LANGUAGE FLUENCY AND EARNINGS: ESTIMATION WITH MISCLASSIFIED LANGUAGE INDICATORS

Christian Dustmann and Arthur van Soest*

Abstract—We use panel data to analyze the determinants of speaking fluency and wages of immigrants. Our model takes account of two problems that may bias OLS estimates of the impact of speaking fluency on earnings. First, subjective variables on an ordinal discrete scale, such as self-reported language ability, can suffer from misclassification errors. The model decomposes misclassification errors into a time-persistent and a time-varying component. Second, the model accounts for correlated unobserved heterogeneity in language and earnings equation. The main finding is that these two generalizations of the standard model both lead to substantial changes in the estimated effect of speaking fluency on earnings.

I. Introduction

LINGUISTIC skills are an important component of host country-specific human capital of migrant workers. In the economic literature, two issues related to language of immigrants have attracted the attention of researchers: what are the determinants of language proficiency of migrants, and what is the relation between fluency in the host country's dominant language and labor market performance? Studies by, for instance, Carliner (1981), McManus, Gould, and Welch (1983), Grenier (1984), Kossoudji (1988), Rivera-Batiz (1990), Chiswick (1991), Dustmann (1994), and Chiswick and Miller (1995) analyze these issues for various countries. Most of these studies conclude that language proficiency is positively related to education level and negatively related to age at arrival, and that it improves substantially with the time spent in the host country. Moreover, the empirical studies have almost unanimously found that language efficiency has a positive effect on earnings.

The empirical work in this area draws nearly exclusively on cross-sectional data. It typically uses self-reported language ability as a measure for language proficiency, but a self-reported variable on language proficiency is likely to suffer from misclassification errors. Individuals may over- or under-evaluate their language fluency. This misclassification may bias the parameter estimates in models for the determinants of language proficiency, because these models are typically nonlinear models for discrete dependent variables. Hausman, Abrevaya, and Scott-Morton (1998) demonstrate that even small probabilities of misclassification may lead to substantially biased parameter estimates in a (binary) probit model.

To study the relationship between earnings and language, most of the studies just cited use ordinary least squares (OLS) in which the earnings variable is regressed on a set of human capital variables and on (an) indicator variable(s) for

language proficiency. Borjas (1994) argues that there may be a positive correlation between unobserved heterogeneity in earnings and speaking fluency equations, leading to an upward bias in the estimated effect of speaking fluency. In other words, the effect of language fluency on earnings may be lower than OLS estimates indicate.

On the other hand, unobserved heterogeneity may also lead to downward-biased estimates. This is the case if foregone earnings of individuals who engage in language education increase with their unobserved ability. (See Willis and Rosen (1979).) This explanation is consistent with the findings of Chiswick and Miller (1995), who use instrumental variables (IV) estimation to correct for the unobserved heterogeneity bias. They compare biased OLS and asymptotically unbiased IV estimates, using data for various countries. Although their results show substantial variation and the IV estimates often have large standard errors, most of their estimates lead to the conclusion that OLS leads to a substantial negative bias in the estimate of the speaking fluency effect on earnings.

An alternative explanation for the negative bias is misclassification error in the language variable. With categorical variables based on subjective evaluations such as self-reported speaking fluency, there are two types of misclassification errors: errors that are purely random and independent over time, and, errors that are time persistent, in the sense that certain individuals always have the same tendency to over- or under-report. Using cross-sectional data only, these two types of misclassification cannot be disentangled. For example, Hausman et al. (1998) model job changes, and they mention recall error and misunderstanding of survey questions as potential sources of misclassification. Whereas the first may lead to errors that are independent over time, the latter may largely reflect a time-persistent characteristic of certain respondents. In the cross-sectional study of Hausman et al. (1998), the distinction between these two types of errors is unidentified. For panel data, however, misclassification errors that are independent over time and those that are time persistent are identified.

The methodological contribution of this paper is that we develop a panel data model that explicitly takes account of the sources of the OLS bias just discussed. The model consists of two equations: an ordered response equation for language proficiency, and an earnings equation in which language proficiency is one of the explanatory variables. We generalize the approach of Lee and Porter (1984) and Hausman et al. (1998) to incorporate misclassification errors, and we distinguish between time-varying and time-persistent misclassification errors. Moreover, we allow for correlated unobserved heterogeneity in the form of

correlated random individual effects in the two equations. Here, we use flexible, bivariate nonparametric specifications, following Heckman and Singer (1984).

We apply our model to male immigrants in (West) Germany, using panel data for 1984–1993, with information on speaking fluency in seven waves. Both the variation within individuals in the panel data and the rich set of background variables available in the survey have given us a privileged position from which to estimate a richer model of the relationship between language and earnings than previous studies.

By comparing different specifications, we determine the consequences of both types of bias separately. Apart from random individual effects reflecting unobserved heterogeneity, both equations in our model also contain residual error terms. If the residual error terms in the two equations are assumed to be uncorrelated, the model is identified without exclusion restrictions. Our most general specification allows for correlation of these residual error terms in the two equations. For nonparametric identification of the earnings equation in this specification, we need instruments in the speaking fluency equation which do not directly affect wages. Here, we draw on the rich set of background variables available in our survey.

Our results indicate that the OLS bias on the effect of speaking fluency on earnings is significant and substantial. We establish both a positive bias due to correlated unobserved heterogeneity, and a negative bias due to misclassification errors. All our estimates of the effect of speaking fluency on earnings are positive, and our largest estimate is approximately eight times larger than the smallest one.

The paper is organized as follows. Section II presents the data. Section III presents the model for language ability and earnings. Section IV discusses the empirical results, and section V concludes.

II.   Data

The data we use is drawn from the German Socio-Economic Panel (GSOEP), an annual panel that started in 1984.[1] We use the boost subsample of the GSOEP consisting of households with a foreign-born head who immigrated to West Germany from Turkey, Yugoslavia, Italy, Greece, or Spain. The first wave of immigrants includes approximately 1,500 households. All adults in this sample answer questions about their economic behavior, as well as about their economic and social integration. We use the seven waves (1984–1987, 1989, 1991, and 1993) in which questions on language fluency are included. Language information is not reported in the 1988, 1990, and 1992 surveys.

Speaking fluency is reported on a five-point scale, with possible answers "very bad" (1), "bad" (2), "intermediate" (3), "good" (4), and "very good" (5). In our analysis, we consider males only. The first wave of the sample includes

1,613 men who provide information on self-assessed language fluency. Due to missing information on explanatory variables, 83 of these could not be used in the analysis, leaving 1,530 observations in the first wave. Due to attrition, the panel is unbalanced. Approximately 15% of the observations are lost between waves 1 and 2. Attrition is smaller in later waves. The numbers of observations used for the analysis are 1,530 in 1984; 1,299 in 1985; 1,237 in 1986; 1,210 in 1987; 1,069 in 1989; 1,024 in 1191; and 958 in 1993.

We use the standard regressors in these models to model language proficiency.[2] The years-since-migration variable picks up the effect of exposure to the host country's language. We include the year of entry to incorporate potential differences between groups of migrants who came to Germany in different years. We also include age at entry and total years of education, as well as dummy variables indicating the immigrants' nationality (Turkish, Yugoslavian, Greek, Italian, or Spanish). In all these countries, German is neither the dominant language nor the first foreign language taught at school. It is therefore likely that the individuals in our sample spoke little or no German upon immigration.

In addition, we include several dummy variables that refer to the education level of the immigrant's father. This information is drawn from the third wave of the panel, which contains information on several parental characteristics. Definitions and summary statistics of all the independent variables we use are displayed in table A1 in appendix A.

Our earnings variable is the natural logarithm of gross monthly earnings. In the earnings regressions, we include only individuals who are in full-time employment during the month to which the earnings information refers.

Table 1 presents bivariate frequency distributions of self-reported speaking fluency in consecutive years for the first four waves. The nondiagonal cells refer to changes in reported speaking fluency. There are many transitions from good to intermediate, from intermediate to bad or very bad, and so on. Although some deterioration of speaking fluency is in principle possible, the large number of below-diagonal observations strongly suggests that the self-reported language ability measure suffers from misclassification errors that vary over time.

Table 2 summarizes the changes in the speaking fluency variable (treated as a cardinal variable with values 1, 2, 3, 4, and 5) between two consecutive years, again for the first four waves. These numbers illustrate the magnitude of potential misclassification in this type of data. The distribution of the changes is nearly symmetric, with similar numbers of deteriorations and improvements. Overall, about 57% of individuals do not report any changes, and 19% report a deterioration by one category and 2.2% by more than one category. The large number of respondents whose self-reported fluency deteriorates suggests that many are

---

[1] See Wagner, Burkhauser, and Behringer (1993) for details on the GSOEP.

[2] Chiswick and Miller (1995) provide a systematic discussion of the determinants of language fluency.

TABLE 1.—CROSS-TABULATIONS OF REPORTED SPEAKING FLUENCY, 1984–1987

	1	2	3	4	5	Total
vertical: 1984; horizontal: 1985						
1	4	9	4	0	1	18
2	7	87	68	19	1	182
3	3	78	253	137	14	485
4	1	21	108	250	59	439
5	0	2	17	67	109	195
Total	15	197	450	473	184	1319
vertical: 1985; horizontal: 1986						
1	4	6	3	1	0	14
2	6	99	61	8	0	174
3	1	58	235	102	10	406
4	0	8	120	259	47	434
5	1	1	8	62	95	167
Total	12	172	427	432	152	1195
vertical: 1986; horizontal: 1987						
1	4	6	1	2	0	13
2	3	95	61	8	1	168
3	0	50	258	104	8	420
4	1	8	93	277	42	421
5	0	0	9	57	92	158
Total	8	159	422	448	143	1180

1: very bad; 2: bad; 3: intermediate; 4: good; 5: very good.

either too optimistic in the first year, or too pessimistic in the second year.

The total variance in the language indicator (on the cardinal 1-to-5 scale, all years) is 0.891. This overall variance can be decomposed in a within-individual variance of 0.253, and a between-individual variance of 0.638. To illustrate the potential importance of misclassification, assume that all reported deterioration is misclassification, that misclassification errors u_t are nonnegatively correlated over time with a time independent variance, and that the distribution of $u_t - u_{t-1}$ is symmetric around zero. Then the variance of the measurement error satisfies $V(u_t) \geq P(y_t - y_{t-1} \leq -1)$, where y_t is observed speaking fluency.[3] Table 2 gives an estimate for this lower bound on the variance of the measurement error of $\hat{P}(y_t - y_{t-1} \leq -1) = 0.214$. Thus, under the assumption that deterioration is impossible, and some auxiliary assumptions, most of the within-individuals variance, and at least 24% of the total variance, is explained by measurement error.

A possible explanation of misclassification might be that immigrants have problems in evaluating their speaking

[3] The assumptions and Chebyshev's rule imply $V(u_t) \geq V(u_t) -$ Cov$(u_t, u_{t-1}) = 0.5V(u_t - u_{t-1}) \geq 0.5P(|u_t - u_{t-1}| \geq 1) = P(u_t - u_{t-1} \leq -1) \geq P(y_t - y_{t-1} \leq -1)$.

fluency during the first year(s) after their arrival in the host country. This would lead to a negative relationship between the probability of misclassification and years since migration. To investigate this, we also present the classification changes separately for those with years since migration above and including the median, and below the median (which is fifteen years). Results are displayed in the last two rows of table 2. They do not suggest that the number of misclassifications falls with years since migration. If anything, the opposite seems to be true: the number of people who report a deterioration of speaking fluency is larger among those with years of residence above the median than among the more recent arrivals. The reason that we do not find any evidence of evaluation problems of more recent immigrants may be that this effect exists only shortly after immigration, and our sample contains hardly any recent immigrants.

In addition to the potential misclassification errors revealed by the tables, some people may persistently over- or under-report their language ability. For example, a respondent who always reports "good" may indeed always have good proficiency. He may also be on an "intermediate" level only, and persistently over-report. This type of time persistent misclassification error is not shown by the cross-tabulations.

III. The Model

In section II, we demonstrated that misclassification in self-reported language variables is substantial. In fact, most of the within-individual variation in this variable is due to misclassification. The cross-tabulations do not help to detect a second source of misclassification (time-persistent over- or under-evaluation of the true language proficiency), which may add to the overall misclassification problem in our data.

In this section, we first develop a panel data model explaining self-reported speaking fluency on a discrete ordinal scale that explicitly incorporates misclassification probabilities. The panel data nature of the model allows us to distinguish between time-persistent misclassification errors and misclassification errors that are independent over time, the two types of classification errors previously discussed. We choose a flexible but nevertheless tractable way to parameterize these two sources of misclassification. We also allow for individual specific heterogeneity, which will be captured by a nonparametric mass point distribution.

We then combine this model with a wage equation that is used to determine the effect of speaking fluency on wages.

TABLE 2.—CATEGORY CHANGES BY YEARS SINCE MIGRATION, 1984–1987

Changes	Deterioration					Improvement			
	−4	−3	−2	−1	0	1	2	3	4
All respondents	0.03	0.14	2.03	19.19	57.42	19.00	2.03	0.14	0.03
Years since migration less than 15	0.05	0.05	2.42	18.59	59.50	17.63	1.56	0.15	0.05
Years since migration at least 15	0.06	0.29	4.39	20.04	54.77	20.68	2.58	0.12	0.00

All numbers are percentages, pooled over the pairs of waves 1984–1985, 1985–1986, and 1986–1987. Observations with years since migration less than 15: 1985. Observations with years since migration at least 15: 1707. Years since migration is measured at the time of the first of the two waves.

Correlation between the unobserved heterogeneity terms in the two equations is allowed for by a bivariate mass point distribution. The model allows us to study the bias due to correlated unobserved heterogeneity as well as measurement (that is, misclassification) error in estimates of language proficiency on earnings. The most general version of this model also allows for correlation in the residual error terms.

A. An Ordered Response Model with Time-Independent and Time-Persistent Misclassification

Speaking fluency is observed on an ordinal scale with five categories. Because of the small number of observations in the extreme categories, we combined levels 1 and 2 and levels 4 and 5, retaining categories "bad" ($y_{it} = 1$), "intermediate" ($y_{it} = 2$), and "good" ($y_{it} = 3$), where i is the individual and t the time period. In the sequel, we will use y_{it} for the observed variables and z_{it} for the underlying true categorical variables. But first we present the standard random-effects ordered probit model, in which z_{it} and y_{it} coincide:

$$y^*_{it} = x'_{it}\beta + \alpha_i + \epsilon_{it}, \tag{1}$$

$$z_{it} = j \quad \text{if} \quad m_{j-1} < y^*_{it} < m_j, \quad j = 1, 2, 3, \tag{2}$$

$$\epsilon_{it} \quad \text{i.i.d.} \quad N(0, \sigma^2_\epsilon), \tag{3}$$

$$\alpha_i \quad \text{i.i.d.} \quad N(0, \sigma^2_\alpha), \tag{4}$$

$$\epsilon_{it}, \alpha_i \quad \text{and} \quad x_{it} \quad \text{independent.} \tag{5}$$

Here x_{it} denotes the vector of explanatory variables, including a constant. Some of the x_{it} are constant over time (country-of-origin dummies, year of entry, age at entry), others vary over time (years of education, family composition and marital status, years since migration), but not much or in a systematic way (years since migration, for example). Due to the lack of time variation in x_{it}, the data do not allow for estimating fixed-effects models or random-effects models in which the individual effects α_i are correlated with x_{it}. We do, however, relax the normality assumption on α_i. Following Heckman and Singer (1984), we replace equation (4) by the assumption that α_i follows a discrete distribution with M mass points:

$$P[\alpha_i = a_k] = p_k, \quad k = 1, \ldots, M. \tag{6}$$

The error term ϵ_{it} is i.i.d. white noise reflecting random variation in speaking fluency. In a model without explicit misclassification errors, this term picks up measurement errors that are independent over time. If misclassification errors are explicitly incorporated, there is less scope for a meaningful interpretation of ϵ_{it}, and we would expect its impact (that is, σ_ϵ) to be smaller. By means of normalization, the category bounds are set to $m_0 = -\infty$, $m_1 = 0$, $m_2 = 10$, and $m_3 = \infty$.

In the model with misclassification errors, we distinguish between the reported category y_{it} and the true category z_{it}. The latter is defined by the latent index function (2). The link between y_{it} and z_{it} is modeled generalizing existing models in the literature that explicitly allow for misclassification errors, such as in Lee and Porter (1984), Hausman et al. (1998) and Douglas, Smith Conway, and Ferrier (1995). The former two studies distinguish only two regimes, and thus work with two misclassification probabilities (the probability that the second regime is observed given that the first is true and vice versa), which are both treated as fixed parameters independent of everything else. Douglas, Smith Conway, and Ferrier (1995) work with three (ordered) regimes, but they impose the restriction that two misclassification probabilities are equal to zero, leaving them with four additional parameters to be estimated.

All three studies analyze cross-sectional models that do not distinguish between time-varying and time-persistent misclassification error. Our interpretation of misclassification is essentially the same: we assume that there is some (unobserved) "true" classification scale and a true (unobserved) continuous score y^*_{it}. Together with the fixed cutoff points m_1 and m_2,[4] this determines what someone's evaluation on a discrete scale should look like, that is, z_{it}. Misclassification then implies that the reported fluency y_{it} differs from the true fluency z_{it}. Following the three studies just referred to, we assume that the probabilities of misclassification $p_{j,k} = P[y_{it} = k | z_{it} = j]$ ($k \neq j$) depend on only k and j, and (conditional on these) not on respondent characteristics. This is a common assumption in this type of model, which is restrictive, but necessary for identification without relying on functional form assumptions, and without additional information on z_{it} (such as an alternative, objective measurement of language proficiency).

The difference from the existing cross-sectional studies is that, in the panel data context, not only the probabilities of misclassification in one specific period play a role, but also the correlation between classifications in different time periods of the same respondent. The two extreme assumptions would be:

- Time independence: Whether a given respondent misclassifies in one period is independent of whether he misclassifies in any other period.
- Persistence: A respondent who over-reports (or under-reports) once will always tend to over-report (or under-report).

Our model captures both of these extremes in a parsimonious way, and lets the data decide which one is more relevant. Our approach is based on the following assumptions.

---

[4] In a cross-sectional context, we have also looked at models that relax the assumption that the thresholds are the same for all individuals. This did not change the conclusions about the importance of the misclassification probabilities. See Dustmann and van Soest (2000).

- We identify four subpopulations: those who never misclassify ($(0,0)$, fraction π_{00}); those who sometimes under-report but never over-report ($(1,0)$; fraction π_{10}); those who over-report but never under-report ($(0,1)$, fraction π_{01}); and those who under- as well as over-report ($(1,1)$, fraction $\pi_{11} = 1 - \pi_{00} - \pi_{01} - \pi_{10}$).
- The distributions of x_{it}, α_i, and ϵ_{it} are the same in each of the four subpopulations.
- Given the subpopulation and conditional on the true speaking fluencies (z_{it}), misclassification events in different periods are mutually independent and independent of the x_{it}.
- The probabilities of under-reporting in the subpopulations $(1,0)$ and $(1,1)$ do not depend on t and are given by $p_{21} = P[y_{it} = 1|z_{it} = 2]$, $p_{31} = P[y_{it} = 1|z_{it} = 3]$, and $p_{32} = P[y_{it} = 2|z_{it} = 3]$.
- The probabilities of over-reporting for the subpopulations $(0,1)$ and $(1,1)$ do not depend on t and are given by $p_{12} = P[y_{it} = 2|z_{it} = 1]$, $p_{13} = P[y_{it} = 3|z_{it} = 1]$, and $p_{23} = P[y_{it} = 3|z_{it} = 2]$.

These assumptions imply, for example, that the probability that an individual in subpopulation $(1,0)$ with $z_{i1} = z_{i2} = 2$ gives answers $y_{i1} = 1$ and $y_{i2} = 2$ is given by $p_{21}(1 - p_{21})$. For someone in subpopulation $(1,1)$, this probability is $p_{21}(1 - p_{21} - p_{23})$. For the other subpopulations, the probability is zero because these subpopulations never under-report. Probabilities that are not conditional upon z_{it} can be written as weighted means of the probabilities given above, weighting with the probability distribution of the z_{it}. These probabilities still take the subpopulation as given, however. In practice, we do not observe in which subpopulation the respondents are. Likelihood contributions are therefore obtained by taking the weighted mean of the probabilities for the subpopulations, using the probabilities π_{00}, π_{01}, π_{10} and π_{11} as weights.[5]

Obviously, this is not the only way to model misclassification explicitly. Compared to other ways, however, our model, has the advantage that it is parsimonious (misclassification is modelled using nine parameters: six $p_{jk}(j, k = 1, 2, 3, j \neq k)$ and π_{00}, π_{01} and π_{10}), but still comprises the two extreme cases of time-independent and time-persistent misclassification. The former is obtained if $\pi_{11} = 1$; in this case, conditional upon the values of true speaking fluency z_{it}, events of misclassification are independent over time. The latter is obtained if, for example, $p_{21} = p_{31} = 1$. In this case, a fraction $(\pi_{10} + \pi_{11})$ always reports "bad" speaking fluency whatever their real speaking fluency.

In general, our model allows for any correlation between misclassification in two different time periods (conditional upon true speaking fluency). For example, the probability that someone is fluent on an intermediate level in both time periods and reports bad fluency twice is given by $(\pi_{10} +$

$\pi_{11})p_{21}^2$. The probability that this happens in one wave is given by $(\pi_{10} + \pi_{11})p_{21}$. If $(\pi_{10} + \pi_{11}) = 1$, the two-wave probability is the product of the two one-wave probabilities, and the misclassification events in the two waves are independent. If $p_{21} = 1$, someone whose fluency is at an intermediate level and who once reports "bad" fluency will always report bad fluency as long as the true fluency remains at the intermediate level. Thus, in this case, misclassification is time persistent. For other values of the parameters, intermediate positive correlation structures are obtained. Negative correlation is not possible and is not plausible in the current context.

The panel is unbalanced. We assume that whether information on individual i is available in wave t or not is independent of $\{\epsilon_{it}, t = 1, \ldots, T\}$ and α_i. This implies that we do not allow for sample selection bias or attrition bias.

The model can be estimated by maximum likelihood. The assumptions just given imply that computing the likelihood contribution for each individual requires numerical integration in one dimension if the specification with normally distributed individual effects in equation (4) is used, as in the binary response case of Butler and Moffitt (1982). If the discrete distribution in equation (6) is used instead, no numerical integration is required.

B. Speaking Fluency and Earnings

To analyze how speaking fluency affects earnings of full-time workers, we add the following equation explaining log monthly earnings w_{it}:

$$w_{it} = x_{it}'\beta^w + \gamma y_{it}^* + \alpha_i^w + \epsilon_{it}^w. \tag{7}$$

We have included the underlying latent speaking fluency variable, y_{it}^*, instead of the discrete variables, z_{it} or y_{it}. We think that y_{it}^* better reflects the impact of speaking fluency on earnings, which should not depend upon the categories that were used in the questionnaire.

As before, we assume that all errors are mean zero, and we neither allow for correlation between individual effects and idiosyncratic errors nor for correlation between the error terms and the x_{it}:

$$\epsilon_{it}^w, \ \alpha_i^w \ \text{ and } \ x_{it} \ \text{ independent.} \tag{8}$$

For the individual heterogeneity terms α_i^w, we again use a Heckman-Singer specification. We distinguish two cases:

Uncorrelated unobserved heterogeneity (α_i and α_i^w are independent):

$$P[(\alpha_i, \alpha_i^w) = (\alpha_l, \alpha_m^w)] = p_l p_m^w, \quad l, m = 1, \ldots, M. \tag{9}$$

Correlated unobserved heterogeneity (α_i and α_i^w are not necessarily independent):

$$P[(\alpha_i, \alpha_i^w) = (\alpha_k, \alpha_k^w)] = p_k, \quad k = 1, \ldots, K. \tag{10}$$

---

[5] Examples and the FORTRAN code with all the likelihood contributions are available upon request from the second author.

According to equation (9), the bivariate distribution of (α_i, a_i^w) has M^2 mass points, obtained by combining the mass points of the two marginal distributions. On the other hand, equation (10) allows for K arbitrary mass points. Equation (9) is a special case of (10) if $K = M^2$. The results we present are based upon $K = 9$ and $M = 3$. Comparing the results with equation (9) imposed with those imposing (10) shows how allowing for correlated (time-persistent) unobserved heterogeneity in speaking fluency and earnings equations affects the estimated impact of language fluency on earnings.

If the explicit misclassification errors included in the speaking fluency model are the only source of measurement error, measurement error is automatically accounted for by including y_{it}^* as a right-hand variable. In this case, there seems to be no reason to allow for correlation between the idiosyncratic errors, ϵ_{it} and ϵ_{it}^w. Comparing the results of the model in which these misclassification errors are and are not included shows how they affect the estimates of the impact of language fluency on earnings.

If our stylized model of misclassification errors does not encompass all time-varying measurement error in observed speaking fluency, then ϵ_{it} may still contain measurement error. This would mean that y_{it}^* suffers from measurement error. Following the standard argument in linear regression models, and assuming that $\gamma > 0$, this would lead to a negative correlation between ϵ_{it}^w and ϵ_{it}.

Therefore, in the most general version of our model, we also allow for correlation between ϵ_{it}^w and ϵ_{it}. We assume that

$$E(\epsilon_{it}^w \epsilon_{it}) = \rho \sigma_\epsilon \sigma_\epsilon^w. \tag{11}$$

We estimate models in which ρ is an unknown parameter, as well as models in which ρ is set equal to zero.

The model that imposes equation (9) and $\rho = 0$ implies that y_{it}^*, the true (unobserved) speaking fluency on a continuous scale, is strictly exogenous in the wage equation. Even in this case we cannot estimate the wage equation separately, because y_{it}^* is not observed. Using equation (10) instead of (9) relaxes the exogeneity assumption through the individual effects. Allowing for a nonzero ρ relaxes this assumption further.

Without exclusion restrictions, the general model with correlated individual effects and correlated error terms in speaking fluency and wage equations is not identified.[6] For identification, we need to exclude variables from the earnings equation that are in the speaking fluency equation. Our identification mainly relies on the father's education level and we assume that the father's education level has no direct effect on earnings. One reason why this assumption has been criticized in the wage literature is that networking by the father may help the child's earnings prospects. The

immigrants in our sample, however, are first-generation immigrants, with their parents typically residing in the home countries. Migration is very likely to cut links with parental networks. On the other hand, there may still be family-specific unobservable effects that are transmitted between generations and that may invalidate the exclusion restriction. Because we also condition on the respondent's own education level, however, we do not expect this to be a large problem. In the restricted models, exclusion restrictions are not necessary for identification. To make the results comparable, however, we impose the exclusion restrictions in the restricted models as well.

To summarize, our model encompasses both correlated unobserved heterogeneity and misclassification and measurement errors. Unobserved heterogeneity is included through the random individual effects. It induces a bias on the OLS estimate of speaking fluency in the earnings equation that has the same sign as the correlation coefficient. As explained in section I, there are economic arguments for a positive (Borjas, 1994) as well as for a negative (Willis & Rosen, 1979) sign of the correlation coefficient and the bias. The data will have to show which of the two is relevant.

On the other hand, misclassification errors and measurement errors always lead to a negative bias on the OLS estimate. These are incorporated in the most general version of our model in three ways: time-independent misclassification, time-persistent misclassification, and correlation between the residual errors ϵ_{it} and ϵ_{it}^w. It is not a priori clear which of the three is most important for the bias; the empirical analysis has to determine this.

IV.   Results

We first present the results for the speaking fluency equation. We then discuss the estimates of the earnings equation in the simultaneous model for speaking fluency and earnings.

A.   Results Speaking Fluency

We have estimated a large variety of specifications: with linear and nonlinear effects of year of entry and years since migration, with and without explicit time-independent or time-persistent misclassification, and with normally distributed random effects and with Heckman-Singer type random effects. Four selected specifications are presented in table 3. They all incorporate Heckman-Singer type random effects, based upon the discrete distribution in equation (6), with four mass points. In terms of goodness of fit, the models with this type of random effects performed better than the models with normally distributed random effects. Both types of models lead to similar estimates of the other parameters.

The first two specifications in table 3 (models 1 and 2) are standard panel data models with random effects, with no explicit misclassification errors. The first comes closest to

---

[6] This is easy to see by substituting equation (1) in (7). The model is linear because y_{it}^* is included and not y_{it}, so that functional form or distributional assumptions do not help.

TABLE 3.—ESTIMATION RESULTS, SPEAKING FLUENCY

	Model 1		Model 2		Model 3		Model 4	
	Coef	StdE	Coef	StdE	Coef	StdE	Coef	StdE
Constant	−1.386	1.042	−9.622	3.095	−10.811	3.004	−9.433	3.336
Year entry	—	—	0.090	0.040	0.087	0.038	0.081	0.051
Age entry	−0.471	0.019	−0.449	0.019	−0.445	0.024	−0.477	0.041
d turkish	0.612	0.577	0.161	0.605	0.118	0.558	−0.122	0.619
d yugos	4.642	0.618	4.379	0.638	4.226	0.613	4.511	0.702
d greek	2.196	0.636	1.397	0.623	1.412	0.568	1.409	0.642
d italian	0.083	0.577	0.235	0.588	0.332	0.551	0.526	0.615
f educ l 2	—	—	0.781	0.447	0.787	0.411	0.761	0.458
f educ l 3	—	—	2.162	0.476	2.221	0.442	2.193	0.490
f educ l 4	—	—	4.489	0.958	4.719	0.860	4.878	0.973
f educ l 5	—	—	4.715	1.676	4.624	1.391	4.672	1.473
f educ l 6	—	—	0.817	0.573	1.001	0.523	1.089	0.610
yrs s migr	0.120	0.019	0.159	0.026	0.157	0.026	0.165	0.028
yrs educ	0.790	0.075	0.735	0.083	0.740	0.079	0.732	0.089
married	−0.455	0.369	−0.297	0.372	−0.075	0.337	−0.166	0.362
σ_ϵ	5.223	0.085	5.215	0.085	4.063	0.298	4.264	0.270
p12					0.141	0.037	0.293	0.067
p13					0.016	0.011	0.029	0.020
p21					0.036	0.011	0.448	0.082
p23					0.084	0.029	0.114	0.044
p31					0.001	0.001	0.006	0.011
p32					0	—	0	—
π_{00}							0.177	0.118
π_{01}							0.700	—
π_{10}							0.123	0.038
p_1	0.187	0.019	0.190	0.021	0.181	0.019	0.174	0.019
a_1	20.961	0.631	20.688	0.646	20.502	0.867	20.881	1.117
p_2	0.405	0.035	0.399	0.034	0.407	0.031	0.395	0.036
a_2	12.593	0.476	12.892	0.498	13.236	0.636	13.193	0.865
p_3	0.307	0.035	0.314	0.034	0.309	0.031	0.331	0.036
a_3	7.369	0.404	7.693	0.404	8.154	0.514	8.005	0.694
log lik		−5261.30		−5242.65		−5220.90		−5204.11

the cross-sectional specifications in existing studies. Neither the year of entry nor the father's education level dummies are included. All these variables are included in model 2. The final two specifications use the same explanatory variables as model 2, but explicitly allow for misclassification. Model 3 allows for time-independent classification errors, and model 4 also allows for time-persistent misclassification. Model 4 is the most general model; it encompasses the other three.

Most of the estimates and significance levels of the slope coefficients β are robust across the four specifications, and also across alternative specifications that are not presented. Age at entry has a significant negative impact, as in other studies. There are two explanations for this. First, learning a foreign language becomes more difficult with age, leading to slower acquisition of language capital among those who immigrate later in life. Second, older migrants have a shorter payoff period on country-specific human capital, and this creates a disincentive effect.

The country dummies reflect distance in culture and language between home and host country. They also reflect different degrees of self-selection from different origin countries. The base category consists of immigrants with Spanish origin. The estimates indicate that Yugoslavian immigrants are more fluent than the other groups, ceteris paribus. Greek immigrants are less fluent than Yugoslavians,

but more fluent than the other three groups. Differences among individuals from the other three origin countries (Turkish, Italian, and Spanish immigrants) are insignificant.

Years of education has a significant positive effect that is similar in all specifications. The more highly educated speak the host country language more fluently than those with a lower education level. This is in line with the existing empirical evidence.

In model 2, we include dummy variables for the education level of the immigrant's father, and the year of entry into Germany. A likelihood ratio test indicates that, overall, model 2 is significantly better than model 1. The father's education level is significant, and speaking fluency increases with the father's educational degree. (The excluded category is fathers with no education.) An explanation is that immigrants from families with a higher educational background may be more likely to develop an interest in all those goods to which language proficiency gives access. They may also grow up in a more open-minded environment, which reduces barriers to contacts to foreign cultures later in life. Furthermore, they have probably been more exposed to foreign languages during their childhood.

The coefficient on the year of entry variable is significantly positive in models 2, 3, and 4. Later cohorts of immigrants speak German more fluently, conditional upon

years since migration, age at entry, country of origin, and so on.[7]

In all specifications, years since migration has a significant positive effect on language proficiency. Comparing models 1 and 2 shows that including the additional variables (in particular, year of entry) increases the estimated effect. The average marginal effect of one additional year of residence on the probability of being fluent or very fluent is 0.5 percentage points and 0.7 percentage points according to models 1 and 2, respectively. This is larger than the earlier finding of Dustmann (1994) for Germany, but still rather small compared to findings for other countries.[8]

The introduction of additional variables also changes the coefficients on the country-of-origin dummies. A possible reason is that the migration density of different origin countries changed over time, so that the country-of-origin dummies and the year of entry are correlated.

Models 3 and 4 explicitly account for misclassification error. The differences between likelihood values of models 2, 3, and 4 show that this improves the fit of the model. In model 3, the estimated misclassification probabilities have small standard errors, indicating that they are estimated rather precisely. This specification allows only for misclassification errors that are independent over time. The probabilities of over-reporting (p_{12} and p_{23}) are substantial, and their confidence intervals do not contain zero.[9] The estimate for p_{12} of 0.141 indicates that someone with bad speaking fluency has a 14% probability of reporting reasonable fluency in a given wave. The under-reporting probability p_{21} is smaller, but its confidence interval still excludes zero. The other three misclassification probabilities are closer to zero. In particular, the estimates imply that individuals with good speaking fluency in German hardly ever misclassify.

In model 4, the complete misclassification framework introduced in the previous section is used. The estimate of π_{11} is zero. This implies that there are individuals who sometimes over-report (70.0%) and individuals who sometimes under-report (12.3%), but there is no evidence of individuals who under-report as well as over-report; 17.7% of all individuals would never under- or over-report. The estimates of these group probabilities π_{rs} are not very precise, however. Although the model is identified in theory, it is hard to distinguish the individual effects, α_i, and the idiosyncratic errors, ϵ_{it}, from the π_{rs} and the p_{jk} in practice.

Still, the likelihood value of model 4 is much higher than that of models 2 and 3.[10]

Some of the p_{jk} in model 4 seem quite large, suggesting that probabilities of misreporting could be substantial for the groups with a tendency to over- or under-report. To compare them with those in model 3, however, we should look at marginal probabilities of misclassification, taking into account that we never observe to which of the three groups ((0,0), (0,1), or (1,0)) a respondent belongs. For example, the probability that someone with bad fluency reports fluency on an intermediate level is $0.70*0.293 = 0.205$ in model 4, compared to 0.141 in model 3. The probability that a randomly drawn individual with bad fluency in two waves reports fluency on an intermediate level twice is $0.70*0.293^2 = 0.059$ in model 4, and $0.141^2 = 0.026$ in model 3. The probability that someone with reasonable fluency under-reports in one wave is $0.123*0.448 = 0.055$ for model 4, and 0.036 in model 3. The probability that this happens twice is $0.123*0.448^2 = 0.027$ for model 4, and $0.036^2 = 0.0013$ for model 3. Thus, model 4 implies larger misclassification probabilities than model 3.

The estimate of σ_ϵ reflects the importance of the idiosyncratic shocks. As expected, this is reduced in models 3 and 4, as compared to models 1 and 2, in which the ϵ_{it} also pick up time-independent misclassification errors. Still, the reduction in σ_ϵ is small. Apparently, there is either more idiosyncratic noise than just misclassification errors, or our stylized model for misclassification is not able to pick up all misclassification errors.

The individual effects, α_i, are assumed to follow a distribution with four mass points. By means of normalization, one mass point is set equal to zero. We have estimated models with five mass points, but the estimated probability for the fifth mass point is close or equal to zero. The implied standard deviations of the α_i are 5.98, 5.84, 5.75, and 5.81 in models 1, 2, 3, and 4, respectively. Thus, the models with explicit misclassification probabilities imply a somewhat smaller role for α_i. We would have expected that the time-persistent heterogeneity in terms of misclassification behavior would reduce the role of the time-persistent heterogeneity in α_i, but comparing models 3 and 4 shows that this is not the case.

To gain further insight into the differences between the four sets of estimates and the implications of the classification errors, we compute some summary statistics of y_{it}^*, z_{it}, and y_{it}. The mean value of y_{it}^* (over individuals and time periods) is approximately 10.2 in models 1 and 2, and approximately 9.2 in models 3 and 4.[11] The misclassification probability estimates of models 3 and 4 imply that over-reporting is more likely than under-reporting. Thus, models

---

[7] Controlling for country of origin may be important: in the United States, cohort effects are largely explained by changes in the country of origin composition (Borjas, 1987).

[8] For instance, Chiswick and Miller (1995) find for Australia the effect of residence to range between 1 and 3.5 percentage points per year, depending on the country of origin. Chiswick (1997) finds an effect of about 2.6 percentage points per year for Israel, evaluated at ten years of residence. For low-skilled workers in the United States, Chiswick (1991) finds that an additional year of residence increases fluency by approximately three percentage points.

[9] Because these probabilities are by definition nonnegative, standard t-tests or likelihood ratio tests on $p_{jk} = 0$ are inappropriate (Shapiro, 1985).

[10] Again, a formal *chi*-squared test is not appropriate, due to the one-sided nature of the alternative.

[11] Note that the normalization of y_{it}^* is the same in all models: the cutoff points are set to zero and ten.

Table 4.—Estimation Results, Wage Equation

	Model W1		Model W2		Model W3		Model W4	
	Coef	StdE	Coef	StdE	Coef	StdE	Coef	StdE
Const wage	7.0846	0.0228	7.3224	0.0275	7.4149	0.0320	7.4205	0.0342
d turkish	−0.0220	0.0107	−0.0370	0.0116	−0.0315	0.0132	−0.0251	0.0136
d yugos	−0.0213	0.0113	−0.0124	0.0133	−0.0200	0.0157	−0.0411	0.0218
d greek	0.0021	0.0113	−0.0184	0.0141	−0.0188	0.0158	−0.0263	0.0170
d italian	0.0214	0.0114	−0.0133	0.0123	0.0069	0.0139	0.0049	0.0152
0.01 yrs s migr	0.0766	0.0710	0.6076	0.0948	0.4882	0.1141	0.1634	0.2134
Exp	0.0319	0.0012	0.0241	0.0013	0.0241	0.0015	0.0266	0.0022
0.01 exp sq	−0.0555	0.0022	−0.0457	0.0025	−0.0444	0.0027	−0.0443	0.0027
Yrs educ	0.0312	0.0014	0.0160	0.0019	0.0134	0.0023	0.0111	0.0028
Married	0.1025	0.0086	0.1141	0.0097	0.1143	0.0093	0.1188	0.0102
Year 85	−0.0279	0.0116	−0.0240	0.0108	−0.0229	0.0106	−0.0234	0.0105
Year 86	0.0251	0.0109	0.0221	0.0101	0.0236	0.0099	0.0228	0.0099
Year 87	0.0311	0.0107	0.0279	0.0096	0.0303	0.0093	0.0288	0.0094
Year 89	0.0987	0.0118	0.0877	0.0113	0.0908	0.0112	0.0886	0.0115
Year 91	0.1138	0.0133	0.1024	0.0125	0.1065	0.0126	0.1036	0.0132
Year 93	0.1167	0.0129	0.0985	0.0126	0.1042	0.0124	0.1008	0.0132
0.01 sp fl	0.3121	0.0458	0.0949	0.0695	0.2858	0.1248	0.8617	0.3627
$\sigma(\epsilon_w)$	0.1988	0.0012	0.1873	0.0010	0.1841	0.0010	0.1862	0.2433
$\rho(\epsilon, \epsilon_w)$							−0.2268	0.1337
$\sigma(\alpha^w)$*	0.1728		0.1933		0.1955		0.1931	
$\rho(\alpha, \alpha^w)$*			0.4519		0.6784		0.8878	
log lik		−5082.37		−4877.91		−4805.54		−4804.22

*$\sigma(\alpha^w)$ and $\rho(\alpha, \alpha^w)$ are computed from the estimates of the parameters in equation (9) (model W1) or equation (10) (models W2 through W4), given in table A2 in appendix A.

3 and 4 predict lower genuine speaking fluency than models 1 and 2 (that is, lower means of y_{it}^* and z_{it}) but, due to over-reporting, models 3 and 4 predict that, on average, reported fluency (y_{it}) exceeds true fluency (z_{it}). The marginal distribution of the reported speaking fluency variable, y_{it}, is similar for all models, and similar to that in the data.

The estimated variance of y_{it}^* is 90.5, 90.9, 82.7, and 87.6 in models 1 to 4, respectively. It is smaller in models 3 and 4 than in models 1 and 2, indicating that the misclassification mechanism adds to the variance of the reported fluency variable.

According to models 3 and 4, the distribution of the true discrete speaking fluency variable z_{it} is different from that of the reported variable y_{it}. In model 4, the marginal probability that the true fluency is bad is 0.155, whereas the probability that bad fluency is reported is only 0.141. The probability that the true fluency is good is 0.470, and the probability that the respondent reports that his fluency is good is 0.502. This again shows that over-reporting is more likely than under-reporting. Similar results are found for model 3. In all models, the mean of y_{it}^* per time period and the probability of good fluency (true or reported) gradually increase over time, due to the increasing sample average of years since migration.

B. Results Earnings Equations

In the earnings equation, we use the same regressors as the existing literature. (See Chiswick and Miller (1995), for example.) We include years of education, potential labor market experience and its square, a marital status dummy, and country-of-origin dummies. All specifications include year dummies to account for macroeconomic effects. Because potential experience is driven by age and education,

we cannot identify cohort effects in the earnings equation separately.

Following the existing literature in this field, we do not address potential selectivity bias because we use earnings of only full-time workers. We thus implicitly assume that whether someone has a full-time job is independent of the error terms in the model, conditional on the covariates.

The model is estimated with maximum likelihood, jointly with the speaking fluency equation. We use the same regressors for the fluency equation as in model 2 in table 3. Thus, all our wage equation estimates impose the same exclusion restrictions: education-level dummies of the immigrant's father are included in the speaking fluency equation but not in the wage equation. We do not present the estimates of the speaking fluency equation when estimated jointly with the earnings equation, because these are very similar to those when the speaking fluency equation is estimated alone.

In table 4, we present the estimation results for the earnings equation for four different specifications. In model W1, we use the specification of the speaking fluency equation that does not allow for misclassification errors (cf. model 2 in table 3). Individual heterogeneity is specified through equation (9), which does not allow for correlation between individual effects in the two equations. The idiosyncratic errors, ϵ_{it} and ϵ_{it}^w, are assumed to be independent (that is, $\rho = 0$ in equation (11)). Thus, this model corrects for neither correlated unobserved heterogeneity nor for measurement errors.

We find that speaking fluency has a positive and significant effect on earnings. The estimated standard deviation of y_{it}^* across individuals in this model is 9.1, so the point estimate of 0.31 implies that a one-standard-deviation increase of y_{it}^* leads to a wage increase of approximately 2.8 percentage points.

In Model W2, we allow for correlated unobserved heterogeneity, using equation (10) instead of (9). A likelihood ratio test suggests that this is a significant improvement: model W1 is rejected against model W2. The estimates of model W2 imply a strong positive correlation between α_i and α_i^w, and the implied estimate for the correlation coefficient is 0.49.[12] This positive correlation implies a positive bias in the estimated effect of speaking fluency in model W1, which is removed in model W2. As a consequence, the effect of speaking fluency is smaller in model W2 than it is in model W1. The effect remains significantly positive, and an increase of y_{it}^* by one standard deviation leads to a rise in earnings of 0.9 percentage points. These results confirm the conjecture by Borjas (1994): ignoring the presence of unobserved heterogeneity leads to an upward bias in the estimated effect of speaking fluency on earnings.

Model W2 has not taken any account of possible misclassification in the language variable. In model W3, the misclassification probabilities are added to the speaking fluency equation (cf. model 4 in table 3). This removes the negative bias in the speaking fluency coefficient due to time-persistent and time-independent misclassification. Allowance for misclassification leads to an increase in the estimated effect of language on earnings, compared to model W2. An increase of y_{it}^* by one standard deviation (8.45 according to this model) leads to a wage rise of 2.4%. This is somewhat smaller than the estimate in model W1. Accordingly, these results indicate that the negative bias due to misclassification and the positive bias due to unobserved heterogeneity are of similar magnitude, and almost cancel out.

In model W4, we relax the assumption that $\rho(\epsilon, \epsilon^w) = 0$, and estimate $\rho = \rho(\epsilon, \epsilon^w)$. (See equation (11).) The estimate of ρ is -0.23 and is significant at the 10% level, but not at the 5% level. The negative sign suggests that, conditional on the way we have already accounted for misclassification in the language equation, ϵ_{it} still contains measurement error. Allowance for a nonzero value of ρ leads to a substantial increase in the estimate of the impact of speaking fluency on earnings: an increase of y_{it}^* by one standard deviation (8.47) rises wages by approximately 7.3 percentage points. This point estimate has a larger standard error than the estimates in the more restrictive models, however.

The differences in the results of models W1 through W4 are in line with the biases predicted from econometric theory. They can also explain the differences between OLS and IV estimates in Chiswick and Miller (1995). Allowance for correlated unobserved heterogeneity takes away a positive bias, and thus reduces the estimated effect of speaking fluency (the difference between W2 and W1). Allowance for misclassification error, on the other hand, takes away a negative bias, and thus increases the estimate (W3 versus

W2). Allowance for more general measurement errors than the misclassification errors in our framework further reduces the negative bias, and thus leads to even higher estimates of the language effect (W4 versus W3). Overall, the measurement error corrections are more important than the correction for correlated unobserved heterogeneity, so that the estimate in the most general model exceeds the estimate that corrects for neither of the two sources of bias (W4 versus W1).

Most of the other coefficients vary less across the four specifications, and are in line with the findings in the literature. The experience pattern is quadratic and increasing during most of the career path. Married workers earn significantly more than their unmarried colleagues, and years of education have a strong positive impact on earnings. Only the effect of years since migration varies substantially across specifications. It is always positive, but small and insignificant in models W1 and W4, but larger and significant in models W2 and W3. Years since migration and speaking fluency are positively correlated. As a consequence, a negative bias on the coefficient of speaking fluency induces a positive bias on the coefficient of years since migration. A similar explanation can be given for the (less dramatic) changes in the country-of-origin effects across specifications.

V.   Conclusions

This paper contributes to the large and growing literature on the determinants of immigrants' language proficiency, and the effect of language fluency on earnings. We draw attention to the fact that misclassification error in self-reported language indicators, usually used as a fluency measure, is substantial. This may lead to biased estimates of coefficients in the language fluency equation, and of the effect of language fluency on earnings. In addition, we address the problem of correlated unobserved heterogeneity when estimating language and earnings equations.

Methodologically, we add to the literature on misclassification errors in a discrete response setting. We generalize the approach by Hausman et al. (1998) to a panel data context. We combine a random-effects ordered response model with an explicit mechanism of misclassification probabilities, in which we distinguish between time-varying and time-persistent misclassification. The panel nature of our data makes it possible to identify the two different sources. This distinction seems plausible when using responses that are based on a subjective scale, such as with evaluations of language fluency. We develop a model that is characterized by a relatively small number of additional parameters, but nevertheless encompasses the extreme cases of time-independent misclassification errors, and purely time-persistent misclassification errors.

We demonstrate that self-reported measures of speaking fluency suffer from misclassification errors. In fact, most of the within-individual variation, and approximately 25% of

---

[12] This is computed from the estimated distribution of (α_i, α_i^w), given in table A2 in appendix A.

the total variation in the language response variable, is explained by measurement error. The estimation results of the language determination equation indicate that the probabilities of over-reporting are larger than the probabilities of under-reporting. Some probabilities of under-reporting are virtually zero. We also find evidence of time-persistent misclassification, that is, a positive correlation between misclassification events in different time periods. Neither the way misclassification is modeled, nor the assumed distribution of the individual heterogeneity, has much effect on the slope coefficient estimates in the speaking fluency equation. Thus, the estimates of the determinants of speaking fluency appear to be rather robust.

We then add an earnings equation to the model and estimate it jointly with the speaking fluency equation. We allow for misclassification according to the way we have specified our language equation. We allow for correlated unobserved heterogeneity in the earnings equation and the speaking fluency equation by a bivariate discrete mass-point distribution. Thus, our model allows us to separate the effects of measurement error and unobserved heterogeneity. Estimation results of models that are similar to those usually presented in the literature result in a positive and significant response of earnings to speaking fluency. We find evidence for a nonnegligible positive bias due to ignoring correlated unobserved heterogeneity, which is reflected by a positive correlation between the individual effects in the two equations. Moreover, our findings suggest that correcting for measurement error in self-reported assessments of language proficiency is crucial. Neglecting this error leads to a substantial downward bias of the impact of speaking fluency on earnings. Our estimates suggest that the bias due to misclassification error, and due to unobserved heterogeneity, are roughly of equal size.

Not all of the measurement error in our model is included by the way we model the misclassification in the language equation. This is indicated by the magnitude of the variance of the residual error term in the language equation. We therefore also estimate a more general model, which allows for correlated error terms in language equation and earnings equation. This correlation picks up measurement error that has not been taken into account by our parametric structure. This leads to a substantial increase in the estimated impact of speaking fluency on earnings. It suggests that the downward bias in the language effect due to measurement error is considerably larger than the upward bias due to unobserved heterogeneity.

The model we have introduced requires a number of specific assumptions. The role of the negative correlation between idiosyncratic error terms in speaking fluency and wage equations, conditional on the way we allow for misclassification, is unexpectedly large. It suggests that the way in which we model misclassification might be too restrictive. One of the most important assumptions we would want to relax is that misclassification probabilities are unrelated to the explanatory variables. Although this is a common assumption in the misclassification literature, it would be desirable to test for this in future research. This is possible if alternative (objective) ways of measuring language proficiency become available. Another crucial assumption in our models concerns the exclusion restrictions. Although we have reason to believe that our main instruments (the education level of the immigrant's father) are appropriate in our context, testing the robustness of our results for alternative instruments remains for future research.

REFERENCES

Borjas, George J., "Self-Selection and the Earnings of Immigrants," *American Economic Review* 77:4 (1987), 531–553.
———, "The Economics of Immigration," *Journal of Economic Literature* 32:4 (1994), 1667–1717.
Butler, J. S., and Robert Moffitt, "A Computationally Efficient Quadrature Procedure for the One-Factor Multinomial Probit Model," *Econometrica* 50:3 (1982), 761–764.
Carliner, Geoffrey, "Wage Differences by Language Group and the Market for Language Skills in Canada," *Journal of Human Resources* 16:3 (1981), 384–399.
Chiswick, Barry R., "Speaking, Reading, and Earnings among Low-Skilled Immigrants," *Journal of Labor Economics* 9:2 (1991), 149–170.
———, "Hebrew Language Usage: Determinants and Effects on Earnings Among Immigrants in Israel," Maurice Falk Institute discussion paper no. 97.09 (1997).
Chiswick, Barry R., and Paul W. Miller, "The Endogeneity between Language and Earnings: International Analyses," *Journal of Labor Economics* 13:2 (1995), 246–288.
Douglas, Stratford M., Karen Smith Conway, and Gary D. Ferrier, "A Switching Frontier Model for Imperfect Sample Separation: With an Application to Constrained Labor Supply," *International Economic Review* 36:2 (1995), 503–526.
Dustmann, Christian, "Speaking Fluency, Writing Fluency and Earnings of Migrants," *Journal of Population Economics* 7:2 (1994), 133–156.
Dustmann, Christian, and Arthur van Soest, "Parametric and Semiparametric Estimation in Models with Misclassified Categorical Dependent Variables," Tilburg University mimeograph (2000).
Grenier, Gilles, "The Effect of Language Characteristics on the Wages of Hispanic American Males," *Journal of Human Resources* 19:1 (1984), 35–52.
Hausman, Jerry A., Jason Abrevaya, and Fiona M. Scott-Morton, "Misclassification of the Dependent Variable in a Discrete Response Setting," *Journal of Econometrics* 87:2 (1998), 239–269.
Heckman, James J., and Burton Singer, "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica* 52:2 (1984), 271–320.
Kossoudji, Sherrie A., "English Language Ability and the Labor Market Opportunities of Hispanic and East Asian Immigrant Men," *Journal of Labor Economics* 6:2 (1988), 205–228.
Lee, Lung-fei, and Robert H. Porter, "Switching Regression Models with Imperfect Sample Separation Information: With an Application on Cartel Stability," *Econometrica* 52:2 (1984), 391–418.
McManus, Walter, William Gould, and Finis Welch, "Earnings of Hispanic Men: The Role of English Language Proficiency," *Journal of Labor Economics* 1:2 (1983), 101–130.
Rivera-Batiz, Francisco L., "English Language Proficiency and the Economic Progress of Immigrants," *Economics Letters* 34:3 (1990), 295–300.
Shapiro, Alexander, "Asymptotic Distribution of Test Statistics in the Analysis of Moment Structures under Inequality Constraints," *Biometrica* 72:1 (1985), 133–144.
Wagner, Gert, Richard V. Burkhauser, and Friederike Behringer, "The English Public Use File of the German Socio Economic Panel," *Journal of Human Resources* 28:2 (1993), 429–433.
Willis, Robert J., and Sherwin Rosen, "Education and Self-Selection," *The Journal of Political Economy* 87:5(part 2) (1979), 57–536.

APPENDIX

TABLE A1.—DESCRIPTION AND SUMMARY STATISTICS FOR WAVE 1

Code	Mean	StD	Explanation
yrs s migr	17.722	5.817	Years of residence in Germany
age	41.234	10.769	Age
age entry	23.512	8.648	Age at entry
d turkish	0.317	0.465	Dummy; 1 if Turkish
d yugos	0.197	0.398	Dummy; 1 if Yugoslavian
d greek	0.139	0.346	Dummy; 1 if Greek
d italian	0.208	0.406	Dummy; 1 if Italian
d spanish	0.136	0.343	Dummy; 1 if Spanish
yrs edu	9.941	2.042	Years of schooling
mar	0.84	0.408	Dummy; 1 if married
n children	1.206	1.248	Number of children
f educ l 1	0.226	0.418	Father no education
f educ l 2	0.332	0.471	Father primary education
f educ l 3	0.292	0.455	Father basic education
f educ l 4	0.038	0.191	Father intermediate education
f educ l 5	0.005	0.073	Father secondary education
f educ l 6	0.104	0.305	Father education missing

TABLE A2.—DISTRIBUTION OF (α_i, α_i^w) IN MODELS W1–W4

	Model W1		Model W2		Model W3		Model W4	
	Coef	StdE	Coef	StdE	Coef	StdE	Coef	StdE
p_1	0.3014	0.0195						
p_2	0.5555	0.0207						
α_1	17.1047	0.3945						
α_2	8.6285	0.2568						
p_1^w	0.1975	0.0151						
p_2^w	0.1359	0.0129						
α_1^w	−0.3009	0.0064						
α_2^w	0.3019	0.0071						
p_1			0.0972	0.0131	0.0757	0.0135	0.0728	0.0131
p_2			0.0096	0.0031	0.0087	0.0030	0.0082	0.0030
p_3			0.1443	0.0162	0.0981	0.0147	0.0985	0.0148
p_4			0.0257	0.0058	0.0256	0.0058	0.0256	0.0059
p_5			0.0414	0.0082	0.0279	0.0066	0.0281	0.0065
p_6			0.3508	0.0209	0.3687	0.0243	0.3689	0.0245
p_7			0.1108	0.0130	0.1150	0.0136	0.1154	0.0136
p_8			0.1177	0.0144	0.1378	0.0160	0.1395	0.0162
α_1			15.9872	0.5146	11.0237	0.7183	11.3831	0.7296
α_2			17.8976	2.2235	13.0770	2.7147	13.1153	2.8539
α_3			18.2689	0.5819	16.6731	0.9770	16.8821	0.9795
α_4			10.0385	0.6005	6.6656	0.7729	6.7511	0.7846
α_5			19.5036	0.8707	9.5354	0.7742	9.8033	0.7748
α_6			9.5964	0.3414	6.0511	0.4760	6.2100	0.4795
α_7			4.7356	0.3880	−1.1294	0.6457	−0.8950	0.6508
α_8			9.3753	0.3980	9.6016	0.5861	9.6192	0.5911
α_1^w			−0.2482	0.0179	−0.3581	0.0226	−0.4205	0.0445
α_2^w			0.7772	0.0325	0.7138	0.0364	0.6495	0.0607
α_3^w			0.0058	0.0180	−0.1350	0.0271	−0.2265	0.0630
α_4^w			−0.6685	0.0173	−0.7448	0.0182	−0.7778	0.0286
α_5^w			0.3063	0.0204	0.3276	0.0323	0.2798	0.0469
α_6^w			−0.0624	0.0126	−0.1485	0.0132	−0.1805	0.0245
α_7^w			−0.2682	0.0133	−0.3210	0.0142	−0.3143	0.0160
α_8^w			0.1733	0.0144	0.0761	0.0199	0.0252	0.0379

Parameters are defined in equation (9) (model W1) or equation (10) (models W2 through W4). See table 4 for parameters of interest.