Modelling Replacement Demand: A Random Coefficient Approach

ROA-RM-1999/2E

Ed Willems

Research Centre for Education and the Labour Market

Faculty of Economics and Business Administration
Maastricht University

Maastricht, August 1999

ISBN-90-5321-267-1
Sec99.053/RvdV

Contents

Abstract

Replacement demand due to regular and early retirement, disablement, temporary withdrawals of women owing to birth and child raising, *et cetera*, constitutes an important element of the future demand for newcomers on the labour market. It is, however, often neglected in manpower forecasting studies. In this paper, we develop a flow model with which we can forecast the replacement demand at a detailed level of types of education. The essential element of the model is the determination of the yearly outflow coefficients for each type of education, distinguished by age category and gender. Instead of applying a fixed coefficient approach (Willems and De Grip (1993), we will introduce the random coefficient estimation technique in the model. This technique takes the reliability of the flow coefficients explicitly into account. If a specific flow coefficient is unreliable, the random coefficient estimate will tend towards the mean value for that age category over the types of education. This means that *ad hoc* corrections based on judgmental forecasting can be considerably reduced. Moreover, it will be shown that the random coefficient estimates of labour market outflow are more reliable than the estimates of the Willems/De Grip model.

Acknowledgements

This study is part of the research programme ‘Project Education and the Labour Market’, subsidised by the Ministry of Education, Culture and Science, LDC National Centre for Career Issues, the Ministry of Social Affairs and Employment, the Ministry of Agriculture, Conservation and Fisheries and the Central Employment Board (CBA). I would like to thank Lex Borghans and Andries de Grip for their helpful comments.

1 Introduction

The match between the educational system and the labour market has been a subject of policy concern for many years. It is in this context that the ROA information system on education and the labour market was developed (see *e.g.* Heijke and De Grip, 1991). The main aim of this system is to provide information to students with respect to the labour market consequences of the various educational or occupational choices. This information should refer to a detailed level of occupational groups and types of education and must be useful for educational and vocational guidance. The most important parts of the information system are the medium-term forecasts. In this forecasting model, the future demand and supply of newcomers by educational category are confronted in order to obtain insight in the future labour market situation.

We can distinguish two possible causes of the demand for newcomers: (i) the expansion demand which represents the additional demand due to employment growth and (ii) the replacement demand, which occurs when people leave the work force and their places are re-filled. This outflow can be the result of retirement, or temporary or permanent withdrawal from the labour market. The supply of newcomers onto the labour market mainly consists of the outflow of school-leavers from the educational system.

This paper will focus on the replacement demand by educational category. Replacement demand is an important element of the total demand for newcomers on the labour market (see De Grip, Meijboom and Willems, 1994), although it is often neglected in the manpower forecasting literature (see Willems, 1996). ROA has developed a replacement demand model in the context of the information system (see Willems and De Grip, 1993). The key point of this model, which is based upon a demographic theory, is that it only requires stock data. This is a large advantage, as flow data at this low level of aggregation is not available for the Netherlands.

The replacement demand model of Willems and De Grip (1993) is based upon a large number of flow coefficients. It distinguishes 20 categories (age by gender) for each type of education. With a level of aggregation of about 100 categories, this implies 2,000 coefficients. The main disadvantage of this model is that these flow coefficients cause some instabilities in the results, because the coefficients are frequently based on only a few observations in the survey.[1] This makes the results more vulnerable to sample error, although it is possible that the errors compensate each other to some extent if they are aggregated over the various cohorts.

This paper aims to develop a flow model that can obviate the instability of the flow

1. Note that this problem is not only caused by applying the demographic model. If real flow data were available, a similar problem would occur.

coefficients. We try to obtain more reliable estimates of the real labour market flows. In Section 2, a brief overview is given of the demographic replacement demand model. This model is the starting point for the analysis in this paper. Section 3 specifies the new model, which makes use of the random coefficient estimation technique. Subsequently, in Section 4 the results are presented. A comparison will be made between the observed labour market outflow (based on the demographic model) and the estimated outflow (based on the random coefficient model). The paper will be summarised in Section 5.

2 The Demographic Replacement Demand Model

Willems and De Grip (1993) have developed a *net* flow model, which is based on data on the age and gender structure of employment by type of education. The key point for this model, which is known in the demographic literature as the cohort components method (see *e.g.* Shryock and Siegel, 1980), is the derivation of the 'cohort-change rates'. These rates refer to the number of workers with a certain educational background at two specific points of time. The cohort-change rates can be written down in symbols as:[2]

$$\ell_{a,i,t-n\backslash t} = \frac{L_{a+n,i,t}}{L_{a,i,t-n}} \tag{2.1}$$

where:

$\ell_{a,i,t-n\backslash t}$	=	cohort-change rate of the workers with age a and educational background i at time t–n during the period (t–n, t);
$L_{a,i,t}$	=	number of workers with age a and educational background i at time t.

In economics it is common practice to speak of annual growth (inflow or outflow) rates instead of cohort-change rates:

$$f_{a,i,t-n\backslash t} = \sqrt[n]{\frac{L_{a+n,i,t}}{L_{a,i,t-n}}} - 1 \tag{2.2}$$

where:

$f_{a,i,t-n\backslash t}$	=	average annual net inflow or outflow rate of the workers with age a and educational background i at time t–n during the period (t–n, t). If the value is less than zero, it refers to a net outflow and if the value is greater than zero, it refers to a net inflow.

2. The gender index is left out for simplicity reasons.

For each education and cohort we can derive such a cohort-change rate, or similarly the net inflow or outflow:

$$F_{a,i,t-n\backslash t} = L_{a+n,i,t} - L_{a,i,t-n} = L_{a,i,t-n}\left(1+f_{a,i,t-n\backslash t}\right)^{n} - L_{a,i,t-n} \qquad (2.3)$$

where:

$F_{a,i,t-n\backslash t}$ = net inflow or outflow of workers with age *a* and educational background *i* at time *t–n* during the period (*t–n*, *t*).

Summing over all age categories and both genders gives the total net outflow for educational category *i*:

$$O_{i,t-n\backslash t} = \sum_{a} -\min\left\{0; F_{a,i,t-n\backslash t}\right\} \qquad (2.4)$$

where:

$O_{i,t-n\backslash t}$ = net outflow for educational category *i* during the period (*t–n*, *t*).

The replacement demand for educational category *i* can then be defined as the net outflow, insofar as the resulting vacancies will be filled. In other words, if the employment for the educational category concerned shrinks, this must be balanced in the replacement demand. In symbols:

$$RD_{i,t-n\backslash t} = O_{i,t-n\backslash t} + \min\left\{0; \Delta L_{i,t-n\backslash t}\right\} \qquad (2.5)$$

with:

$$\Delta L_{i,t-n\backslash t} = L_{i,t} - L_{i,t-n} \qquad (2.6)$$

where:

$RD_{i,t-n\backslash t}$ = replacement demand for educational category *i* during the period *(t–n, t)*;

$L_{i,t}$ = number of workers with educational background *i* at time *t*.

This method implies that the replacement demand that will be filled by entrants and re-entrants of the same cohort cannot be determined. However, as the aim of the labour market forecasts in the context of the ROA information system is to provide insight in the employment prospects for *newcomers*, this is not a real restriction.

To forecast the future replacement demand, we can use the historically observed cohort-change rates, together with the most recent information of the age and gender decomposition of the employment for the educational category concerned. The future net outflow will then be:

$$\hat{O}_{i,t\backslash t+n} = \sum_a -\min\left\{0;\hat{F}_{a,i,t\backslash t+n}\right\} \quad (2.7)$$

where O_i and $F_{a,i}$ are defined similarly as before[3] and:

$$\hat{F}_{a,i,t\backslash t+n} = L_{a,i,t}\left(1+f_{a,i,t-n\backslash t}\right)^n - L_{a,i,t} \quad (2.8)$$

However, for their forecasts Willems and De Grip (1993) made two corrections on the historically observed cohort-change rates. Firstly, this concerns a correction for the business cycle. If unemployment is growing during the observation period the outflow measured will be higher than the 'structural' outflow. Therefore the net inflow or outflow is made neutral for cyclical variations by means of the difference between the growth of the labour force on the one hand and the growth of the number of working persons on the other. This correction is cohort specific and can be written as:

$$f^s_{a,t-n\backslash t} - f_{a,t-n\backslash t} \quad (2.9)$$

with:

$$f^s_{a,t-n\backslash t} = \sqrt[n]{\frac{L^s_{a+n,t}}{L^s_{a,t-n}}} - 1 \quad (2.10)$$

and:

$$\begin{aligned} f_{a,t-n\backslash t} &= \sqrt[n]{\frac{L_{a+n,t}}{L_{a,t+n}}} - 1 \\ &= \sqrt[n]{\frac{\sum_i L_{a+n,i,t}}{\sum_i L_{a,i,t+n}}} - 1 \end{aligned} \quad (2.11)$$

where:

$L^s_{a,t}$	=	number of persons in the labour force (labour supply) with age *a* at time *t*;
$f^s_{a,t-n\backslash t}$	=	growth rate of the labour force with age *a* at time *t–n* during the period (*t–n*, *t*);
$L_{a,t}$	=	total number of workers with age *a* at time *t*;
$f_{a,t-n\backslash t}$	=	growth rate of the total number of workers with age *a* at time *t–n* during

3. A hat (^) above the symbols indicates a forecast.

the period (t–n, t).

The second correction refers to the degree of participation. If the general degree of participation grows during the forecasting period, for instance due to institutional reasons, i.e. increasing participation of women, the future net outflow will become smaller. So, the second correction term incorporates this autonomous development of labour participation (independent of educational category):

$$\hat{f}^{s}_{a,t\backslash t+n} - f^{s}_{a,t-n\backslash t} \tag{2.12}$$

The forecasted net inflow or outflow rates are then:

$$\begin{aligned}\hat{f}_{a,i,t\backslash t+n} &= f_{a,i,t-n\backslash t} + f^{s}_{a,t-n\backslash t} - f_{a,t-n\backslash t} + \hat{f}^{s}_{a,t\backslash t+n} - f^{s}_{a,t-n\backslash t} \\ &= f_{a,i,t-n\backslash t} - f_{a,t-n\backslash t} + \hat{f}^{s}_{a,t\backslash t+n}\end{aligned} \tag{2.13}$$

Subsequently these forecasts of the future flow rates are projected to the age and gender structure of the workers with the educational background concerned at the beginning of the forecasting period. This yields the forecasts of the future net outflow (compare equation (2.7)-(2.8)).

As indicated above, the future replacement demand for educational categories with a growing employment is equal to the net outflow. For shrinking types of education, however, the replacement demand equals the number of job openings that will be filled, *i.e.* the net inflow, which is similar to the net outflow discounted for the (future) shrinking of employment for the educational category concerned. For this correction Willems and De Grip do not use the employment forecasts available within the information system, but an implicit forecast made in the replacement demand model:

$$\hat{f}_{i,t\backslash t+n} = f_{i,t-n\backslash t} - f_{t-n\backslash t} + \hat{f}^{s}_{t\backslash t+n} \tag{2.14}$$

with:

$$f_{i,t-n\backslash t} = \sqrt[n]{\frac{L_{i,t}}{L_{i,t-n}}} - 1 \tag{2.15}$$

$$f_{t-n\backslash t} = \sqrt[n]{\frac{\sum_i L_{i,t}}{\sum_i L_{i,t-n}}} - 1 \tag{2.16}$$

$$\hat{f}^{s}_{t\backslash t+n} = \sqrt[n]{\frac{L^{s}_{t+n}}{L^{s}_{t}}} - 1 \qquad (2.17)$$

where:

$f_{i,t-n\backslash t}$ =growth rate of total employment for educational category *i* during the period (*t*–*n*, *t*);

$f_{t-n\backslash t}$ = growth rate of total employment during the period (*t*–*n*, *t*);

$\hat{f}^{s}_{t\backslash t+n}$ = forecast growth rate of the total labour force during the period *(t, t+n)*.

The replacement demand forecast can then be written as (compare equations (2.5)-(2.8)):

$$R\hat{D}_{i,t\backslash t+n} = \hat{O}_{i,t\backslash t+n} + \min\left\{0; \Delta\hat{L}_{i,t\backslash t+n}\right\} \qquad (2.18)$$

with:

$$\Delta\hat{L}_{i,t\backslash t+n} = L_{i,t}\left(1+\hat{f}_{i,t\backslash t+n}\right)^{n} \qquad (2.19)$$

3 The Random Coefficient Model

The replacement demand model as described, has an important disadvantage. For the Netherlands, we have to use a large amount of highly detailed information from the Labour Force Survey (LFS). If we distinguish 100 types of education, 10 5-band age classes, and the two sexes, we have 2,000 flow coefficients. This implies that, using a sample of about 100,000 working individuals, many of these coefficients are based on a very small number of observations, which means that disturbances (‘noise’) can easily occur.

However, this is not a problem that is specific for the model developed, which makes use of stock data. When flow data is available, for instance based on retrospective information, one also has to distinguish such a number of flow coefficients. Inflow or outflow rates split up by demographic category have to be available, as the demographic structure of the working population is an important explaining factor for the outflow of workers out of a job. For almost every type of education, there is an outflow from the oldest cohorts, due to (early) retirement *et cetera* and a net inflow for the younger age groups. For women we see an additional outflow around the age of 30 to 35 because of the withdrawal from the labour force linked to the birth and raising of children.

An easy way to incorporate these similar flow patterns into a model is by ‘pooling’ the data over types of education. This multiplies the number of observations for each cohort distinguished. Using this method filters out the effect of ‘outliers’ in the observations, enabling us to obtain more robust forecasts of the flow coefficients.

However, an important disadvantage of this pooling method is that each type of education will have the same (average) flow coefficients.[4] The variation between the various educational categories is completely lost, which means that the differences in labour market flows between these categories can only be due to different age and gender structure of employment. This does not seem to be a very realistic assumption.

The problem can be overcome by specifying a so-called random coefficient model.[5] Such a model can be characterised by a position in between the pooling model and the education-by-education model specification. The parameters of the random coefficient model are the weighted average of the least squares estimates by educational category and the average 'pooled' estimates. If the specific estimate is less reliable, more weight will be attributed to the average estimate; if the specific estimate is relatively 'sure', the parameters will tend towards these single estimates.

Econometric theory on random coefficient models

Specify the standard linear model for education *i*:

$$y_i = X_i \beta_i + \varepsilon_i \qquad (3.1)$$

where:

y_i	=	vector of observations of the dependent variable for education *i*;
X_i	=	matrix of observations of the explanatory variables for education *i*;
β_i	=	vector of parameters for *i*;
ε_i	=	disturbance vector for *i*.

In the pooled model the vector β will be the same for each educational category. However, as said above, parameter β in the random coefficient model can vary between the types of education distinguished. Vector β_i can be seen as a random vector drawn from a probability distribution with mean $\bar{\beta}$ and covariance matrix Δ. Equation (3.1) can then be written as:

$$y_i = X \left(\bar{\beta} \quad {}_i\right) + \varepsilon_i \qquad (3.2)$$

μ_i and ε_i are:

4. A dummy variable for the educational type distinguished can of course be included. This however implies a shift effect that is the same for all cohorts.

5. See Swamy (1970 and 1971). Borghans and Heijke (1994) applied the random coefficient model to explain the occupational structure within occupations.

$$E(\mu_i) = 0\,,\; E(\mu_i \mu_i') = \Delta\,, \text{ and } E(\mu_i \mu_{i'}') = 0 \text{ for } i \neq i' \tag{3.3a}$$

$$E(\varepsilon_i) = 0\,,\; E(\varepsilon_i \varepsilon_i') = \sigma_i^2 I\,, \text{ and } E(\varepsilon_i \varepsilon_{i'}') = 0 \text{ for } i \neq i' \tag{3.3b}$$

$$E(\mu_i \varepsilon_{i'}') = 0 \tag{3.3c}$$

where I is the identity matrix.

This random coefficient model can be estimated in the following way (see Judge *et al.*, 1980). First estimate equation (3.1) with Ordinary Least Squares (OLS):

$$b_i = \left(X_i' X_i\right)^{-1} X_i' y_i \tag{3.4}$$

where:
b_i = least squares estimator of β_i.

Subsequently, the variances of ε_i and μ_i can be estimated by making use of the residuals of the least squares estimates:

$$\hat{\varepsilon}_i = y_i - X_i b_i \tag{3.5}$$

An unbiased estimator of σ_i^2 is:

$$\hat{\sigma}_i^2 = \frac{\hat{\varepsilon}_i' \hat{\varepsilon}_i}{M - K} \tag{3.6}$$

where:
M = number of observations for education i;
K = number of parameters.

An unbiased estimator for the matrix Δ is:

$$\hat{\Delta} = \frac{S_b}{N-1} - \frac{1}{N} \sum_{i=1}^{N} \hat{\sigma}_i^2 \left(X_i' X_i\right)^{-1} \tag{3.7}$$

with:

$$S_b = \sum_{i=1}^{N} b_i b_i' - \frac{1}{N}\sum_{i=1}^{N} b_i \sum_{i=1}^{N} b_i'$$

$$= \sum_{i=1}^{N} \left(b_i - \bar{b}\right)\left(b_i - \bar{b}\right) \qquad (3.8)$$

where:

N	=	number of educational categories distinguished;
$\bar{b}$	=	mean of the least squares estimates.

However, the difficulty with the estimator $\hat{\Delta}$ is that it may be negative definite, which of course is not allowed for a covariance matrix. To overcome this problem, De Crombrugghe and Dhaene (1991) suggest an alternative estimator for Δ that is always semi-positive definite. This estimator is obtained by means of an interactive process:

$$\Delta^{(1)} = \frac{S_b}{N-1} \qquad (3.9)$$

$$\Delta^{(n)} = \Delta^{(n-1)}\left(\Delta^{(n-1)} + \frac{1}{N}\sum_{i=1}^{N} \sigma_i^2 \left(X_i' X_i\right)^{-1}\right) \qquad (3.10)$$

Convergence gives an estimator for Δ:

$$\hat{\Delta} = \lim_{n \to \infty} \Delta^{(n)} \qquad (3.11)$$

In the case of a negative definite matrix $\hat{\Delta}$, on the other hand, Judge *et al.* (1980) simply suggest:

$$\hat{\Delta} = \frac{S_b}{N-1} \qquad (3.12)$$

In both cases, the estimator of the variance matrix of the composite disturbance term is:

$$\hat{\Phi}_i = E\left(X_i \mu_i + \varepsilon_i\right)\left(X_i \mu_i + \varepsilon_i\right)'$$

$$= X_i \hat{\Delta} X_i' + \sigma_i^2 I \qquad (3.13)$$

The Estimated Generalised Least Squares (EGLS) estimator for $\bar{\beta}$ is then:

$$\hat{\beta} = \left(\sum_{i=1}^{N} X_i' \hat{\Phi}_i^{-1} X_i \right)^{-1} \sum_{i=1}^{N} X_i' \hat{\Phi}_i^{-1} y_i$$

$$= \sum_{i=1}^{N} \hat{W}_i b_i \qquad (3.14)$$

with:

$$\hat{W}_i = \left(\sum_{i=1}^{N} \left(\hat{\Delta} + \hat{\sigma}_i^2 \left(X_i' X_i \right)^{-1} \right)^{-1} \right)^{-1} \left(\hat{\Delta} + \hat{\sigma}_i^2 \left(X_i' X_i \right)^{-1} \right)^{-1} \qquad (3.15)$$

The covariance matrix of $\hat{\beta}$ is:

$$E\left(\hat{\beta}\,\hat{\beta}' \right) = \left(\sum_{i=1}^{N} X_i' \hat{\Phi}_i^{-1} X_i \right)^{-1}$$

$$= \left(\sum_{i=1}^{N} \left(\hat{\Delta} + \hat{\sigma}_i^2 \left(X_i' X_i \right)^{-1} \right)^{-1} \right)^{-1} \qquad (3.16)$$

Finally, the estimator of the random parameter vector for *i* is:

$$\hat{\beta}_i = \left(\hat{\Delta}^{-1} + \hat{\sigma}_i^{-2} \left(X_i' X_i \right) \right)^{-1} \left(\hat{\sigma}_i^{-2} \left(X_i' X_i b_i \right) + \hat{\Delta}^{-1} \hat{\beta} \right) \qquad (3.17)$$

This means that the parameter estimates can be written as the weighted average of the individual OLS estimator (b_i) and the estimate of the mean parameter values ($\hat{\beta}$).

Model specification

Within the general framework of random coefficient models, we can now specify a model for explaining and estimating labour market flows. We opt for the following simple specification for educational category *i*:[6]

$$f_i = f + D_a \alpha_i + \varepsilon_i \qquad (3.18)$$

6. Separate equations are estimated for males and females. Instead of using the specification described in equation (3.18), we could also use different specifications. However, the number of parameters that we can include in the model is restricted due to the lack of observations. Moreover, with the variation allowed by educational category, gender, and age group, most of the variance that could be expected theoretically is dealt with. Labour market flows mainly vary between ages and men and women.

where:

f_i = vector of net inflow or outflow rates for educational category i, with observations per age category and time period;

f = vector of net inflow or outflow rates for the total work force, with observations per age category and time period;

D_a = matrix of dummy variables; elements are equal to 1 for age category a and 0 otherwise;

α_i = vector of random parameters.

Equation (3.18) states that the labour market flows for a certain educational category per age category are equal to the average labour market flow for that age group, apart from an education-specific variation, which can differ between age groups. In equation (3.18), the vector f_i has the following elements:

$$f_i = \begin{pmatrix} f_{1,i,t-T-1\backslash t-T} \\ \vdots \\ f_{1,i,t-1\backslash t} \\ f_{2,i,t-T-1\backslash t-T} \\ \vdots \\ f_{2,i,t-1\backslash t} \\ \vdots \\ f_{A,i,t-T-1\backslash t-T} \\ \vdots \\ f_{A,i,t-1\backslash t} \end{pmatrix} \tag{3.19}$$

where T and A indicate the total number of time periods and age categories distinguished, respectively. As in the LFS, information per educational category – at a detailed level – has only been available since 1992, T = 4 (1992-1993, [...], 1995-1996).[7] With a distinction in 5-year groups of ages, 10 categories are available: 15-19, 20-24, *et cetera*, to 60-64.[8] However, the (net) inflow into the younger age groups will not be estimated. This inflow consists almost entirely of school-leavers. The labour market flows that we take into account refer to the age of 25 and older, *i.e.* the first age category incorporated in the

7. By now, data is available on 1997. For the empirical applications in this study, this data is not used.

8. As we use annual data, comparisons are made with the age categories 16-20, 21-25, *et cetera*, within the cohort components method. If no single-age data is available, imaginary cohorts could be used. In this case, the flows for category a during the period (t–1, t) can be calculated by comparing the number of workers with age category a at time t–1 with the workers of the imaginary cohort 0.8a + 0.2(a+1) at time t.

model is 25-29. This implies that 8 age groups are considered (A = 8). Hence, for each sex we have 4 x 8 = 32 observations per educational category. Obviously, vector f has the same form as f_i; matrix D_a and vector α_i have the following format:

$$D_a \alpha_i = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \\ 1 & 0 & & \\ 0 & 1 & & \vdots \\ & \vdots & & \\ & 1 & & \\ \vdots & & \ddots & 0 \\ & & 0 & 1 \\ & & \vdots & \vdots \\ 0 & \cdots & 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha_{1,i} \\ \alpha_{2,i} \\ \vdots \\ \alpha_{A,i} \end{pmatrix} \tag{3.20}$$

As pointed out above, the parameters are random coefficients, which implies that α_i can be written as:

$$\alpha_i = \overline{\alpha} + \mu_i \tag{3.21}$$

Model estimation

The model can be estimated following the procedure described above. However, instead of estimating the mean of the random parameters, we can derive these parameters on theoretical grounds. It may be obvious that the age-specific net inflow and outflow, aggregated over the educational categories distinguished, have to be equal to the total net inflow into or outflow from the work force. This implies that the parameters α have to be zero, on average: $\overline{\alpha} = 0$. The parameter estimates can then be interpreted as follows. If $\hat{\alpha}_{a,i} > 0$, then $\mu_{a,i} > 0$, the inflow or outflow rate for the educational category concerned (*i*) will be greater for that specific age group (*a*) than the average flow rate for the total work force, and *vice versa* if $\hat{\alpha}_{a,i} < 0$.

First, let us write equation (3.18) as:

$$df_i = f_i - f = D_a \alpha_i + \varepsilon_i \tag{3.22}$$

However, we know in advance that some observations of the flow rates will be more reliable than others, because they are based upon more respondents in the survey. A general solution for this is to weight the observations; those observations that are more reliable, *i.e.* that are based on more respondents, obtain a higher weight than the observations that are more doubtful. This leads to the following weight matrix for education *i*:

$$
\Omega_i = \begin{pmatrix} \sqrt{L_{1,i,t-T-1}} & 0 & \cdots & & & & 0 \\ 0 & \ddots & & & & & \\ & & \sqrt{L_{1,i,t-1}} & & & & \vdots \\ \vdots & & & \ddots & & & \\ & & & & \sqrt{L_{A,i,t-T-1}} & & \\ & & & & & \ddots & 0 \\ 0 & & \cdots & & & 0 & \sqrt{L_{A,i,t-1}} \end{pmatrix} \qquad (3.23)
$$

Both the left-hand side and the right-hand side of equation (3.22) will be multiplied with the matrix Ω_i, which gives:

$$df_i^* = \Omega_i df_i \qquad (3.24a)$$

$$D_a^* = \Omega_i D_a \qquad (3.24b)$$

$$\varepsilon_i^* = \Omega_i \varepsilon_i \qquad (3.24c)$$

Equation (3.22) can then be written as:

$$df_i^* = D_a^* \alpha_i + \varepsilon_i^* \qquad (3.25)$$

This equation can be estimated with random coefficients.

Forecasting future outflow of workers

With the estimated parameter values for α, the model specified in equation (3.18) can also be used for forecasting purposes. However, in the specification used, we need forecasts of the total net inflow or outflow rates by gender and age category (f). In general, these will not be available. Forecasts of the future size of the labour force are available, however.[9] By definition, the following is true:

$$L_{a,t}^s \equiv L_{a,t} + UE_{a,t} \qquad (3.26)$$

where:

$UE_{a,t}$ = number of unemployed people with age category a at time t.

9. These can be obtained by combining the population forecasts of Statistics Netherlands with the forecasts of the degree of participation of the Central Planning Bureau.

In terms of forecasts for t+1, equation (3.26) is:

$$\hat{L}^{s}_{a,t+1} \equiv \hat{L}_{a,t+1} + \hat{U}E_{a,t+1} \tag{3.27}$$

To make flow forecasts, we start from the assumption that unemployment will have relatively the same development as the labour force. In terms of flow ratios, we can then write:

$$\hat{f}_{a,t\backslash t+1} = \hat{f}^{s}_{a,t\backslash t+1} \tag{3.28}$$

This assumption can be interpreted as follows. Suppose 10% of the labour force of age 50-54 in 1995 will not supply on the labour market in 1996. There may be several reasons for this withdrawal from the labour market. Firstly, people may have died. Secondly, individuals may have become disabled or migrated to other countries. Thirdly, other reasons may have caused their withdraw from the labour market. These reasons could be infused by changing preferences or changing wage potentials of individuals. The assumption made says that the extent to which individuals behave in a certain way, does not differ between workers and non-workers. In other words, in the example given, the 10% withdrawal from the labour market refers to both workers and unemployed individuals.

It can be shown that within the random coefficient model such an assumption is consistent with the correction factors that are suggested within the demographic model presented in Section 2. Substituting the mean for α_i ($\alpha_i = 0$), equations for age category a and the periods (t–1, t) and (t, t+1) of model (3.18) are:

$$f_{a,i,t-1\backslash t} = f_{a,t-1\backslash t} \tag{3.29a}$$

$$\hat{f}_{a,i,t\backslash t+1} = \hat{f}_{a,t\backslash t+1} \tag{3.29b}$$

Subtracting equation (3.29a) from equation (3.29b) and substituting equation (3.28) yields:

$$\hat{f}_{a,i,t\backslash t+1} = f_{a,i,t-1\backslash t} + \hat{f}^{s}_{a,t\backslash t+1} - f_{a,t-1\backslash t} \tag{3.30}$$

which is equal to equation (2.14) in Section 2.[10] Combining these model predictions of the future net inflow or outflow rates with the current gender and age structure of employment, enables us to forecast future net labour market outflow by educational category:

$$\hat{O}_{i,t\backslash t+1} = -\sum_{a} \min\left\{0; \hat{f}_{a,i,t\backslash t+1} L_{a,i,t}\right\} \tag{3.31}$$

---

10. In equation (2.14), the periods (t–n, t) and (t, t+n) were distinguished. As we have annual data available, we here use (t–1, t) and (t, t+1) instead.

For medium-term forecasting, we will first predict the future demographic structure of employment, resulting if employment were to develop equally over time:[11]

$$\begin{aligned}
\hat{L}_{a,i,t+1} &= L_{a,i,t}\left(1+\hat{f}_{a,i,t\backslash t+1}\right) \\
&\vdots \\
\hat{L}_{a,i,t+n} &= \hat{L}_{a,i,t+n-1}\left(1+\hat{f}_{a,i,t+n-1\backslash t+n}\right)
\end{aligned} \tag{3.32}$$

Future net labour market outflow is then:

$$\hat{O}_{i,t\backslash t+n} = \sum_{h=1}^{n}\left(-\sum_{a}\min\left\{0;\hat{L}_{a,t+h}-\hat{L}_{a,t+h-1}\right\}\right) \tag{3.33}$$

Replacement demand is then:

$$\hat{R}D_{i,t\backslash t+n} = \hat{O}_{i,t\backslash t+n} + L_{i,t}\min\left\{0;\left(1+f_{i,t-m\backslash t}\right)^{n}-1\right\} \tag{3.34}$$

with:

$$f_{i,t-m\backslash t} = \sqrt[m]{\frac{L_{i,t}}{L_{i,t-m}}}-1$$

or in words the annual relative employment development during the observation period (*t*–*m*, *t*).

4 Results

The model described in the previous section was estimated for the 113 educational types distinguished by Heijke, Matheeuwsen and Willems (1998). For these estimates, time series were used for the period 1992-1996. As said, separate models have been estimated for males and females, allowing for the fact that the flow patterns of men and women differ.[12] We have excluded those types of education from the model that had fewer than 2,000 workers for the sex concerned. This means that for women, 40 types of education

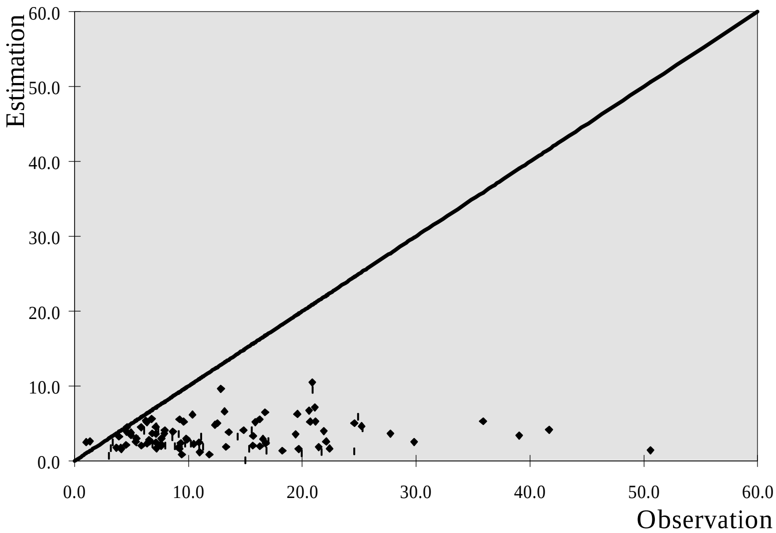
11. The estimated net inflow or outflow rates only distinguish age categories of 5 years (for instance 40-44, 45-49). The net inflow or outflow rate for 42-46 year old people can be approximated by 0.6 times the flow rate for 40-44 year old and 0.4 times the flow rate for 45-49 year old.

12. As the original estimator for Δ was negative definite, we applied the Judge estimator as presented in equation (3.12). It hardly differs from the De Crombrugghe/Dhaene estimator.

are not incorporated in the model. For men, only 12 types of education had to be excluded. For these categories, we assume that the net inflow or outflow percentages are equal to the average for all working persons of that sex for the age group concerned. However, the weight of such a category in the total net outflow for the type of education concerned will be small.[13] This means that the assumption is not very restrictive.

Throughout this section, we will 'predict' the net outflow[14] for the period 1995-1996, using the model estimates based on data for 1992-1996. Comparing these estimated flows with the 'observations' for 1995-1996, derived from the Willems and De Grip model described in Section 2[15], makes it possible to give an evaluation of the quality of the random coefficient approach. In figure 4.1, a graphic presentation is given of the estimated (random coefficient model) and observed (demographic model) net outflow by educational category. For both the observation and the estimation, we have expressed the net outflow as a percentage of the number of workers in 1995. In addition, the 45°-line is drawn in this diagram, which indicates the points with equal observation and estimation.

Figure 4.1
Estimation and observation of net outflow by educational category, 1995-1996 in percentages of employment 1995



Remarkably, for all but two educational categories the estimated value of the net outflow is smaller than the observed net outflow.[16] This seems to indicate that the model developed in

13. Mostly, this refers to educational types that are strongly segregated by gender, such as technical studies (almost all men) or nursing studies (almost all women).

14. In the remaining part of this paper, we will focus on labour market outflow and not on replacement demand.

15. We will refer to this model as the 'demographic model'.

16. These two are IVE electrical engineering and IVE Business Administration. The latter is not a full-time course. For both educational categories the estimated and observed outflow is small.

this paper would underestimate to a high extent the real net outflow (and therefore also the replacement demand). However, if we take a closer look at the observations derived from the demographic model, it can be seen that these values are quite high. For 57 of the 113 educational categories distinguished, the observed net outflow is over 10% in one year. For 22 categories it is even more than 20%. This implies that within only 5 years the entire work force with this educational background will be renewed, which seems to be very unrealistic.

Hence, the demographic model seems to overestimate the annual net outflow.[17] The small sample by education, age category and gender, causes disturbances. Such measurement errors in the stock data obviously result in measurement errors for the flows. In a situation with a stable number of workers, this will always implicate an overestimation of the outflow. An example may illustrate this. Suppose the real number of workers for a certain category is 100 for both 1995 and 1996. If the number of workers in 1996 is incorrectly measured as 90, then we overestimate the outflow by 10. The same occurs if the number of workers in 1995 is incorrectly measured as 110. However, if the opposite is the case – the number of workers is measured as 110 in 1996 or 90 in 1995 – then the outflow is measured correctly (0).

To further analyse the difference between the observations and estimates of the net outflow per type of education, we first look at the total net outflow, aggregated by level of education. These results are presented in Table 4.1. The average net labour market outflow, directly computed from the data, is 7.9% for the period 1995-1996. The model estimates a net outflow of 3.4% in this period, which is, as noted above, a remarkable difference.

Table 4.1
Estimation and observation of net outflow by level of education, 1995-1996, in percentages of employment 1995

Level of education	estimation %	observation %
Primary education	5.4	6.3
PVE, LGSE	3.8	7.8
IVE, HGSE	2.8	6.2
HVE	3.3	11.5
AE	3.1	12.0
Total	3.4	7.9

PVE Preparatory Vocational Education

17. The annual observed outflow rates, derived from the demographic model, are much higher than those reported by Willems and De Grip (1993). The reasons for this is that in the latter paper, the overestimation of flows is averaged over 5 years, resulting in smaller annual overestimation.

LGSE	Lower General Secondary Education
IVE	Intermediate Vocational Education
HGSE	Higher General Secondary Education
HVE	Higher Vocational Education
AE	Academic Education

As might expected from the results presented in Figure 4.1, the differences between estimated and observed labour market outflows can be seen for all educational levels. However, the degree of difference is not equal over the levels of education. For primary education, the estimated and observed outflows are relatively close. On the other hand, the observed outflow for HVE and UE is about four times the estimated outflow. As the higher types of education on average employ fewer workers, resulting in less reliable data, this confirms the hypothesis of measurement error in the observed outflow.

Although it could be argued that the observed outflow is too high, this does not guarantee that the model estimations of net outflow are closer to reality. We have some points of reference, however. Firstly, we have also observed the total net labour market outflow (not distinguished by educational category). This is the f -vector in equation (3.18). For the period 1995-1996, this is 2.2%, 1.2 percentage point less than the average estimated outflow. There are, however, two explanations for this difference. The first is that the net outflow for the total work force is also measured by the cohort components method. This can be explained by the fact that, apart from sample errors, the measurement of flows is better if we distinguish more categories, as compensating flows are not measured in this way. In our case, this implies that the inflow of 25-year old individuals with an academic degree can compensate the outflow of 25-29 year old persons with only primary education. If we distinguish between educational categories these flows are observed. Secondly, part of the difference may be the result of the fact that workers obtain a different educational background during their working careers. This further training or re-training will be measured as inflow for the one educational category and outflow for the other.

A second point of reference is the directly observed outflow from the work force. Statistics Netherlands recently started publishing some selected figures of flows on the labour market. For the total work force, the (*gross*) outflow counts 5.7% for the period 1995-1996 (CBS, 1997). The difference with our estimate of the *net* labour market outflow could refer to the number of re-entrants to the labour market as far as this inflow compensates the measured outflow of workers. The variation in the estimates of labour market flows makes the availability of real and reliable data on labour market dynamics more urgent. However, we may conclude that the random coefficient model results in sufficiently reliable estimates of *net* labour market outflow.

Given this interpretation of the estimated net flow percentages, we will look at the results in greater detail. Table 4.2 gives an overview of the estimated and observed inflow and outflow percentages by age category and gender. As in Table 4.1, the averages by level of education are presented. Again we see that the variances between the estimated flow rates are much smaller than the variances between the observed flows. Obviously this is the result of the random coefficient approach.

When we look at the estimated flow percentages, the table shows some clear, although not very surprising results. In general, men enter the labour market until the age of 35 and start leaving the work force from about this age. This holds for all educational levels. Remarkably, men in the age of 40-44 on balance enter the work force. It is not clear what the reason for this is.[18] The men's flow balance differs considerably by level of education. The net labour market inflow of 25-29 year old men increases with the level of education. Especially the higher educated frequently enter the work force at this age, obviously due to the longer educational routes. Regarding the outflow of men, we also see different patterns for the educational levels. Workers with primary education or PVE/LGSE often leave the labour market relatively early (in the ages 35-49). For men with only primary education the outflow of 50-54 year old is also relatively high. These higher outflow rates are probably related to the negative employment development for most of these educational categories. Lastly it can be observed that the outflow of the elderly is relatively large for all educational levels. These figures reflect the low participation rates for the Netherlands.

The net inflow and outflow pattern of women is somewhat different. Lower educated women withdraw from the labour market at relatively young ages. For the age categories 30-34 and 35-39, we observe a net inflow of women. This is due to the re-entrance of women onto the labour market. The inflow of re-entrants probably compensates the outflow of women with the same ages. Remarkably, we do not observe a net outflow of young women with only primary education.[19] Probably the outflow of women with this educational background happens before the age of 25 and will then be compensated by late labour market inflow (or early re-entrants). The second peak in the outflow rates from the age of 45 seems to be somewhat higher for women than for men. Apparently, women leave the labour market earlier, even if they stayed on the labour market until the age of 45 or have re-entered the labour market. The outflow rates of 55-59 year old women seem to be a little lower than the corresponding outflow percentages for men. However, one has to realise that most women leave the labour market before that, resulting in a selected group that stays in the labour market.

18. This is, however, not a stable result over the years. For 1992-1993, 1993-1994 and 1994-1995, we do not observe such net inflow for this age category. However, for 1992-1993 we see a small (0.1%) net inflow of 50-54 year old men; for 1993-1994 and 1994-1995 a net inflow of 45-49 year old men of about 1% is observed.

19. For 1992-1993 we do observe a net outflow of women with this age.

Table 4.2
Average estimated (E) and observed (O) net inflow and outflow by level of education, age category and gender, 1995-1996 in percentages of employment 1995

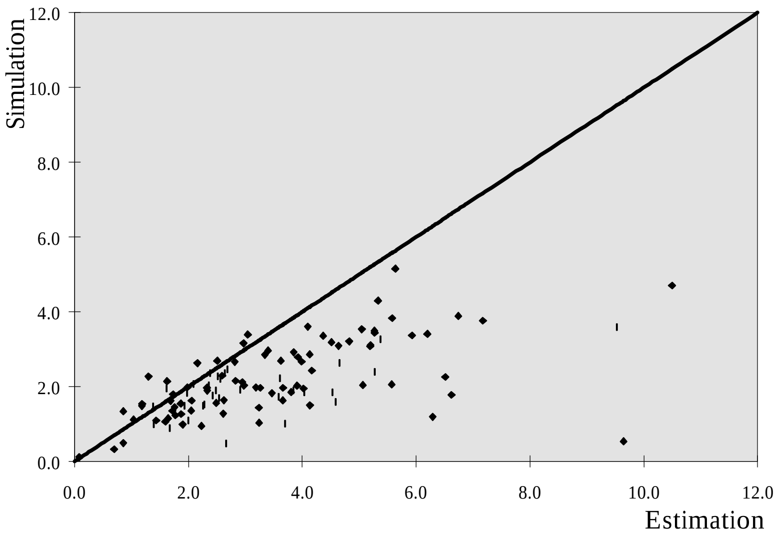
Level of education		25-29 %	30-34 %	35-39 %	40-44 %	45-49 %	50-54 %	55-59 %	60-64 %
Men									
Primary education	E	0.9	1.3	-6.9	-0.7	-6.9	-4.6	-21.8	-34.9
	O	7.0	3.4	-6.0	-9.6	-7.4	-1.3	-26.7	-50.3
PVE, LGSE	E	1.0	1.6	-3.6	1.3	-5.6	-1.2	-27.0	-27.5
	O	-6.7	2.5	-3.7	10.0	-9.3	12.2	-16.2	-45.8
IVE, HGSE	E	4.6	1.3	-1.6	1.0	-2.4	-2.2	-22.3	-30.1
	O	5.2	-0.7	-2.4	2.8	-0.9	-5.3	-18.1	-12.6
HVE	E	10.0	3.2	-1.5	0.7	-0.7	-0.5	-25.2	-34.5
	O	15.0	2.4	3.2	-2.8	0.2	-6.1	-11.8	-36.5
UE	E	15.8	3.4	-3.7	2.1	-3.3	-0.9	-23.0	-32.3
	O	26.1	15.1	-9.2	-4.8	-3.9	-8.2	1.1	23.7
Total	E	5.2	1.9	-2.6	0.9	-3.4	-1.9	-23.7	-30.9
	O	5.8	2.2	-2.6	1.0	-3.4	-1.9	-16.5	-30.0
Women									
Primary education	E	3.9	1.9	-1.1	-2.2	-9.1	-5.6	-19.7	-40.8
	O	2.0	9.9	11.9	9.5	-5.9	5.1	-29.3	-55.8
PVE, LGSE	E	-4.3	2.1	2.9	-2.4	-6.9	-10.4	-11.2	-36.6
	O	-9.5	1.4	-3.0	-8.9	-10.9	-23.3	6.2	-31.3
IVE, HGSE	E	-1.3	4.8	2.3	3.3	-2.3	-7.1	-21.2	-42.2
	O	1.6	0.8	0.8	5.4	0.5	0.6	-21.4	-31.3
HVE	E	3.8	0.5	2.1	2.9	-2.3	-8.3	-20.8	-38.1
	O	-4.9	-0.8	7.7	6.4	-7.5	-12.0	-11.7	-33.1
UE	E	12.0	3.0	1.7	1.4	-5.0	-4.6	-17.9	-36.4
	O	28.2	25.0	-4.5	-3.3	-11.6	-10.4	-37.4	-82.2
Total	E	0.5	3.1	2.1	1.2	-4.3	-8.0	-17.8	-39.2
	O	0.2	2.8	2.1	1.9	-5.5	-9.4	-13.4	-39.5

Demographic effects or flow effects?

Differences in the net outflow of workers can essentially be the result of two effects. First, the *demographic effect*: older workers withdraw from the labour market and youngsters enter the labour market. Moreover, we have seen differences in flow patterns between men and women. Secondly, we can mention an *educational effect*. Apart from differences in the demographic structure of the work force, labour market flows may differ between educational categories.

In the remaining part of this section, we will examine both effects. It is important because it can provide more insight in the causes of the labour market flows. This may help forecasting future flows and therefore future replacement demand. In the extreme case that labour market flows are only determined by demographic factors, it makes no sense to specify a flow model that distinguishes a large number of educational categories. However, if educational effects are more important, such distinctions should be made. Although labour supply behaviour differs over educational categories, we do not expect that this – corrected for demographic differences – will lead to a considerable variation in labour market flows. Particularly if we incorporate differences between levels of education (in other words taking into account that the degree of participation is generally higher for the higher educated), the remaining variance, *i.e.* an educational field effect, suggests that labour market flows are affected by demand factors.

Figure 4.2
Simulated net outflow by educational category with average net inflow and outflow coefficients, compared to estimated outflow, 1995-1996 in percentages of employment 1995

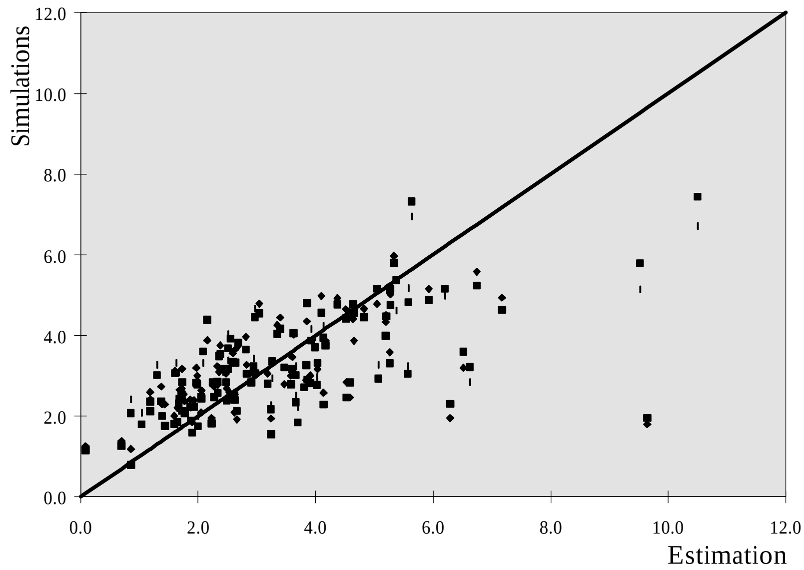


One way of separating demographic effects from educational effects is by applying the average net inflow and outflow percentages per age category and gender to all educational categories distinguished. Figure 4.2 gives the results, comparing the simulated outflow with the estimated outflow as presented in Figure 4.1. The figure shows that for most educational categories the simulated outflow is smaller than the 'real' estimated outflow. This indicates there is an important educational effect.

For only 15 of the 113 types of education the simulation shows a higher outflow. Moreover, for these 15 categories the differences are small, whereas for the other educational types sometimes substantial differences occur. The fact that the simulated outflow is generally smaller than the estimated flows is quite obvious. In the average flow coefficients, which are the input for the simulation, inflow and outflow can compensate. As argued above, if an individual with a certain age (category) and sex and a specific educational background leaves the work force and at the same time another person with the same age and sex, but a different educational background enters the labour market, this is not observed in the average outflow rate as adopted in the simulation.

Although the example clearly shows the relevancy of educational effects and the importance of distinguishing educational categories in the measurement of flows (when adopting the cohort components method), a clearer simulation would be to abstract from these compensating flows. This is possible by averaging over the net *out*flow for each age category and gender. The result is presented with a "   " in Figure 4.3. Again a comparison is made with the estimated flows. The figure shows that there is still an important educational effect. The average simulated outflow is now equal to the average estimated outflow. However, a considerable dispersion exists between the single values.

Figure 4.3
Simulated net outflow by educational category with average net outflow coefficients, compared to estimated outflow, 1995-1996 in percentages of employment 1995



We have also carried out a simulation by averaging the outflow rates by level of education, in other words in this simulation relative outflow by demographic group is equal for all educational categories within a level of education. This should to a certain extent correct for the differences in labour supply behaviour between the educational levels. As we have seen, higher educated people enter the labour market later, due to their longer stay in the educational system. Lower educated people generally leave the labour market more early.[20]

20. It has to be noted that this is not necessarily a supply effect.

The results of this simulation are also presented in Figure 4.3, indicated with a “ ”. Remarkably the results are not very different from the previous simulation, again implying that apart from the demographic effect a serious educational effect determines the outflow of workers.

5 Conclusions

In this paper, we have refined the demographic model of Willems and De Grip (1993) for measuring labour market flows. An important drawback of the Willems and De Grip model is that it needs the determination of numerous flow coefficients, based on observations at two different times and at a detailed level. This implies that there is a great chance of measurement errors.

A solution to this problem is found in the application of a random coefficient estimation technique to the flows as observed with the demographic model. This enables us to make a statistically founded balance of the observed flows by educational category and the average for the whole labour force. We concluded that this method is sufficiently reliable, although it has to be noted that – as a consequence of the demographic model – only *net* flows can be measured. This means that the compensating inflow of (re-)entrants and the outflow of workers are not observed. Recently published flow figures for the total work force indicate that this leads to an underestimation of the total outflow of about 2 percentage points.

With the random coefficient model, estimates of the net inflow or outflow by educational type, age category, and gender can be obtained. In general, men start leaving the work force from the age of 35, with a major outflow from about 55 years old. Lower educated women start leaving the labour market before their 30s, undoubtedly because of the birth and care of their children. A second peak in the outflow of women starts from the age of 45, where the outflow of 50-54 year old women is also relatively higher than the outflow of men at that age.

We concluded this paper by examining whether the labour market outflow is mainly due to demographic reasons, *i.e.* the demographic composition of the work force, or whether educational effects can also be observed. Although demographic reasons are undoubtedly relevant, it can be concluded that educational differences are important too. Several explanations can be given for this. First, an individual’s labour supply behaviour differs over educational categories due to the sunk costs in human capital investments. However, if we correct for the level of education – which, apart from demographic variation, seems the most important source of differences in labour supply – the educational effect remains. This suggests that labour demand behaviour is also relevant. If fewer workers with a certain educational background are needed in the labour market, this probably results in a higher outflow of the current stock of workers with this background. The higher outflow rates for the lower educational categories, which generally face decreasing employment levels, points in that direction. However, such changing employment needs could also result in a lower inflow of youngsters. This is a point of attention for further research.

References

Borghans, L., H. Heijke (1994), *Een random coëfficiëntenmodel voor het voorspellen van de beroepenstructuur van bedrijfstakken*, ROA-W-1994/1, Maastricht.

CBS (1997), *Sociaal-Economische Dynamiek 1997*, Voorburg/Heerlen.

Crombrugghe, D. de, G. Dhaene (1991), *Constrained Estimation of the Covariance Matrix of Random Coefficients*, Research Memorandum 91-025, Rijksuniversiteit Limburg, Maastricht.

Grip, A. de, P. Meijboom, E. Willems (1994), Vacancies, Employment Growth and the Demand for Newcomers on the Labour Market, in: J. Muijsken (ed.), *Measurement and Analysis of Job Vacancies*, Avebury Publishers, Aldershot, pp. 101-125.

Heijke, J.A.M., A. de Grip, *The Information System on Education and the Labour Market*, ROA-W-1991/2E, Maastricht.

Heijke, H., A. Matheeuwsen, E. Willems (1998), *Clustering Educational Categories in a Heterogeneous Labour Market*, ROA-RM-1998/2^{E}, Maastricht.

Judge, G.G., W.E. Grifiths, R. Carter Hill, T-C. Lee (1980), *The Theory and Practice of Econometrics*, John Wiley & Sons, New York.

ROA (1995a), *The Labour Market by Education and Occupation to 2000*, ROA-R-1995/3E, Maastricht.

ROA (1995b), *The Labour Market by Education and Occupation to 2000. Statistical Appendix*, ROA-R-1995/3BE, Maastricht.

Shryock, H.S., J.S. Siegel (1980), *The Methods and Materials of Demography*, U.S. Bureau of the Census, U.S. Government Printing Office, Fourth Printing (rev.), Washington D.C.

Swamy, P.A.V.B. (1970), Efficient Inference in a Random Coefficient Regression Model, *Econometrica* 38, pp. 311-323.

Swamy, P.A.V.B. (1971), *Statistical Inference in Random Coefficient Models*, New York.

Willems, E. (1996), *Manpower Forecasting and Modelling Replacement Demand: an Overview*, ROA-W-1996/4E, Maastricht.

Willems, E.J.T.A., A. de Grip, (1993) Forecasting Replacement Demand by Occupation and Education, *International Journal of Forecasting,* vol. 9, pp. 173-185.