

Low-wage employment in Portugal: a mixed logit approach

Carlos Pestana Barros
Instituto Superior de Economia e Gestão
Technical University of Lisbon
Rua Miguel Lupi, 20
1249-078 Lisbon, Portugal
cbarros@iseg.utl.pt

Isabel Proença
Instituto Superior de Economia e Gestão
Technical University of Lisbon
Rua do Quelhas, 2
1200-781 Lisbon, Portugal
isabelp@iseg.utl.pt

José Cabral Vieira
Department of Economics and Management
University of Azores
Rua Mãe de Deus
9500 Ponta Delgada
josevieira@notes.uac.pt

Abstract: In this paper, we examine the determinants of low-wage employment in Portugal. For this purpose, we use a data file of the European Community Household Panel (ECHP) for the years 1998 and 1999. In order to take into account unobserved heterogeneity in the data, a random-parameter logit model is used to analyse the probability of a worker receiving a low wage. The results indicate that the consideration that the effects of the explanatory variables are the same across all individuals, such as is assumed in most of the literature may be misleading. From the policy perspective, this implies that the use of a single instrument in order to combat low-wage employment is inappropriate to satisfy the whole population. In view of this, policies tailored by clusters would be more appropriate.

Key words: low-wage employment, random-parameter logit model, public policy.

JEL Code: C25, J31, J38

Financial support from the Fundação para a Ciência e Tecnologia/MCT under FCT/POCTI, partially funded by FEDER (EU Regional Development Fund), is gratefully appreciated. We are also grateful to the Portuguese National Institute of Statistics (INE) for providing access to the European Community Household Panel (ECHP) for Portugal. Finally, we thank João Santos Silva for his helpful comments. The usual disclaimer applies.

1. Introduction

This paper examines the profile of low-wage receivers in the Portuguese labour market. For this purpose, we use a data file from the European Community Household Panel (ECHP) for the years 1998 and 1999. The motivation for this research stems mainly from three critical issues associated with low-wage receivers. First, reducing the incidence of low-wage employment should contribute significantly to the alleviation of poverty, since wages are the main income source of many families. Second, low earnings prospects may discourage individuals from entering formally into the labour market, owing to the existence of alternatives when they are unemployed, such as income support programs. Finally, low wage prospects may create incentives to engage in illegal activities or work in the informal sector (see OECD, 1999).

Indeed, the incidence and the persistence of low-paid work has become a matter of great concern in many developed economies as a result of increasing wage inequality (see OECD, 1996 and 1997a, Asplund et al., 1998, Stewart and Swaffield, 1999, Cappellari, 2000, Lucifora and Salverda, 1998, and Salverda *et al.* 2001). This paper contributes to the literature in two ways. Firstly, the panel nature of the data is associated with the existence of unobserved individual heterogeneity constant in time that has to be carefully assessed in order to guarantee consistency of the parameter estimators under flexible and realistic assumptions. To do that we consider the random-parameters logit model (RPL), used by Train (1998) and Revelt and Train (1998). Previous studies on the determinants of low-wage employment have relied mainly on pooled standard Logit or Probit models (see, for instance, Cardoso *et al.*, 2000). However, this procedure may lead to misleading conclusions in the event of unobserved heterogeneity. Such a shortcoming is overcome with the RPL model by allowing that the latter depends on individual characteristics and simultaneously to estimate coefficients of variables that are constant in time. Secondly, the paper seeks to contribute to a

better understanding of the low-wage determinants. In particular, we intend to determine which type of individual and job-related attributes increase the risk of low-wage employment. The value of such research is that it can contribute to policy formulation with regard to the labour market.

The paper is organised as follows. The next section briefly describes the contextual setting. Section 3 comprises a literature review. Section 4 presents the mixed logit model. Section 5 includes the theoretical framework. Section 6 describes the data and presents the estimation results. Finally, section 7 summarises and concludes.

2. Contextual Setting

Interest in the Portuguese labour market has broadened over the last decade. This interest was boosted by a good labour market performance after the mid-1980s compared with other western economies. High wage flexibility has been pointed as a particular feature of this market (OECD, 1994). In addition, it is well known that wages in Portugal are low when compared with other western economies. It has been argued that these lower wages reflect lower labour productivity, which itself may indicate reduced levels of physical and human capital (see Branco and Mello, 1992).

Moreover, the economy has been under a process of modernisation and restructuring, particularly since joining the European Union in 1986. As a consequence, demand for skilled workers has increased and overall wage inequality has expanded, mainly because wages at the top of the wage distribution grew at a faster rate than those at the lower end (Cardoso, 1997). This has been associated with the evolution of the rates of return on education and job requirements, and is consistent with the skill-biased technological change hypothesis (see Hartog *et al.*, 2001). Wage inequality in Portugal is high and reaches a level similar to that of

the United Kingdom, that is, slightly lower than the United States, which is usually taken as the paradigm of an unequal labour market, but higher than Canada, and much higher than the former West Germany or Sweden (see Cardoso, 1998).

Despite the increasing international interest in low wages, very little is known about the phenomenon in Portugal, where research into the issue has only recently begun. Cardoso *et al.* (2000) analysed the profile of the low-wage worker and concludes that low-pay is more likely among females, youngsters and less-educated workers. They also find evidence that industry and firm size have an impact on the probability of low pay. In particular, those working in small firms, retail, food and beverages, hotels and restaurants as well as in exported-oriented industries such as clothing and textiles, footwear and timber and cork are more likely to be found in the low wage segment. The incidence of low wages amounted to 13% of the workers in 1986, 18% in 1994, and 15% in 1997. These figures align at an intermediate level between those which have been reported for Northern European countries, such as Sweden, and those reported for countries such as the USA and the United Kingdom (see Cardoso, 2000).

More recently, Vieira (2005) examined low-wage mobility in Portugal. The results suggest that there is a high persistence rate and that for some types of workers, the low wage is an enduring phenomenon. In particular, females and less-educated workers are less likely to escape from such a situation. He also reports that 17,5% of the workers in the private sector and public firms were in the low pay segment in 1996. Moreover, 68.3% of the low-paid in 1996 remained in this position four years later, thus revealing a high persistence.

3. Literature Review

Most of the studies on low-pay employment have concentrated on the examination of its incidence and persistence. More recently, a new stream of research has attempted to relate low-pay employment and job quality. At this level, Leontaridi and Sloane (2001) conclude that low-pay employment in the United Kingdom is not necessarily associated with jobs of low quality, since low-paid workers report a level of satisfaction with their jobs as high as those in the high pay segment, or even higher.

With respect to the incidence of low pay, the literature indicates that the risk of such a situation is higher among females, young workers, low-educated workers, hotel and catering workers, retailing workers, part-timers and workers with low seniority (Salverda *et al.* 2001 OECD, 1999, Cardoso, 1997, and Vieira 2000). Furthermore, empirical evidence indicates that there are substantial differences in the incidence of low-pay employment across countries (see OECD, 2001 and Kesse and Swaim, 1997). In particular, the OECD (1999) reports that in countries such as the United States, Canada, Korea and Hungary, more than 20% of full-time workers came within the low-pay category in the mid-1990s. At the same time, the figures amounted to less than 10% in countries such as Sweden, Finland and Belgium. These results suggest that the proportion of low-pay employment relates to the indices of corporatism/centralisation, which have been presented in the literature (see, for instance, Bruno and Sachs, 1985, Calmfors and Driffill, 1988, and OECD, 1997b). In particular, it seems that the higher the level of corporatism/centralisation of a country, the lower the incidence of low-wage receivers. Indeed, this is a reasonable hypothesis, since corporatism apparently compresses the wage structure (see Teulings and Hartog, 1998).

4. Mixed Logit

The choice of the model depends on the characteristics of our data. In this particular case, we have a panel with two waves and attrition, because in the second year, some individuals have left the panel while others have entered it. Therefore, we have a percentage of individuals who are observed only once, which makes it impossible to apply some of the known panel data models for binary choice-dependent.

For cross-sectional studies, the standard logit is a highly popular model to deal with binary outcomes. Mcfadden (1974), Ben-Akiva and Lerman (1985) and Train (1986) use the logit model to relate the probability of making a choice to a set of variables reflecting decision-maker preferences. In the present paper, we use the model to relate the probability of an individual having a low wage to a set of variables which reflect the characteristics of the individual in the context of the panel data. The logit is usually used as a behavioural model. In the present paper, we use the model in a slightly different approach, which is statistical rather than behavioural, because we are more interested in discriminating in the population of workers between those who are low paid and those who are not, given the particular characteristics of the worker.

Define $y_{it} = 1$ if the individual i at time t (with $t = 1, 2$) is a low-wage receiver and $y_{it} = 0$ otherwise. The outcome y_{it} is determined by an equation which depends on a set of explanatory variables, x_{it} , associated with individual i and his job. We can define a non-observable latent variable y_{it}^* , such that $y_{it}^* = \beta'x_{it} + \nu_i + \varepsilon_{it}$, where β is a vector of unknown parameters, ν_i and ε_{it} are unobserved random variables allowing that individuals with the same characteristics x_{it} have different outcomes. The variable ν_i contains all effects that determine whether the individual belongs or not to the low wage group, which do not

vary over time and are not included in the set of regressors x_{it} , reflecting the heterogeneous behaviour of individuals. This equation can be thought of as explaining a certain salary threshold of an individual with given characteristics x_{it} . The individual has a low wage when it is below a certain threshold. To use the general framework of binary choice models, let us simply suppose that the individual has a low wage, $y_{it} = 1$, when $y_{it}^* > 0$.

A simple solution to deal with such a short in time panel data with attrition (with persons observed only once) is to perform a pooled analysis, applying a traditional procedure for cross-sectional data, such as the standard logit. However, it neglects the unobserved heterogeneity and may lead to inconsistent parameter estimates.

The standard pooled logit model ignores the unobserved effect v_i and assumes that ε_{it} are iid with logistic distribution. Therefore, the probability that a low wage is received by the worker i at time t is equal to:

$$P(\beta, x_{it}) = \frac{e^{\beta' x_{it}}}{1 + e^{\beta' x_{it}}} . \quad (1)$$

The Fixed Effects Probit considers v_i as parameters to be estimated. With only two time periods it will lead to an incidental parameters problem, resulting in inconsistent estimation of β . The Random Effects Probit assumes that v_i is independent of the regressors and is normally distributed. In our application, this may be inappropriate because it is likely that heterogeneity in individual behaviour regarding the probability of having a low wage will depend on the individual's characteristics, for instance age, the region where he or she lives and other variables related to the individual and the job environment. We will however test this assumption in the empirical part of this paper.

Other methods that allow a more flexible set of hypotheses about v_i , such as the Fixed Effects Logit, Chamberlain's Random Effects Probit and the semi-parametric Manski's

Maximum Score estimator, cannot be applied to our data, since at least two observations in time for each individual are required.

A more convenient approach is the random coefficient logit (RCL) model (also called mixed logit), which considers individual heterogeneity, relaxing the assumption that the coefficients are the same for all individuals. Therefore, it assumes that the coefficient of individual i on some characteristic j , β_{ji} , is a random draw from some distribution where the family of the distribution is specified, but the mean and variance are unknown and have to be estimated. We consider $\beta_i \equiv \mu + \eta_i$ with $\eta_i \sim F(0, \Omega)$, independent of ε_{it} . When $F(\bullet)$ is symmetric, it is usually considered to be the normal, and less often, the uniform or triangular distributions. If, for example, the coefficient can only assume positive values with asymmetric distribution, usually $F(\bullet)$ is lognormal. The latent variable equation can be written as:

$$y_{it}^* = \beta_i' x_{it} + \varepsilon_{it} = \mu' x_{it} + \eta_i' x_{it} + \varepsilon_{it}, \quad (2)$$

where the random unobserved component, $\eta_i' x_{it} + \varepsilon_{it}$, is heterogeneous, with heterogeneity depending on the explanatory variables.

The RPL probability that a worker receives a low wage is the integral of the standard logit probability in (1) over the density of the parameters,

$$P_{it} = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} P(\beta, x_{it}) f(\beta | \mu, \Omega) d\beta. \quad (3)$$

The model estimates the coefficients mean, μ , and the covariance between them, Ω . Exact maximum-likelihood estimation is not possible, since the integral cannot be calculated analytically and requires simulation. Recently developed techniques for simulating probabilities (Train, 2003) have made it feasible to estimate such models. Applications include Train (1998), Revelt and Train (1998), Mcfadden and Train (2000) and Rouwendal and Meijer (2001).

Observe that P_{it} is the expectation of $P(\beta, x_{it})$ so that it can be calculated by summing over R simulated $P(\beta_i, x_{it})$ with β_i drawn from $F(\beta | \mu, \Omega)$. These draws can be obtained randomly using a pseudo-random generator but more recently, systematic methods, such as Halton draws, have proved to be more efficient (see Train, 2003, for further details). The simulated probability is:

$$SP_{it} = \frac{1}{R} \sum_{r=1}^R P(\beta_i^r, x_{it}), \quad (4)$$

where β_i^r is the β from the r th draw from $F(\beta | \mu, \Omega)$ for individual i . Thus the simulated log-likelihood function for the RPL is:

$$SL = \sum_{i=1}^N \sum_{t=1}^T \log \left[SP_{it}^{y_{it}} (1 - SP_{it})^{1 - y_{it}} \right], \quad (5)$$

which depends on μ and Ω . The maximum-likelihood estimates of those parameters (given their chosen initial values) are obtained with iterative numerical optimisation procedures. See Train (2003) and Hensher and Greene (2003) for further explanations. Additionally, we remark that the mixed logit allows for heteroskedasticity in the error term, depending on the explanatory variables.

5. Model Specification

In this paper, we estimate a mixed logit model to analyse the low wage determinants in Portugal for the years 1998 and 1999. The choice of years is based on the availability of the data.

Before defining the empirical model based on the RCL framework, it is necessary to define which of the coefficients are random, that is, which variables have a heterogeneous behaviour in the population for the determination of a low wage. One simple approach is to consider first all the parameters as random estimating for each the mean and the standard deviation

and then whenever the last is not rejected to be null consider the respective parameter to be fixed. However, when models have many variables, this procedure may be difficult to implement because it becomes computationally very demanding. Therefore, we have carried out a prior analysis, based on a computationally simple procedure suggested by Chesher and Santos-Silva (2002) that allows us to identify the variables that explain the heterogeneous behaviour and the test proposed by Mcfadden and Train (2000) (from now on referred as M&T). Next, in estimating the RCL, we have considered as having random coefficients only those variables which lead to equation (6) presented below.

In our problem, the latent variable equation (2), mentioned in section 4, for the it individual at time t is:

$$\begin{aligned}
y_{it}^* = & \beta_0 + \mu_1 Male_{it} + \mu_2 Hours_{it} + \beta_3 Age_{it} + \beta_4 Age_{it}^2 + \beta_5 Educ1_{it} + \beta_6 Educ2_{it} + \\
& \mu_7 Full_{it} + \mu_8 Size1_{it} + \mu_9 Size2_{it} + \beta_{10} Size3_{it} + \beta_{11} Size4_{it} + \mu_{12} Size5_{it} + \\
& \mu_{13} Pcontr_{it} + \mu_{14} Public_{it} + \beta_{15} Region2_{it} + \mu_{16} Region3_{it} + \beta_{17} Region11_{it} + \\
& \beta_{18} Region12_{it} + \beta_{19} Region14_{it} + \beta_{20} Region15_{it} + \beta_{21} Industry1_{it} + \\
& \beta_{22} Industry2_{it} + \beta_{23} Year99 + \lambda_{it} + \varepsilon_{it}
\end{aligned} \tag{6}$$

where,

$$\begin{aligned}
\lambda_{it} = & \eta_{1i} Male_{it} + \eta_{2i} Hours_{it} + \eta_{7i} Full_{it} + \eta_{8i} Size1_{it} + \eta_{9i} Size2_{it} + \eta_{12i} Size5_{it} + \\
& \eta_{13i} Pcontr_{it} + \eta_{14i} Public_{it} + \eta_{16i} Region3_{it} .
\end{aligned} \tag{7}$$

Observe that $\beta_{ji} = \mu_j + \eta_{ji}$, $j = 1, 2, 7, 8, 9, 12, 13, 14, 16$. The meaning of the variables can be seen in table 1.

The model estimates the coefficients β_j , μ_j and the standard deviations of the random coefficients β_{ji} specified above.

6. Data and Empirical Results

The data used in this paper was gathered from the European Community Household Panel (ECHP) for the years of 1998 and 1999. This data set includes such information about the individuals as gender, education, age and wages, among others. It also includes information on variables such as the type of employment contract, employer size and the number of hours of work in the main activity. A low-wage worker is defined as an individual who earns less than two thirds of the median hourly wage (which amounted to 2,81 euros in 1998 and to 2,96 euros in 1999).

Table 1 presents the descriptive statistics of the data, as well as the definition of the variables. As we can observe, 13,1% of the workers fall into the low pay segment. This figure is slightly lower than the one reported by Vieira (2005). This may be because the data used by Vieira (2005) only includes workers in private and public firms and thus excludes civil servants.

Most of the explanatory variables to be used comprise dummy variables, with only two continuous variables: *hours* and *age*. Furthermore, the data is an unbalanced panel set, with 4852 respondents, of whom 670 have been observed only in 1998 and 705 only in 1999, giving the total of 8329 observations. As we can see from the information included in the same table, more than 50% of the low-paid workers are females. Moreover, the low-paid group is nearly two years younger than the higher-paid counterparts and the incidence of low-wage employment decreases with the level of education and with the employer size. Finally, the weight of public-sector workers is much lower in the low-wage group (nearly 3%) than among the higher-paid (27%), indicating that low-wage employment is mainly concentrated in the private sector.

[insert Table 1 about here]

We start the empirical analysis with the estimation of a standard logit model. To take into consideration heterogeneity, we also estimate the Random Effects (RE) Probit and the mixed logit.¹ Estimation results are included in Table 2. In addition, Table 3 presents the results of the application of some specification tests.

[insert Table 2 and Table 3 about here]

The RESET test gives evidence of mis-specification of the standard logit. This may be due to the presence of unobserved heterogeneity. The RE Probit considers this heterogeneity as being independent of the explanatory variables. Since the estimate of ρ is statistically significant, we have evidence of the presence of unobserved heterogeneity. Moreover, analysing the magnitude of this estimate we find that this heterogeneity is important, since its conditional variance is estimated to be around 80% of the conditional variance of the latent variable. There is no evidence of mis-specification in the RE Probit from the application of the RESET test, as can be observed in Table 3.

The application of the M&T test to the data shows evidence of random coefficients on the variables *Male*, *Hours*, *Fulltime*, *Pcontract*, *Public*, *Size1*, *Size2*, *Size5*, and *Reg3* leading to the specification of the mixed logit in (6). As we have mentioned before, it is likely that unobserved heterogeneity also depends on the explanatory variables. To test this hypothesis, we applied the HAL test of Chesher and Santos-Silva (2002), using likelihood ratio procedures. First, we considered that it was dependent on all the explanatory variables. Next,

¹ To estimate the mixed logit model, we used a simulator for RPL in the Gauss programming language, available in Kenneth Train's home page (<http://elsa.berkeley.edu/~train/Ps.html>).

we applied a classic selection procedure. The results show evidence of unobserved heterogeneity depending on the same variables as identified by the M&T test except that *Age* was detected instead of *Male*. Given that the random effect for *REG3* was not statistically significant we have restricted to the model in Table 2.

As usual, the magnitude of the coefficients is not directly comparable across models. However, they lead to similar conclusions with regard to the signs and the statistical significance of the variable effects. We observe that the probability of receiving a low hourly wage is lower for men, and higher for women; it is higher for workers who work more hours; it decreases with age for workers aged less than nearly 43, increasing for the others; it is lower for more educated workers; it is lower for full-time workers; it is lower for workers with a permanent contract; and it is lower for those working in the public sector. On the other hand, it is higher for those working in small firms, it is higher for those working in agriculture and those working in the Azores islands, as well as those in the central regions of mainland Portugal. However, the probability is lower for those working in the Algarve.

Interpreting the random effects in framework of RCL, we conclude that the effect of *Male*, *Hours*, *Fulltime*, *Pcontract*, *Public*, *Size1*, *Size2*, and *Size5* measured by the corresponding coefficients of the mixed logit, is not the same for all individuals (they are random). The estimated mean and standard deviation of these coefficients can be seen in Table 4. The same table includes the estimated probability of these coefficients being positive in the population of individuals, assuming they are normally distributed. We can conclude that the effect of *Hours* is always positive for all individuals, while the effect of *Full-time* is always negative, despite the fact that they vary from individual to individual. With respect to the other coefficients, we observe that for less than 10% of the individuals being male rises the probability of having a low wage, while for nearly 83% of the individuals, having a permanent contract lowers the referred probability and for around 90% of the population, to

work in the public sector also lowers the same probability. With reference to the size of the firm, we conclude that the effect on the probability of receiving a low wage for those working in the lowest category is positive for around 89% of the individuals, while for the second category this figure amounts to 93% and for the highest to nearly 47%.

The implications for policy of the random effects have to take into account these interpretations. For example, we estimate that for nearly 83% of the population to have a permanent contract will lower the probability to be a low wage worker but there are nearly 17% of workers where the effect is the opposite. If a policy based in the type of contract would be eventually devised then a further analysis on the causes that discriminate both behaviors would have to be considered.

7. Conclusions

This paper has examined the determinants of low-wage employment in Portugal. The results suggest that a low-wage worker in Portugal is normally a woman with a low level of education, working in agriculture with a part-time contract. Furthermore, she works in small enterprises and particularly in the region of the Azores or in central mainland. However, the results indicate that any assumption that the impact of such determinants is the same for all individuals may be a serious limitation, because of the presence of unobserved heterogeneity. In particular, we found that the effects of hours of work, full-time employment, gender, the type of contract (permanent or not), the public sector and the two lowest categories of employer size (Size 1 and Size 2) together with the highest (Size 5) are random.

In such a context, a general policy to overcome the problem of low-wage employment would be appropriate for the case of the non-random variables. However, a policy to act on

the random variables should take into account that their impact varies along the sample and therefore, a single policy instrument will not be sufficient to bring satisfaction to the entire working population. Policies tailored by clusters, which exist in the sample, would be more suitable.

References

- Asplund, R., Sloane, P. and Theodossiou, I. 1998. *Low pay and earnings mobility in Europe*. Cheltenham: Edward Elgar.
- Ben-Akiva, M. and Lerman, S. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA: MIT Press.
- Branco, F. and Mello, A. 1992. Why are wages in Portugal lower than elsewhere in the EEC. In J. F. Amaral, D. Lucena e A. S. Mello (eds.) *The Portuguese Towards 1992*, Kluwer: Boston, 131-151.
- Bruno, M. and Sachs, J. 1985. *Economics of worldwide stagflation*. Cambridge: Harvard University Press.
- Calmfors, L. and Driffill J. 1988. Centralization of wage bargaining and economic performance. *Economic Policy*, 6: 13-61.
- Cappellari, L. 2000. Low-wage mobility in the Italian labour market. *International Journal of Manpower*, 21: 264-290.
- Cardoso, A. 1997. *Earnings inequality in Portugal: The relevance and the dynamics of employer behaviour*. Ph.D Dissertation, European University Institute, Florence.
- Cardoso, A. 1998. Earnings inequality in Portugal: high and rising?. *Review of Income and Wealth*, 44: 325-343.
- Cardoso, A., Sousa, R., Castro, V. and Ferreira, P. 2000. Perfil do trabalhador e da empresa de baixos salários em Portugal. *Economia*, XXIV: 53-66.
- Chesher, A. and Santos-Silva, J. 2002. Taste Variation in Discrete Choice Models, *Review of Economic Studies*, 69: 147-168.

- Hartog, J., Pereira, P. and Vieira, J. 2001. Changing Returns to Education in Portugal during the 1980s and Early 1990s: OLS and Quantile Regression Analysis”, *Journal of Applied Economics*, 33: 1021-1037.
- Hensher, D.A. and Greene, W.H. 2003. The Mixed Logit Model: The State of Practice. *Transportation*. 30: 133-176.
- Kesse, M., and Swaim, P. 1997. The incidence and dynamics of low-paid employment in OECD countries”, paper presented at the LOWER Conference on Problems of Low-wage Employment, Bordeaux, France.
- Leontaridi, R. and Sloane, P. 2001. Measuring the quality of jobs. Working paper N° 7, European Low-wage Employment Research Network, Amsterdam.
- Lucifora, C. and Salverda, W. (1998) *Policies for low-wage employment and social exclusion in Europe*, FrancoAngeli: Milan.
- Mcfadden, D. 1974. Conditional Logit Analysis of Qualitative Choice Behavior. In *Frontiers in Econometrics*, edited by P. Zarembka, 105-142. New York: Academic Press.
- Mcfadden, D. and Train, K. 2000. Mixed MNL models of discrete response. *Journal of Applied Econometrics*, 15: 447-470.
- Mincer, J. 1974. *Schooling, experience and earnings*, New York: Columbia University Press.
- OECD (Organization for economic co-operation and development). 1999. Earnings distribution and poverty. In *Implementing OECD Jobs Strategy: assessing performance and Policy*, Paris: 89-106.
- OECD (Organization for economic co-operation and development).1997a. Earnings mobility: taking a longer run view. In *Employment Outlook 1997*, Paris: 27-61.
- OECD (Organization for economic co-operation and development).1997b. Economic performance and the structure of collective bargaining. *Employment Outlook 1997*, Paris: 63-92.

- OECD (Organization for economic co-operation and development).1996. Earnings inequality, low-paid employment and earnings mobility”, *Employment Outlook 1996*, Paris: 59-108.
- Revelt, D. and Train, K. 1998. Mixed logit with repeated choices. *Review of Economics and Statistics*, 80: 647-657.
- Rouwendal, J. and Meijer, E. 2001. Preferences for Housing, Jobs and Commuting: A Mixed Logit Analysis. *Journal of Regional Science*, 41 (3): 475-505.
- Salverda, W., Bazen, S. and Gregory, M. 2001. *The European-American employment gap, wage inequality, earnings mobility and skill: a study for France; Germany, the Netherlands, the United Kingdom and the United States*. University of Amsterdam, European Low-wage Employment Research Network (LOWER), Final Report.
- Stewart, M. and Swaffield, J. 1999. Low pay dynamics and transition probabilities. *Economica*, 66: 23-42.
- Teulings, C. and Hartog, J. 1998. *Competition or corporatism? Labour contracts, institutions and wage structures in international comparison*. Cambridge: Cambridge University Press.
- Train, K. 2003. *Discrete Choice Methods with Simulation*. Cambridge University Press.
- Train, K. 1998. Recreational demand models with taste variation. *Land Economics*, 74: 230-239.
- Train, K. 1986. *Qualitative Choice Analysis*. Cambridge, MA: MIT Press.
- Vieira, J. .2005. Low-wage mobility in the Portuguese labour market”, *Portuguese Economic Journal*, 4, 1-14.

Table 1 – Variables: Definition and Mean Values

Variable	Description	All workers	Low paid	Higher paid
Low Wage (Dependent)	Equals 1 if worker earns less than two-thirds of the median wage, 0 otherwise	0.131	-	-
Male	Equals 1 if the worker is a male, 0 otherwise	0.571	0.456	0.591
Hours	Number of hours worked per month	163.7	181.3	161.0
Age	Age of the worker (years)	37.09	35.5	37.3
Educ1	Equals 1 if the worker has a recognized third level education, 0 otherwise	0.115	0.006	0.132
Educ2	Equals 1 if the worker has second stage of secondary education, 0 otherwise	0.144	0.081	0.153
Full	Equals 1 when the individual is a full-time worker, 0 otherwise	0.967	0.921	0.974
Pcontr	Equals 1 if the worker has a permanent contract, 0 otherwise	0.795	0.615	0.822
Public	Equals 1 if the worker is a civil servant, 0 otherwise	0.238	0.028	0.270
Size1	Equals 1 if the employer has 1 – 4 workers, 0 otherwise	0.210	0.509	0.165
Size2	Equals 1 if the employer has 5 – 19 workers, 0 otherwise	0.344	0.305	0.350
Size3	Equals 1 if the employer has 20 – 49 workers, 0 otherwise	0.159	0.088	0.169
Size4	Equals 1 if the employer has 50 – 99 workers, 0 otherwise	0.101	0.052	0.108
Size5	Equals 1 if the employer has 100 – 499 workers, 0 otherwise	0.123	0.032	0.137
Industry1	Equals 1 if the individual works in agricultural activities, 0 otherwise	0.053	0.167	0.036
Industry2	Equals 1 if individual works in the	0.352	0.313	0.358

	industry,0 otherwise			
Region2	Equals 1 if the individual works in Azores, 0 otherwise	0.138	0.179	0.131
Region3	Equals 1 if the individual works in Madeira, 0 otherwise	0.091	0.054	0.097
Region11	Equals 1 if the individual works in the North, 0 otherwise	0.193	0.171	0.196
Region12	Equals 1 if the individual works in central mainland, 0 otherwise	0.220	0.264	0.214
Region14	Equals 1 the individual works in Alentejo, 0 otherwise	0.102	0.104	0.102
Region15	Equals 1 if the individual works in Algarve, 0 otherwise	0.138	0.140	0.138
Year99	Equals 1 if the observation refers to 1999, 0 otherwise	0.502	0.521	0.499

Table 2: Parameter Estimates and t-statistics (Dependent variable: Low wage)

Variables	Stand. Logit		RE Probit		Mixed Logit	
	Coeff.	t-stat	Coeff.	t-stat	Coeff.	t-stat
Intercept	-1.1152	-2.16	-0.4243	-1.44	-1.4798	-1.21
Male	-1.3991	-15.74	-0.7135	-13.31	-4.3691	-9.51
Hours	0.0278	20.52	0.0147	19.77	0.0726	11.69
Age	-0.1783	-10.11	-0.1031	-9.30	-0.4401	-7.60
Age ²	0.0021	9.59	0.0012	8.79	0.0052	7.13
Age* (turn. point)	42.5		42.7		42.3	
Educ1	-3.0561	-6.62	-1.3562	-6.86	-6.6011	-5.86
Educ2	-0.8633	-6.35	-0.4173	-5.40	-2.1275	-5.75
Fulltime	-2.6582	-13.29	-1.3177	-11.35	-6.7316	-10.74
Pcontr	-0.6901	-7.92	-0.3984	-8.13	-2.0309	-6.90
Public	-1.5062	-7.39	-0.6165	-6.39	-6.5465	-3.92
Size1	2.3292	7.32	1.0851	6.82	4.2218	0.68
Size2	1.4874	4.71	0.6774	4.33	2.5689	4.09
Size3	0.9958	3.03	0.4214	2.60	1.9292	3.19
Size4	0.9823	2.87	0.4343	2.54	1.9550	3.03
Size5	0.3421	0.96	0.1126	0.65	-0.1356	-0.14
Ind1	0.9330	6.47	0.4482	5.06	2.0988	5.46
Ind2	0.0850	0.89	-0.0022	-0.04	0.3949	1.58
Reg2	0.7107	4.26	0.3336	3.35	1.5216	3.38
Reg3	0.0799	0.39	-0.0511	-0.41	-0.3552	-0.63
Reg11	0.2205	1.34	0.1304	1.35	0.0779	0.18
Reg12	0.6529	4.20	0.3007	3.27	1.0345	2.57
Reg14	-0.2567	-1.39	-0.1001	-0.95	-0.7032	-1.53
Reg15	0.1715	1.01	0.0906	0.90	-0.0015	-0.00
Year99	0.0749	0.97	0.0290	0.93	0.1819	1.24
Random Effects						
Male					3.3002	7.81
Hours					0.0093	3.39
Full					-1.3186	-2.74
Pcontr					2.1847	6.052
Public					4.9525	4.38
Size1					3.4599	7.12
Size2					1.7113	3.25
Size5					1.9733	1.76
Rho			0.8032	36.29		
Observations		8329		8329		8329
LogLikelihood		-2227.57		-2022.92		-2014.43

Table 3: Specification Tests

Test	Statistic	p-value
RESET on standard Logit	-2.0075	0.045
standard Logit vs HAL	175.52	0.000
M&T	111.47	0.000
RESET on RE Probit	-0.1556	0.876

RESET was performed with $\hat{\beta}'x_i^2$;

HAL: heterogeneity depends on *hours*, *age*, *fulltime*, *pcontract*, *public*, *size1*, *size2*, *size5* and *reg3*.

M&T: Mcfadden and Train (2000) to test for random parameters. The null is the Logit.

Table 4: Mean, Standard Deviation and Probability of Random Coefficients being Positive

Coefficient	Mean	St. Deviation	P(coeff. > 0)
Male	-4.3691	3.3002	0.0928
Hours	0.0726	0.0093	1.0000
Full	-6.7316	1.3186	0.0000
Pcont	-2.0309	2.1847	0.1763
Public	-6.5465	4.9525	0.0931
Size1	4.2218	3.4599	0.8888
Size2	2.5689	1.7113	0.9333
Size5	-0.1356	1.9733	0.4726