An Analysis of Speaking Fluency of Immigrants Using Ordered Response Models With Classification Errors

Christian DUSTMANN
University College London, London, U.K., and Institute for Fiscal Studies, (*c.dustmann@ucl.ac.uk*)
Arthur VAN SOEST
RAND Corporation, Santa Monica, CA, and Tilburg University (*vansoest@rand.org*)

We develop parametric models that incorporate misclassification error in an ordered response model and compare them with a semiparametric model that nests the parametric models. We apply these estimators to the analysis of English-speaking fluency of immigrants in the United Kingdom, focusing on Lazear's theory that due to learning or self-selection, there is a negative relation between speaking fluency and the ethnic minority concentration in the region. Specification tests show that the model allowing for misclassification errors outperforms ordered probit. All models lead to similar qualitative conclusions, but there is substantial variation in the size of the marginal effects.

KEY WORDS: Immigrant language proficiency; Measurement error; Misclassification; Semiparametric estimation.

1. INTRODUCTION

Many empirical studies in economics and other social sciences are concerned with the analysis of ordered categorical dependent variables, such as banded data on earnings, income, or hours worked. This data, often retrieved from surveys, have a true objective underlying scale but can be affected by misclassification error. Another type of categorical data that has become increasingly popular in applied econometrics is based on subjective evaluations. Examples include data on job satisfaction (see, e.g., Clark and Oswald 1996), satisfaction with health (Kerkhofs and Lindeboom 1995), future expectations of household income (Das and van Soest 1997), or subjective evaluations of English-speaking fluency of immigrants in the United Kingdom (e.g., Chiswick 1991; Chiswick and Miller 1995; Dustmann 1994), which we analyze in this article. Such data may suffer from the same misclassification problem. Moreover, the bounds used to distinguish, for example, good from reasonable, reasonable from bad, and so on, may be specific to the person doing the evaluation (the respondent or the interviewer).

In applied work, ordered categorical dependent variables are typically analyzed with ordered probit or ordered logit models. In these nonlinear models, misclassification can lead to biased estimates of the parameters of interest. To deal with this problem in the binary choice case, several parametric models have been introduced that explicitly incorporate misclassification probabilities as additional parameters. Lee and Porter (1984) estimated an exogenous switching regression model for market prices of grain, distinguishing regimes in which firms are cooperative and noncooperative. They observed an imperfect indicator of the actual regime and extended the standard probit model with two misclassification probabilities for the events that regime A is observed given that regime B is active or vice versa. They estimated these probabilities jointly with the parameters of the price equations in both regimes. Hausman, Abrevaya, and Scott-Morton (1998) estimated binary choice models for job changes. Using parametric models, they found significant probabilities of misclassifying in both directions. Using semiparametric models, they obtained estimates of the slope coefficients of interest that are similar to the estimates in the parametric model allowing for misclassification.

In this article we follow Lee and Porter (1984) and Hausman et al. (1998) and incorporate misclassification errors in an ordered response model. Standard tests cannot be used to test for the presence of misclassification errors, because the null hypothesis puts the parameters on the boundary of the parameter space. We apply a simulation-based testing procedure recently developed by Andrews (2001). We use the same type of test to test our model against a model that also allows for the possibility that different evaluators use different thresholds, generalizing the random thresholds model introduced by Das (1995).

In addition, we consider a semiparametric model that nests all parametric models and avoids distributional assumptions on the error terms. Because this is a single-index model, the slope parameters of interest can be estimated using the semiparametric least squares estimator of Ichimura (1993).

The main issue in our application is the relationship between host country language proficiency of immigrant minorities and the regional concentration of the minority group. Understanding the assimilation and adaptation of minority and immigrant groups is an important and growing area of research in economics, which is becoming ever more relevant as societies are increasingly characterized by a mix of individuals with different cultural backgrounds. Speaking a common language is a key factor in this process. In an influential recent study, Lazear (1999) developed a model in which trade between different groups requires the ability to communicate with one another. To enhance trading possibilities, minority individuals may learn the language of the majority group. The incentive to learn the language is larger the smaller the relative size of the minority group. Moreover, minority individuals with low proficiency in the majority language may sort themselves into communities in which individuals speaking their own minority language are concentrated. As Lazear pointed out, the two processes both

Journal of Business & Economic Statistics
July 2004, Vol. 22, No. 3
DOI 10.1198/073500104000000163

lead to a negative association between minority concentration and fluency in the majority language. If the effect of minority concentration on language is created primarily through learning, then the interaction between minority concentration and years of residence should contribute to explaining language proficiency. But if sorting is the only relevant mechanism, then this interaction should not be significant. Comparing data from the U.S. census for 1900 and 1990, Lazear concluded that only sorting mattered in 1990, whereas learning was important in 1900.

We investigate the same issue for the United Kingdom, using cross-sectional data on immigrants from ethnic minority communities drawn in 1994. Our parameters of interest are, as in Lazear's study, the effects of the regional minority concentration and its interaction with years of residence on English-language proficiency of immigrants.

In survey data, language proficiency is typically evaluated by the respondent or the interviewer on a four- or five-point scale, ranging from bad or very bad to very good. It seems likely that evaluators differ in terms of the perceived difference between bad and reasonable, reasonable and good, and so on. In addition, the reported variable may suffer from the same misclassification error as objective variables, such as the job change variable investigated by Hausman et al. (1998). Dustmann and van Soest (2002) focused on the latter type of error, comparing answers to identical survey questions on self-reported speaking fluency in the host country language by the same immigrants at different points in time. They found that, under the assumption that a decrease in language capacity is not possible, more than one-fourth of the total variance in the observed speaking fluency variable is due to random misclassification.

Our main empirical question is whether generalizing the ordered response model to allow for misclassification affects the answers to the economic questions concerning the relation between language proficiency, minority concentration, and years of residence. The results of our empirical analysis show that allowing for classification errors is a clear improvement over the standard ordered probit model. In particular, the estimated probabilities of misclassification into the extreme categories are large. A formal test based on work of Andrews (2001) clearly rejects the null hypothesis that all misclassification probabilities are 0. Moreover, the model with misclassification errors cannot be rejected against a more general model that also allows for random threshold variation across evaluators. Allowing for misclassification also leads to substantially different estimates of some of the slope coefficients of the regressors compared with ordered probit.

The qualitative conclusions on the effect of minority concentration on speaking fluency do not change if misclassification is allowed for. The effect is significantly negative. This is confirmed by the semiparametric estimates. The estimates of the size of the marginal effects, however, are substantially biased if misclassification is ignored, particularly at low values of the concentration index. The interaction term between years of residence and minority concentration is significant at the 10% level only in the parametric models and is insignificant in the semiparametric model, suggesting that for our particular application, self-selection is a better explanation for the negative relation between minority concentration and speaking fluency than learning.

The article is organized as follows. In Section 2 we present the models and their estimators. In Section 3 we briefly describe the data. We provide semiparametric and parametric estimates in Sections 4 and 5. In Section 6 we compare predictions of the two parametric models and the semiparametric model and test the parametric specifications. We provide some concluding remarks in Section 7.

2. CATEGORICAL DATA AND MISCLASSIFICATION

We assume that the dependent variable is observed on an ordinal scale with three levels, coded 1, 2, and 3. In our application, these levels correspond to speaking English slightly or not at all, reasonably well, and very well. The models that we discuss extend straightforwardly to the case of more than three categories, but the parametric models will lead to more auxiliary parameters and more intricate expressions for the likelihood function. The starting point is the ordered probit model, not allowing for classification errors. It relates observed categorical information for respondent i to an underlying latent index y_i^* as follows:

$$y_i^* = \mathbf{x}_i'\boldsymbol{\beta} + u_i; \tag{1}$$

$$y_i = j \quad \text{if} \quad m_{j-1} < y_i^* \le m_j, \qquad j = 1, 2, 3; \tag{2}$$

and

$$u_i | \mathbf{x}_i \sim \mathrm{N}(0, \sigma^2). \tag{3}$$

Here $\mathbf{x}_i$ is a vector of explanatory variables including a constant term, $\boldsymbol{\beta}$ is the vector of parameters of interest, and u_i is the error term. We assume that $m_0 = -\infty$, $m_1 = 0$, and $m_3 = \infty$. The variance σ^2 and the bound m_2 can be seen as nuisance parameters. We fix σ^2 to 100 to identify the scale. Throughout, we assume that the observations $(y_i, \mathbf{x}_i)$ are a random sample from the population of interest.

2.1 A Parametric Misclassification Model

For the binary choice case, Hausman et al. (1998) showed that the bias in estimates of $\boldsymbol{\beta}$ can be substantial if some observations on the endogenous variable are misclassified. They proposed a generalization of the binary probit model to take into account classification errors. We extend this model to the ordered probit case.

We assume that the reported category is y_i but the (unobserved) true category is z_i, which is related to the latent variable y_i^* as in the ordered probit model

$$z_i = j \quad \text{if} \quad m_{j-1} < y_i^* \le m_j, \qquad j = 1, 2, 3. \tag{4}$$

The probabilities of misclassification are given by

$$\Pr(y_i = j | z_i = k, \mathbf{x}_i) = p_{k,j}, \qquad j, k = 1, 2, 3, j \ne k. \tag{5}$$

Thus $p_{k,j}$ is the probability that an observation belonging to category k is classified in category j. If $p_{k,j} = 0$ for all j, k with $j \ne k$, then there is no misclassification, and the model simplifies to the ordered probit model. The model with three categories has six misclassification probabilities, $p_{k,j}$.

In this model, the latent variable y_i^* can be seen as a perfect indicator of speaking fluency on a continuous scale, something like the score on the ideal objective speaking fluency test.

The "true" category z_i is the categorical outcome based on this score. Misclassification means that the wrong outcome is reported. It should be acknowledged that this is only one way to model misclassification. For example, another source of misclassification would be measurement error in y_i^*, but a normally distributed measurement error would be captured in u_i and would not be identified. A third source would be individual variation in cutoff points, which we test for in Section 5.

The main identifying assumption in the model is that $p_{k,j}$ does not depend on $\mathbf{x}_i$ (except through z_i). This is the common identifying assumption in this literature, used by Hausman et al. (1998), Lee and Porter (1984), and Douglas, Smith Conway, and Ferrier (1995), among others. Lewbel (2000) showed that the binary choice model is still identified under the weaker assumption that one continuous variable with nonzero $\boldsymbol{\beta}$ does not affect the misclassification probabilities, but no obvious candidate for such an exclusion restriction is available. Assumptions like this could be avoided if a different measurement could be used as a benchmark, such as, in our empirical example, objective measurement of language proficiency (see Charette and Meng 1994).

For the binary choice case (with categories denoted by 0 and 1), Hausman et al. (1998) showed that identification of $p_{k,j}$, $j, k = 0, 1$, does not rely on the normality assumption, as long the support of $\mathbf{x}_i'\boldsymbol{\beta}$ is the whole real line; that is, as long as there are sufficient observations with very low and very high values of $\mathbf{x}_i'\boldsymbol{\beta}$. The probabilities of misclassification are then given by

$$p_{1,0} = \lim_{\mathbf{x}_i'\boldsymbol{\beta}\to\infty} \Pr(y_i = 0|\mathbf{x}_i)$$

and

$$p_{0,1} = \lim_{\mathbf{x}_i'\boldsymbol{\beta}\to-\infty} \Pr(y_i = 1|\mathbf{x}_i).$$

Hausman et al. (1998) showed that their model satisfies the single-index property that $E\{y_i|\mathbf{x}_i\}$ depends on $\mathbf{x}_i$ via $\mathbf{x}_i'\boldsymbol{\beta}$ only. Therefore, $\boldsymbol{\beta}$ is identified up to location, scale, and sign. The additional condition required for identification is that $p_{0,1}$ and $p_{1,0}$ not be too large,

$$p_{1,0} + p_{0,1} < 1. \tag{6}$$

This guarantees that $E\{y_i|\mathbf{x}_i\}$ increases with $\mathbf{x}_i'\boldsymbol{\beta}$. Accordingly, the sign of $\boldsymbol{\beta}$ is also identified, and (5) implies that $p_{0,1}$ and $p_{1,0}$ are nonparametrically identified.

For the ordered probit case with categories 1, 2, and 3 and six misclassification probabilities, we get

$$\begin{aligned} E\{y_i|\mathbf{x}_i\} &= 2 - p_{2,1} + p_{2,3} \\ &\quad - \Phi\big((m_1 - \mathbf{x}_i'\boldsymbol{\beta})/\sigma\big)(1 - p_{1,2} - p_{2,1} + p_{2,3} - 2p_{1,3}) \\ &\quad + \big[1 - \Phi\big((m_2 - \mathbf{x}_i'\boldsymbol{\beta})/\sigma\big)\big] \\ &\quad \times (1 - p_{3,2} - p_{2,3} + p_{2,1} - 2p_{3,1}). \end{aligned} \tag{7}$$

Thus the condition that $E\{y_i|\mathbf{x}_i\}$ increases with $\mathbf{x}_i'\boldsymbol{\beta}$ for every value of $\mathbf{x}_i'\boldsymbol{\beta}$ implies that [instead of (6) for the binary choice case]

$$\begin{aligned} &p_{1,2} + p_{2,1} - p_{2,3} + 2p_{1,3} < 1 \quad \text{and} \\ &p_{2,3} + p_{3,2} - p_{2,1} + 2p_{3,1} < 1. \end{aligned} \tag{8}$$

This condition is satisfied for sufficiently small values of the misclassification probabilities. A sufficient condition for (8) has been given by Abrevaya and Hausman (1999),

$$p_{1,1} > p_{2,1} > p_{3,1} \quad \text{and} \quad p_{3,3} > p_{2,3} > p_{1,3}. \tag{9}$$

This condition is stronger than (8) but easier to understand intuitively.

The argument for nonparametric identification in the binary choice case applies to $p_{1,2}, p_{1,3}, p_{3,1}$, and $p_{3,2}$, but not to $p_{2,1}$ or $p_{2,3}$. Identification of these is achieved in this parametric model by imposing normality of the error terms. The model can be straightforwardly estimated by maximum likelihood (ML), where the $p_{k,j}$'s are estimated jointly with the slope parameters $\boldsymbol{\beta}$.

2.2 A Semiparametric Approach

The parametric ML estimates of the slope parameters $\boldsymbol{\beta}$ in the models introduced earlier require distributional assumptions and may not be robust to misspecification. If we are interested in $\boldsymbol{\beta}$ only and consider the $p_{k,j}$ nuisance parameters, then semiparametric estimation seems to be a good alternative.

Consider the model with misclassification probabilities. The conditional mean of the observed categorical variable y_i in model (1)–(5) given $\mathbf{x}_i$ is given by (7). It depends on $\mathbf{x}_i$ only through the index $\mathbf{x}_i'\boldsymbol{\beta}$. Thus (1)–(5) is a special case of the single-index model given by

$$E\{y_i|\mathbf{x}_i\} = H(\mathbf{x}_i'\boldsymbol{\beta}), \tag{10}$$

where H is an unknown link function. If we relax the normality assumption (3) and replace it by the assumption that

$$u_i \text{ is independent of } \mathbf{x}_i, \tag{11}$$

then we get the following expression instead of (7):

$$\begin{aligned} E\{y_i|\mathbf{x}_i\} &= 2 - p_{2,1} + p_{2,3} \\ &\quad - G(m_1 - \mathbf{x}_i'\boldsymbol{\beta})(1 - p_{1,2} - p_{2,1} + p_{2,3} - 2p_{1,3}) \\ &\quad + [1 - G(m_2 - \mathbf{x}_i'\boldsymbol{\beta})] \\ &\quad \times (1 - p_{3,2} - p_{2,3} + p_{2,1} - 2p_{3,1}), \end{aligned} \tag{12}$$

where G is the distribution function of the error term u_i ($G(t) = \Pr[u_i \le t]$).

Again, the right-hand side depends on $\mathbf{x}_i$ only through $\mathbf{x}_i'\boldsymbol{\beta}$, so that (1), (2), (4), (5), and (11) lead to the single-index model (10) with link function H given by (12). As stated before, the crucial assumption here is that the misclassification probabilities in (4)–(5) do not depend on $\mathbf{x}_i$.

Moreover, under the same assumptions, it is straightforward to show that the conditional variance $V\{y_i|\mathbf{x}_i\}$ also depends on $\mathbf{x}_i$ through the same index $\mathbf{x}_i'\boldsymbol{\beta}$ only. This implies that the model for y_i is heteroscedastic, but the heteroscedasticity has a special form. Finally, it is easy to show that the inequalities in (8) imply that H can be chosen to be nondecreasing.

Thus the model discussed earlier is a special case of the general single-index model (10) for some (unknown) link function H. In this model the vector $\boldsymbol{\beta}$ of slope parameters is identified up to scale; the constant term is not identified. A number of asymptotically normal root n-consistent estimators for $\boldsymbol{\beta}$

in this model have been discussed in the literature, requiring various assumptions on the distribution of the explanatory variables $\mathbf{x}_i$ and regularity conditions on the link function H. Ichimura (1993) used nonlinear least squares combined with nonparametric estimation of H. This estimator requires numerical minimization of a nonconvex objective function. Hausman et al. (1998) used the maximum rank correlation estimator of Han (1987). This also requires numerical optimization. We experimented with applying this estimator but ran into convergence problems with the Han estimator, possibly due to the relatively large number of explanatory variables.

Attractive from a computational standpoint is the class of (weighted or unweighted) average derivative estimators (see, e.g., Powell, Stock, and Stoker 1989). They require that the distribution of $\mathbf{x}$ be absolutely continuous and thus are not directly applicable to our empirical example. Horowitz and Haerdle (1996) have developed an estimator that allows for discrete variables, but not for interaction terms of continuous variables. Because interaction terms are important in our particular application, the Horowitz and Haerdle (1996) estimator cannot be applied. We therefore focus on Ichimura's semiparametric least squares (SLS) estimator.

Ichimura's SLS estimator minimizes the sum of squares $S_n(\boldsymbol{\beta})$ over $\boldsymbol{\beta}$, where

$$S_n(\boldsymbol{\beta}) = 1/n \sum \left(y_i - \hat{E}[y_i|\mathbf{x}_i'\boldsymbol{\beta}]\right)^2. \qquad (13)$$

Here $\hat{E}[y_i|\mathbf{x}_i'\boldsymbol{\beta}]$ is a univariate kernel regression estimate of y_i on the index $\mathbf{x}_i'\boldsymbol{\beta}$ (for given $\boldsymbol{\beta}$). Finding the $\boldsymbol{\beta}$ at which (13) is minimized requires an iterative procedure. If smooth kernel weights are used, then the function to be minimized is smooth in $\boldsymbol{\beta}$, and a Newton–Raphson technique can be used to find the optimal $\boldsymbol{\beta}$, that is, $\hat{\boldsymbol{\beta}}_{\text{SLS}}$. Ichimura (1993) showed that under appropriate regularity conditions, this yields a $\sqrt{n}$-consistent asymptotically normal estimator. He also derived the asymptotic covariance matrix of this estimator and showed how it can be estimated consistently.

Ichimura (1993) also indicated how to design an asymptotically efficient weighted semiparametric least squares (WSLS) estimator that uses SLS as the first step. For the general case, this requires nonparametric regression of the squared SLS residuals on $\mathbf{x}$ and leads to problems if $\mathbf{x}$ contains interaction terms or discrete variables. In our case, however, we have demonstrated that the natural generalization of the parametric models implies that $V[y_i|\mathbf{x}_i]$ depends on $\mathbf{x}_i$ only through $\mathbf{x}_i'\boldsymbol{\beta}$, and for this special case Ichimura showed that the efficient WSLS estimator requires weighting with $\hat{V}[y_i|\mathbf{x}_i'\hat{\boldsymbol{\beta}}_{\text{SLS}}]^{-1}$, obtained by a nonparametric regression of the squared SLS residuals on the index $\mathbf{x}_i'\hat{\boldsymbol{\beta}}_{\text{SLS}}$.

Implementing the SLS and WSLS estimators in practice requires choosing a kernel and a bandwidth. We work with the Gaussian kernel. For consistency, the bandwidth should tend to 0 if $n \to \infty$ at a sufficiently slow rate. Theoretical results for similar problems suggest that undersmoothing will be optimal; that is, the optimal bandwidth will be smaller than the optimal bandwidth for the nonparametric regression of y_i on $\mathbf{x}_i'\boldsymbol{\beta}$. The common approach for choosing a bandwidth in a situation like this is to experiment with the bandwidth that would be optimal for the nonparametric regression problem (given a value of $\boldsymbol{\beta}$) and with smaller bandwidth values (to under smooth). We will present results for several choices of the bandwidth.

Once $\boldsymbol{\beta}_{\text{SLS}}$ (or $\boldsymbol{\beta}_{\text{WSLS}}$) is obtained, the link function H can be estimated by a nonparametric (kernel) regression of y_i on the estimated index $\mathbf{x}_i'\hat{\boldsymbol{\beta}}_{\text{SLS}}$. The usual asymptotic properties of a kernel estimator apply because $\hat{\boldsymbol{\beta}}_{\text{SLS}}$ converges at a faster rate than the nonparametric estimator.

3. DATA

We apply the models and techniques discussed earlier to analyze the effect of minority concentration on immigrants' proficiency in the host country language. The empirical analysis is based on the Fourth National Survey on Ethnic Minorities (FNSEM), a cross-sectional survey carried out in the United Kingdom in 1993 and 1994. Individuals included are age 16 or older. There are 5,196 observations in the minority sample. We focus on a homogeneous sample of 1,471 men of Indian ethnicity (from India, Bangladesh, Pakistan, or Uganda). The FNSEM contains information on the concentration of the individual's own minority group at the ward level, which has been matched to the survey from the 1991 Census. (A ward is the smallest geographical area identified in the Population Census, with a mean population of 5,459 individuals in 1991.) The language information in the survey is based on the interviewer's evaluation of the individual's language ability in English, with categorical answers (speaks English) very well, fairly well, slightly, and not at all. For the empirical analysis, we have combined the categories slightly and not at all and recoded the three categories as 3 (very well), 2 (fairly well), and 1 (slightly or not at all).

Summary statistics on the resulting categorical speaking fluency variable and on other individual characteristics are presented in Table 1. About 47% of the 1,471 men in the survey data are reported to speak English very well. Only 4.3% are assigned to the category not at all; this group is merged with the 22.6% in the category slightly.

On average, the concentration of minorities of the same ethnicity as the respondent is about 16.2%, with substantial variation in the sample and a sample standard deviation of 15.2%. There is a clear negative correlation between language proficiency and minority concentration. Average minority concentration in the subsample of people with low speaking fluency is about 20.8%, in the subsample of the most fluent speakers it is only 13.7%. The rank correlation coefficient is −.215 (with p-value .000).

Table 1. Variable Definitions and Sample Statistics

Variable	*Code*	*Mean*	*Standard deviation*
Speaks English slightly or not at all	SPF = 1	.2685	
Speaks English fairly well	SPF = 2	.2624	
Speaks English very well	SPF = 3	.4691	
Age (years)	age	42.38	14.27
Years since migration	ysm	19.58	9.35
Country of birth: Africa	afroas	.2271	
Country of birth: Bangladesh	bangladesh	.1788	
Country of birth: India	indian	.2998	
Country of birth: Pakistan	pakistan	.2944	
Minority concentration (%)	conc index	16.20	15.20

NOTE: Source: FNSEM, 1,471 observations.

4. SEMIPARAMETRIC ESTIMATES

Some of the SLS and WSLS estimates explained in Section 2.3 are presented in Table 2. The first column presents SLS estimates with the bandwidth set equal to $1.06\hat{\sigma}(\mathbf{x}'\hat{\boldsymbol{\beta}})n^{-.2}$, where n is the number of observations and $\hat{\sigma}(\mathbf{x}'\hat{\boldsymbol{\beta}})$ is the estimated standard deviation of the single index. This is the rule-of-thumb estimate for the optimal bandwidth in the kernel regression (Silverman 1986). Because undersmoothing typically gives more efficient estimates for the single index (Powell 1994), we also present the results for a bandwidth that is half as large (third column). The differences between the two sets of estimates or their standard errors are small, confirming the general finding in this literature that the SLS results are not sensitive to the choice of the bandwidth (see, e.g., Bellemare, Melenberg, and van Soest 2002). The second column presents the WSLS estimates, using the same bandwidth as in the first column. These estimates are very similar to those in the first column. Estimated standard errors are somewhat smaller in most cases, in line with the fact that WSLS is asymptotically efficient but SLS is not, but there are also two parameters for which the estimated standard error is slightly larger for WSLS than for SLS. Results for the smaller bandwidth (not presented) tell the same story.

Standard errors are based on the asymptotic distribution of the estimator. Bootstrapped standard errors give the same economic conclusions and thus are not presented. They are larger than the asymptotic standard errors for some parameters and smaller for others.

The constant term is not estimated. The coefficient of YSM (years since migration) is normalized to .9634, its estimate in the ordered probit model (see below). This normalization makes it easy to compare semiparametric and parametric results. The variable YSM has a significant positive effect with a large absolute t value in all parametric models, which justifies the assumption that the coefficient is nonzero, the (only) necessary condition for using this normalization.

The estimation results are qualitatively in line with those reported by Lazear (1999). Because not only YSM itself, but also YSM squared and YSM interacted with the minority concentration index, are included among the regressors, the effect of an increase of YSM on expected speaking fluency varies across observations. Nonetheless, the marginal effect of increasing YSM on expected fluency is positive at almost all observations. The negative sign of YSM squared implies that this effect is smaller for those with longer years of residence. Conditional on YSM, older immigrants are less fluent in English than younger immigrants. The country of origin dummies indicate that, keeping other characteristics constant, immigrants from Pakistan and Bangladesh are significantly less fluent than immigrants from India, whereas the individuals of Afro-Asian origin are the most fluent.

Speaking fluency falls with minority concentration at a declining rate, confirming Lazear's finding for the United States. One explanation for this is that individuals who live in areas with high concentrations of residents of their own minority have lower incentives to learn the majority language. Another explanation is that individuals select their area of residence according to their language proficiency. As Lazear pointed out, a significant negative effect of the concentration variable on speaking fluency is consistent with both explanations. In both cases, the individual's (location or learning) choice is determined by the objective of maximizing interaction with individuals with whom they share a common language.

To distinguish between the two explanations, Lazear added an interaction term between minority concentration and years of residence (YSM). An insignificant interaction term favors the self-selection hypothesis, because the learning argument would imply a negative interaction effect—a larger learning rate, (i.e., a higher effect of YSM) when learning pays off more (i.e., when minority concentration is lower). In Table 2 the coefficient on the interaction term of YSM and minority concentration is negative but insignificant and close to 0, favoring the self-selection hypothesis. Interestingly, this result is similar to what Lazear found for the 1990 U.S. census.

Figure 1 illustrates the estimated link function H in (10) for the first set of results in Table 2. (The quartic kernel was used, with bandwidth chosen by visual inspection.) The figure looks very similar for the other results. The figure also shows 95% uniform confidence bounds (based on Haerdle and Linton 1994). The estimated link function is increasing on its full domain except at very low values of the index, for which the estimates are imprecise due to the small number of observations in that region. In an ordered response model without misclassification, the value of the link function should tend to 1 if the index value tends to $-\infty$. The figure suggests that this is not the case, however. This could be due to misclassification of some respondents with low speaking fluency.

Table 2. Semiparametric Estimation Results

	SLS; $h = 1.5470^a$		WSLS; $h = 1.5470^a$		SLS; $h = .7563^b$	
Bandwidth	Coefficient	Standard error	Coefficient	Standard error	Coeffecient	Standard error
ysm	.9634		.9634		.9634	
age	−.9617	.1071	−.9923	.1201	−.9840	.1094
conc index	−25.1826	6.4955	−26.5351	6.4125	−20.2509	6.1338
afroas	4.2826	.9689	4.1344	.9171	4.1996	.9296
pakistan	−4.2943	.8178	−4.4190	.7726	−3.7830	.7825
bangladesh	−4.3825	.8960	−4.6807	.8785	−4.1162	.8716
age sq	.0061	.0010	.0063	.0010	.0064	.0010
ysm sq	−.0140	.0011	−.0139	.0010	−.0147	.0010
conc ind sq	32.3767	9.2184	34.0762	8.8466	24.2073	8.6987
ysm * conc ind	−.0655	.1420	−.0656	.1524	−.0872	.1460

[a] $1.06\sqrt{\hat{V}(x'\hat{\beta})}n^{-.2}$ (Silverman's rule of thumb).

[b] $.53\sqrt{\hat{V}(x'\hat{\beta})}n^{-.2}$.

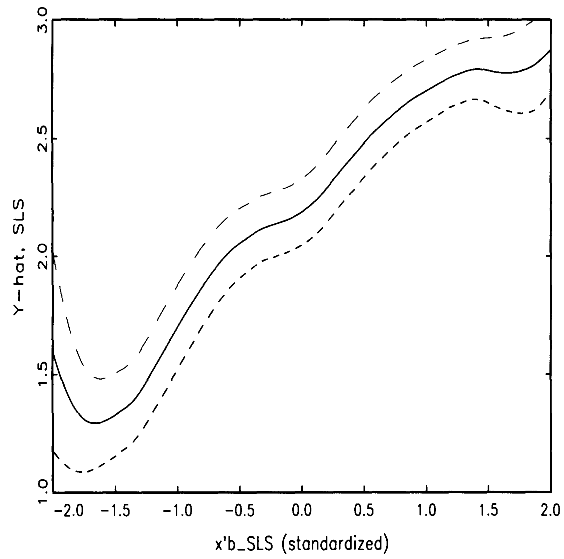

Figure 1. Link Function of the Semiparametric Model. This is a nonparametric kernel regression of speaking fluency (1, slightly or not at all; 2, fairly well; 3, very well) on the semiparametric estimate of the index. The broken curves are uniform 95% confidence bands. The link function is monotonically increasing but does not range from 1 to 3, suggesting misclassification.

5. PARAMETRIC ESTIMATES

Estimates for several parametric models are presented in Table 3. The first column gives the results of the standard ordered probit model. The second column incorporates misclassification probabilities (see Sec. 2.1).

The two sets of parametric estimates of the slope coefficients are generally in line with each other in terms of signs and significance levels, but there are substantial differences in magnitude. We discuss the magnitude of the most important estimates later when we look at predicted marginal effects on the probabilities of good and reasonable speaking fluency. The coefficients all have the same sign as in the semiparametric model. Fluency increases with YSM at a decreasing rate. Immigrants from Pakistan and Bangladesh are less fluent than immigrants from India, whereas Afro-Asian immigrants have the highest fluency, ceteris paribus. Speaking fluency is lower in regions where the concentration of immigrants from the same country of origin is larger.

The estimated coefficient on the interaction term of minority concentration and YSM is negative and significant at approximately the two-sided 10% level in both models. This differs from the semiparametric estimates, which were negative but smaller in magnitude and not significant at all. Whereas the semiparametric evidence suggested that the negative effect of minority concentration on speaking fluency is due to self-selection into local areas and not to the effort at learning the language, the parametric results suggest that learning could play a role as well. But t values are not sufficiently high to draw any final conclusions on this. For those with 0 years of residence, the estimated pattern of speaking fluency as a function of minority concentration is decreasing up to about the 88th percentile of minority concentration according to the model with misclassification and up to the 94th percentile for the semiparametric models. This suggests that already shortly after entry, immigrants in low minority concentration areas speak better English, a finding that can be explained only by self-selection.

The misclassification probabilities in the second column are by definition nonnegative, implying that standard t tests or likelihood ratio tests on $p_{k,j} = 0$ are inappropriate (see, e.g., Shapiro 1985). Still, the estimates of the $p_{k,j}$ and their standard errors imply that 0 is not contained in the one-sided 95% confidence intervals of four of them, suggesting that adding the probabilities of misclassification is an improvement compared to the standard ordered probit model. A formal test of the hypothesis $p_{k,j} = 0$ for all $j \neq k$ can be based on the likelihood ratio (LR) using the method proposed by Andrews (2001). The LR test statistic does not have the usual chi-squared distribution under the null, because the test is one-sided and because under the null, the parameter vector is not in the interior of the parameter

Table 3. Estimation Results Parametric Models

	Ordered probit		*Misclassification model*	
	Coefficient	*Standard error*	*Coefficient*	*Standard error*
Constant	28.4750	3.3542	55.6333	13.2788
ysm	.9634	.1342	2.0196	.4168
age	−.8258	.1411	−1.6342	.4246
conc index	−34.0386	7.7247	−64.1300	17.8579
afroas	3.7520	.9326	7.9896	2.5771
pakistan	−6.0868	.8292	−9.6401	2.2333
bangladesh	−6.0094	.9649	−10.0340	2.4232
age sq	.0041	.0015	.0082	.0034
ysm sq	−.0152	.0032	−.0314	.0079
conc ind sq	48.2251	10.8978	93.2072	24.1164
ysm * conc ind	−.3918	.2307	−.6923	.4164
m_2	8.7001	.3913	23.2845	5.3590
Probability 2 if 1			0	—*
Probability 3 if 1			.1029	.0458
Probability 1 if 2			.2725	.0473
Probability 3 if 2			.2450	.0570
Probability 1 if 3			.0095	.0146
Probability 2 if 3			.1042	.0381
Log-likelihood	−1,317.646		−1,309.332	

*Estimate at the lower bound.

space. Andrews (2001) demonstrated that the LR test statistic can still be used and showed how to compute the appropriate asymptotic critical values, using a quadratic approximation to the likelihood. In the Appendix we give the algorithm used for our case. We find a 5% critical value of 9.04 and a 1% critical value of 12.88. Because the realization of the LR test statistic is 16.72, the null hypothesis is rejected at the 1% level. This confirms that allowing for misclassification errors significantly improves the fit of the model.

Another way to investigate the validity of the model is to test it against a more general parametric model that allows the threshold parameters to vary across evaluators. Evaluators (in our case, the interviewers) are not precisely instructed on how to construct their scores. This makes allowing for heterogeneity in the threshold values intuitively attractive, because it implies that two evaluators who perceive the same proficiency y_i^* may still give different answers on the ordinal scale. Das (1995) allowed for unobserved heterogeneity in the bounds. His approach can be straightforwardly extended to also allow for misclassification. (Details and estimation results of this model are available on request from the authors.) The Andrews test could not reject the model with classification errors only against the model with classification errors as well as random thresholds (test statistic, 2.12; 10% critical value, 2.90). This supports the model with misclassification errors and implies that extending the model with random variation in thresholds is not necessary.

The estimates of the misclassification probabilities in Table 3 amply satisfy the inequalities of Abrevaya and Hausman (1999) that are sufficient for identification and imply monotonicity of the link function. The estimates of $p_{2,1}$ and $p_{2,3}$ have the largest standard errors, reflecting the problem that these are more difficult to identify. Compared with the ordered probit model, most slope coefficients and the estimate of the category bound m_2 have increased by approximately a factor of 2. Due to the normalization, this can also be seen as a reduction of the standard deviation of the error term u by about 50%. The interpretation is that part of the unsystematic variation in observed speaking fluency is now explained by classification errors.

The results of the parametric models can be used to analyze the size of the effects of concentration of immigrants of a certain language minority on true speaking fluency, not affected by misclassification error. Table 4 summarizes the results. It presents the estimated marginal effects of minority concentration on the probabilities of at least slight fluency and very good fluency according to each of the models in Table 3 at the first, second, and third quartiles of the sample distribution of the concentration index. Other regressors have been set to their sample means. The estimated marginal effects are functions of the estimates of $\boldsymbol{\beta}$ and m_2. Misclassification probabilities are discarded; the marginal effects refer to the true classification, not to the reported classification.

The table shows some substantial differences in the estimated marginal effects. For example, let us compare two otherwise identical immigrants in a region with approximately median ethnic concentration. If the area of the one immigrant has a 1-percentage point higher ethnic concentration than the area of the other immigrant, then the ordered probit model predicts a 1.33-percentage point lower probability of speaking English very well for the immigrant in the lower concentration area. According to the misclassification model, the difference has the same sign but is much larger, about 2.27-percentage points (with standard error .44-percentage points).

Model 2 allows for misclassification and significantly outperforms the ordered probit model. But it leads to much larger standard errors on the estimated marginal effects. As an intermediate case, we also estimated a model that allows for misclassification in an adjacent category, but not in nonadjacent categories. In other words, we imposed $p_{1,3} = p_{3,1} = 0$ in model 2. We do not present detailed results for this model, because this model is formally rejected against model 2. Nonetheless, most of the estimation results are similar to those of model 2. The estimates of the misclassification probabilities are, for example, $\hat{p}_{1,2} = 0$ (the lower bound), $\hat{p}_{2,1} = .2528$ (standard error .0468), $\hat{p}_{2,3} = .3023$ (standard error .0384), and $\hat{p}_{3,2} = .0877$ (standard error .0379)—values that are similar to those given in Table 3. The estimated marginal effects are also similar to those of model 2, but with standard errors that are about 20% smaller on average.

5.1 Comparing Two Parametric Models

Figures 2 and 3 compare the predictions of two parametric models, ordered probit and the misclassification model. We look at the estimated probabilities that true fluency is (at least) good and that reported fluency is good. In the ordered probit model, observed and true speaking fluency (y and z) coincide, but in the model with misclassification errors they do not.

Figure 2 presents a scatterplot of the predicted probabilities of good speaking fluency according to the two parametric models. For the misclassification model (vertical axis), the figure shows the predictions of the true speaking fluency variable z.

Table 4. Marginal Effects of Minority Concentration, Parametric Models

Quantile of minority concentration	Ordered probit: Effect	Ordered probit: Standard error	Misclassification model: Effect	Misclassification model: Standard error
P(fairly or very fluent)				
At 1st quartile	−.8811	.0972	−.1857	.1878
At median	−.9255	.1140	−.3733	.2614
At 3rd quartile	−.7895	.0928	−.6286	.2422
P(very fluent)				
At 1st quartile	−1.5310	.1983	−2.8962	.5999
At median	−1.3293	.1638	−2.2688	.4407
At 3rd quartile	−.8728	.0860	−1.0649	.1985

NOTE: Marginal effect of an increase of ethnic concentration by 1-percentage point on the probability (in percentage points) of speaking English fairly or very well (top) or very well (bottom).

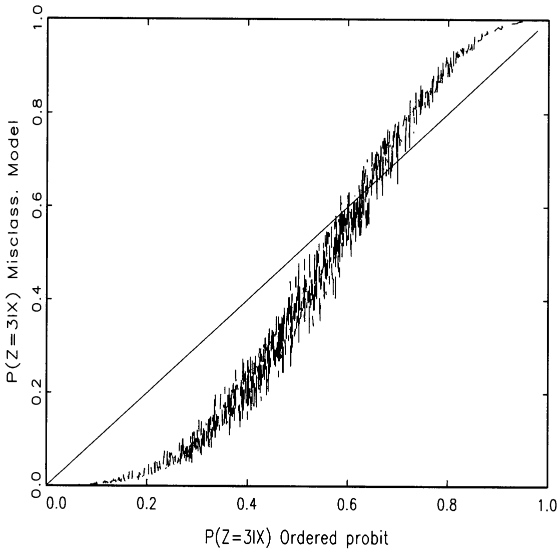

*Figure 2. Comparing P[z = 3|**x**] for the Ordered Probit and the Misclassification Model. Scatterplot of the predicted probabilities that someone speaks English very well according to the ordered probit (horizontal axis) and the misclassification model (vertical axis). The solid line is the 45-degree line. The models give clearly different predictions of true speaking fluency.*

For the ordered probit model (horizontal axis and 45-degree line), predictions of y and z coincide. We find that the misclassification model leads to more probability estimates close to 0 or 1 than the ordered probit model, leading to a larger dispersion in $\hat{P}[z = 3|\mathbf{x}]$ according to the misclassification model than according to ordered probit. Still, the correlation between the two sets of predictions is quite large (with a sample correlation coefficient of .97).

Figure 3 compares predictions of the probability that individuals report good or very good speaking fluency. In the misclassification model, the probability of reporting good or very good fluency is never close to 1 or 0. For most observations with predicted probabilities not close to 1 or 0, the predictions according to ordered probit and misclassification models are similar. The correlation coefficient is almost .99.

The substantial differences between true and reported fluency in the misclassification model confirm the conclusion from the misclassification probabilities in Table 4. Generalizing the ordered probit model by incorporating misclassification probabilities is useful in this empirical example. The same conclusion is obtained for the probability of bad or very bad speaking fluency (figures not reported).

5.2 Misspecification Tests of Parametric Models

In principle, the parametric models could be tested against the semiparametric model using a Hausman test. Under the null that the parametric model is correct, the parametric ML estimates are asymptotically efficient and the SLS estimates are consistent. Under the alternative that the semiparametric model is correctly specified but the parametric model is not, only the SLS estimates are consistent. Thus a chi-squared test can be based on the difference between parametric and semiparametric estimates. Unfortunately, however, the estimated standard errors of the SLS estimates are not always larger than those of the parametric ML estimates. This implies that the Hausman test statistic cannot be computed. This problem remains if bootstrapped standard errors are used for the semiparametric model. The procedure of Newey (1985) cannot be used, because it does not apply to the semiparametric estimator.

An alternative, graphical, specification test of parametric models has been introduced by Horowitz (1993). The null hypothesis is that the parametric model is correctly specified. The result for the parametric model with misclassification is illustrated in Figure 4. The figure presents two functions of the index

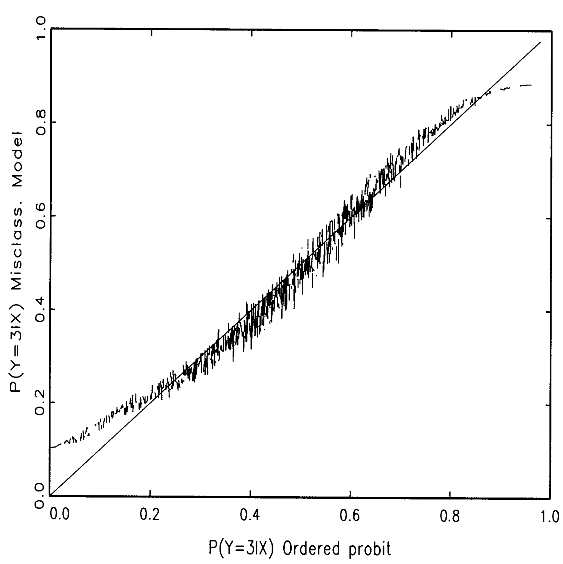

*Figure 3. Comparing P[y = 3|**x**] for the Ordered Probit and the Misclassification Model. Scatterplot of the predicted probabilities that someone is reported to speak English very well according to the ordered probit (horizontal axis) and the misclassification model (vertical axis). The solid line is the 45-degree line. The models give similar predictions of reported speaking fluency except in the tails.*

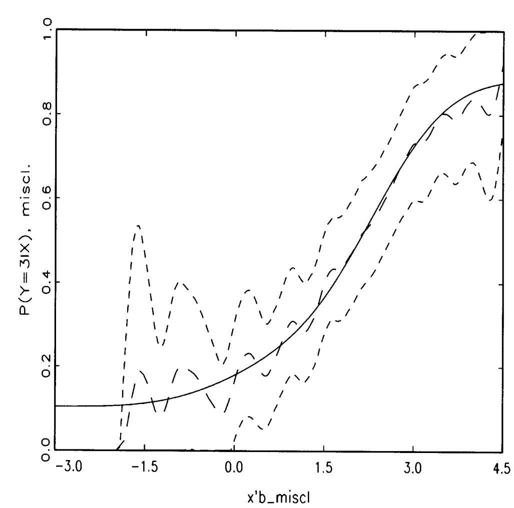

Figure 4. Misspecification Test for the Misclassification Model. The solid curve is the predicted probability of reported very good speaking fluency as a function of the estimated index for the misclassification model. The dashed lines are a nonparametric regression of a dummy for very good reported speaking fluency on the estimated index, with 95% uniform confidence bounds. The fact that the solid curve lies between the lower and upper confidence band (except where there are hardly any data) implies that the null hypothesis that the misclassification model is correct is not rejected.

estimate $\mathbf{x}'\mathbf{b}/s$, where $\mathbf{b}$ and s are the parametric estimates of $\boldsymbol{\beta}$ and σ in Table 3. The solid curve gives the predicted probabilities $\hat{P}[y_i = 3|\mathbf{x}_i] = \hat{P}[y_i = 3|\mathbf{x}_i'\mathbf{b}]$ according to the parametric model, as a function of $\mathbf{x}_i'\mathbf{b}$. The dashed curves represent nonparametric kernel regression estimates of the observed dummy indicator variable $I(y_i = 3)$ on the same index $\mathbf{x}_i'\mathbf{b}$ with uniform 95% confidence bands. Because the estimator $\mathbf{b}$ converges to $\boldsymbol{\beta}$ at rate root n, which is a faster rate than the rate of convergence of the nonparametric estimator, the standard errors of $\mathbf{b}$ are asymptotically negligible, and confidence bands are calculated as if $\mathbf{b}$ were known.

Under the null that the parametric model is specified correctly, $\mathbf{b}$ is consistent for $\boldsymbol{\beta}$, and the parametric expression for the predicted probability $\hat{P}[y_i = 3|\mathbf{x}_i]$ is consistent for $P[y_i = 3|\mathbf{x}_i]$. But the null hypothesis also implies that $P[y_i = 3|\mathbf{x}_i]$ is a single-index function of $\mathbf{x}_i'\boldsymbol{\beta}$ and that $\mathbf{b}$ is a consistent estimate of this single index (up to scale). The nonparametric curve is the estimated link function, which also will be consistent for $P[y_i = 3|\mathbf{x}_i]$. Thus under the null, both curves are consistent for the same function and should be similar. The null hypothesis will be rejected if the nonparametric (dashed) curve is significantly different from the parametric (solid) curve. Because the parametric curve is based on estimates that converge at rate $\sqrt{n}$, whereas the nonparametric curve converges at the lower rate $n^{.4}$, the imprecision in the former curve can be neglected compared with that in the latter curve, and the test can be based on the uniform confidence bands around the nonparametric curve.

The result is that the solid curve is everywhere between the uniform confidence bands, so that the parametric model cannot be rejected. This can be seen as evidence in favor of the parametric misclassification model. It should be admitted, however, that the same test cannot reject the ordered probit model either, whereas we already saw that the Andrews test rejects this model against the model with misclassification errors. This casts some doubt on the power of this type of test. The same conclusions are obtained if $P[y_i = 1|\mathbf{x}_i'\mathbf{b}]$ is used instead of $P[y_i = 3|\mathbf{x}_i'\mathbf{b}]$.

6. SUMMARY AND CONCLUSIONS

In models with ordered categorical dependent variables where the categorical assignment is based on subjective evaluations, misclassification may have two sources: classical misclassification due to simple reporting errors and misclassification due to a subjective choice of scale. Both sources can lead to seriously biased parameter estimates and predictions. Parametric estimators that incorporate and estimate misclassification probabilities, as well as semiparametric estimators, are an alternative to standard parametric models. Extending the work of Lee and Porter (1984) and Hausman et al. (1998), we have introduced a parametric model that incorporates misclassification probabilities for the case of more than two ordered categories and that allows for scale heterogeneity. We have shown that this model is a special case of a semiparametric single-index model that can be estimated with semiparametric least squares.

Using these models, we analyzed the association between minority concentration and speaking fluency of immigrants, using data for the United Kingdom. We found that the misclassification model is a significant improvement over the standard probit model. Allowing for random thresholds does not lead to further improvements. The qualitative effects of minority concentration are similar, supporting Lazear's finding for the United States that speaking fluency falls with minority concentration. However, marginal effects show that the size of the correlation and the shape of the relationship between fluency and minority concentration are quite different according to the two models. The models both show weak evidence in favor of a learning effect, reflected by a negative interaction effect of minority concentration and YSM that is significant at the one-sided 10% level. The evidence in favor of self-selection of more fluent immigrants into areas with lower minority density is much stronger and insensitive to the chosen model. Semiparametric estimates in a model that nests all parametric models considered confirm the qualitative conclusions, although the evidence of a learning effect is even weaker.

A shortcoming of the model is that probabilities of misclassification in intermediate categories are not precisely estimated, because their identification relies on parametric assumptions. Better estimates of all misclassification probabilities would require additional data, such that alternative measurements (Charette and Meng 1994) or panel data. This is on our research agenda.

ACKNOWLEDGMENTS

The authors thank an anonymous referee, an associate editor, the editor, Bertrand Melenberg, Frank Windmeijer, Marcel Das, and Charles Bellemare for useful comments.

APPENDIX: ANDREWS TEST

Here we explain how to test the null hypothesis $H_0: p_{jk} = 0$, $j, k = 1, 2, 3$, $j \neq k$, against the alternative $p_{jk} > 0$ for at least one pair $j \neq k$. (Tests for random thresholds against fixed thresholds are constructed in the same way.) Because the model is not defined for $p_{jk} < 0$, the parameter vector is not an internal point of the parameter space under the null hypothesis, implying that standard asymptotic theory of the ML estimator does not apply. It also imples that alternative tests for inequality constraints, such as those of Andrews (1998) or Szroeter (1997), cannot be applied. Andrews (1999) derived the asymptotic distribution of a class of a general class of estimators including ML when the true parameter value lies on the boundary of the parameter space. Andrews (2001) later applied the earlier results (Andrews 1999) to derive the asymptotic distribution of the quasi-LR test statistic, which is what we need. (Andrews also allowed for nuisance parameters, which play a role under the alternative only. Such parameters do not appear in our case; see thm. 4 in Andrews 2001.) The special case without nuisance parameters that are not identified under the null also follows from theorem 3 of Andrews (1999). It is straightforward to check that the regularity assumptions required for this theorem are satisfied in our example, because observations are iid, ML estimation is used, the log-likelihood has continuous right partial derivatives of second order, and the parameter space has the form of a convex cone. Checking the regularity conditions is basically the same as for the example of a random coefficients model of Andrews (1999).

Let LR represent the LR test statistic, $2(\ln L_1 - \ln L_0)$, where L_1 is the unrestricted maximum of the likelihood (allowing for all $p_{j,k} \geq 0$) and L_0 is the restricted maximum (imposing $p_{j,k} = 0$ for all $j, k = 1, 2, 3$). The parameter vector can be written as $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')'$, where $\boldsymbol{\theta}_2$ contains the six misclassification probabilities $p_{1,2}, \ldots, p_{3,2}$ and $\boldsymbol{\theta}_1$ contains the other 12 (unrestricted) parameters of the model. The parameter space can be written as $V = (-\infty, \infty)^{12} \times [0, \infty)^6$, and the null hypothesis is $\boldsymbol{\theta} \in V_0 = (-\infty, \infty)^{12} \times \{0\}^6$. (We ignore the obvious lower bound on the threshold m_2, because it is not binding and is irrelevant for the local approximations.) Let $\mathbf{J}$ be minus the expected value of the Hessian of the log-likelihood contribution of a random observation at the true parameter values, which under the null can be consistently estimated in the usual way by $\hat{\mathbf{J}}$, the sample mean of the matrix of second-order partial derivatives at each observation, evaluated at the restricted ML estimates. Similarly, let $\mathbf{I}$ be the expected value of the outer product of the gradient of the log-likelihood contribution of a random observation and let $\hat{\mathbf{I}}$ be its natural estimate under the null. The only difference with the usual case of an internal point of the parameter space is that right partial derivatives are used for the parameters $p_{j,k}$.

Theorem 4 of Andrews (2001) now implies that LR has the same asymptotic distribution as

$$\mathrm{Inf}_{[\boldsymbol{\theta} \in V_0]}\, q(\boldsymbol{\theta}) - \mathrm{Inf}_{[\boldsymbol{\theta} \in V]}\, q(\boldsymbol{\theta}), \qquad \text{(A.1)}$$

with

$$q(\boldsymbol{\theta}) = (\boldsymbol{\theta} - \mathbf{Z})'\mathbf{J}(\boldsymbol{\theta} - \mathbf{Z}), \qquad \mathbf{Z} \sim \mathrm{N}(\mathbf{0}, \mathbf{J}^{-1}\mathbf{I}\mathbf{J}^{-1}). \qquad \text{(A.2)}$$

The asymptotic distribution of LR is thus be obtained by the following simulation procedure:

- Plug in the estimates $\hat{\mathbf{J}}$ for $\mathbf{J}$ and $\hat{\mathbf{I}}$ for $\mathbf{I}$. (As in the usual ML case, $\mathbf{I}$ and $\mathbf{J}$ coincide under the null, so an asymptotically equivalent procedure is to use an estimate for only one of them.)
- Generate multivariate normal draws of $\mathbf{Z}$.
- Solve the two quadratic programming problems in (A.1) for each draw.
- Consider the obtained simulated distribution of the difference between the two minimum values.

[Received September 2001. Revised October 2003.]

REFERENCES

Abrevaya, J., and Hausman, J. A. (1999), "Semi-Parametric Estimation With Mismeasured Dependent Variables: An Application to Duration Models for Unemployment Spells," *Annales d'Economie et de Statistique*, 55–56, 243–276.

Andrews, D. W. K. (1998), "Hypothesis Testing With a Restricted Parameter Space," *Journal of Econometrics*, 84, 155–199.

——— (1999), "Estimation When a Parameter Is on a Boundary," *Econometrica*, 67, 1341–1384.

——— (2001), "Testing When a Parameter Is on the Boundary of the Maintained Hypothesis," *Econometrica*, 69, 683–734.

Bellemare, C., Melenberg, B., and van Soest, A. (2002), "Semi-Parametric Models for Satisfaction With Income," *Portuguese Economic Journal*, 1, 181–203.

Borjas, G. J. (1987), "Self-Selection and the Earnings of Immigrants," *American Economic Review*, 77, 531–553.

Charette, M. F., and Meng, R. (1994), "Explaining Language Proficiency," *Economics Letters*, 44, 313–321.

Chernoff, H. (1954), "On the Distribution of the Likelihood Ratio," *The Annals of Mathematical Statistics*, 54, 573–578.

Chiswick, B. R. (1991), "Reading, Speaking, and Earnings Among Low-Skilled Immigrants," *Journal of Labor Economics*, 9, 149–170.

Chiswick, B. R., and Miller, P. W. (1995), "The Endogeneity Between Language and Earnings: International Analyses," *Journal of Labor Economics*, 13, 246–288.

Clark, A. E., and Oswald, A. J. (1996), "Satisfaction and Comparison Income," *Journal of Public Economics*, 61, 359–381.

Das, M. (1995), "Extensions of the Ordered Response Model Applied to Consumer Valuation of New Products," CentER DP Series No. 9515, Tilburg University.

Das, M., and van Soest, A. (1997), "Expected and Realized Income Changes: Evidence From the Dutch Socio-Economic Panel," *Journal of Economic Behavior and Organization*, 32, 137–154.

Douglas, S. M., Smith Conway, K., and Ferrier, G. D. (1995), "A Switching Frontier Model for Imperfect Sample Separation Information, With an Application to Labor Supply," *International Economic Review*, 36, 503–527.

Dustmann, C. (1994), "Speaking Fluency, Writing Fluency and Earnings of Migrants," *Journal of Population Economics*, 7, 133–156.

Dustmann, C., and van Soest, A. (2002), "Language and the Earnings of Immigrants," *Industrial and Labor Relations Review*, 55, 473–492.

Haerdle, W., and Linton, O. B. (1994), "Applied Nonparametric Methods," in *Handbook of Econometrics*, Vol. 4, eds. R. Engle and D. McFadden, Amsterdam: North-Holland, pp. 2295–2339.

Han, A. K. (1987), "Non-Parametric Analysis of a Generalized Regression Model: The Maximum Rank Correlation Estimator," *Journal of Econometrics*, 35, 303–316.

Hausman, J. A., Abrevaya, J., and Scott-Morton, F. M. (1998), "Misclassification of a Dependent Variable in a Discrete Response Setting," *Journal of Econometrics*, 87, 239–269.

Horowitz, J. L. (1993), "Semi-Parametric Estimation of a Work-Trip Mode Choice Model," *Journal of Econometrics*, 58, 49–70.

Horowitz, J. L., and Haerdle, W. (1996), "Direct Semi-Parametric Estimation of Single-Index Models With Discrete Covariates," *Journal of the American Statistical Association*, 91, 1632–1640.

Ichimura, H. (1993), "Semi-Parametric Least Squares (SLS) and Weighted SLS Estimation of Single-Index Models," *Journal of Econometrics*, 58, 71–120.

Kerkhofs, M., and Lindeboom, M. (1995), "Subjective Health Measures and State Dependent Reporting Errors," *Health Economics*, 4, 221–235.

Lazear, E. P. (1999), "Culture and Language," *Journal of Political Economy*, 107, S95–S126.

Lee, L. F., and Porter, R. H. (1984), "Switching Regression Models With Imperfect Sample Information With an Application on Cartel Stability," *Econometrica*, 52, 391–418.

Lewbel, A. (2000), "Identification of the Binary Choice Model With Misclassification," *Econometric Theory*, 16, 603–609.

Newey, W. K. (1985), "Generalized Method of Moments Specification Testing," *Journal of Econometrics*, 29, 229–256.

Powell, J. L. (1994), "Estimation of Semiparametric Models," in *Handbook of Econometrics*, Vol. 4, eds. R. Engle and D. McFadden, Amsterdam: North-Holland, pp. 2444–2523.

Powell, J. L., Stock, J. H., and Stoker, T. M. (1989), "Semi-Parametric Estimation of Index Coefficients," *Econometrica*, 57, 1403–1430.

Shapiro, A. (1985), "Asymptotic Distribution of Test Statistics in the Analysis of Moment Structures Under Inequality Constraints," *Biometrica*, 72, 133–144.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman & Hall.

Szroeter, J. (1997), "Standard Normal Tests of Multiple Inequality Constraints: First-Order Large Sample Theory," mimeo, University College London.