A NOTE ON THE LINEAR, LOGIT AND PROBIT FUNCTIONAL FORM OF THE LABOUR FORCE PARTICIPATION RATE EQUATION

J. Paul Elhorst

Department of Economics, University of Groningen,

P.O. Box 800, 9700 AV Groningen, the Netherlands.

tel. 31 50 3633893, fax. 31 50 3637337,

e-mail: J.P.Elhorst@eco.rug.nl

Abstract

The commonly used specification in regional economic research on labour force participation is the linear probability function. An important alternative recommended in the *Handbook of Regional and Urban Economics* in the contribution of Isserman et al. (1986) on 'Regional Labor Market Analysis' is the logit probability function. Their argument for the logit probability function is as follows. Given that economic theory on labour force participation does not suggest to pick one functional form over another and that the parameters of the logit probability function are estimable by OLS under the usual assumptions about the error term, the benefit of the logit probability function is that any estimated value for L lies within the logical bounds [0,1]. This feature is particularly desirable in a forecasting context when out of sample data might otherwise potentially yield absurd labour force participation rates. In this note two counter-arguments are gathered against using the logit probability function which are lacking in the *Handbook of Regional and Urban Economics.* Furthermore, it is shown that the logit probability function in this discourse can be replaced by the probit probability function equally well.

Keywords: logit, probit, labour force participation rate.

Paper prepared for the 38th Congress of the European Regional Science Association

28 August- 1 September 1998 in Vienna

1. INTRODUCTION

The commonly used specification in regional economic research on labour force participation is the linear probability function in which the participation rate L measured on the interval [0,1] is explained by a vector of unknown parameters β and a vector of independent variables X: $L = \beta'X$. An important alternative recommended by Isserman et al. (1986: 561-562) in the *Handbook of Regional and Urban Economics* is the logit probability function: $L=\exp(\beta'X)/1+\exp(\beta'X)$. Their argument for the logit probability function is as follows. Given that economic theory on labour force participation does not suggest to pick one functional form over another and that the parameters of the logit probability function are estimable by OLS under the usual assumptions about the error term - this is possible by rewriting the logit probability function and allowing for an additive error term $\ln(L/1-L)=\beta'X+\varepsilon$ -, the benefit of the logit probability function is that any estimated value for L lies within the logical bounds [0,1]. This feature is particularly desirable in a forecasting context when out of sample data might otherwise potentially yield absurd labour force participation rates. Another reason to adopt the logit probability function is that it might better approximate the empirical relationship which we want to examine, though it is to be noted that this reason has not been mentioned by Isserman et al. (1986).

In this note we argue against the notion that the logit probability function automatically produces better results than the linear one for the simple fact that the linear probability function might not be a sensible way to model probabilities. We will gather two notable counter-arguments which Isserman et al. did not take into account and so would like to complete the *Handbook of Regional and Urban Economics* on this particular point. For that purpose this note is set up as follows. First, we outline the theoretical background of the labour force participation rate equation. Except for the linear and logit probability function, we also discuss the probit probability function; any estimated value for L from this functional form also lies within the logical bounds [0,1] and therefore the probit probability function might be used equally well. Next, we discuss a test procedure which can be used to compare the fit of these non-nested probability functions. Finally, the probability functions are made the subject of an empirical analysis. Starting from regional data across different countries of the European Union over time, participation rate equations are investigated for the total working age population, for males, for females, and for males and females both in five different age categories. We conclude by expressing the contents of the two counter-arguments.

2. THEORETICAL BACKGROUND

At the micro level the decision to participate in the labour market is a dichotomous random variable L which takes the value of 1 if the market wage rate w exceeds the individual's reservation wage w^* and 0 if it does not. If we start from data observed at regional units instead of individual data, we have binary response data pooled into grouped data (Amemiya, 1981: 1493-1494; Aldrich and Nelson, 1984; Cramer, 1991: 27-28; Greene, 1993: 653-655). Within this context the observed dependent variable consists of a proportion L_j of n_j individuals belonging to the working age population in region j (j=1,...,m), who respond with $L_{ij}=1$ (i=1,...,n_j). Let $f(w_j^*)$ be the density function describing the distribution of reservation wages across individuals in region j and $F(w_j^*)$ the cumulative distribution function corresponding to the density function. Furthermore, let X_j be a vector of variables which in addition to the wage rate affect the participation rate observed in region j. Then the participation rate L_j in region j is the cumulative distribution of w_j^* evaluated at $w_j^*=w_j$

$$L(w_j,X_j,\beta) = F(w_j| X_j,\beta) + \varepsilon_j, \quad j=1,...,m, \tag{1}$$

where ε_j is an error term.

Functional forms of F most frequently used in applications are the linear, logit and probit probability function

$$F(\beta' X_j) = \beta' X_j, \tag{2}$$

$$F(\beta' X_j) = \frac{\exp(\beta' X_j)}{1+\exp(\beta' X_j)}, \tag{3}$$

$$F(\beta' X_j) = \int_{-\infty}^{\beta' X_j} \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)\ dt, \qquad (4)$$

where it has been assumed that X_j includes w_j. The linear probability function has an obvious defect in that $\beta' X_j$ is not constrained to lie in the interval [0,1] as a probability should. Although this defect can be corrected by defining $L_j=1$ if $F(\beta' X_j)>1$ and $L=0$ if $F(\beta' X_j)<0$, this procedure produces unrealistic kinks at the truncation points. For this reason the logit and probit probability function might be superior to the linear probability function, especially if a large number of the observations are close to the bounds.

Simple least squares regression based on equation (1) would be unbiased but inefficient since it ignores the properties of the error structure. Under the assumption of independent samples from a binomial population, the observed proportion L_j is asymptotically normally distributed with mean $\pi_j=F(\beta' X_j)$ and variance $\pi_j(1-\pi_j)/n_j$ - based on the De Moivre-Laplace limit theorem (see Mood et al., 1974: 120; Dobson, 1990: 116-118)

$$L_j = F(\beta' X_j)+\varepsilon_j = \pi_j+\varepsilon_j, \quad E(\varepsilon_j)=0, \quad Var(\varepsilon_j)=\pi_j(1-\pi_j)/n_j. \qquad (5)$$

In case of the linear probability function one can use a linear weighted least squares method to estimate β using as weights $w_j=\sqrt{n_j/[\pi_j(1-\pi_j)]}$. This method is known as the minimum chi-squared method. It has been found that this estimator has the same asymptotic properties as the maximum likelihood estimator (Amemiya, 1980, 1985: 275-280). Berkson (1980) has argued that the performance of the minimum chi-squared method is even better. Since the weights are functions of the unknown parameters β, $\pi_j=\pi_j(\beta)$, an iterative two-step procedure is called for. A direct estimate of π_j can also be obtained by replacing π_j with the observed proportion L_j, but the former method is preferable.

In case of the logit and probit probability function one can use a nonlinear weighted least squares method to estimate β. But there is a simpler way to proceed. Since the function $F(\beta' X_j)$ is strictly monotonic, it has an inverse. Expanding $F^{-1}(L_j)=F^{-1}(\beta' X_j+\varepsilon_j)$ around $\varepsilon_j=0$ in a Taylor series and omitting higher-order terms, we get

$$F^{-1}(L_j)=F^{-1}(\beta' X_j+\varepsilon_j)\approx F^{-1}(\beta' X_j)+\frac{dF^{-1}(\beta' X_j)}{d\beta' X_j}\ \varepsilon_j=\beta' X_j+\frac{\varepsilon_j}{f(\beta' X_j)}. \qquad (6)$$

This again produces a heteroscedastic regression

$$F^{-1}(L_j)=\beta' X_j+u_j, \quad E(u_j)=0, \quad Var(u_j)=\frac{F(\beta' X_j)(1-F(\beta' X_j))}{n_j\ f(\beta' X_j)^2} = \frac{\pi_j(1-\pi_j)}{n_j\ f(\beta' X_j)^2}. \tag{7}$$

In case of the logit probability function the inverse $F^{-1}(L_j)$ is easy to obtain $F^{-1}(L_j)= \ln(L_j/1-L_j)$, while $var(u_j)$ reduces to $var(u_j)=1/[n_j\pi_j(1-\pi_j)]$. For the probit probability function the inverse function $F^{-1}(L_j)$ must be approximated by a ratio of polynomials (see Abramowitz and Stegun, 1965: 931-932). Again one can use a weighted least squares method to estimate β using as weights $w_j=\sqrt{[n_j\pi_j(1-\pi_j)]}$ for the logit probability function and $w_j=\sqrt{[n_j f(\beta' X_j)^2/\pi_j(1-\pi_j)]}$ for the probit probability function. These methods are known as the minimum logit and the minimum probit chi-squared method.

Generally, there are more applications of the linear probability function than of the logit probability function in empirical research. Applications of the former can be found in Bowen and Finegan (1969), Fleisher and Rhodes (1976), Van der Veen and Evers (1983), Molho and Elias (1984), Lillydahl and Singell (1985), Baumann et al. (1988), Nord (1989), Clark and Anker (1990), and Gallaway et al. (1991), and applications of the latter in Schubert (1982), Siegers (1983), and Ward and Dale (1992).[1] Only one of these studies (Schubert, 1982: 1242) has really adjusted the estimation model for the above type of heteroscedasticity, while in two studies (Siegers, 1983: 403; Baumann et al., 1988: 1090) the authors admit that the application of simple least squares leads to inefficient parameter estimates. So regional applications generally ignore the heteroscedastic nature of the error terms. On the other hand, the question arises whether the unobserved individual responses which underlie a proportion should really be considered as single random drawings from a binomial distribution. This issue would be unambiguous in experimental data sets but it is less clear with regional labour force participation data, especially since in this case the n_j individuals represent all of the potential respondents in region j rather than a random sample of respondents. Within this latter context it seems appropriate to use as weights just the square root of the size of the working age population on which each regional observation is based, albeit this type of heteroscedasticity is not very popular either. Most authors have treated all regions equally irrespective of their size or do not mention whether they have weighted the regional observations. The only exception is Siegers (1983: 403). Bowen and Finegan (1969: 776-778) admit that a good case can be made for using weighted regression but they did not apply it after they found out that the general contour of relations in the weighted regressions was quite similar to the pattern of coefficients in the unweighted runs.

3. CHOOSING BETWEEN PROBABILITY FUNCTIONS

The linear, logit and probit probability function can be considered as three competing models we want to choose from. The difference between the logit and the probit probability function is negligible. Although the logit probability function does have some slightly heavier tails than the probit probability function, both functions are quite similar in shape, especially in the mid-range (Amemiya, 1985: 1487; Cramer, 1991: 14-17). So in this section we restrict our attention to whether the linear or the logit/probit probability function is more appropriate for a given analysis. Consider the hypotheses

$$H_0\colon L_j = \beta_0' X_j + \varepsilon_{0j}, \quad E(\varepsilon_{0j})=0, \quad E(\varepsilon_{0j}^2)=\sigma_0^2\tau_{0j}, \quad j=1,...,m, \tag{8a}$$

$$H_1\colon L_j = F(\beta_1' X_j) + \varepsilon_{1j}, \quad E(\varepsilon_{1j})=0, \quad E(\varepsilon_{1j}^2)=\sigma_1^2\tau_{1j}, \quad j=1,...,m, \tag{8b}$$

where $F(\beta_1' X_j)$ denotes either the logit or probit probability function and τ is a skedastic function. The econometric literature has produced two widespread methods to test for non-nested regression models[2]: tests based on artificial nesting of regression equations (Davidson and MacKinnon, 1993: 381-388) and tests which elaborate on the classic work of Cox (Pesaran, 1974; Pesaran and Deaton, 1978; among others). Generally, these tests are only spelled out for choosing between two possible sets of regressors and between a linear or log-linear model. We explicate the first type of test for choosing between a linear or logit/probit probability functional form. Furthermore, we allow for heteroscedasticity.

Following Fisher and McAleer (1981), we first consider an artifical nesting of the two hypotheses into the combined regression model

$$L_j = \frac{(1-\alpha)\tau_{1j}\sigma_1^2}{(1-\alpha)\tau_{1j}\sigma_1^2+\alpha\tau_{0j}\sigma_0^2}\beta_0' X_j + \frac{\alpha\tau_{0j}\sigma_0^2}{(1-\alpha)\tau_{1j}\sigma_1^2+\alpha\tau_{0j}\sigma_0^2}F(\beta_1' X_j) + \epsilon_j, \quad j=1,...,m, \tag{9a}$$

where $\alpha \in [0,1]$ and the variance of ε is

$$\tau_j\sigma^2 = \frac{\tau_{0j}\sigma_0^2\tau_{1j}\sigma_1^2}{(1-\alpha)\tau_{0j}\sigma_0^2+\alpha\tau_{1j}\sigma_1^2}, \quad j=1,...,m. \tag{9b}$$

If we take the view that the hypotheses under test relate only to the expected value of L_j, we may quite reasonably impose the restriction $\tau_{0j}\sigma_0^2=\tau_{1j}\sigma_1^2=\tau_j\sigma^2$ (cf. Fisher and McAleer, 1981: 106). The combined regression model then simplifies to

$$L_j = (1\text{-}\alpha)\beta_0'X_j + \alpha F(\beta_1'X_j) + \varepsilon_j, \quad E(\varepsilon_j)=0, \quad E(\varepsilon_j^2)=\tau_j\sigma^2, \quad j=1,...,m, \tag{10}$$

which is the starting point for the analysis of Davidson and MacKinnon (1993). As written, a test of $\alpha=0$ would be a test against H_1, but the problem is that α cannot be estimated in this model. One solution to this problem, originally suggested in Davidson and MacKinnon (1981), is to replace equation (10) by a model in which the unknown parameters of the model that is not being tested are replaced by estimates of those parameters that would be consistent with H_1. In other words, replacing β_1 with its consistent estimate under H_1 is a practical way of obtaining a t-ratio for the least squares estimate of α. The parameters of this new compound model (dropping the j-index for ease of notation)

$$L = (1-\alpha)\beta_0'X+\alpha F(\hat{\beta}_1'X) = (1-\alpha)\beta_0'X+\alpha\hat{F}, \tag{11}$$

can be obtained by nonlinear estimation. To avoid the computational problems involved with estimating this nonlinear regression, this model can further be linearized in a Taylor series around the point ($\beta_0=\hat{\beta}_0,\alpha=0$) to yield

$$L-\hat{L}_{\beta_0=\hat{\beta}_0} = \beta_{00}'X+\alpha(\hat{F}-\hat{L}_{\beta_0=\hat{\beta}_0})+\epsilon_{00}. \tag{12}$$

This artificial regression is called the Gauss-Newton regression (see Davidson and MacKinnon, 1993: 381-388). The logic of the test is thus first to estimate the linear and the logit/probit probability function and to compute their predictions

$$1.\ L=\beta_0'X, \quad \hat{L}_0=\hat{\beta}_0'X, \quad \hat{L}_{0\to1}=F^{-1}(\hat{\beta}_0'X) \tag{13a}$$

$$2.\ F^{-1}(L)=\beta_1'X, \quad \hat{L}_1=\hat{\beta}_1'X, \quad \hat{L}_{1\to0}=F(\hat{\beta}_1'X), \tag{13b}$$

and next to estimate the linear probability function including the computed predictions according to the Gauss-Newton regression

$$3.\ L-\hat{L}_0=\beta_{00}'X+\alpha[\hat{L}_{1\to 0}-\hat{L}_0]+\varepsilon_{00}. \tag{13c}$$

A conventional t-test on the parameter α can now be used to test the H_0-hypothesis, since the t-ratio of $\hat{\alpha}$, called the J-statistic[3], is asymptotically distributed N(0,1) under H_0 (see Davidson and MacKinnon, 1981, 1993). So if $\hat{\alpha}$ is statistically different from zero, we may conclude that the logit/probit probability function adds significant fit to the linear one, thus arguing against the linear probability function. Rejection of H_0, however, does not automatically mean that H_1 is true, because the t-statistic on the parameter $\hat{\alpha}$ is conditional on the truth of H_0, and not on the truth of H_1. If one wants to test H_1, the simplest procedure is to reverse the roles of H_0 and H_1 and to carry out the test again. This can be done by a fourth regression

$$4.\ F^{-1}(L)-\hat{L}_1=\beta_{11}'X+\alpha[\hat{L}_{0\to 1}-\hat{L}_1]+\varepsilon_{11}. \tag{13d}$$

If $\hat{\alpha}$ is again statistically different from zero we may conclude that the linear probability function also adds significant fit to the logit/probit one, thus arguing against the logit/probit probability function. The possibility of simultaneous rejection of both hypotheses is one of the problems of the J-test, since it may lead to conflicting conclusions. On the other hand, Pesaran and Deaton (1978: 678) and McAleer et al. (1982) argue that tests between non-nested hypotheses or models should encompass the possibility of rejecting both. Two competing and observationally non-equivalent models cannot both be "true" when applied to the same data, although it is certainly the case that both can be "false". For those cases in which both hypotheses are rejected, we also adopt the Cox-test as a secondary test. As the Cox-test for nonlineair regression models is very well documented and its application to the linear and logit/probit probability function is comparable to that of the J-test, it suffices to refer to Pesaran and Deaton (1978).

Finally, it should be stressed that the J-test has also been criticized. According to Godfrey and Pesaran (1983: 144), the J-test is less powerful than other testing procedures; its small and finite sample significance levels are often too high, especially when one or more of the following features is present: (i) a poor fit of the true model, (ii) low or moderate correlations between the regressors of the two models, and (iii) the false model includes more regressors than the correct specification. However, none of these features seem to bother our empirical analysis in the next section, while the number of observations in the empirical analysis is reasonable.

4. EMPIRICAL ANALYSIS

In this section the linear, logit and probit probability function are made the subject of an empirical analysis. We experimented with the regional participation rate of different population groups: the total working age population, males, females, as well as males and females both in five different age categories (15-24, 25-34, 35-44, 45-54, 55-64). Each regression equation was estimated with the help of an unbalanced panel of 146 regions across the twelve member states of the European Union during the period 1983-1989. These observations were obtained from the Eurostat file called "Regions". Below we first explain the background of the selected regressors and then we discuss the estimation results.

In an overview paper in which 17 empirical studies on the regional participation rate have been surveyed (Elhorst, 1996a), the following general model of the regional participation rate has been inferred. The participation rate reflects the proportion of people who want to work at the current real wage controlled for the frequency with which socio-economic characteristics can be observed among the population and the probability of finding a job. The most widely used indicators of the socio-economic variables are the sex and age structure and the educational attainment of the population, and of the probability of finding a job the unemployment rate and the sectoral composition of employment. In view of this general model, each regional participation rate has been taken to depend on the following set of variables

WAGE	the wage rate measured as average hourly earnings of manual and non-manual workers in manufacturing after tax and social security allowances and converted to 1985 ECUs with the help of Purchasing Power Parities developed by Eurostat,
BIRTH	the birth rate (number per 1000 population),
EDUCATION	the educational attainment of the working age population (%),
UNEMPLOYMENT	the unemployment rate (%),

SERVICES	the share of employment in services (%),
DUMMIES	country and annual time dummies.

The variables BIRTH, EDUCATION and the DUMMIES require some further explanation. As the regional data file of Eurostat does not offer data on the family structure at regional level, the birth rate has been submitted as a proxy for having young children. It is well-known from previous research that the presence of young children restrains the female participation rate and sustains the male participation rate (Pott-Buter, 1993: 287). In addition, the birth rate has been submitted as a proxy for the ageing of the population. Since the birth rate changes over time only gradually, the difference between the birth rate and the ageing of the population is not very great. Generally, if the birth rate is high (low), the population tends to be young (ageing), appearing from the fact that under this circumstance the share of the population under 25 years of age is rather large (small) and the shares of the population aged between 25-55 and over 55 years of age are rather small (large). From previous research (Elhorst, 1996b) it appeared that the less a region suffers from an ageing population, the higher the male and the lower the female participation rate will be, as is the case in Spain, Greece and Ireland; conversely, the more a region suffers from an ageing population, the lower the male and the higher the female participation rate will be, as is the case in Germany, Italy, France and the Netherlands. In sum, the birth rate is expected to have a positive sign in the equations for males and a negative sign in the equations for females.

Following the OECD (1989, 1992), the educational attainment of the population of working age has been determined by distinguishing four levels of education. Each level has been completed by a certain percentage of the adult population. These percentages have then been summed according to the following equation

$$\begin{aligned} \text{EDUCATION} = & - 0{,}017 * \text{\%primary education} \\ & + 0{,}031 * \text{\%lower secondary education} \\ & + 0{,}020 * \text{\%upper secondary or post-secondary education} \\ & + 0{,}137 * \text{\%university degree.} \end{aligned} \tag{14}$$

The figures in this equation have been determined by applying principal component analysis. As the four indicators of the educational attainment of the population are strongly correlated with each other, it has been decided to reduce them to one single component in order to avoid multicollinearity (see Greene, 1993: 271-273). This single principal component accounts for 63.7 per cent of the variation in its four indicators.

Finally, country dummies have been added to give way to country-specific circumstances which affect the level around which regional participation rates within one country vary. Not accounting for this country heterogeneity runs the risk of obtaining biased results - see the theory on panel data models (Hsaio, 1986; Baltagi, 1995). Annual time dummies have been added in order to prevent that trends along the observations over time, either linear or cyclical, might bias the actual cross-sectional relation which we want to examine.

Since a static regression equation would suffer from high serial autocorrelation, and the precise dynamic structure of the regression equation is not known, we have adopted a first-order autoregressive lag model (see for a recent explanation, Hendry, 1995: 231-308; Mizon, 1995)

$$L_{jt} = \tau L_{jt-1} + \beta_0' D_{jt} + \beta_1' X_{jt} + \beta_2' X_{jt-1} + \varepsilon_{jt} \quad (L_{jt} = \pi_{jt} + \varepsilon_{jt}), \quad \varepsilon_{jt} \sim N(0, \pi_{jt}(1-\pi_{jt})/n_{jt}), \tag{15}$$

where L_{jt} is the participation rate for the j^{th} region in the t^{th} time period; D_{jt} is a vector containing the country and annual time dummies; X_{jt} is a vector containing the explanatory variables for the j^{th} region in the t^{th} time period; τ, β_0, β_1 and β_2 are vectors of parameters to be estimated. Equation (15) is the regression equation that has been estimated. When, after its estimation, the non-stochastic part of this equation is reformulated as

$$L_{jt} = -\frac{\tau}{1-\tau}\Delta L_{jt} + \frac{\beta_0'}{1-\tau} D_{jt} + \frac{\beta_1' + \beta_2'}{1-\tau} X_{jt} - \frac{\beta_2'}{1-\tau}\Delta X_{jt} = \tau^* \Delta L_{jt} + \beta_0^{*'} D_{jt} + \beta_1^{*'} X_{jt} + \beta_2^{*'} \Delta X_{jt}. \tag{16}$$

it can be seen that so-called short-run dynamics have been added to the static equation. There still exists a static long-run equilibrium relationship between L and X, but short-run dynamics of how an equilibrium is approached are explicitly taken into account. β_1^* reflects the long-run effect of X on L, while β_2^* reflects the short-run or immediate response of L to a change in X.

Against this background we discuss the main results of our analysis. First, the long-run coefficients[4] have generated a plausible model structure. The participation rates appear to be positively related to the wage rate and the educational attainment of the population, and negatively related to the regional unemployment rate. The latter effect is remarkably stable; it is statistically different from zero for all population groups even at the 1% significance level. Therefore, it might be considered as further evidence that the discouragement effect dominates the labour market in the European Union (see also Elhorst, 1996b). The share of employment in services has a positive effect on the participation rate of females aged between 25-34, 35-44 and 45-54, indicating that especially prime-aged women have benefited from a growing services sector. The effect on the participation rate of all other population groups is negative, though insignificant. As has been expected, the birth rate has a positive effect on the participation rate of male population groups, and a negative effect on the participation rate of female population groups.

Second, the linear probability model did produce not one single prediction outside the interval [0,1], even not for males aged between 25-34, 35-44 and 45-54. So from a forecasting point of view there is no reason to reject the linear probability function in favour of the logit or probit probability function. It is clear that this result hinges strongly on the completeness of the model, which in this study is acceptable. The R-squared ranges from 0.53 and 0.54 for males aged between 35-44 and 45-54, to over 0.65 for males aged between 25-34 and females aged between 55-64, and to over 0.75 for all other population groups.

Third, the application of the iterative two-step minimum chi-squared method produced almost the same parameter estimates as the one-step least squares method using as weights the square root of the size of the working age population on which each regional observation is based. The null hypothesis that the two estimates have the same expected value could be rejected not once.[5] Consequently, there is no need to use this iterative estimation method and it suffices to compute the J and Cox test for the one-step WLS regressions only.

Fourth, the logit as well as the probit probability function appears to be superior to the linear one for males aged between 25-34, 35-44 and 45-54, on the basis of the J test as well as on the basis of the Cox test. For all other population groups the logit or probit probability function cannot said to be superior. For males aged between 15-24 and 55-64 and females aged between 25-34, 35-44 and 45-54, the linear probability function could not be rejected, neither on the basis of the J-test nor on the basis of the Cox test. For females as a whole, the linear probability function could only be rejected on the basis of the J test (double rejection on the basis of the Cox test). For the remaining population groups - the total working age population, males aged between 15-64 and females aged between 15-24 and 55-64 - the linear probability function has been rejected on the basis of the J test, but the logit/probit probability function as well. By contrast, for all these population groups the linear probability function could not be rejected on the basis of the Cox test.

Fourth, although the differences between the long-run coefficients of the linear probability function and the comparable marginal effects[6] of the logit/probit probability function were perceptible, they were not impressive. To illustrate this we report the estimation results obtained for males aged between 35-44 in table 2, one of the three population groups for which the logit/probit probability function is superior and so different estimates would be most natural.[7]

5. CONCLUSIONS

From our analysis it is possible to deduce two arguments against using the logit or probit probability function for estimating the labour force participation rate equation. The first argument challenges the proposition that the parameters of the logit or probit probability model are estimable by OLS under the usual assumptions about the error term. The second argument challenges the benefit of the logit or probit probability function that any estimated value for L lies within the logical bounds [0,1] and that the logit or probit probability function may better approximate the empirical relationship that determines aggregate labour force behaviour.

1. Estimating the logit or probit probability function is not simply a matter of rewriting the regression equation and allowing for an additive error term. The thing is that the estimation model should be adjusted for heteroscedasticity, although it is to be noted that this also holds for the linear probability function. Regional applications generally ignore the heteroscedastic nature of the error terms. From a theoretical viewpoint, one

may give preference to the minimum chi-squared method which calls for a complicated iterative two-step procedure using as weights $w_j=\sqrt{[n_j f(\beta' X_j)^2/\pi_j(1-\pi_j)]}$ or to a simple one-step least squares procedure using as weights the square root of the size of the working age population on which each regional observation is based. This choice depends on the view whether or not the unobserved individual responses which underlie a regional participation rate should be considered as random drawings from a binomial distribution. From an empirical viewpoint, this choice does not really matter, since the difference between the parameter estimates of both procedures appeared to be statistically insignificant.

2. From a forecasting viewpoint, the necessity to use the logit or probit probability function is questionable. Although the linear probability function has an obvious defect in that it is not constrained to the interval [0,1] as a probability should, it does not give any problems in practice. First, because the participation rate of most population groups is far from the bounds. Table 1 showed that only the regional participation rate of males aged between 25-34, 35-44 and 45-54 might give problems, because their mean value is greater than 0.90. But in our empirical analysis the linear probability function did produce not one single prediction outside the interval [0,1], even not for males aged between 25-34, 35-44 and 45-54. Second, because population groups become interesting not until their participation rate is much lower than the upper bound. In this respect it should be stressed that most authors only analyse the regional participation rate of females (Van der Veen and Evers, 1983; Molho and Elias, 1984; Ward and Dale, 1992) or of older males (Clark and Anker, 1990), for this has really important implications for both the size and the composition of the labour force. Others analyse the participation rate of broad population groups not broken down by age, that is the participation rate of males and females both as one group (Fleisher and Rhodes, 1976; Siegers, 1983; Lillydahl and Singell, 1985) or of the total working age population (Baumann et al., 1988; Nord, 1989; Gallaway et al., 1991). So in practice the participation rate of prime-aged males is hardly analysed. This is probably not only because the participation rate of this population group hardly changes over time and between regions, but also because it is much harder to find a reasonable economic explanation for the rather small differences within this group. In this respect it should be stressed that the R-squared of

males aged between 25-34, 35-44 and 45-54 indeed appeared to be relatively low. From the evident practice to analyse only those population groups whose participation rate is much lower than the upper bound two important implications follow. First, the potential problem that the tails of probability function might be S-shaped automatically dissolves, since there are no observations at the tails. Second, it is no longer necessary to apply the logit or probit probability function because under these circumstances, and starting from the results produced by the test statistics for choosing between non-nested models, it does not give any better fit than the linear one.

NOTES

1. Molho and Elias (1984: 167) and Baumann et al. (1988: 1090) did consider the logit probability function as well, but they did not apply it.
2. Two regressions are said to be non-nested when one is not a restricted form of the other, or may not be obtained as a limiting form of a suitable approximation of the other.
3. The term ‘J-statistic’ actually refers to the compound model in (11), as α and β are estimated *jointly*, while the statistic referring to the Gauss-Newton regression is originally called the P-statistic due to its *projection* (Davidson and MacKinnon, 1993: 383).
4. The long-run coefficients have been obtained from the estimates of the original coefficients in equation (15). Similarly, the variance of the long-run coefficients have been obtained from the covariance matrix of equation (15) (cf. Mood et al., 1974: 179-191).
5. Let V_0 and V_1 be the covariance matrices of the estimates of β_0 and β_1. A test that the difference between the parameters is zero can then be based on the Wald statistic

   $$W = (\hat{\beta}_0 - \hat{\beta}_1)'(V_0 + V_1)^{-1}(\hat{\beta}_0 - \hat{\beta}_1),$$

   which has a chi-squared distribution with k degrees of freedom.
6. The interpretation of the parameters of the linear probability function is relatively straightforward, as they express the resulting change in the measurement scale of the participation rate for a unit change in the independent variables. By contrast, the

interpretation of the parameters of the logit and probit probability function is more difficult, as these parameters express the resulting change in the logit or the normit of the participation rate for a unit change in the independent variables. To make the parameters comparable to those of the linear probability functio, we have computed marginal effects. The formulas for the marginal effects and the corresponding asymptotic covariance matrx of the logit and probit probability function can be found in Greene (1993: 636-648).

7. The estimation results obtained for the other population groups show similar patterns and therefore are not reported for reasons of parsimony and efficiency.

REFERENCES

Abramowitz M. and Stegun I.A. (1965) *Handbook of mathematical functions*. New York, Dover Publications.

Aldrich J.H. and Nelson F.D. (1984) *Linear probability, logit, and probit models*. Beverly Hills, California, Sage Publications.

Amemiya T. (1980) The n^{-2}-order mean squared errors of the maximum likelihood and the minimum logit chi-square estimator. *Annals of statistics* 8: 488-505.

Amemiya T. (1981) Qualitative response models: A survey. *Journal of Economic Literature* 19: 1483-1536.

Amemiya T. (1985) *Advanced econometrics*. Basil Blackwell, Oxford.

Baltagi B.H. (1995) *Econometric analysis of panel data*. Wiley, Chichester, England.

Baumann J., Fischer M.M. and Schubert U. (1988) A choice-theoretical labour-market model: empirical tests at the meso-level. *Environment and Planning A* 20: 1085-1102.

Berkson J. (1980) Minimum chi-square, not maximum likelihood. *Annals of statistics* 8: 457-487 (including discussion).

Bowen W.G. and Finegan T.A. (1969) *The economics of labor force participation*. University Press, Princeton.

Clark R.L and Anker R. (1990) Labour force participation rates of older persons: an international comparison. *International Labour Review* 129: 255-271.

Cramer J.S. (1991) *The logit model*. Edward Arnold, London.

Davidson R. and MacKinnon J.G. (1981) Several tests for model specification in the presence of alternative hypotheses. *Econometrica* 49: 781-793.

Davidson R., MacKinnon J.G. (1993) *Estimation and inference in econometrics*. University Press, Oxford.

Dobson A.J. (1990) *An introduction to generalized linear models*. Chapman and Hall, London.

Elhorst J.P. (1996a) Regional labour market research on participation rates. Tijdschrift voor Economische en Sociale Geografie 87: 209-221.

Elhorst J.P. (1996b) A regional analysis of labour force participation rates across the member states of the European Union. *Regional Studies* 30.5: 455-465.

Fisher G.R., McAleer M. (1981) Alternative procedures and associated tests of significance for non-nested hypotheses. *Journal of Econometrics* 16: 103-119.

Fleisher B.M. and Rhodes G. (1976) Unemployment and the labor force participation of married men and women: a simultaneous model. *The Review of Economics and Statistics* 58: 398-406.

Gallaway L., Vedder R. and Lawson R. (1991) Why people work: an examination of interstate variations in labor force participation. *Journal of Labor Research* 12: 47-59.

Godfrey L.G., Pesaran M.H. (1983) Tests of non-nested regression models. *Journal of Economietrics* 21: 133-154.

Greene W.H. (1993) *Econometric analysis*, 2nd edition. MacMillan, New York.

Hendry D.F. (1995) Dynamic econometrics. University Press, Oxford.

Hsiao C. (1986) *Analysis of panel data*. Cambridge, University Press.

Isserman A., Taylor C., Gerking S. and Schubert U. (1986) Regional labor market analysis. In: Nijkamp P. (ed.) *Handbook of regional and urban economics*, volume 1: 543-580. Elsevier, Amsterdam.

Lillydahl J.H. and Singell L.D. (1985) The spatial variation in unemployment and labour force participation rates of male and female workers. *Regional Studies* 19: 459-469.

McAleer M., Fisher G. and Volker P. (1982) Separate misspecified regressions and the U.S. long-run demand for money function. *The Review of Economcs and Statistics* 64: 572-583.

Mizon G.E. (1995) A simple message for autocorrelation correctors: Don't. *Journal of Econometrics* 69: 267-288.

Molho I. and Elias P. (1984) A study of regional trends in the labour force participation of married women in the UK, 1968-1977. *Applied Economics* 16: 163-173.

Mood A.M., Graybill F.A. and Boes D.C. (1974) *Introduction to the theory of statistics*, 3rd edition. Tokyo, McGraw-Hill.

Nord S. (1989) The relationships among labor-force participation, service-sector employment, and underemployment. *Journal of Regional Science* 29, 407-421.

Pesaran M. (1974) On the general problem of model selection. *Review of Economic Studies* 41: 153-171.

Pesaran M., Deaton A. (1978) Testing non-nested nonlinear regression models. *Econometrica* 46: 677-694.

Schubert U. (1982) REMO - an interregional labor market model of Austria. *Environment and Planning A* 14: 1233-1249.

Siegers J.J. (1983) An economic-demographic ten-equation model. *De Economist* 131: 400-443.

Veen A. van der and Evers G.H.M. (1983) A simultaneous model for regional labor supply, incorporating labor force participation, commuting and migration. *Socio Economic Planning Sciences* 17: 239-250.

Ward C. and Dale A. (1992) Geographical variation in female labour force participation: An application of multilevel modelling. *Regional Studies* 26: 243-255.

Table 1 Mean, standard deviation and range of regional labour force participation rates by sex and age in 146 regions across the twelve member states of the European Union over the period 1983-1989

males 15-24	0.529	0.093	0.286 - 0.774
males 25-34	0.945	0.031	0.802 - 0.985
males 35-44	0.967	0.020	0.919 - 0.994
males 45-54	0.917	0.039	0.823 - 0.984
males 55-64	0.549	0.112	0.236 - 0.794
males 15-64	0.788	0.038	0.659 - 0.877
females 15-24	0.475	0.099	0.176 - 0.704
females 25-34	0.637	0.100	0.322 - 0.888
females 35-44	0.587	0.130	0.194 - 0.898
females 45-54	0.488	0.150	0.123 - 0.812
females 55-64	0.231	0.093	0.048 - 0.574
females 15-64	0.486	0.092	0.231 - 0.754
total population of working age (15-64)	0.631	0.059	0.493 - 0.808

Table 2 Estimation* and test** results obtained for males aged between 35-44

Explanatory variables	Linear probability function				Logit probability function				Probit probability function			
	one-step WLS		min. chi-squared		one-step WLS		min. chi-squared		one-step WLS		min. chi-squared	
	par.	T-value	par.	T-value	par.	T-value	par.	T-value	par.	T-value	par.	T-value
UNEMPLOYMENT	-0.195	-7.24	-0.188	-7.46	-0.163	-3.11	-0.163	-3.14	-0.168	-3.51	-0.168	-3.54
WAGE	0.026	3.22	0.029	3.68	0.017	2.70	0.022	3.20	0.019	3.00	0.022	3.41
EDUCATION	0.531	6.34	0.509	6.27	0.488	2.70	0.465	2.80	0.490	3.09	0.490	3.09
BIRTH	0.048	0.90	0.044	0.92	0.029	0.72	0.030	0.77	0.031	0.75	0.031	0.75
SERVICES	-0.013	-0.85	-0.013	-0.92	-0.009	-0.58	-0.009	-0.63	-0.009	-0.61	-0.009	-0.66
R^2	0.530				0.539				0.534			
out of range***	0											
α (J-test)****												
H_1 linear					-0.317	-0.84			-0.282	-0.64		
H_1 logit	1.064	2.26										
H_1 probit	1.664	2.95										
T_0 (Cox test)												
H_1 linear					-0.442	-1.09			-0.414	-1.27		
H_1 logit	-7.192	-3.67										
H_1 probit	-7.129	-4.18										

* Actual parameters of the linear probability function, marginal effects of the logit and probit probability function
** Test results based on one-step WLS regressions
*** Number of predictions outside the interval [0,1]
**** Negative or large positive values for α can come up since α is only asymptotically distributed N(0,1) and only if H_0 is in fact true