.072***	-0.172***	-0.197***	-0.127***	-0.030***	0.086***	0.775***
basic industry	0.244***	0.272***	0.269***	0.312***	0.296***	0.195***	0.175***	0.074***	0.074***	0.109***	0.381***
clothes, paper & print	0.354***	0.367***	0.342***	0.347***	0.367***	0.373***	0.381***	0.376***	0.395***	0.393***	0.704***
food industry	0.403***	0.406***	0.444***	0.502***	0.469***	0.489***	0.516***	0.515***	0.537***	0.591***	0.781***
construction	1.048***	1.054***	1.066***	1.068***	1.052***	1.020***	1.035***	1.012***	1.012***	1.065***	1.253***
trade	0.941***	1.016***	1.003***	1.030***	1.022***	1.022***	1.025***	0.987***	1.061***	1.096***	1.175***
transport & communication	0.949***	0.975***	0.913***	1.006***	0.979***	1.008***	0.990***	0.985***	1.086***	1.117***	1.314***
business services	0.984***	0.981***	0.932***	0.969***	0.886***	0.836***	0.858***	0.822***	0.929***	0.959***	1.216***
consumer services	1.531***	1.518***	1.513***	1.514***	1.534***	1.632***	1.630***	1.615***	1.647***	1.697***	1.295***
education	-0.158***	-0.112***	-0.084***	0.017***	0.034***	-0.038***	0.028***	0.029***	0.061***	0.042***	0.590***
public administration	-0.308***	-0.466***	-0.432***	-0.365***	-0.402***	-0.468***	-0.533***	-0.673***	-0.787***	-0.779***	-0.274***
age	0.748***	0.679***	0.617***	0.517***	0.479***	0.535***	0.740***	0.789***	0.731***	0.799***	0.920***
age_sq	-0.075***	-0.068***	-0.062***	-0.051***	-0.048***	-0.054***	-0.078***	-0.086***	-0.080***	-0.086***	-0.100***
3–6 months	-0.043***	0.125***	-0.088***	-0.203***	-0.081***	-0.222***	0.110***	-0.253***	-0.194***	-0.259***	-0.045***
6–12 months	0.134***	0.229***	-0.018***	0.008***	0.274***	0.054***	0.176***	-0.096***	-0.094***	-0.008***	0.091***
12–24 months	0.188***	0.256***	0.151***	0.043***	0.174***	0.041***	0.154***	-0.074***	-0.066***	-0.128***	0.137***
2–5 years	0.051***	0.111***	0.046***	0.035***	0.169***	0.012***	0.108***	-0.145***	-0.177***	-0.139***	-0.095***
5–10 years	-0.007***	0.021***	-0.168***	-0.149***	-0.032***	-0.215***	-0.151***	-0.361***	-0.376***	-0.331***	-0.258***
>10 years	-0.443***	-0.418***	-0.557***	-0.538***	-0.460***	-0.678***	-0.659***	-0.910***	-0.913***	-0.925***	-0.857***
spell length	-0.504***	-0.526***	-0.457***	-0.461***	-0.549***	-0.509***	-0.406***	-0.501***	-0.397***	-0.410***	-0.352***
intercept	-1.733***	-1.449***	-0.870***	-0.482***	-0.484***	-0.402***	-0.955***	-0.742***	-0.291***	-0.383***	-0.444***
McFadden's R^2	0,118	0,119	0,120	0,118	0,121	0,122	0,130	0,133	0,139	0,143	0,135

Table 10: Estimates for Weighted Case Deletion

	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
foreign	-0.054***	-0.049***	-0.047***	-0.053***	-0.052***	-0.047***	-0.043***	-0.037***	-0.047***	-0.042***	-0.047***
farmer	-0.117***	-0.110***	-0.120***	-0.112***	-0.109***	-0.114***	-0.160***	-0.168***	-0.195***	-0.184***	-0.103***
service worker	0.008***	0.014***	0.015***	0.014***	0.016***	0.022***	0.002***	-0.011***	-0.021***	-0.019***	-0.006***
sales worker	0.262***	0.271***	0.274***	0.282***	0.293***	0.295***	0.317***	0.321***	0.324***	0.332***	0.346***
clerical worker	0.214***	0.223***	0.221***	0.232***	0.241***	0.252***	0.271***	0.274***	0.279***	0.298***	0.300***
admin worker	0.310***	0.313***	0.310***	0.317***	0.325***	0.325***	0.349***	0.352***	0.352***	0.371***	0.362***
agriculture	0.010***	0.008***	-0.001***	0.008***	0.008***	0.001***	-0.024***	-0.050***	-0.044***	-0.039***	-0.215***
basic industry	0.000***	0.015***	0.004***	0.004***	0.003***	0.001***	-0.007***	-0.018***	-0.014***	-0.014***	-0.090***
clothes, paper & print	-0.071***	-0.063***	-0.081***	-0.080***	-0.085***	-0.090***	-0.100***	-0.111***	-0.107***	-0.116***	-0.113***
food industry	-0.098***	-0.088***	-0.117***	-0.113***	-0.112***	-0.124***	-0.130***	-0.144***	-0.142***	-0.142***	-0.179***
construction	-0.036***	-0.024***	-0.055***	-0.060***	-0.078***	-0.096***	-0.109***	-0.131***	-0.142***	-0.152***	-0.194***
trade	-0.153***	-0.150***	-0.169***	-0.166***	-0.171***	-0.181***	-0.194***	-0.199***	-0.199***	-0.214***	-0.255***
transport & communication	-0.105***	-0.120***	-0.145***	-0.146***	-0.154***	-0.169***	-0.189***	-0.208***	-0.216***	-0.232***	-0.269***
business services	-0.066***	-0.067***	-0.092***	-0.091***	-0.099***	-0.110***	-0.146***	-0.164***	-0.169***	-0.169***	-0.251***
consumer services	-0.299***	-0.292***	-0.322***	-0.315***	-0.314***	-0.329***	-0.367***	-0.380***	-0.404***	-0.424***	-0.423***
education	-0.050***	-0.056***	-0.065***	-0.082***	-0.084***	-0.077***	-0.080***	-0.093***	-0.085***	-0.096***	-0.135***
public administration	-0.149***	-0.158***	-0.165***	-0.160***	-0.167***	-0.176***	-0.166***	-0.172***	-0.166***	-0.173***	-0.196***
ND	-0.114***	-0.113***	-0.113***	-0.114***	-0.111***	-0.113***	-0.123***	-0.124***	-0.123***	-0.123***	-0.117***
HS	0.119***	0.137***	0.123***	0.122***	0.120***	0.118***	0.099***	0.119***	0.107***	0.146***	0.133***
HSVT	0.175***	0.176***	0.169***	0.170***	0.169***	0.172***	0.196***	0.192***	0.199***	0.204***	0.198***
TC	0.284***	0.286***	0.282***	0.276***	0.281***	0.281***	0.319***	0.315***	0.325***	0.349***	0.320***
UD	0.445***	0.454***	0.448***	0.416***	0.419***	0.461***	0.522***	0.514***	0.527***	0.560***	0.536***
age	0.244***	0.224***	0.213***	0.192***	0.171***	0.165***	0.182***	0.160***	0.184***	0.206***	0.174***
age_sq	-0.026***	-0.023***	-0.022***	-0.020***	-0.017***	-0.017***	-0.019***	-0.017***	-0.020***	-0.023***	-0.020***
3–6 months	0.036***	0.072***	0.060***	0.066***	0.039***	0.081***	0.076***	0.067***	0.055***	0.041***	0.013***
6–12 months	0.066***	0.130***	0.126***	0.148***	0.103***	0.144***	0.165***	0.064***	0.138***	0.082***	0.127***
12–24 months	0.153***	0.178***	0.181***	0.203***	0.184***	0.177***	0.225***	0.194***	0.211***	0.182***	0.188***
2–5 years	0.288***	0.315***	0.308***	0.337***	0.312***	0.316***	0.337***	0.280***	0.332***	0.306***	0.360***
5–10 years	0.362***	0.416***	0.433***	0.468***	0.461***	0.465***	0.492***	0.432***	0.468***	0.450***	0.524***
>10 years	0.505***	0.553***	0.567***	0.607***	0.608***	0.619***	0.659***	0.603***	0.645***	0.632***	0.716***
reportmiss	-0.101***	-0.106***	-0.110***	-0.118***	-0.117***	-0.125***	-0.159***	-0.166***	-0.177***	-0.189***	-0.181***
intercept	3.430***	3.368***	3.436***	3.461***	3.527***	3.555***	3.512***	3.662***	3.614***	3.614***	3.665***
σ	0.289***	0.292***	0.296***	0.297***	0.300***	0.313***	0.365***	0.376***	0.390***	0.400***	0.374***

Table 11: Estimates for IP1 Rule-Based Improved Education Variable

	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
foreign	-0.035***	-0.033***	-0.028***	-0.034***	-0.032***	-0.025***	-0.018***	-0.011***	-0.013***	-0.012***	-0.019***
farmer	-0.135***	-0.123***	-0.135***	-0.127***	-0.120***	-0.124***	-0.178***	-0.182***	-0.198***	-0.190***	-0.112***
service worker	0.003***	0.010***	0.011***	0.011***	0.012***	0.020***	0.004***	-0.005***	-0.011***	-0.009***	0.001***
sales worker	0.209***	0.221***	0.227***	0.237***	0.250***	0.255***	0.273***	0.275***	0.277***	0.284***	0.298***
clerical worker	0.170***	0.180***	0.182***	0.193***	0.202***	0.211***	0.223***	0.225***	0.229***	0.242***	0.247***
admin worker	0.267***	0.269***	0.267***	0.275***	0.282***	0.284***	0.302***	0.304***	0.303***	0.317***	0.311***
agriculture	0.013***	0.011***	0.003***	0.012***	0.010***	-0.001***	-0.028***	-0.051***	-0.046***	-0.040***	-0.214***
basic industry	0.007***	0.021***	0.010***	0.008***	0.006***	0.002***	-0.005***	-0.015***	-0.013***	-0.014***	-0.091***
clothes, paper & print	-0.064***	-0.056***	-0.077***	-0.076***	-0.084***	-0.090***	-0.098***	-0.107***	-0.105***	-0.114***	-0.117***
food industry	-0.093***	-0.086***	-0.114***	-0.111***	-0.113***	-0.127***	-0.133***	-0.148***	-0.144***	-0.145***	-0.185***
construction	-0.032***	-0.019***	-0.051***	-0.056***	-0.073***	-0.090***	-0.098***	-0.118***	-0.125***	-0.134***	-0.179***
trade	-0.152***	-0.149***	-0.167***	-0.165***	-0.173***	-0.184***	-0.194***	-0.198***	-0.198***	-0.212***	-0.251***
transport & communication	-0.089***	-0.102***	-0.127***	-0.126***	-0.139***	-0.157***	-0.171***	-0.187***	-0.187***	-0.201***	-0.239***
business services	-0.067***	-0.068***	-0.093***	-0.095***	-0.107***	-0.121***	-0.149***	-0.167***	-0.168***	-0.168***	-0.242***
consumer services	-0.280***	-0.279***	-0.296***	-0.291***	-0.290***	-0.304***	-0.329***	-0.341***	-0.349***	-0.361***	-0.388***
education	-0.060***	-0.066***	-0.075***	-0.093***	-0.096***	-0.095***	-0.102***	-0.117***	-0.112***	-0.123***	-0.169***
public administration	-0.143***	-0.151***	-0.159***	-0.156***	-0.163***	-0.174***	-0.162***	-0.168***	-0.164***	-0.170***	-0.200***
ND	-0.100***	-0.098***	-0.098***	-0.100***	-0.099***	-0.100***	-0.106***	-0.107***	-0.107***	-0.109***	-0.105***
HS	0.083***	0.113***	0.105***	0.103***	0.068***	0.049***	-0.012***	0.003***	-0.006***	0.010***	0.002***
HSVT	0.132***	0.137***	0.136***	0.137***	0.139***	0.143***	0.161***	0.164***	0.167***	0.170***	0.169***
TC	0.277***	0.284***	0.278***	0.277***	0.285***	0.289***	0.317***	0.314***	0.319***	0.331***	0.314***
UD	0.414***	0.426***	0.420***	0.400***	0.408***	0.440***	0.489***	0.486***	0.493***	0.519***	0.505***
age	0.143***	0.129***	0.108***	0.097***	0.073***	0.066***	0.088***	0.082***	0.095***	0.102***	0.069***
age_sq	-0.015***	-0.013***	-0.010***	-0.009***	-0.006***	-0.005***	-0.008***	-0.008***	-0.010***	-0.011***	-0.007***
3–6 months	0.050***	0.062***	0.050***	0.071***	0.047***	0.066***	0.075***	0.095***	0.079***	0.095***	0.071***
6–12 months	0.101***	0.115***	0.110***	0.132***	0.107***	0.131***	0.166***	0.167***	0.177***	0.176***	0.170***
12–24 months	0.168***	0.162***	0.162***	0.186***	0.173***	0.184***	0.252***	0.252***	0.256***	0.263***	0.250***
2–5 years	0.290***	0.286***	0.270***	0.299***	0.289***	0.300***	0.362***	0.367***	0.376***	0.389***	0.393***
5–10 years	0.380***	0.391***	0.388***	0.419***	0.421***	0.429***	0.490***	0.493***	0.488***	0.507***	0.526***
>10 years	0.498***	0.511***	0.511***	0.547***	0.558***	0.572***	0.645***	0.649***	0.647***	0.666***	0.689***
underreport	-0.126***	-0.125***	-0.129***	-0.127***	-0.128***	-0.134***	-0.154***	-0.157***	-0.159***	-0.159***	-0.156***
reportmiss	-0.094***	-0.094***	-0.101***	-0.107***	-0.111***	-0.122***	-0.154***	-0.166***	-0.173***	-0.186***	-0.184***
intercept	0.010***	3.658***	3.728***	3.707***	3.737***	3.745***	3.645***	3.669***	3.648***	3.619***	3.711***
σ	0.001***	0.276***	0.278***	0.279***	0.285***	0.297***	0.340***	0.349***	0.356***	0.361***	0.343***

Table 12: Estimates for IP2 Rule-Based Improved Education Variable

	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
foreign	-0.036***	-0.035***	-0.030***	-0.036***	-0.034***	-0.028***	-0.021***	-0.015***	-0.017***	-0.017***	-0.025***
farmer	-0.135***	-0.123***	-0.135***	-0.126***	-0.119***	-0.123***	-0.177***	-0.182***	-0.199***	-0.190***	-0.112***
service worker	0.003***	0.009***	0.010***	0.010***	0.012***	0.019***	0.004***	-0.006***	-0.012***	-0.009***	0.000***
sales worker	0.209***	0.220***	0.226***	0.236***	0.249***	0.254***	0.273***	0.275***	0.276***	0.284***	0.297***
clerical worker	0.170***	0.181***	0.182***	0.193***	0.202***	0.212***	0.225***	0.227***	0.231***	0.245***	0.249***
admin worker	0.267***	0.270***	0.267***	0.275***	0.282***	0.284***	0.303***	0.305***	0.304***	0.317***	0.311***
agriculture	0.013***	0.011***	0.003***	0.012***	0.010***	-0.001***	-0.028***	-0.051***	-0.046***	-0.040***	-0.215***
basic industry	0.006***	0.021***	0.010***	0.007***	0.006***	0.001***	-0.006***	-0.016***	-0.014***	-0.015***	-0.092***
clothes, paper & print	-0.063***	-0.056***	-0.076***	-0.076***	-0.084***	-0.090***	-0.098***	-0.107***	-0.105***	-0.114***	-0.117***
food industry	-0.093***	-0.086***	-0.114***	-0.111***	-0.113***	-0.128***	-0.134***	-0.149***	-0.144***	-0.146***	-0.186***
construction	-0.031***	-0.019***	-0.051***	-0.055***	-0.073***	-0.090***	-0.097***	-0.117***	-0.125***	-0.134***	-0.178***
trade	-0.151***	-0.148***	-0.166***	-0.164***	-0.172***	-0.182***	-0.193***	-0.197***	-0.196***	-0.211***	-0.249***
transport & communication	-0.088***	-0.101***	-0.126***	-0.125***	-0.138***	-0.156***	-0.170***	-0.186***	-0.186***	-0.200***	-0.238***
business services	-0.066***	-0.068***	-0.093***	-0.094***	-0.106***	-0.121***	-0.148***	-0.165***	-0.166***	-0.167***	-0.242***
consumer services	-0.279***	-0.278***	-0.296***	-0.290***	-0.290***	-0.304***	-0.329***	-0.341***	-0.350***	-0.362***	-0.389***
education	-0.061***	-0.067***	-0.076***	-0.093***	-0.096***	-0.096***	-0.103***	-0.118***	-0.113***	-0.124***	-0.169***
public administration	-0.143***	-0.150***	-0.159***	-0.156***	-0.163***	-0.174***	-0.162***	-0.168***	-0.164***	-0.170***	-0.200***
ND	-0.101***	-0.099***	-0.099***	-0.101***	-0.100***	-0.100***	-0.106***	-0.105***	-0.106***	-0.107***	-0.100***
HS	0.078***	0.104***	0.115***	0.107***	0.083***	0.066***	0.019***	0.020***	0.015***	0.027***	0.047***
HSVT	0.148***	0.154***	0.150***	0.153***	0.156***	0.160***	0.179***	0.181***	0.187***	0.192***	0.193***
TC	0.285***	0.290***	0.285***	0.282***	0.292***	0.295***	0.324***	0.324***	0.330***	0.345***	0.328***
UD	0.427***	0.440***	0.434***	0.412***	0.419***	0.453***	0.504***	0.502***	0.507***	0.536***	0.524***
age	0.142***	0.128***	0.108***	0.097***	0.073***	0.067***	0.088***	0.083***	0.097***	0.104***	0.073***
age_sq	-0.015***	-0.013***	-0.010***	-0.009***	-0.006***	-0.005***	-0.008***	-0.008***	-0.010***	-0.011***	-0.008***
3–6 months	0.049***	0.062***	0.050***	0.070***	0.048***	0.068***	0.078***	0.095***	0.082***	0.097***	0.077***
6–12 months	0.098***	0.113***	0.108***	0.128***	0.106***	0.131***	0.167***	0.168***	0.178***	0.174***	0.180***
12–24 months	0.161***	0.154***	0.157***	0.180***	0.168***	0.182***	0.249***	0.249***	0.254***	0.257***	0.257***
2–5 years	0.284***	0.278***	0.263***	0.290***	0.281***	0.297***	0.359***	0.362***	0.374***	0.382***	0.399***
5–10 years	0.374***	0.383***	0.381***	0.410***	0.413***	0.425***	0.486***	0.487***	0.483***	0.498***	0.531***
>10 years	0.492***	0.503***	0.504***	0.538***	0.550***	0.568***	0.640***	0.643***	0.642***	0.657***	0.693***
underreport	-0.122***	-0.120***	-0.122***	-0.120***	-0.122***	-0.131***	-0.149***	-0.157***	-0.157***	-0.156***	-0.147***
overreport	-0.064***	-0.062***	-0.042***	-0.060***	-0.053***	-0.023***	-0.038***	-0.020***	-0.012***	-0.009***	0.106***
reportmiss	-0.088***	-0.087***	-0.093***	-0.098***	-0.103***	-0.111***	-0.142***	-0.152***	-0.158***	-0.170***	-0.166***
intercept	3.674***	3.667***	3.736***	3.716***	3.743***	3.748***	3.648***	3.670***	3.647***	3.624***	3.697***
σ	0.272***	0.276***	0.278***	0.279***	0.285***	0.297***	0.340***	0.349***	0.355***	0.361***	0.343***

Table 13: Estimates for IP2b Rule-Based Improved Education Variable

	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
foreign	-0.037***	-0.035***	-0.030***	-0.036***	-0.034***	-0.028***	-0.020***	-0.015***	-0.017***	-0.016***	-0.023***
farmer	-0.136***	-0.124***	-0.136***	-0.127***	-0.121***	-0.124***	-0.178***	-0.183***	-0.199***	-0.191***	-0.111***
service worker	0.002***	0.009***	0.010***	0.010***	0.012***	0.019***	0.004***	-0.006***	-0.012***	-0.010***	0.000***
sales worker	0.211***	0.223***	0.228***	0.239***	0.251***	0.257***	0.276***	0.277***	0.280***	0.287***	0.301***
clerical worker	0.171***	0.182***	0.184***	0.194***	0.204***	0.213***	0.226***	0.227***	0.232***	0.245***	0.250***
admin worker	0.268***	0.271***	0.269***	0.277***	0.284***	0.286***	0.305***	0.307***	0.306***	0.320***	0.313***
agriculture	0.013***	0.011***	0.004***	0.012***	0.010***	-0.001***	-0.027***	-0.051***	-0.045***	-0.039***	-0.214***
basic industry	0.006***	0.020***	0.009***	0.007***	0.005***	0.001***	-0.006***	-0.016***	-0.014***	-0.015***	-0.092***
clothes, paper & print	-0.064***	-0.057***	-0.077***	-0.077***	-0.084***	-0.091***	-0.099***	-0.108***	-0.106***	-0.114***	-0.118***
food industry	-0.094***	-0.086***	-0.115***	-0.112***	-0.114***	-0.128***	-0.134***	-0.149***	-0.144***	-0.146***	-0.186***
construction	-0.031***	-0.019***	-0.051***	-0.055***	-0.073***	-0.090***	-0.097***	-0.116***	-0.124***	-0.134***	-0.178***
trade	-0.152***	-0.149***	-0.167***	-0.165***	-0.173***	-0.184***	-0.194***	-0.198***	-0.198***	-0.212***	-0.251***
transport & communication	-0.089***	-0.101***	-0.127***	-0.126***	-0.139***	-0.156***	-0.171***	-0.186***	-0.187***	-0.201***	-0.240***
business services	-0.066***	-0.068***	-0.093***	-0.094***	-0.106***	-0.120***	-0.148***	-0.166***	-0.167***	-0.167***	-0.242***
consumer services	-0.279***	-0.278***	-0.296***	-0.290***	-0.290***	-0.303***	-0.329***	-0.341***	-0.349***	-0.361***	-0.389***
education	-0.061***	-0.067***	-0.075***	-0.092***	-0.096***	-0.095***	-0.103***	-0.118***	-0.112***	-0.124***	-0.169***
public administration	-0.143***	-0.151***	-0.159***	-0.156***	-0.164***	-0.175***	-0.163***	-0.169***	-0.165***	-0.171***	-0.200***
ND	-0.097***	-0.096***	-0.096***	-0.098***	-0.097***	-0.098***	-0.104***	-0.103***	-0.103***	-0.105***	-0.101***
HS	0.083***	0.109***	0.108***	0.103***	0.069***	0.049***	-0.008***	0.007***	0.005***	0.020***	0.022***
HSVT	0.128***	0.132***	0.128***	0.130***	0.133***	0.138***	0.156***	0.159***	0.163***	0.166***	0.165***
TC	0.280***	0.285***	0.279***	0.276***	0.286***	0.289***	0.319***	0.318***	0.323***	0.336***	0.317***
UD	0.424***	0.436***	0.429***	0.407***	0.414***	0.448***	0.499***	0.497***	0.502***	0.530***	0.515***
age	0.141***	0.127***	0.107***	0.096***	0.072***	0.065***	0.086***	0.081***	0.093***	0.100***	0.067***
age_sq	-0.015***	-0.013***	-0.010***	-0.009***	-0.006***	-0.005***	-0.008***	-0.008***	-0.010***	-0.011***	-0.007***
3–6 months	0.048***	0.062***	0.049***	0.069***	0.047***	0.065***	0.076***	0.093***	0.080***	0.096***	0.073***
6–12 months	0.095***	0.110***	0.103***	0.127***	0.102***	0.126***	0.163***	0.163***	0.174***	0.173***	0.174***
12–24 months	0.158***	0.151***	0.151***	0.175***	0.161***	0.175***	0.243***	0.241***	0.247***	0.253***	0.249***
2–5 years	0.278***	0.273***	0.256***	0.286***	0.274***	0.289***	0.350***	0.354***	0.365***	0.376***	0.390***
5–10 years	0.368***	0.377***	0.374***	0.405***	0.406***	0.417***	0.478***	0.480***	0.477***	0.495***	0.524***
>10 years	0.486***	0.497***	0.497***	0.533***	0.543***	0.561***	0.633***	0.636***	0.636***	0.655***	0.688***
underreport	-0.116***	-0.114***	-0.116***	-0.113***	-0.115***	-0.124***	-0.142***	-0.149***	-0.149***	-0.147***	-0.139***
overreport	-0.142***	-0.136***	-0.125***	-0.144***	-0.136***	-0.103***	-0.124***	-0.111***	-0.112***	-0.114***	-0.018***
reportmiss	-0.090***	-0.089***	-0.096***	-0.101***	-0.106***	-0.116***	-0.147***	-0.158***	-0.165***	-0.178***	-0.174***
intercept	3.680***	3.672***	3.743***	3.721***	3.751***	3.757***	3.658***	3.681***	3.659***	3.634***	3.714***
σ	0.273***	0.276***	0.278***	0.279***	0.285***	0.297***	0.340***	0.349***	0.356***	0.362***	0.343***

Table 14: Estimates for IP3 Rule-Based Improved Education Variable

	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
foreign	-0.036***	-0.034***	-0.029***	-0.035***	-0.033***	-0.026***	-0.018***	-0.012***	-0.014***	-0.013***	-0.020***
farmer	-0.133***	-0.122***	-0.134***	-0.125***	-0.118***	-0.121***	-0.176***	-0.180***	-0.196***	-0.188***	-0.110***
service worker	0.004***	0.010***	0.012***	0.011***	0.013***	0.021***	0.005***	-0.005***	-0.010***	-0.008***	0.002***
sales worker	0.210***	0.222***	0.227***	0.238***	0.250***	0.255***	0.274***	0.276***	0.278***	0.286***	0.300***
clerical worker	0.171***	0.182***	0.183***	0.194***	0.203***	0.213***	0.225***	0.226***	0.230***	0.244***	0.249***
admin worker	0.269***	0.272***	0.270***	0.278***	0.285***	0.287***	0.305***	0.307***	0.306***	0.320***	0.314***
agriculture	0.012***	0.010***	0.003***	0.011***	0.009***	-0.002***	-0.028***	-0.052***	-0.046***	-0.040***	-0.214***
basic industry	0.007***	0.021***	0.010***	0.008***	0.006***	0.002***	-0.005***	-0.015***	-0.013***	-0.014***	-0.091***
clothes, paper & print	-0.063***	-0.056***	-0.076***	-0.076***	-0.083***	-0.090***	-0.098***	-0.107***	-0.105***	-0.113***	-0.117***
food industry	-0.093***	-0.086***	-0.114***	-0.110***	-0.113***	-0.127***	-0.133***	-0.148***	-0.144***	-0.146***	-0.185***
construction	-0.031***	-0.018***	-0.050***	-0.055***	-0.073***	-0.090***	-0.097***	-0.117***	-0.124***	-0.134***	-0.178***
trade	-0.151***	-0.149***	-0.167***	-0.165***	-0.173***	-0.184***	-0.194***	-0.198***	-0.198***	-0.212***	-0.251***
transport & communication	-0.089***	-0.102***	-0.127***	-0.126***	-0.139***	-0.157***	-0.172***	-0.187***	-0.187***	-0.202***	-0.240***
business services	-0.067***	-0.068***	-0.093***	-0.094***	-0.107***	-0.121***	-0.149***	-0.167***	-0.167***	-0.167***	-0.242***
consumer services	-0.280***	-0.279***	-0.296***	-0.291***	-0.290***	-0.304***	-0.329***	-0.341***	-0.350***	-0.362***	-0.389***
education	-0.061***	-0.066***	-0.075***	-0.093***	-0.096***	-0.095***	-0.102***	-0.117***	-0.112***	-0.123***	-0.169***
public administration	-0.144***	-0.151***	-0.159***	-0.156***	-0.164***	-0.175***	-0.163***	-0.169***	-0.165***	-0.171***	-0.200***
ND	-0.097***	-0.096***	-0.097***	-0.098***	-0.098***	-0.099***	-0.106***	-0.105***	-0.106***	-0.108***	-0.104***
HS	0.104***	0.127***	0.124***	0.116***	0.077***	0.057***	0.005***	0.013***	0.004***	0.017***	0.015***
HSVT	0.140***	0.142***	0.142***	0.144***	0.146***	0.150***	0.166***	0.167***	0.171***	0.174***	0.171***
TC	0.280***	0.285***	0.279***	0.278***	0.286***	0.291***	0.318***	0.316***	0.322***	0.333***	0.315***
UD	0.421***	0.432***	0.426***	0.405***	0.412***	0.444***	0.494***	0.493***	0.498***	0.523***	0.509***
age	0.145***	0.130***	0.109***	0.098***	0.074***	0.068***	0.089***	0.083***	0.096***	0.104***	0.071***
age_sq	-0.015***	-0.013***	-0.011***	-0.009***	-0.006***	-0.006***	-0.008***	-0.008***	-0.010***	-0.011***	-0.008***
3–6 months	0.050***	0.062***	0.050***	0.071***	0.047***	0.067***	0.075***	0.096***	0.081***	0.097***	0.073***
6–12 months	0.102***	0.116***	0.111***	0.133***	0.108***	0.132***	0.168***	0.169***	0.180***	0.178***	0.173***
12–24 months	0.169***	0.163***	0.163***	0.188***	0.174***	0.186***	0.254***	0.254***	0.258***	0.266***	0.254***
2–5 years	0.292***	0.288***	0.271***	0.301***	0.290***	0.302***	0.365***	0.370***	0.380***	0.392***	0.397***
5–10 years	0.382***	0.393***	0.390***	0.421***	0.423***	0.431***	0.493***	0.496***	0.491***	0.511***	0.531***
>10 years	0.500***	0.512***	0.512***	0.549***	0.559***	0.574***	0.648***	0.652***	0.650***	0.669***	0.693***
underreport	-0.136***	-0.133***	-0.136***	-0.134***	-0.135***	-0.143***	-0.160***	-0.165***	-0.167***	-0.165***	-0.160***
overreport	0.050***	0.056***	0.059***	0.058***	0.061***	0.061***	0.080***	0.064***	0.084***	0.114***	0.138***
reportmiss	-0.095***	-0.094***	-0.101***	-0.107***	-0.112***	-0.122***	-0.153***	-0.166***	-0.173***	-0.186***	-0.184***
intercept	3.661***	3.654***	3.725***	3.704***	3.734***	3.741***	3.642***	3.664***	3.641***	3.613***	3.703***
σ	0.273***	0.276***	0.278***	0.279***	0.285***	0.297***	0.340***	0.349***	0.355***	0.361***	0.343***

Table 15: Estimates for IP1 improved Education Variable via "EM by the method of weights"

	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
foreign	-0.035***	-0.033***	-0.028***	-0.034***	-0.032***	-0.025***	-0.018***	-0.011***	-0.013***	-0.012***	-0.019***
farmer	-0.135***	-0.123***	-0.135***	-0.127***	-0.120***	-0.124***	-0.178***	-0.182***	-0.198***	-0.190***	-0.112***
service worker	0.003***	0.010***	0.011***	0.011***	0.012***	0.020***	0.004***	-0.005***	-0.011***	-0.009***	0.001***
sales worker	0.209***	0.221***	0.227***	0.237***	0.250***	0.255***	0.273***	0.275***	0.277***	0.284***	0.298***
clerical worker	0.170***	0.180***	0.182***	0.193***	0.202***	0.211***	0.223***	0.225***	0.229***	0.242***	0.247***
admin worker	0.267***	0.269***	0.267***	0.275***	0.282***	0.284***	0.302***	0.304***	0.303***	0.317***	0.311***
agriculture	0.013***	0.011***	0.003***	0.012***	0.010***	-0.001***	-0.028***	-0.051***	-0.046***	-0.040***	-0.214***
basic industry	0.007***	0.021***	0.010***	0.008***	0.006***	0.002***	-0.005***	-0.015***	-0.013***	-0.014***	-0.091***
clothes, paper & print	-0.064***	-0.056***	-0.077***	-0.076***	-0.084***	-0.090***	-0.098***	-0.107***	-0.105***	-0.114***	-0.117***
food industry	-0.093***	-0.086***	-0.114***	-0.111***	-0.113***	-0.127***	-0.133***	-0.148***	-0.144***	-0.145***	-0.185***
construction	-0.032***	-0.019***	-0.051***	-0.056***	-0.073***	-0.090***	-0.098***	-0.118***	-0.125***	-0.134***	-0.179***
trade	-0.152***	-0.149***	-0.167***	-0.165***	-0.173***	-0.184***	-0.194***	-0.198***	-0.198***	-0.212***	-0.251***
transport & communication	-0.089***	-0.102***	-0.127***	-0.126***	-0.139***	-0.157***	-0.171***	-0.187***	-0.187***	-0.201***	-0.239***
business services	-0.067***	-0.068***	-0.093***	-0.095***	-0.107***	-0.121***	-0.149***	-0.167***	-0.168***	-0.168***	-0.242***
consumer services	-0.280***	-0.279***	-0.296***	-0.291***	-0.290***	-0.304***	-0.329***	-0.341***	-0.349***	-0.361***	-0.388***
education	-0.060***	-0.066***	-0.075***	-0.093***	-0.096***	-0.095***	-0.102***	-0.117***	-0.112***	-0.123***	-0.169***
public administration	-0.143***	-0.151***	-0.159***	-0.156***	-0.163***	-0.174***	-0.162***	-0.168***	-0.164***	-0.170***	-0.200***
ND	-0.100***	-0.098***	-0.098***	-0.100***	-0.099***	-0.100***	-0.106***	-0.107***	-0.107***	-0.109***	-0.105***
HS	0.083***	0.113***	0.105***	0.103***	0.068***	0.049***	-0.012***	0.003***	-0.006***	0.010***	0.002***
HSVT	0.132***	0.137***	0.136***	0.137***	0.139***	0.143***	0.161***	0.164***	0.167***	0.170***	0.169***
TC	0.277***	0.284***	0.278***	0.277***	0.285***	0.289***	0.317***	0.314***	0.319***	0.331***	0.314***
UD	0.414***	0.426***	0.420***	0.400***	0.408***	0.440***	0.489***	0.486***	0.493***	0.519***	0.505***
age	0.143***	0.129***	0.108***	0.097***	0.073***	0.066***	0.088***	0.082***	0.095***	0.102***	0.069***
age_sq	-0.015***	-0.013***	-0.010***	-0.009***	-0.006***	-0.005***	-0.008***	-0.008***	-0.010***	-0.011***	-0.007***
3–6 months	0.050***	0.062***	0.050***	0.071***	0.047***	0.066***	0.075***	0.095***	0.079***	0.095***	0.071***
6–12 months	0.101***	0.115***	0.110***	0.132***	0.107***	0.131***	0.166***	0.167***	0.177***	0.176***	0.170***
12–24 months	0.168***	0.162***	0.162***	0.186***	0.173***	0.184***	0.252***	0.252***	0.256***	0.263***	0.250***
2–5 years	0.290***	0.286***	0.270***	0.299***	0.289***	0.300***	0.362***	0.367***	0.376***	0.389***	0.393***
5–10 years	0.380***	0.391***	0.388***	0.419***	0.421***	0.429***	0.490***	0.493***	0.488***	0.507***	0.526***
>10 years	0.498***	0.511***	0.511***	0.547***	0.558***	0.572***	0.645***	0.649***	0.647***	0.666***	0.689***
underreport	-0.126***	-0.125***	-0.129***	-0.127***	-0.128***	-0.134***	-0.154***	-0.157***	-0.159***	-0.159***	-0.156***
reportmiss	-0.094***	-0.094***	-0.101***	-0.107***	-0.111***	-0.122***	-0.154***	-0.166***	-0.173***	-0.186***	-0.184***
intercept	0.010***	3.658***	3.728***	3.707***	3.737***	3.745***	3.645***	3.669***	3.648***	3.619***	3.711***
σ	0.001***	0.276***	0.278***	0.279***	0.285***	0.297***	0.340***	0.349***	0.356***	0.361***	0.343***

Table 16: Estimates for IP2 improved Education Variable via “EM by the method of weights”

	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
foreign	-0.036***	-0.035***	-0.030***	-0.036***	-0.034***	-0.028***	-0.021***	-0.015***	-0.017***	-0.017***	-0.025***
farmer	-0.135***	-0.123***	-0.135***	-0.126***	-0.119***	-0.123***	-0.177***	-0.182***	-0.199***	-0.190***	-0.112***
service worker	0.003***	0.009***	0.010***	0.010***	0.012***	0.019***	0.004***	-0.006***	-0.012***	-0.009***	0.000***
sales worker	0.209***	0.220***	0.226***	0.236***	0.249***	0.254***	0.273***	0.275***	0.276***	0.284***	0.297***
clerical worker	0.170***	0.181***	0.182***	0.193***	0.202***	0.212***	0.225***	0.227***	0.231***	0.245***	0.249***
admin worker	0.267***	0.270***	0.267***	0.275***	0.282***	0.284***	0.303***	0.305***	0.304***	0.317***	0.311***
agriculture	0.013***	0.011***	0.003***	0.012***	0.010***	-0.001***	-0.028***	-0.051***	-0.046***	-0.040***	-0.215***
basic industry	0.006***	0.021***	0.010***	0.007***	0.006***	0.001***	-0.006***	-0.016***	-0.014***	-0.015***	-0.092***
clothes, paper & print	-0.063***	-0.056***	-0.076***	-0.076***	-0.084***	-0.090***	-0.098***	-0.107***	-0.105***	-0.114***	-0.117***
food industry	-0.093***	-0.086***	-0.114***	-0.111***	-0.113***	-0.128***	-0.134***	-0.149***	-0.144***	-0.146***	-0.186***
construction	-0.031***	-0.019***	-0.051***	-0.055***	-0.073***	-0.090***	-0.097***	-0.117***	-0.125***	-0.134***	-0.178***
trade	-0.151***	-0.148***	-0.166***	-0.164***	-0.172***	-0.182***	-0.193***	-0.197***	-0.196***	-0.211***	-0.249***
transport & communication	-0.088***	-0.101***	-0.126***	-0.125***	-0.138***	-0.156***	-0.170***	-0.186***	-0.186***	-0.200***	-0.238***
business services	-0.066***	-0.068***	-0.093***	-0.094***	-0.106***	-0.121***	-0.148***	-0.165***	-0.166***	-0.167***	-0.242***
consumer services	-0.279***	-0.278***	-0.296***	-0.290***	-0.290***	-0.304***	-0.329***	-0.341***	-0.350***	-0.362***	-0.389***
education	-0.061***	-0.067***	-0.076***	-0.093***	-0.096***	-0.096***	-0.103***	-0.118***	-0.113***	-0.124***	-0.169***
public administration	-0.143***	-0.150***	-0.159***	-0.156***	-0.163***	-0.174***	-0.162***	-0.168***	-0.164***	-0.170***	-0.200***
ND	-0.101***	-0.099***	-0.099***	-0.101***	-0.100***	-0.100***	-0.106***	-0.105***	-0.106***	-0.107***	-0.100***
HS	0.078***	0.104***	0.115***	0.107***	0.083***	0.066***	0.019***	0.020***	0.015***	0.027***	0.047***
HSVT	0.148***	0.154***	0.150***	0.153***	0.156***	0.160***	0.179***	0.181***	0.187***	0.192***	0.193***
TC	0.285***	0.290***	0.285***	0.282***	0.292***	0.295***	0.324***	0.324***	0.330***	0.345***	0.328***
UD	0.427***	0.440***	0.434***	0.412***	0.419***	0.453***	0.504***	0.502***	0.507***	0.536***	0.524***
age	0.142***	0.128***	0.108***	0.097***	0.073***	0.067***	0.088***	0.083***	0.097***	0.104***	0.073***
age_sq	-0.015***	-0.013***	-0.010***	-0.009***	-0.006***	-0.005***	-0.008***	-0.008***	-0.010***	-0.011***	-0.008***
3–6 months	0.049***	0.062***	0.050***	0.070***	0.048***	0.068***	0.078***	0.095***	0.082***	0.097***	0.077***
6–12 months	0.098***	0.113***	0.108***	0.128***	0.106***	0.131***	0.167***	0.168***	0.178***	0.174***	0.180***
12–24 months	0.161***	0.154***	0.157***	0.180***	0.168***	0.182***	0.249***	0.249***	0.254***	0.257***	0.257***
2–5 years	0.284***	0.278***	0.263***	0.290***	0.281***	0.297***	0.359***	0.362***	0.374***	0.382***	0.399***
5–10 years	0.374***	0.383***	0.381***	0.410***	0.413***	0.425***	0.486***	0.487***	0.483***	0.498***	0.531***
>10 years	0.492***	0.503***	0.504***	0.538***	0.550***	0.568***	0.640***	0.643***	0.642***	0.657***	0.693***
underreport	-0.122***	-0.120***	-0.122***	-0.120***	-0.122***	-0.131***	-0.149***	-0.157***	-0.157***	-0.156***	-0.147***
overreport	-0.064***	-0.062***	-0.042***	-0.060***	-0.053***	-0.023***	-0.038***	-0.020***	-0.012***	-0.009***	0.106***
reportmiss	-0.088***	-0.087***	-0.093***	-0.098***	-0.103***	-0.111***	-0.142***	-0.152***	-0.158***	-0.170***	-0.166***
intercept	3.674***	3.667***	3.736***	3.716***	3.743***	3.748***	3.648***	3.670***	3.647***	3.624***	3.697***
σ	0.272***	0.276***	0.278***	0.279***	0.285***	0.297***	0.340***	0.349***	0.355***	0.361***	0.343***

Table 17: Estimates for IP2B improved Education Variable via "EM by the method of weights"

	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
foreign	-0.037***	-0.035***	-0.030***	-0.036***	-0.034***	-0.028***	-0.020***	-0.015***	-0.017***	-0.016***	-0.023***
farmer	-0.136***	-0.124***	-0.136***	-0.127***	-0.121***	-0.124***	-0.178***	-0.183***	-0.199***	-0.191***	-0.111***
service worker	0.002***	0.009***	0.010***	0.010***	0.012***	0.019***	0.004***	-0.006***	-0.012***	-0.010***	0.000***
sales worker	0.211***	0.223***	0.228***	0.239***	0.251***	0.257***	0.276***	0.277***	0.280***	0.287***	0.301***
clerical worker	0.171***	0.182***	0.184***	0.194***	0.204***	0.213***	0.226***	0.227***	0.232***	0.245***	0.250***
admin worker	0.268***	0.271***	0.269***	0.277***	0.284***	0.286***	0.305***	0.307***	0.306***	0.320***	0.313***
agriculture	0.013***	0.011***	0.004***	0.012***	0.010***	-0.001***	-0.027***	-0.051***	-0.045***	-0.039***	-0.214***
basic industry	0.006***	0.020***	0.009***	0.007***	0.005***	0.001***	-0.006***	-0.016***	-0.014***	-0.015***	-0.092***
clothes, paper & print	-0.064***	-0.057***	-0.077***	-0.077***	-0.084***	-0.091***	-0.099***	-0.108***	-0.106***	-0.114***	-0.118***
food industry	-0.094***	-0.086***	-0.115***	-0.112***	-0.114***	-0.128***	-0.134***	-0.149***	-0.144***	-0.146***	-0.186***
construction	-0.031***	-0.019***	-0.051***	-0.055***	-0.073***	-0.090***	-0.097***	-0.116***	-0.124***	-0.134***	-0.178***
trade	-0.152***	-0.149***	-0.167***	-0.165***	-0.173***	-0.184***	-0.194***	-0.198***	-0.198***	-0.212***	-0.251***
transport & communication	-0.089***	-0.101***	-0.127***	-0.126***	-0.139***	-0.156***	-0.171***	-0.186***	-0.187***	-0.201***	-0.240***
business services	-0.066***	-0.068***	-0.093***	-0.094***	-0.106***	-0.120***	-0.148***	-0.166***	-0.167***	-0.167***	-0.242***
consumer services	-0.279***	-0.278***	-0.296***	-0.290***	-0.290***	-0.303***	-0.329***	-0.341***	-0.349***	-0.361***	-0.389***
education	-0.061***	-0.067***	-0.075***	-0.092***	-0.096***	-0.095***	-0.103***	-0.118***	-0.112***	-0.124***	-0.169***
public administration	-0.143***	-0.151***	-0.159***	-0.156***	-0.164***	-0.175***	-0.163***	-0.169***	-0.165***	-0.171***	-0.200***
ND	-0.097***	-0.096***	-0.096***	-0.098***	-0.097***	-0.098***	-0.104***	-0.103***	-0.103***	-0.105***	-0.101***
HS	0.083***	0.109***	0.108***	0.103***	0.069***	0.049***	-0.008***	0.007***	0.005***	0.020***	0.022***
HSVT	0.128***	0.132***	0.128***	0.130***	0.133***	0.138***	0.156***	0.159***	0.163***	0.166***	0.165***
TC	0.280***	0.285***	0.279***	0.276***	0.286***	0.289***	0.319***	0.318***	0.323***	0.336***	0.317***
UD	0.424***	0.436***	0.429***	0.407***	0.414***	0.448***	0.499***	0.497***	0.502***	0.530***	0.515***
age	0.141***	0.127***	0.107***	0.096***	0.072***	0.065***	0.086***	0.081***	0.093***	0.100***	0.067***
age_sq	-0.015***	-0.013***	-0.010***	-0.009***	-0.006***	-0.005***	-0.008***	-0.008***	-0.010***	-0.011***	-0.007***
3–6 months	0.048***	0.062***	0.049***	0.069***	0.047***	0.065***	0.076***	0.093***	0.080***	0.096***	0.073***
6–12 months	0.095***	0.110***	0.103***	0.127***	0.102***	0.126***	0.163***	0.163***	0.174***	0.173***	0.174***
12–24 months	0.158***	0.151***	0.151***	0.175***	0.161***	0.175***	0.243***	0.241***	0.247***	0.253***	0.249***
2–5 years	0.278***	0.273***	0.256***	0.286***	0.274***	0.289***	0.350***	0.354***	0.365***	0.376***	0.390***
5–10 years	0.368***	0.377***	0.374***	0.405***	0.406***	0.417***	0.478***	0.480***	0.477***	0.495***	0.524***
>10 years	0.486***	0.497***	0.497***	0.533***	0.543***	0.561***	0.633***	0.636***	0.636***	0.655***	0.688***
underreport	-0.116***	-0.114***	-0.116***	-0.113***	-0.115***	-0.124***	-0.142***	-0.149***	-0.149***	-0.147***	-0.139***
overreport	-0.142***	-0.136***	-0.125***	-0.144***	-0.136***	-0.103***	-0.124***	-0.111***	-0.112***	-0.114***	-0.018***
reportmiss	-0.090***	-0.089***	-0.096***	-0.101***	-0.106***	-0.116***	-0.147***	-0.158***	-0.165***	-0.178***	-0.174***
intercept	3.680***	3.672***	3.743***	3.721***	3.751***	3.757***	3.658***	3.681***	3.659***	3.634***	3.714***
σ	0.273***	0.276***	0.278***	0.279***	0.285***	0.297***	0.340***	0.349***	0.356***	0.362***	0.343***

Table 18: Estimates for IP3 improved Education Variable via "EM by the method of weights"

	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003
foreign	-0.036***	-0.034***	-0.029***	-0.035***	-0.033***	-0.026***	-0.018***	-0.012***	-0.014***	-0.013***	-0.020***
farmer	-0.133***	-0.122***	-0.134***	-0.125***	-0.118***	-0.121***	-0.176***	-0.180***	-0.196***	-0.188***	-0.110***
service worker	0.004***	0.010***	0.012***	0.011***	0.013***	0.021***	0.005***	-0.005***	-0.010***	-0.008***	0.002***
sales worker	0.210***	0.222***	0.227***	0.238***	0.250***	0.255***	0.274***	0.276***	0.278***	0.286***	0.300***
clerical worker	0.171***	0.182***	0.183***	0.194***	0.203***	0.213***	0.225***	0.226***	0.230***	0.244***	0.249***
admin worker	0.269***	0.272***	0.270***	0.278***	0.285***	0.287***	0.305***	0.307***	0.306***	0.320***	0.314***
agriculture	0.012***	0.010***	0.003***	0.011***	0.009***	-0.002***	-0.028***	-0.052***	-0.046***	-0.040***	-0.214***
basic industry	0.007***	0.021***	0.010***	0.008***	0.006***	0.002***	-0.005***	-0.015***	-0.013***	-0.014***	-0.091***
clothes, paper & print	-0.063***	-0.056***	-0.076***	-0.076***	-0.083***	-0.090***	-0.098***	-0.107***	-0.105***	-0.113***	-0.117***
food industry	-0.093***	-0.086***	-0.114***	-0.110***	-0.113***	-0.127***	-0.133***	-0.148***	-0.144***	-0.146***	-0.185***
construction	-0.031***	-0.018***	-0.050***	-0.055***	-0.073***	-0.090***	-0.097***	-0.117***	-0.124***	-0.134***	-0.178***
trade	-0.151***	-0.149***	-0.167***	-0.165***	-0.173***	-0.184***	-0.194***	-0.198***	-0.198***	-0.212***	-0.251***
transport & communication	-0.089***	-0.102***	-0.127***	-0.126***	-0.139***	-0.157***	-0.172***	-0.187***	-0.187***	-0.202***	-0.240***
business services	-0.067***	-0.068***	-0.093***	-0.094***	-0.107***	-0.121***	-0.149***	-0.167***	-0.167***	-0.167***	-0.242***
consumer services	-0.280***	-0.279***	-0.296***	-0.291***	-0.290***	-0.304***	-0.329***	-0.341***	-0.350***	-0.362***	-0.389***
education	-0.061***	-0.066***	-0.075***	-0.093***	-0.096***	-0.095***	-0.102***	-0.117***	-0.112***	-0.123***	-0.169***
public administration	-0.144***	-0.151***	-0.159***	-0.156***	-0.164***	-0.175***	-0.163***	-0.169***	-0.165***	-0.171***	-0.200***
ND	-0.097***	-0.096***	-0.097***	-0.098***	-0.098***	-0.099***	-0.106***	-0.105***	-0.106***	-0.108***	-0.104***
HS	0.104***	0.127***	0.124***	0.116***	0.077***	0.057***	0.005***	0.013***	0.004***	0.017***	0.015***
HSVT	0.140***	0.142***	0.142***	0.144***	0.146***	0.150***	0.166***	0.167***	0.171***	0.174***	0.171***
TC	0.280***	0.285***	0.279***	0.278***	0.286***	0.291***	0.318***	0.316***	0.322***	0.333***	0.315***
UD	0.421***	0.432***	0.426***	0.405***	0.412***	0.444***	0.494***	0.493***	0.498***	0.523***	0.509***
age	0.145***	0.130***	0.109***	0.098***	0.074***	0.068***	0.089***	0.083***	0.096***	0.104***	0.071***
age_sq	-0.015***	-0.013***	-0.011***	-0.009***	-0.006***	-0.006***	-0.008***	-0.008***	-0.010***	-0.011***	-0.008***
3–6 months	0.050***	0.062***	0.050***	0.071***	0.047***	0.067***	0.075***	0.096***	0.081***	0.097***	0.073***
6–12 months	0.102***	0.116***	0.111***	0.133***	0.108***	0.132***	0.168***	0.169***	0.180***	0.178***	0.173***
12–24 months	0.169***	0.163***	0.163***	0.188***	0.174***	0.186***	0.254***	0.254***	0.258***	0.266***	0.254***
2–5 years	0.292***	0.288***	0.271***	0.301***	0.290***	0.302***	0.365***	0.370***	0.380***	0.392***	0.397***
5–10 years	0.382***	0.393***	0.390***	0.421***	0.423***	0.431***	0.493***	0.496***	0.491***	0.511***	0.531***
>10 years	0.500***	0.512***	0.512***	0.549***	0.559***	0.574***	0.648***	0.652***	0.650***	0.669***	0.693***
underreport	-0.136***	-0.133***	-0.136***	-0.134***	-0.135***	-0.143***	-0.160***	-0.165***	-0.167***	-0.165***	-0.160***
overreport	0.050***	0.056***	0.059***	0.058***	0.061***	0.061***	0.080***	0.064***	0.084***	0.114***	0.138***
reportmiss	-0.095***	-0.094***	-0.101***	-0.107***	-0.112***	-0.122***	-0.153***	-0.166***	-0.173***	-0.186***	-0.184***
intercept	3.661***	3.654***	3.725***	3.704***	3.734***	3.741***	3.642***	3.664***	3.641***	3.613***	3.703***
σ	0.273***	0.276***	0.278***	0.279***	0.285***	0.297***	0.340***	0.349***	0.355***	0.361***	0.343***

Table 19: Estimates for IP1 Rule-Based Improved Education Variable without interns, working students and partly retired workers

	1999	2000	2001	2002	2003
foreign	-0.021***	-0.014***	-0.016***	-0.015***	-0.020***
farmer	-0.179***	-0.184***	-0.199***	-0.192***	-0.114***
service worker	0.006***	-0.004***	-0.010***	-0.007***	0.002***
sales worker	0.272***	0.274***	0.276***	0.283***	0.295***
clerical worker	0.226***	0.228***	0.233***	0.245***	0.249***
admin worker	0.301***	0.304***	0.302***	0.316***	0.310***
agriculture	-0.028***	-0.052***	-0.046***	-0.040***	-0.213***
basic industry	-0.005***	-0.014***	-0.013***	-0.013***	-0.091***
clothes, paper & print	-0.098***	-0.107***	-0.105***	-0.113***	-0.116***
food industry	-0.134***	-0.148***	-0.143***	-0.145***	-0.185***
construction	-0.098***	-0.118***	-0.125***	-0.134***	-0.178***
trade	-0.195***	-0.199***	-0.198***	-0.212***	-0.251***
transport & communication	-0.172***	-0.187***	-0.187***	-0.202***	-0.240***
business services	-0.147***	-0.165***	-0.166***	-0.166***	-0.240***
consumer services	-0.328***	-0.339***	-0.348***	-0.359***	-0.383***
education	-0.102***	-0.116***	-0.112***	-0.123***	-0.169***
public administration	-0.164***	-0.169***	-0.165***	-0.171***	-0.200***
ND	-0.106***	-0.108***	-0.108***	-0.109***	-0.106***
HS	0.077***	0.101***	0.110***	0.127***	0.116***
HSVT	0.166***	0.171***	0.174***	0.175***	0.175***
TC	0.316***	0.314***	0.319***	0.333***	0.315***
UD	0.486***	0.485***	0.492***	0.519***	0.506***
age	0.086***	0.078***	0.089***	0.096***	0.065***
age_sq	-0.008***	-0.008***	-0.009***	-0.010***	-0.007***
3–6 months	0.084***	0.104***	0.085***	0.106***	0.083***
6–12 months	0.162***	0.174***	0.184***	0.178***	0.181***
12–24 months	0.241***	0.259***	0.254***	0.267***	0.255***
2–5 years	0.340***	0.357***	0.366***	0.383***	0.392***
5–10 years	0.466***	0.479***	0.472***	0.494***	0.518***
>10 years	0.620***	0.636***	0.632***	0.654***	0.682***
underreport	-0.152***	-0.156***	-0.158***	-0.159***	-0.155***
reportmiss	-0.151***	-0.162***	-0.171***	-0.183***	-0.181***
intercept	3.674***	3.688***	3.673***	3.644***	3.728***
σ	0.337***	0.345***	0.351***	0.357***	0.339***

Table 20: Estimates for IP2 Rule-Based Improved Education Variable without interns, working students and partly retired workers

	1999	2000	2001	2002	2003
foreign	-0.024***	-0.018***	-0.020***	-0.019***	-0.026***
farmer	-0.179***	-0.183***	-0.199***	-0.192***	-0.114***
service worker	0.005***	-0.004***	-0.010***	-0.008***	0.001***
sales worker	0.272***	0.274***	0.276***	0.282***	0.295***
clerical worker	0.228***	0.231***	0.236***	0.248***	0.252***
admin worker	0.303***	0.305***	0.303***	0.316***	0.310***
agriculture	-0.028***	-0.052***	-0.046***	-0.040***	-0.214***
basic industry	-0.006***	-0.015***	-0.014***	-0.014***	-0.092***
clothes, paper & print	-0.098***	-0.107***	-0.105***	-0.113***	-0.116***
food industry	-0.135***	-0.149***	-0.144***	-0.146***	-0.186***
construction	-0.098***	-0.117***	-0.125***	-0.134***	-0.178***
trade	-0.193***	-0.197***	-0.196***	-0.211***	-0.250***
transport & communication	-0.170***	-0.186***	-0.186***	-0.201***	-0.239***
business services	-0.146***	-0.164***	-0.164***	-0.165***	-0.241***
consumer services	-0.329***	-0.339***	-0.348***	-0.360***	-0.383***
education	-0.103***	-0.117***	-0.113***	-0.124***	-0.168***
public administration	-0.164***	-0.169***	-0.166***	-0.171***	-0.201***
ND	-0.106***	-0.106***	-0.107***	-0.108***	-0.100***
HS	0.099***	0.111***	0.121***	0.125***	0.135***
HSVT	0.181***	0.185***	0.192***	0.196***	0.196***
TC	0.322***	0.323***	0.329***	0.345***	0.329***
UD	0.501***	0.499***	0.506***	0.535***	0.523***
age	0.086***	0.080***	0.091***	0.097***	0.068***
age_sq	-0.008***	-0.008***	-0.010***	-0.011***	-0.007***
3–6 months	0.086***	0.103***	0.087***	0.107***	0.089***
6–12 months	0.162***	0.174***	0.184***	0.175***	0.190***
12–24 months	0.237***	0.256***	0.251***	0.260***	0.259***
2–5 years	0.336***	0.353***	0.363***	0.375***	0.396***
5–10 years	0.460***	0.473***	0.466***	0.484***	0.520***
>10 years	0.615***	0.630***	0.626***	0.644***	0.683***
underreport	-0.148***	-0.155***	-0.156***	-0.155***	-0.146***
overreport	-0.039***	-0.022***	-0.019***	-0.015***	0.094***
reportmiss	-0.138***	-0.148***	-0.155***	-0.166***	-0.163***
intercept	3.677***	3.690***	3.675***	3.651***	3.718***
σ	0.336***	0.345***	0.351***	0.357***	0.339***

Table 21: Estimates for IP2b Rule-Based Improved Education Variable without interns, working students and partly retired workers

	1999	2000	2001	2002	2003
foreign	-0.023***	-0.018***	-0.020***	-0.019***	-0.025***
farmer	-0.179***	-0.184***	-0.199***	-0.192***	-0.113***
service worker	0.005***	-0.005***	-0.011***	-0.008***	0.001***
sales worker	0.275***	0.277***	0.279***	0.286***	0.298***
clerical worker	0.229***	0.231***	0.236***	0.248***	0.253***
admin worker	0.305***	0.306***	0.305***	0.318***	0.312***
agriculture	-0.028***	-0.051***	-0.045***	-0.039***	-0.214***
basic industry	-0.006***	-0.015***	-0.014***	-0.014***	-0.092***
clothes, paper & print	-0.099***	-0.108***	-0.105***	-0.114***	-0.118***
food industry	-0.135***	-0.149***	-0.144***	-0.146***	-0.187***
construction	-0.098***	-0.116***	-0.124***	-0.134***	-0.178***
trade	-0.195***	-0.198***	-0.198***	-0.212***	-0.251***
transport & communication	-0.171***	-0.186***	-0.187***	-0.202***	-0.241***
business services	-0.146***	-0.164***	-0.165***	-0.166***	-0.241***
consumer services	-0.329***	-0.339***	-0.347***	-0.359***	-0.383***
education	-0.103***	-0.117***	-0.113***	-0.124***	-0.168***
public administration	-0.164***	-0.169***	-0.166***	-0.172***	-0.201***
ND	-0.104***	-0.104***	-0.104***	-0.106***	-0.101***
HS	0.079***	0.108***	0.123***	0.134***	0.127***
HSVT	0.160***	0.165***	0.170***	0.171***	0.170***
TC	0.317***	0.317***	0.323***	0.337***	0.318***
UD	0.496***	0.495***	0.501***	0.529***	0.515***
age	0.084***	0.077***	0.087***	0.093***	0.062***
age_sq	-0.008***	-0.008***	-0.009***	-0.010***	-0.007***
3–6 months	0.084***	0.102***	0.085***	0.106***	0.086***
6–12 months	0.159***	0.170***	0.181***	0.174***	0.185***
12–24 months	0.232***	0.248***	0.245***	0.257***	0.251***
2–5 years	0.329***	0.345***	0.355***	0.370***	0.388***
5–10 years	0.454***	0.467***	0.460***	0.481***	0.514***
>10 years	0.609***	0.625***	0.621***	0.643***	0.679***
underreport	-0.140***	-0.148***	-0.148***	-0.147***	-0.138***
overreport	-0.124***	-0.111***	-0.119***	-0.117***	-0.026***
reportmiss	-0.144***	-0.155***	-0.163***	-0.174***	-0.171***
intercept	3.686***	3.700***	3.687***	3.661***	3.735***
σ	0.337***	0.345***	0.351***	0.357***	0.339***

Table 22: Estimates for IP3 Rule-Based Improved Education Variable without interns, working students and partly retired workers

	1999	2000	2001	2002	2003
foreign	-0.021***	-0.015***	-0.017***	-0.015***	-0.021***
farmer	-0.177***	-0.182***	-0.197***	-0.190***	-0.112***
service worker	0.007***	-0.004***	-0.009***	-0.007***	0.003***
sales worker	0.273***	0.275***	0.277***	0.284***	0.297***
clerical worker	0.227***	0.230***	0.234***	0.246***	0.251***
admin worker	0.305***	0.306***	0.305***	0.319***	0.313***
agriculture	-0.029***	-0.052***	-0.046***	-0.040***	-0.213***
basic industry	-0.005***	-0.014***	-0.013***	-0.013***	-0.091***
clothes, paper & print	-0.098***	-0.107***	-0.104***	-0.113***	-0.116***
food industry	-0.134***	-0.148***	-0.144***	-0.146***	-0.186***
construction	-0.098***	-0.117***	-0.124***	-0.134***	-0.178***
trade	-0.195***	-0.198***	-0.198***	-0.212***	-0.251***
transport & communication	-0.172***	-0.187***	-0.187***	-0.202***	-0.241***
business services	-0.147***	-0.165***	-0.165***	-0.166***	-0.240***
consumer services	-0.329***	-0.339***	-0.348***	-0.359***	-0.383***
education	-0.102***	-0.116***	-0.112***	-0.123***	-0.169***
public administration	-0.164***	-0.169***	-0.166***	-0.172***	-0.201***
ND	-0.106***	-0.106***	-0.107***	-0.109***	-0.105***
HS	0.092***	0.111***	0.122***	0.130***	0.124***
HSVT	0.172***	0.175***	0.178***	0.179***	0.177***
TC	0.317***	0.315***	0.322***	0.335***	0.316***
UD	0.491***	0.491***	0.497***	0.523***	0.509***
age	0.086***	0.079***	0.091***	0.097***	0.066***
age_sq	-0.008***	-0.008***	-0.009***	-0.010***	-0.007***
3–6 months	0.084***	0.105***	0.086***	0.107***	0.085***
6–12 months	0.163***	0.176***	0.186***	0.179***	0.184***
12–24 months	0.242***	0.261***	0.256***	0.270***	0.258***
2–5 years	0.342***	0.360***	0.370***	0.386***	0.396***
5–10 years	0.468***	0.482***	0.475***	0.497***	0.522***
>10 years	0.623***	0.639***	0.635***	0.657***	0.685***
underreport	-0.158***	-0.163***	-0.165***	-0.164***	-0.159***
overreport	0.079***	0.071***	0.078***	0.107***	0.129***
reportmiss	-0.150***	-0.162***	-0.171***	-0.183***	-0.180***
intercept	3.671***	3.684***	3.669***	3.640***	3.722***
σ	0.337***	0.345***	0.351***	0.357***	0.339***

Table 23: Estimates for IP1 improved Education Variable via "EM by the method of weights" without interns, working students and partly retired workers

	1999	2000	2001	2002	2003
foreign	0.010***	0.009***	0.006***	0.010***	0.008***
farmer	-0.188***	-0.202***	-0.214***	-0.212***	-0.133***
service worker	-0.015***	-0.047***	-0.049***	-0.051***	-0.033***
sales worker	0.274***	0.274***	0.275***	0.284***	0.285***
clerical worker	0.212***	0.206***	0.208***	0.218***	0.217***
admin worker	0.298***	0.298***	0.293***	0.304***	0.287***
agriculture	-0.026***	-0.046***	-0.036***	-0.030***	-0.209***
basic industry	0.003***	-0.006***	-0.001***	-0.005***	-0.087***
clothes, paper & print	-0.091***	-0.106***	-0.100***	-0.112***	-0.120***
food industry	-0.132***	-0.149***	-0.141***	-0.145***	-0.182***
construction	-0.097***	-0.116***	-0.119***	-0.135***	-0.183***
trade	-0.200***	-0.212***	-0.207***	-0.225***	-0.260***
transport & communication	-0.148***	-0.160***	-0.152***	-0.172***	-0.216***
business services	-0.152***	-0.187***	-0.177***	-0.183***	-0.242***
consumer services	-0.311***	-0.343***	-0.343***	-0.355***	-0.446***
education	-0.122***	-0.149***	-0.141***	-0.157***	-0.202***
public administration	-0.167***	-0.180***	-0.174***	-0.179***	-0.209***
ND	-0.135***	-0.141***	-0.143***	-0.148***	-0.146***
HS	-0.552***	-0.060***	0.178***	0.241***	0.264***
HSVT	0.140***	0.162***	0.153***	0.160***	0.160***
TC	0.290***	0.323***	0.329***	0.350***	0.336***
UD	0.515***	0.564***	0.579***	0.615***	0.609***
age	0.301***	0.294***	0.296***	0.310***	0.298***
age_sq	-0.030***	-0.030***	-0.031***	-0.032***	-0.030***
3–6 months	0.166***	0.329***	0.238***	0.295***	0.199***
6–12 months	0.251***	0.393***	0.372***	0.405***	0.309***
12–24 months	0.294***	0.466***	0.424***	0.483***	0.409***
2–5 years	0.396***	0.563***	0.528***	0.598***	0.539***
5–10 years	0.543***	0.707***	0.669***	0.737***	0.698***
>10 years	0.679***	0.845***	0.811***	0.882***	0.849***
underreport	-0.270***	-0.328***	-0.349***	-0.380***	-0.390***
overreport	-1.387***	-1.521***	-1.452***	-1.716***	-1.988***
reportmiss	-0.273***	-0.379***	-0.398***	-0.449***	-0.471***
intercept	3.112***	2.979***	3.011***	2.914***	3.009***
σ	0.352***	0.386***	0.384***	0.391***	0.370***

Table 24: Estimates for IP2 improved Education Variable via "EM by the method of weights" without interns, working students and partly retired workers

	1999	2000	2001	2002	2003
foreign	0.007***	0.006***	0.003***	0.007***	0.006***
farmer	-0.188***	-0.200***	-0.214***	-0.212***	-0.130***
service worker	-0.018***	-0.045***	-0.048***	-0.047***	-0.030***
sales worker	0.268***	0.273***	0.275***	0.283***	0.284***
clerical worker	0.204***	0.204***	0.208***	0.218***	0.216***
admin worker	0.290***	0.294***	0.289***	0.300***	0.284***
agriculture	-0.028***	-0.048***	-0.038***	-0.032***	-0.209***
basic industry	0.002***	-0.008***	-0.002***	-0.007***	-0.087***
clothes. paper & print	-0.092***	-0.106***	-0.099***	-0.111***	-0.120***
food industry	-0.132***	-0.149***	-0.141***	-0.145***	-0.182***
construction	-0.094***	-0.115***	-0.119***	-0.135***	-0.181***
trade	-0.197***	-0.209***	-0.204***	-0.223***	-0.258***
transport & communication	-0.150***	-0.159***	-0.150***	-0.171***	-0.214***
business services	-0.159***	-0.190***	-0.180***	-0.186***	-0.244***
consumer services	-0.332***	-0.347***	-0.342***	-0.356***	-0.445***
education	-0.130***	-0.152***	-0.144***	-0.161***	-0.204***
public administration	-0.167***	-0.181***	-0.173***	-0.178***	-0.206***
ND	-0.135***	-0.140***	-0.144***	-0.148***	-0.148***
HS	-0.009***	0.054***	0.113***	0.209***	0.242***
HSVT	0.169***	0.176***	0.172***	0.179***	0.182***
TC	0.321***	0.338***	0.343***	0.361***	0.342***
UD	0.558***	0.584***	0.599***	0.631***	0.620***
age	0.299***	0.295***	0.297***	0.312***	0.297***
age_sq	-0.030***	-0.030***	-0.031***	-0.032***	-0.030***
3–6 months	0.191***	0.353***	0.279***	0.307***	0.199***
6–12 months	0.290***	0.438***	0.419***	0.416***	0.312***
12–24 months	0.337***	0.506***	0.473***	0.496***	0.413***
2–5 years	0.446***	0.606***	0.578***	0.617***	0.547***
5–10 years	0.599***	0.751***	0.717***	0.755***	0.703***
>10 years	0.742***	0.891***	0.861***	0.902***	0.855***
underreport	-0.263***	-0.305***	-0.327***	-0.361***	-0.374***
overreport	-0.259***	-0.341***	-0.297***	-0.411***	-0.626***
reportmiss	-0.291***	-0.366***	-0.384***	-0.437***	-0.454***
intercept	3.059***	2.934***	2.960***	2.892***	3.006***
σ	0.358***	0.386***	0.384***	0.391***	0.369***

Table 25: Estimates for IP2B improved Education Variable via “EM by the method of weights” without interns, working students and partly retired workers

	1999	2000	2001	2002	2003
foreign	0.008***	0.006***	0.003***	0.006***	0.005***
farmer	-0.185***	-0.200***	-0.213***	-0.211***	-0.130***
service worker	-0.014***	-0.046***	-0.048***	-0.048***	-0.031***
sales worker	0.275***	0.273***	0.275***	0.284***	0.285***
clerical worker	0.213***	0.206***	0.209***	0.219***	0.217***
admin worker	0.298***	0.296***	0.291***	0.303***	0.287***
agriculture	-0.027***	-0.048***	-0.038***	-0.031***	-0.211***
basic industry	0.002***	-0.007***	-0.002***	-0.007***	-0.088***
clothes, paper & print	-0.092***	-0.106***	-0.099***	-0.112***	-0.120***
food industry	-0.133***	-0.150***	-0.142***	-0.147***	-0.182***
construction	-0.096***	-0.115***	-0.118***	-0.135***	-0.181***
trade	-0.199***	-0.210***	-0.205***	-0.224***	-0.259***
transport & communication	-0.148***	-0.160***	-0.151***	-0.172***	-0.216***
business services	-0.154***	-0.189***	-0.180***	-0.185***	-0.244***
consumer services	-0.316***	-0.347***	-0.344***	-0.357***	-0.447***
education	-0.125***	-0.151***	-0.143***	-0.160***	-0.204***
public administration	-0.167***	-0.181***	-0.174***	-0.179***	-0.207***
ND	-0.133***	-0.138***	-0.140***	-0.143***	-0.141***
HS	-0.483***	-0.019***	0.101***	0.227***	0.253***
HSVT	0.150***	0.174***	0.167***	0.168***	0.174***
TC	0.303***	0.333***	0.339***	0.357***	0.338***
UD	0.531***	0.577***	0.591***	0.623***	0.614***
age	0.303***	0.295***	0.297***	0.311***	0.296***
age_sq	-0.030***	-0.030***	-0.031***	-0.032***	-0.030***
3–6 months	0.189***	0.348***	0.275***	0.307***	0.199***
6–12 months	0.278***	0.422***	0.405***	0.419***	0.310***
12–24 months	0.324***	0.492***	0.459***	0.498***	0.409***
2–5 years	0.426***	0.589***	0.562***	0.613***	0.541***
5–10 years	0.574***	0.734***	0.701***	0.753***	0.700***
>10 years	0.711***	0.873***	0.845***	0.899***	0.853***
underreport	-0.255***	-0.311***	-0.334***	-0.365***	-0.376***
overreport	-0.259***	-0.394***	-0.394***	-0.516***	-0.760***
reportmiss	-0.270***	-0.370***	-0.386***	-0.437***	-0.459***
intercept	3.077***	2.950***	2.977***	2.896***	3.011***
σ	0.354***	0.387***	0.385***	0.392***	0.370***

Table 26: Estimates for IP3 improved Education Variable via "EM by the method of weights" without interns, working students and partly retired workers

	1999	2000	2001	2002	2003
foreign	0.008***	0.007***	0.005***	0.008***	0.007***
farmer	-0.185***	-0.198***	-0.210***	-0.209***	-0.131***
service worker	-0.015***	-0.046***	-0.049***	-0.050***	-0.033***
sales worker	0.274***	0.273***	0.276***	0.285***	0.286***
clerical worker	0.212***	0.205***	0.209***	0.219***	0.218***
admin worker	0.299***	0.297***	0.293***	0.304***	0.287***
agriculture	-0.027***	-0.047***	-0.038***	-0.031***	-0.210***
basic industry	0.002***	-0.007***	-0.002***	-0.007***	-0.088***
clothes, paper & print	-0.092***	-0.106***	-0.100***	-0.112***	-0.121***
food industry	-0.132***	-0.149***	-0.142***	-0.147***	-0.184***
construction	-0.096***	-0.115***	-0.118***	-0.134***	-0.182***
trade	-0.200***	-0.211***	-0.206***	-0.225***	-0.260***
transport & communication	-0.149***	-0.161***	-0.153***	-0.172***	-0.217***
business services	-0.153***	-0.188***	-0.178***	-0.183***	-0.243***
consumer services	-0.315***	-0.346***	-0.344***	-0.356***	-0.450***
education	-0.124***	-0.150***	-0.141***	-0.159***	-0.204***
public administration	-0.168***	-0.181***	-0.175***	-0.179***	-0.209***
ND	-0.132***	-0.137***	-0.139***	-0.143***	-0.140***
HS	-0.517***	-0.004***	0.197***	0.243***	0.267***
HSVT	0.149***	0.168***	0.156***	0.160***	0.162***
TC	0.299***	0.330***	0.336***	0.355***	0.337***
UD	0.527***	0.576***	0.587***	0.624***	0.613***
age	0.301***	0.295***	0.296***	0.310***	0.296***
age_sq	-0.030***	-0.030***	-0.031***	-0.032***	-0.030***
3–6 months	0.185***	0.350***	0.280***	0.315***	0.203***
6–12 months	0.274***	0.426***	0.415***	0.433***	0.314***
12–24 months	0.324***	0.501***	0.468***	0.511***	0.418***
2–5 years	0.426***	0.601***	0.575***	0.627***	0.548***
5–10 years	0.576***	0.748***	0.718***	0.769***	0.707***
>10 years	0.714***	0.888***	0.862***	0.915***	0.861***
underreport	-0.273***	-0.334***	-0.355***	-0.388***	-0.391***
overreport	-0.055***	-0.119***	-0.158***	-0.097***	-0.322***
reportmiss	-0.273***	-0.376***	-0.395***	-0.446***	-0.467***
intercept	3.079***	2.936***	2.960***	2.880***	3.003***
σ	0.353***	0.387***	0.385***	0.392***	0.371***