JOURNAL OF APPLIED ECONOMETRICS
J. Appl. Econ. **16**: 599–617 (2001)
DOI: 10.1002/jae.620

INCOME DISTRIBUTION AND INCOME DYNAMICS IN THE UNITED KINGDOM

JAYASRI DUTTA,[a] J. A. SEFTON[b]* AND M. R. WEALE[b]
[a] *Faculty of Economics and Politics, Sidgwick Avenue, Cambridge CB3 9DD, UK*
[b] *National Institute of Economic and Social Research, 2 Dean Trench Street, London SW1P 3HE, UK*

SUMMARY

In this paper, we propose a model of income dynamics which takes account of mobility both within and between jobs. The model is a hybrid of the mover-stayer model of income dynamics and a geometric random walk. In any period, individuals face a discrete probability of 'moving', in which case their income is a random drawn from a stationary recurrent distribution. Otherwise, they 'stay' and incomes follow a geometric random walk. The model is estimated on income transition data for the United Kingdom from the British Household Panel Survey (BHPS) and provides a good explanation of observed non-linearities in income dynamics. The steady-state distribution of the model provides a good fit for the observed, cross-sectional distribution of earnings. We also evaluate the impact of tertiary education on income transitions and on the long-run distribution of incomes. Copyright © 2001 John Wiley & Sons, Ltd.

1. INTRODUCTION

In this paper, we propose a new model of income dynamics which takes account of income mobility both within and between jobs. The model is estimated on income transition data for the United Kingdom available from the British Household Panel Survey (BHPS). We show that this model provides an explanation for the observed non-linearities in income dynamics. Expected future incomes increase with current incomes at an increasing rate, and variance of future incomes is a U-shaped function of current incomes. We compute the steady-state distribution of the model and compare its fit of the observed, cross-sectional distribution of earnings with the current best-practice parametric model.

The incomes of individuals may show considerable variation from one year to the next. At the same time, the current income of an individual is likely to be an important predictor of her future income for a variety of reasons. A reasonable model of the income distribution should be consistent with a theory of income dynamics, which allows for the sort of stochastic variation usually observed. The modelling of income dynamics is often done in the context of linear autoregressive models, e.g. by Atkinson, Bourguignon and Morrisson (1992). These are particularly useful to measure the impact of economic variables, and have the useful property of being preserved under aggregation. An alternative approach (Champernowne, 1953; Hart, 1976) models income transitions in terms of Markov chains. This method is often model-free, and mainly intended to summarize the evidence on the frequency with which individuals move across income groups. Our approach draws on both, and on the statistical model of *movers and stayers*, (Goodman, 1961;

* Correspondence to: Dr. J. A. Sefton, National Institute of Economic and Social Research, 2 Dean Street, London SW1P 3HE, UK. E-mail: jsefton@niess.ac.uk
Contract/grant sponsor: ESRC; Contract/grant number: R000237688.

Received 16 January 1998
Revised 24 October 2000

Frydman, 1984; Sampson, 1990). Indeed, in our framework both a linear switching model (**L**), and the mover–stayer model (**M**) appear as special cases, and can be written as testable parametric restrictions.

The dynamic model is explained quite simply, as follows. For many people, incomes in one year are likely to be closely related to those in the previous year. These are the stayers. However, some people face the prospect of a 'jump' : a substantial disruption to their earning process or other sources of income. It may be useful to think of this as moving across jobs or even occupations, and we refer to these individuals as movers. Individuals may move, voluntarily or otherwise. The change in incomes of movers is likely to be rather different from that of stayers, even though it is unlikely to be fully predictable in either case. The *mover–stayer* model builds on this hypothesis, with the additional restriction that the income of stayers is unchanged. We do not impose such a restriction, and propose a parametric model for the distribution of incomes of movers as well as stayers. We assume that a mover's income is drawn randomly from a time-invariant distribution: call this the recurrent income distribution. The result is a mixed model of income dynamics, whose parameters can be estimated directly from data on income transitions.

The widening earnings distribution in both the UK and the US over the past twenty years has reinvigorated the research effort into modelling the dynamics of individual and household incomes. Gottschalk and Moffitt (1994) on US data, and Dickens (1997) on UK data have aimed to decompose the changes in an individual's earnings into a transitory and a permanent component. The permanent component is associated with a random walk process, and the transitory with a stationary one. However, it is often difficult to estimate these type of models on short data histories. In an alternative approach, Atkinson *et al.* (1992) used a simple first-order autoregressive process to model income dynamics in a number of developed economies. They found a degree of mean reversion in all these countries in the income dynamic process. We will later compare the explanatory power of our maintained model with this AR model. Alternatively on US data Abowd and Card (1989) looked at modelling the changes in annual earnings and hours worked, rather than levels. They found that a moving average error process of order at most two adequately described the income data.

The distribution of income has been studied extensively, e.g. by Atkinson (1970) and Atkinson, Rainwater and Smeeding (1995) among others. Typically, the log-normal distribution provides a reasonable fit. However, it fails to match the upper tail, as the proportion of the population earning relatively high incomes is larger than the distribution predicts. Recently, Majumder and Chakravarty (1990) and McDonald and Mantrala (1995), building on the work of Fisk (1961) suggest that the generalized β-distribution provides a better representation. The log-normal distribution is known to be the steady-state distribution generated when the logarithm of individual incomes follow a linear dynamic process with normally distributed innovations. The generalized β, and competing models, summarize the distribution of incomes at a point in time, but have little to say about their dynamic behaviour: there is no known dynamic process which yields these as the steady-state distribution of incomes.

We define the transition model in Section 2, and evaluate its implications for the long-run distribution of incomes in Section 3. In Section 4, we summarize the empirical properties of income transitions in the data. Section 5 reports maximum likelihood estimates with various restrictions. We estimate the income transition parameters for individuals with and without higher education, and can so evaluate the effect of education on income uncertainty. In Section 6, we evaluate the properties of the estimated model, including the predicted steady-state distributions. Section 7 presents conclusions.

2. THE MODEL

Let $Y_{i,t}$ be the income of individual i at time t, and $y_{i,t} = \log Y_{i,t}$. Define a binary variable $m_{i,t} \in \{0, 1\}$ where $m_{i,t} = 1$ denotes the event that individual i 'moves', or changes income state between periods t and $t + 1$, and $m_{i,t} = 0$ denotes the event that this individual 'stays' in the same income state. Further, the income of a stayer is equal to previous income times a random shock; the income of a mover is a random draw from a fixed distribution. Thus:

$$y_{i,t+1} = y_{i,t} + \varepsilon_{i,t} \quad \text{if } m_{i,t} = 0$$
$$y_{i,t+1} = \tilde{y}_{i,t} \quad \text{if } m_{i,t} = 1$$

where $\varepsilon_{i,t}$, $\tilde{y}_{i,t}$ are independent draws from

$$\varepsilon_{i,t} \sim N(\mu, \sigma^2)$$
$$\tilde{y}_{i,t} \sim N(\tilde{\mu}, \tilde{\sigma}^2)$$

We refer to the distribution of $\tilde{y}$ as the recurrent distribution. With the additional assumption that $\varepsilon_{i,t}$, $\tilde{y}_{i,t}$ are independent of current and past incomes, this provides a complete specification of the distribution of $y_{i,t+1}$ conditional on $y_{i,t}$ and $m_{i,t}$. We need to augment this with a model of the distribution of $m_{i,t}$ conditional on $y_{i,t}$, in the form of probabilities $\theta(y) = \Pr(m_{i,t} = 0 | y_{i,t} = y)$.

The conditional distribution of $y_{i,t+1}$ given $y_{i,t}$ is fully specified by the parameters $\mu, \sigma, \tilde{\mu}, \tilde{\sigma}$ and a function $\theta(y)$ such that $0 \leq \theta(y) \leq 1$. The conditional probability distribution of $y_{i,t+1}$ given $y_{i,t}$ is

$$\Pr(y_{i,t+1} \leq y | y_{i,t}) = \theta(y_{i,t})\Phi\left(\frac{y - y_{i,t} - \mu}{\sigma}\right) + (1 - \theta(y_{i,t}))\Phi\left(\frac{y - \tilde{\mu}}{\tilde{\sigma}}\right)$$

where $\Phi(\cdot)$ is the cumulative density function of the normal distribution. Notice that $\theta(y) < 1$ implies that incomes follow a (geometric) random walk for a random length of time. One obtains the mean and variance of the conditional distribution as

$$m(y_{i,t}) \equiv E(y_{i,t+1} | y_{i,t}) = \theta(y_{i,t})(y_{i,t} + \mu) + (1 - \theta(y_{i,t}))\tilde{\mu}$$
$$v(y_{i,t}) \equiv \text{var}(y_{i,t+1} | y_{i,t}) = \theta(y_{i,t})\sigma^2 + (1 - \theta(y_{i,t}))\tilde{\sigma}^2 + \theta(y_{i,t})(1 - \theta(y_{i,t}))(y_{i,t} + \mu - \tilde{\mu})^2$$

We consider a sequence of restrictions on the function $\theta(y)$ which are useful for estimation and inference in this framework. The first restriction, which we maintain as a hypothesis, is particularly useful for grouped data.

Let $G_j; j = 1, \ldots, N$ be N income groups with end-points x_j, i.e.

$$y_{i,t} \in G_j \quad \text{whenever } x_{j-1} \leq y_{i,t} < x_j$$

with $x_0 = -\infty < x_1 < \cdots x_{N-1} < x_N = \infty$.

Restriction 1 (Grouping) The model satisfies restriction (**G**) if, and only if, function $\theta(y)$ is a step function:

$$(\mathbf{G}) y_{i,t} \in G_j \Rightarrow \theta(y) = \theta_j \in [0, 1]$$

for $j = 1, \ldots, N$.

The restriction asserts that the probability of moving is constant within these known income bands. The vector $\Theta = (\theta_1, \ldots, \theta_N)$ summarizes all relevant information about the conditional distribution of $m_{i,t}$. The problem of estimating the optimal number of income bands and break points, the multiple change-point problem, is known to be difficult (Lombard, 1987). We preferred therefore to choose a partition and test for sensitivity of the estimates to this chosen partition. The distribution of $y_{i,t+1}$ conditional on $y_{i,t} \in G_j$ is

$$y_{i,t+1} \sim N(y_{i,t} + \mu, \sigma^2) \text{ with probability } \theta_j$$
$$y_{i,t+1} \sim N(\tilde{\mu}, \tilde{\sigma}^2) \text{ with probability } (1 - \theta_j).$$

This model has $N + 4$ unknown parameters

$$\phi = (\mu, \tilde{\mu}, \sigma^2, \tilde{\sigma}^2, \Theta); \quad \sigma^2, \tilde{\sigma}^2 \geq 0; \Theta \in [0, 1]^N$$

It is useful to note that (**G**) specifies that the conditional expectation $m_G(y_t) = E y_{t+1} | y_t$ is a piecewise linear function. The next restriction is equivalent to the condition that $m(y)$ is linear.

Restriction 2 (Linearity) The model satisfies restriction (**L**) if, and only if, the function $\theta(y)$ is constant, i.e.

$$(\mathbf{L}) \quad \theta(y) = \theta \text{ for all } y; 0 \leq \theta \leq 1$$

This restriction implies

$$m_L(y_{i,t}) = (\theta\mu + (1 - \theta)\tilde{\mu}) + \theta y_{i,t} \tag{1}$$
$$v_L(y_{it}) = \theta\sigma^2 + (1 - \theta)\tilde{\sigma}^2 + \theta(1 - \theta)(y_{i,t} + \mu - \tilde{\mu})^2 \tag{2}$$

The conditional expectation is linear, and the conditional variance is quadratic in current income. This restriction is testable, once the probability function $\theta(y)$ is suitably specified. Specifically, if (**G**) is the maintained hypothesis, the restriction (**L**) imposes $N - 1$ restrictions.

A quite different type of restriction yields the traditional mover-stayer model, which assumes that $y_{i,t+1} = y_{i,t}$ if $m_{i,t} = 0$.

Restriction 3 (Mover–Stayer) The model satisfies restriction (**M**) if and only if

$$(\mathbf{M}) \quad \mu = \sigma^2 = 0$$

We note that this restriction implies

$$m_M(y_{it}) = \theta(y_{i,t})y_{i,t} + (1 - \theta(y_{i,t}))\tilde{\mu}$$
$$v_M(y_{i,t}) = (1 - \theta(y_{i,t}))\tilde{\sigma}^2 + \theta(y_{i,t})(1 - \theta(y_{i,t}))(y_{i,t} - \tilde{\mu})^2$$

3. THE STEADY-STATE DISTRIBUTION

We have described a model which specifies the probability distribution of $y_{i,t+1}$ conditional on $y_{i,t}$, as $\Pr(y_{i,t+1}|y_{i,t}, \phi)$. The result is a non-linear first-order Markov process. The long-run properties of this process are described by the invariant, or steady-state distribution.

The long-run distribution is of interest for more than one reason. As we have mentioned, the cross-sectional distribution of income, and its changes, are of direct interest to economists. It is then useful to ask whether our model is able to explain the shape of the distribution: the relevant object for comparisons is precisely the steady-state distribution. Our approach also allows for a decomposition of factors affecting this distribution into static and dynamic components. Specifically, changes in the mobility pattern—as measured by $\theta(y)$—are likely to affect the inequality of incomes. An increase in inequality due to changes in mobility is likely to have very different welfare implications from one arising from, say, an increase in the variance of the recurrent income distribution $\tilde{\sigma}^2$.

The invariant distribution, $F_*(y)$, satisfies the functional equation

$$F_*(y) = \int_{-\infty}^{\infty} \Pr(y|x, \phi)\mathrm{d}F_*(x) \quad \text{for all } y \in \mathbf{IR}$$

specialized to our model, this is

$$(S) \quad F_*(y) = \int_{-\infty}^{\infty} \theta(x)\Phi\left(\frac{y-x-\mu}{\sigma}\right)\mathrm{d}F_*(x) + \Phi\left(\frac{y-\tilde{\mu}}{\tilde{\sigma}}\right)\int_{-\infty}^{\infty}(1-\theta(x))\mathrm{d}F_*(x)$$

where $\Phi(\cdot)$ is the normal distribution function. The existence, and uniqueness, of such a distribution can be established by appeal to properties of Markov processes. We are particularly interested in the characteristics of this distribution.

Theorem 1 Suppose the model satisfies restriction (**G**). A steady-state distribution F_* exists, and is unique, whenever $\max_{j=1,\ldots,N}\theta_j < 1$, and $\tilde{\sigma}^2 > 0$.

Proof: We appeal to Theorem 11.12 of Stokey and Lucas (1989), noting that $\Pr(y_{i,t+1} \in G_k | y_{i,t} \in G_j) \in (0, 1)$ for each $j, k \in \{1, \ldots, N\}$ whenever $\theta_j < 1$ and $\tilde{\sigma}^2 > 0$. This ensures their condition (M), and guarantees the existence, and uniqueness, of the distribution F_*. □

Unfortunately, the steady-state distribution cannot be solved for analytically in the general case. If, however, either condition (**L**) or condition (**M**) hold, the long-run distribution can be completely characterized, and we now do so.

3.1. Linearity

Suppose, first, that condition (**L**) holds. Equation (S) rewrites as

$$(S_L) \quad F_{*L}(y) = \theta \int \Phi\left(\frac{y-x-\mu}{\sigma}\right)\mathrm{d}F_{*L}(x) + (1-\theta)\Phi\left(\frac{y-\tilde{\mu}}{\tilde{\sigma}}\right)$$

The following theorem states F_{*L}.

Theorem 2 Suppose $\theta(y) = \theta \in (0, 1)$ for all y. The invariant distribution of y is F_{*L}, where

$$F_{*L}(y) = (1-\theta)\sum_{t=0}^{\infty}\theta^t\Phi\left(\frac{y-\tilde{\mu}-t\mu}{\tilde{\sigma}^2 + t(\sigma^2)^{1/2}}\right)$$

Proof: It suffices to check that F_{*L} satisfies (S_L). This is established by the fact that the normal distribution is additive, i.e.

$$\int \Phi\left(\frac{a-x}{s}\right) \mathrm{d}\Phi\left(\frac{x-b}{s'}\right) = \Phi\left(\frac{a-b}{(s^2+s'^2)^{1/2}}\right)$$

this implies

$$\int \Phi\left(\frac{y-x-\mu}{\sigma}\right) \mathrm{d}F_{*L}(x) = (1-\theta)\sum_{t=0}^{\infty}\theta^t \Phi\left(\frac{y-\tilde{\mu}-(t+1)\mu}{(\tilde{\sigma}^2+(t+1)\sigma^2)^{1/2}}\right)$$

Substitution yields the result. □

The invariant distribution F_{*L} is a geometric mixture of normal distributions. An alternative and constructive method for deriving the steady-state distribution is helpful in understanding its properties. First, note that an individual moves with probability $1-\theta$ in any period, independently of income and of past history of movements. The probability that she will stay for exactly n periods is $\theta^{n-1}(1-\theta)$, independently of income and calendar time. Her income, Y_n, is log-normal conditional on n:

$$y_n | n \sim N(\tilde{\mu} + (n-1)\mu, \quad \tilde{\sigma}^2 + (n-1)\sigma^2)$$

The distribution F_{*L} obtains by integrating over the geometric distribution of n.

We can now state some further properties of F_{*L}.

Corollary 1 The moment-generating function of F_{*L} is

$$M_{*L}(\tau) = E_{*L}\exp(\tau y) = (1-\theta)\frac{\exp\left(\tilde{\mu}\tau + \dfrac{\tilde{\sigma}^2}{2}\tau^2\right)}{1-\theta\exp\left(\mu\tau + \dfrac{\sigma^2}{2}\tau^2\right)}$$

Further,

$$m_{*L} = E_{*L}y = \tilde{\mu} + \frac{\theta}{1-\theta}\mu$$

$$v_{*L} = \mathrm{var}_{*L}(y) = \tilde{\sigma}^2 + \frac{\theta}{1-\theta}\sigma^2 + \frac{\theta}{(1-\theta)^2}\tilde{\mu}^2$$

An increase in mobility corresponds to a decrease in θ. We note that increased mobility is equalizing in the sense that it necessarily reduces the variance of the steady-state distribution (as $\partial v_{*L}/\partial\theta > 0$). Its effect on average incomes depends on the sign of μ (i.e. on the expected growth rate of stayer's income): with $\mu > 0$, as we estimate, mobility decreases long-run average incomes.

The variance v_* is increasing in $\tilde{\mu}$ as well $\tilde{\sigma}$. An upward shift in the recurrent distribution necessarily increases inequality in the long run. This increase ($(1+g)\tilde{y}$, say, with positive g) will typically increase individual welfares, unlike a decline in mobility.

Further properties of F_{*L}, such as skewness and kurtosis, can be worked out from the moment-generating function. We demonstrate them in the special case $\mu = 0$, which simplifies derivation. In this case,

$$E_{*L}(y - E_{*L}(y))^3 = \frac{\theta}{1-\theta}\tilde{\mu}\tilde{\sigma}^2$$

$$E_{*L}(y - E_{*L}(y))^4 = 3(\mathrm{var}_{*L}(y))^2 + \frac{\theta\sigma^2}{(1-\theta)^2}(3\sigma^2 - 2(1-\theta)(\tilde{\sigma}^2 + \tilde{\mu}^2)$$

Thus, the steady-state distribution of log income is typically asymmetric; and it has thicker tails than the normal distribution if σ^2 is sufficiently large: this property is widely observed in data.

3.2. The Mover–stayer Model

Recall that condition (**M**) is $\mu = \sigma = 0$. Specializing the defining equation (S) yields

$$(S_M)F_{*M}(y) = \int_{-\infty}^{y}\theta(x)\mathrm{d}F_{*M}(x) + \Phi\left(\frac{y-\tilde{\mu}}{\tilde{\sigma}}\right)\int_{-\infty}^{\infty}(1-\theta(x))\mathrm{d}F_{*M}(x)$$

We state the distribution in the next theorem.

Theorem 3 Let $\mu = \sigma = 0$, and $0 \le \theta(y) < 1$ for $y \in \mathbf{IR}$. The invariant distribution of y is

$$F_{*M}(y) = \frac{\displaystyle\int_{-\infty}^{y}\frac{\phi\left(\frac{x-\tilde{\mu}}{\tilde{\sigma}}\right)}{1-\theta(x)}\mathrm{d}x}{\displaystyle\int_{-\infty}^{\infty}\frac{\phi\left(\frac{x-\tilde{\mu}}{\tilde{\sigma}}\right)}{1-\theta(x)}\mathrm{d}x}$$

where

$$\phi(\mathrm{z}) = \frac{\exp -\dfrac{x^2}{2}}{(2\pi\sigma^2)^{1/2}}$$

is the normal density function.

Proof: We note that

$$(S_M) \Rightarrow \mathrm{d}F_{*M}(y) = \theta(y)\mathrm{d}F_{*M}(y) + \phi\left(\frac{y-\tilde{\mu}}{\tilde{\sigma}}\right)C$$

where C is a constant. It follows that

$$\mathrm{d}F_{*M}(y) = \mathrm{C}\frac{\phi\left(\frac{y-\tilde{\mu}}{\tilde{\sigma}}\right)}{1-\theta(y)}$$

the stated result obtains by integration. □

This distribution can be evaluated for any (bounded) function $\theta(y)$. The step-function formulation corresponding to (**G**) yields a distribution for grouped data, which we state as Corollary 2. Let $G_i = (x_{i-1}, x_i]$, as before. Define

$$F_z(i) = \Pr(z \in G_i) = \Phi\left(\frac{x_i - \tilde{\mu}}{\tilde{\sigma}}\right) - \Phi\left(\frac{x_{i-1} - \tilde{\mu}}{\tilde{\sigma}}\right)$$

Corollary 2 If restrictions (**G**) and (**M**) hold, and $\max_{i=1,\ldots,N} \theta_i < 1$, the invariant distribution of y satisfies

$$F_{*M}(i) = \frac{\dfrac{F_z(i)}{1-\theta_i}}{\displaystyle\sum_{i=1}^{N} \frac{F_z(i)}{1-\theta_i}}$$

4. THE DATA

Our data sample is taken from the first five waves of the British Household Panel survey (BHPS). This panel survey, which began in 1991, collects data among other things on the incomes and educational attainment of each member of the sampled households.

From this dataset we choose to study the income process of the heads of households aged between 20 and 50 in 1991;[1] in this way we could focus on those individuals active in the labour market. We used as our measure of their income all labour income, benefits and pension payments that could be explicitly assigned to the head of the household. We then removed from this sample any individuals without a complete history over the five years, and 30 outliers who had an annual income below £500 or above £200,000 in any year. This left us with a total sample of 2026 individuals. Finally we split this sample into two groups in order to investigate the implications of educational attainment on an individual's income; splitting the group into those with and those without any qualification above an A-level, split the sample into almost two equal sized groups. We refer to the better-qualified group as having tertiary education.

Table I summarizes our data. The incomes of the subgroup with tertiary education are, on average substantially higher; they are also slightly less variable than those of the less well educated. But the larger negative value of skewness and the bigger kurtosis indicate that the population with tertiary education has fatter tails and in particular a fatter upper tail. Figure 1 displays the same information graphically.

In Figures 2 and 3 we have plotted the means and the variances of the sample conditional distributions of log of next year's income given this year's income as a function of this year's income; all incomes are normalized so that they are relative to the sample mean income. Figure 2 demonstrates the data exhibits some mean reversion, and further the relationship appears at first inspection linear. This is consistent with the findings of Atkinson *et al.* (1992) and with predictions of our model as given in equation (1). However Figure 3 shows that the conditional variance has a very distinct U-shape; there is a considerably more income uncertainty at low levels of income

[1] We also estimated a model for evolution of total household income. Though the fit of the model was actually better, we felt it was harder to interpretate these results as the model is also implicitly modelling changes to household structure. These results are available on request.

Table I. Statistics of the observed, cross-sectional distribution of incomes

Characteristic		Without tertiary education $e = 1$	With tertiary education $e = 0$	Total population
Average annual income in £	$\overline{Y}$	10,681	17,183	14,041
Coefficient of variation of income	$cv(Y)$	0.73	0.60	0.69
Gini coefficient of income		0.34	0.31	0.35
Mean of log of normalized income	$m(y)$	−0.46	0.05	−0.19
Variance of log of normalized income	$v(y)$	0.49	0.45	0.54
Skewness of log of normalized income	skewness(y)	−0.57	−1.08	−0.72
Kurtosis of log of normalized income	kurtosis(y)	3.74	5.48	3.96

Normalized income is define as the observed income divided mean income, $Y/\overline{Y}$. Log of normalized income is then naturally $y = \log(Y/\overline{Y})$.

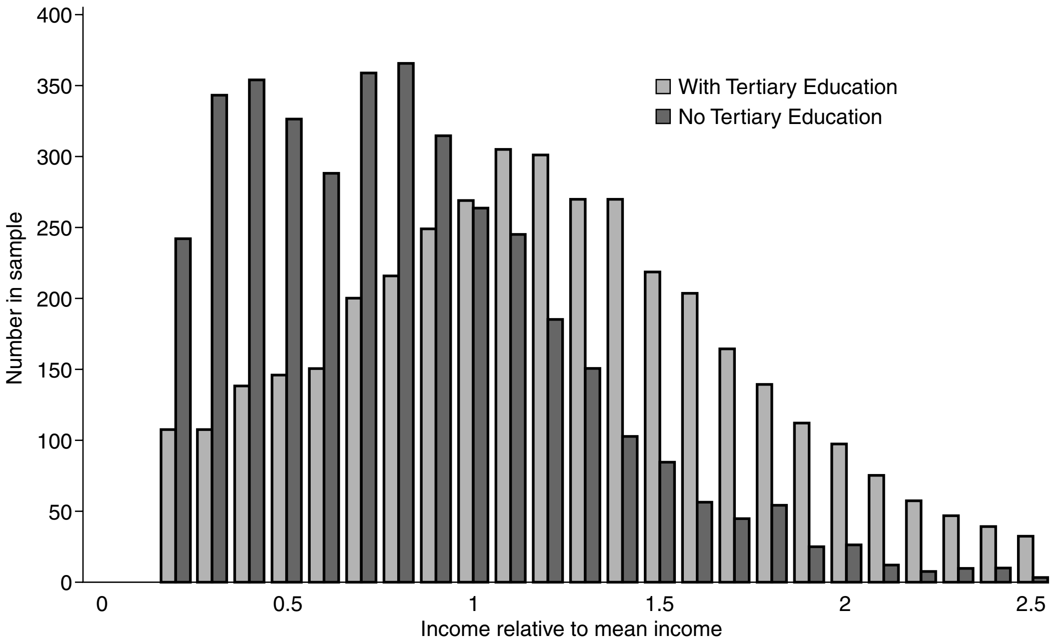

Figure 1. The observed, cross-sectional distribution of incomes for the different educational groups in 1991

than at average income levels, with the uncertainty rising again, though not as sharply, for high levels of income. It is this property of the data that suggests one must look beyond the class of linear or more precisely ARIMA models used by Atkinson *et al.* (1992), Gottschalk and Moffitt (1994) and Dickens (1997); these models all have the property that the variance of the conditional distributions of income will be a constant. Our model, which belongs to the larger class of linear Markov chain or switching models, has the property that the predicted conditional variance is a quadratic function of income; see equation (2). This preliminary investigation of the data therefore

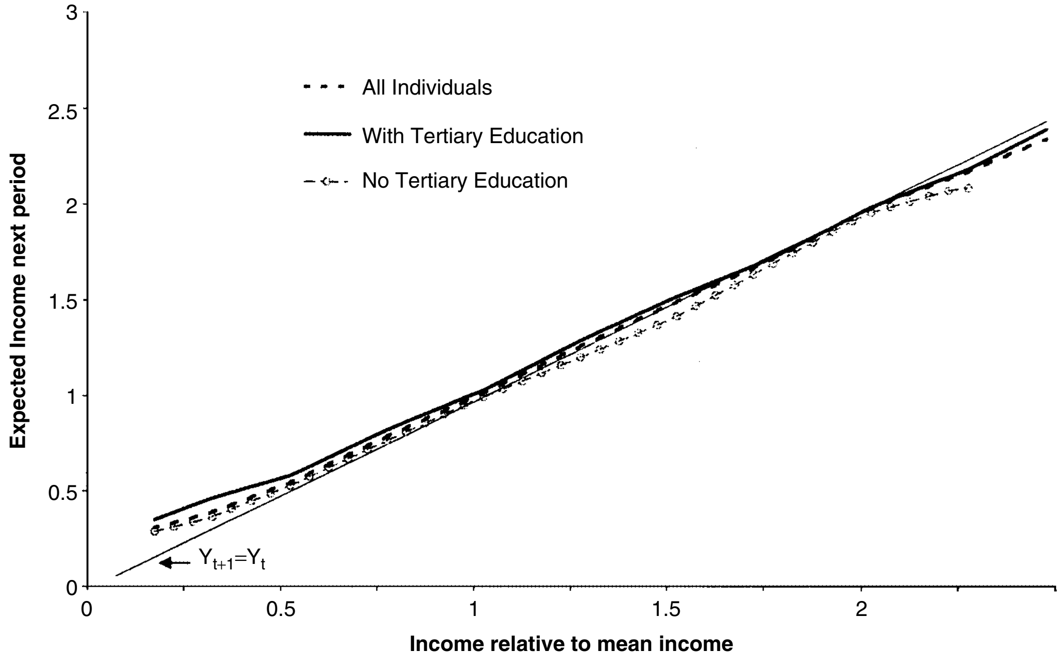

Figure 2. The expected value of the log next year's income given this year's income, $m(y_{t+1}|y_t)$, as a function of this year's income

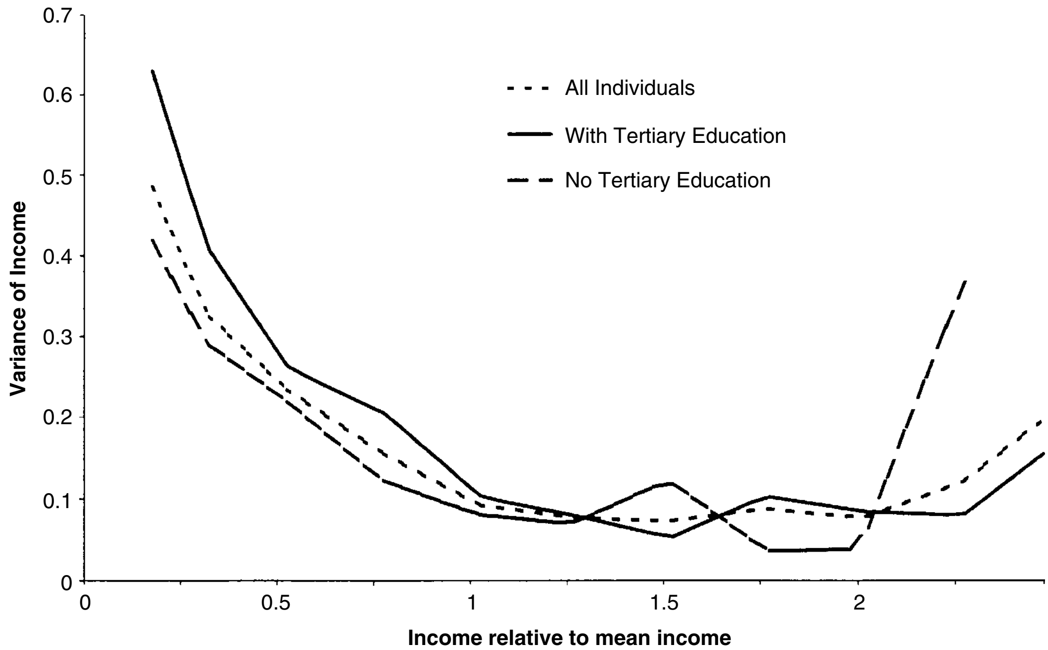

Figure 3. The variance of the log of next year's income given this year's income, $v(y_{t+1}|y_t)$, as a function of this year's income

suggests that our model has the potential to capture more of the properties of the data effectively. This can be regarded as the principle empirical motivation for examining our class of models (the other motivation is that dynamic process described by the model can be given an economic interpretation).

5. ESTIMATION: INCOME TRANSITIONS

We want to estimate the parameters $\phi = (\mu, \sigma^2, \tilde{\mu}, \tilde{\sigma}^2, \Theta)$ subject to alternative restrictions. The model specifies the conditional distribution $F(y_{it+1}|y_{it}, \phi)$, and we estimate the parameters by maximizing the conditional likelihood function

$$L(\mathbf{y}, \phi) = \prod_{i=1,t=1}^{i=N,t=T-1} f(y_{i,t+1}|y_{it}, \phi)$$

with f the density function of F. In our sample, $T = 5$ and $N = 2026$. We chose to estimate the parameters by maximizing this conditional likelihood criterion as opposed to the unconditional likelihood criterion[2] as the final estimated model will better describe the dynamic properties of the income process and further it does not impose the restriction that the distribution of income had achieved its stationary value in any particular year. This has the cost that the final estimated model in this case is likely to be a poorer description of the steady-state income distribution.[3]

We report only the estimates from the conditional likelihood. However we describe, in the text, where there were any significant differences in the parameter estimates if one used the unconditional criterion instead of the conditional one. We also present, in Section 6.1, the implied steady-state distributions calculated from both the models estimated using the conditional and unconditional criterion.

5.1. Estimation with Grouping

In our first set of results, we approximate $\theta(y)$ by a piecewise linear function. The problem of estimating the optimal number of income bands and break points, the multiple change-point problem, is known to be difficult (Lombard, 1987). We therefore chose to divide the income range into 9 regular income groups and test for the sensitivity of the estimates to the grouping. Table II

---

[2] Our model yields a steady-state marginal distribution for y, $F_*(y|\phi)$, with density function f_*, so the unconditional likelihood function can be written

$$L^{UC}(\mathbf{y}, \phi) = \prod_{i=1,t=1}^{i=N,t=T-1} f(y_{i,t+1}|y_{it}, \phi) \prod_{i=1}^{i=N} f_*(y_{i1}|\phi)$$

[3] To illustrate this point, consider the simple AR(1) model, $y_{t+1} = \rho y_t + \varepsilon_t$ where $\varepsilon_t \sim N(0, \sigma)$. The steady-state distribution is then $y \sim N(0, \sigma/\sqrt{(1-\rho^2)})$. Estimating the model simply from the dynamic specification on income data, one will typically estimate ρ to be close to one. The conditional likelihood is unlikely to be particularly sensitive to exact value of ρ and will pick the value that best fits the dynamic data. However the steady-state distribution is very sensitive to the precise value of ρ, if it is close to 1, a small change in its value could easily double the variance of this distribution. Hence estimating this parameter using the unconditional likelihood is likely to choose an estimate of ρ that best fits the steady-state distribution.

Table II. Income group and quantile grouping (1991)

Group	Income grouping $z = Y/\overline{Y}$	Distribution of income grouping (%)	Quantile grouping $z = Y/\overline{Y}$	Distribution of quantile grouping (%)
1	$0 < z \leq 0.2$	4.3	$0 < z \leq 0.31$	11.1
2	$0.2 < z \leq 0.4$	11.9	$0.31 < z \leq 0.48$	11.1
3	$0.4 < z \leq 0.6$	11.5	$0.48 < z \leq 0.61$	11.1
4	$0.6 < z \leq 0.8$	15.5	$0.61 < z \leq 0.80$	11.1
5	$0.8 < z \leq 1.0$	14.0	$0.80 < z \leq 0.96$	11.1
6	$1.0 < z \leq 1.2$	12.8	$0.96 < z \leq 1.11$	11.1
7	$1.2 < z \leq 1.4$	9.9	$1.11 < z \leq 1.33$	11.1
8	$1.4 < z \leq 1.8$	12.1	$1.33 < z \leq 1.67$	11.1
9	$1.8 < z$	8.1	$1.67 < z$	11.1

Table III. Estimates with grouping: $\theta(y) = \theta_i$

Parameter	Estimate	Standard error
$\tilde{\mu}$	−0.67	0.019
$\tilde{\sigma}^2$	0.81	0.013
μ	0.05	0.002
σ^2	0.12	0.002
θ_1	0.22	0.028
θ_2	0.36	0.022
θ_3	0.51	0.023
θ_4	0.71	0.018
θ_5	0.83	0.014
θ_6	0.85	0.013
θ_7	0.90	0.012
θ_8	0.90	0.011
θ_9	0.87	0.012

Log-likelihood/$(N(T-1)) = -0.12$; $N = 2026$;
Estimates for $T = 5$ (1991–1996).

reports the income grouping over which we assume $\theta(y)$ is constant; we also report a quantile grouping for comparison. This restriction corresponds to restriction (**G**), which is our maintained hypothesis.

Table III reports the maximum likelihood estimates. We note that the estimates are all significantly different from zero; and that $\hat{\theta}_1 < \hat{\theta}_2 < \cdots < \hat{\theta}_8$. The increase in θ_i is noticeably sharp at group 4, which contains individuals with incomes just below average. Further, $\theta_9 < \theta_8$, and the difference is significantly different from zero. Hence mobility—as measured by the jump probability $(1 - \theta_i)$—decreases with income levels, stabilizing at about 0.10 for high income levels before displaying a slight increase at the highest income level.

The only significant difference between these conditional estimates and the unconditional estimates is that the unconditional estimate of the parameter θ_9 is significantly smaller at $\theta_9 = 0.75$. As we see later, this reduces the size of the tail of the steady-state distribution of incomes.

5.2. Testing for Restrictions

We note, from Table III, that μ, σ^2 are significantly different from zero; restriction (**M**), the pure mover–stayer model, is therefore rejected.

We now report estimates, and tests, of alternative restrictions on $\boldsymbol{\theta}$. Table IV reports the equation estimated with the linearity restriction

$$(\mathbf{L}) \quad \theta_1 = \cdots = \theta_9$$

This hypothesis is also rejected by a likelihood ratio test, at any reasonable significance level of the chi-squared distribution with 8 degrees of freedom, as the p-value is zero to 10 decimal places. The unconstrained estimates suggest that θ is increasing with y, and quite sharply at around average income $y \sim \ln(\overline{Y})$.

We next assume a smooth parametric form for $\theta(y)$: to mimic the sharp increase, we choose a logistic function

$$\theta(y) = \theta_L + (\theta_H - \theta_L)\frac{\exp(\tau_2(y - \tau_1)}{1 + \exp(\tau_2(y - \tau_1))}$$

This formulation implies $\theta_L \leq \theta(y) \leq \theta_H$ (see Table V).

The logistic hypothesis fits the data as well as the maintained hypothesis (**G**), as the maximized value of the likelihood is virtually identical (note that the hypotheses are non-nested). We observe that the finding $\hat{\theta}_9 < \hat{\theta}_8$, cannot be replicated by this formulation as the logistic function is constrained to be monotonic.

Table IV. Estimates with linearity: constant θ

Parameter	Estimate	Standard error
$\tilde{\mu}$	−0.56	0.020
$\tilde{\sigma}^2$	0.84	0.014
μ	0.05	0.002
σ^2	0.12	0.002
θ	0.73	0.007

Log-likelihood/$N(T-1) = -0.19$; $N = 2026$; $T = 5$.

Table V. Estimates with logistic $\theta(y)$

Parameter	Estimate	Standard error
$\tilde{\mu}$	−0.67	0.019
$\tilde{\sigma}^2$	0.81	0.013
μ	0.05	0.002
σ^2	0.12	0.002
τ_1	0.56	0.022
τ_2	4.21	0.451
θ_L	0.27	0.025
θ_H	0.90	0.008

Log-likelihood/$(N(T-1)) = -0.12$; $N = 2026$; $T = 5$.

It is also desirable to compare our model against the AR model studied by Atkinson *et al.* (1992). Their model is described by the first-order difference equation

$$y_{it} = \rho y_{it-1} + \mu(age_{it}) + \varepsilon_t$$

where log of an individual's income in this period is a linear function of log of income in the last period, a constant which may be a function of the individual's age and a normally distributed innovation. As this model belongs to the same parametric family as ours, it is valid to compare the achieved fit of the two models using Aikake's Information Criteria (AIC) or Schwarz's Bayesian Information Criteria (BIC).[4] For our maintained model, $AIC = 1967.8$ and $BIC = 1029.4$. For the AR model where μ is not a function of age, its $AIC = 8727.3$ and its $BIC = 4409.1$. Therefore both criteria overwhelmingly prefer our model. We then generalized the AR model slightly and allowed μ to be piecewise linear function of age, where a separate μ parameter was estimated for each five-year interval. Thus if an individual was aged between 20 and 24, parameter μ_1 was used as the constant in the dynamic income equation, and if aged between 25 and 29 parameter μ_2 was used, etc. This model is almost identical to the model used in Atkinson *et al.* (1992) study. The AIC and BIC for this model were 8693.6 and 4392.3 respectively. Again our maintained model is preferred.

5.3. Education and Income Transitions

It is possible that individual characteristics affect income dynamics—and therefore, of course, the long-run distribution of incomes. We consider the role of one such observable characteristic: higher education. The dummy variable $e_i \in \{1, 0\}$ measures whether an individual i has further education, beyond A-levels including professional or university qualifications.

Table VI. Estimates showing the effect of tertiary education: $\theta(y) = \theta_i$

Parameter	Estimate $e = 0$	Standard error	Estimate $e = 1$	Standard error
$\tilde{\mu}$	−0.87	0.024	−0.38	0.029
$\tilde{\sigma}^2$	0.74	0.016	0.85	0.020
μ	0.05	0.003	0.05	0.002
σ^2	0.13	0.003	0.10	0.002
θ_1	0.24	0.036	0.18	0.043
θ_2	0.38	0.027	0.25	0.037
θ_3	0.58	0.028	0.43	0.038
θ_4	0.79	0.020	0.60	0.031
θ_5	0.86	0.017	0.78	0.024
θ_6	0.87	0.020	0.86	0.019
θ_7	0.89	0.021	0.89	0.016
θ_8	0.87	0.023	0.86	0.013
θ_9	0.84	0.031	0.85	0.014
Sample size N		979		1047
Log-likelihood/$N(T-1)$		−0.23		0.01

The income groups are the income groups as shown in Table II.

[4] The BIC penalizes more heavily than the AIC a model with a large number of parameters. It has been shown that in lag order selection for liner models, AIC will tend to overestimate the lag order, whereas the BIC will tend to underestimate.

Estimates are reported in Table VI. The most substantial difference between the two sets of estimates is in the value of $\tilde{\mu}$, the mean of the recurrent distribution—the distribution of incomes from which the income of a mover is drawn. Education has no impact on expected growth rates of the income of stayers, though it does reduce the variability of income growth, $\sigma_1^2 < \sigma_0^2$.

We also observe that $\theta_i(1) < \theta_i(0)$ for $i = 1, \ldots, 8$, with the reverse being true for θ_9. However, for $i = 5, \ldots, 9$, the mobility estimates are almost identical for the two different groups. For lower θ_i the difference is larger but is always less than one standard deviation of the estimate. Hence tertiary education may marginally increase mobility at low levels of income.

It follows that the principal difference in the income dynamics of individuals with a tertiary education and those without is not in the way their income progresses from year to year, nor in their relative mobilities but only in their fall-back income—the income they receive if there is a disruption to their career. In this case an individual with a tertiary education is likely to receive a considerably higher income than one without.

6. INCOME UNCERTAINTY AND DISTRIBUTION

The maintained hypothesis **(G)** has substantive implications for the nature of uncertainty about future incomes facing individuals, as well as the long-run or steady-state distribution of incomes in the economy. We examine these properties here.

6.1. Means and Variances

Recall that

$$m(y_t) = Ey_{i,t+1}|y_{i,t} = \theta(y_{it})(y_{it} + \mu) + (1 - \theta(y_{it}))\tilde{\mu}$$

$$v(y_t) = \mathrm{var}(y_{i,t+1}|y_{it}) = \theta(y_{it})\sigma^2 + (1 - \theta(y_{it})\tilde{\sigma}^2 + \theta(y_{it})(1 - \theta(y_{it}))(y_{it} + \mu - \tilde{\mu})^2$$

We write $m_e(y)$, $v_e(y)$ for the moments conditional on education e. Tables VII and VIII report the expected incomes, and income uncertainty, predicted by the maximum likelihood estimates of ϕ (as reported in Tables III and VI).

Table VII. Conditional expected incomes for the different educational groups

Relative income $Y_t/\overline{Y}_t$	Expected income $e = 0$ $\exp(m_0(y_t))$	Expected income $e = 1$ $\exp(m_1(y_t))$	Expected income overall $\exp(m(y_t))$
0.25	0.33	0.49	0.37
0.50	0.49	0.59	0.52
0.75	0.74	0.76	0.74
1.00	1.01	1.02	1.01
1.25	1.26	1.29	1.28
1.50	1.48	1.54	1.53
1.75	1.71	1.79	1.77
2.00	1.90	1.98	1.97
2.25	2.12	2.21	2.20
2.50	2.34	2.44	2.43

Parameter estimates from model (**G**).
$m_e(y_t) \equiv E(y_{t+1}|y_t, e)$; $e \in \{0, 1\}$.

Table VIII. Conditional income uncertainty for the different educational groups

Relative income $Y_t/\overline{Y}_t$	Income uncertainty $e = 0$ $v_0(y_t)$	Income uncertainty $e = 1$ $v_1(y_t)$	Income uncertainty overall $v(y_t)$
0.25	0.328	0.674	0.458
0.50	0.182	0.328	0.235
0.75	0.101	0.196	0.137
1.00	0.077	0.078	0.079
1.25	0.077	0.052	0.063
1.50	0.122	0.054	0.069
1.75	0.142	0.063	0.080
2.00	0.200	0.121	0.135
2.25	0.223	0.137	0.151
2.50	0.244	0.152	0.166

Parameter estimates from model (**G**).
$v_e(y_t) = \text{var}(y_{t+1}|y_t, e)$, $e \in \{0, 1\}$.

Observe, in Table VII that expected incomes are almost linear in current income. Table VIII displays an important prediction of our model. The conditional variance is U-shaped in current income and the U-shape is especially pronounced for those with a tertiary education, $e = 1$. This replicates the property of the data shown in Figure 3. The pronounced U-shape is generated because individuals with low incomes are more mobile and likely to experience an income disruption. Their income in the next period is drawn from the recurrent distribution and is therefore likely to be higher. Those with an income close to the average level face very little income uncertainty whether they are stayers or movers. Those with high incomes are less mobile but should they experience an income disturbance, their incomes next year are likely to be much lower. It is this effect which causes the level of income uncertainty to rise again at high levels of income.

Individuals with tertiary education have higher expected incomes for all levels of income (i.e. $m_1(y) > m_0(y)$ for every y). At low levels of income, they also face more uncertainty. The ranking changes for $Y/\overline{Y} > 1.25$, and incomes of the educated become less uncertain. A person with a high level of income and no further education is less likely to maintain his or her income. Our model suggests that this is not because those with no tertiary education are more mobile but because their fall-back income—the mean of their recurrent distribution—is lower.

In Table IX we report the characteristics of the steady-state earnings distribution F_{*0}, F_{*1}. These are derived from the parameter estimates of the grouped (**G**) models derived earlier. They can be compared with the summary of the data in Table I. We observe the point alluded to in Section 5. The conditionally estimated model, which models only the income dynamics, yields a steady-state distribution slightly different from that observed income distribution in the data (although it provides the better model of income dynamics). The unconditionally estimated model produces a steady-state distribution much closer to the 1991 distribution (which was used as the reference distribution in its estimation). As we described earlier, the only significant difference between the two models was in the estimate of θ_9; this was estimated to be about 0.12 points in the unconditionally estimated model. This lower estimate has the effect of reducing significantly the 'fatness' of the high-income tail is the steady state.

We found the unconditionally estimated model gave a slightly better fit than the generalized β-distribution of the cross-sectional distribution of earnings in 1991 for those people without

Table IX. Properties of the implied steady-state distributions of the different estimated models

	Conditionally estimated model		Unconditionally estimated model		Generalized beta model	
	$e = 0$	$e = 1$	$e = 0$	$e = 1$	$e = 0$	$e = 1$
mean(Y)	1.06	1.69	0.85	1.33	0.78	1.25
$cv(Y)$	0.83	0.7	0.72	0.66	0.64	0.69
Gini(Y)	0.4	0.37	0.35	0.33	0.30	0.28
mean(y)	−0.22	0.35	−0.41	0.08	−0.46	0.05
var(y)	0.73	0.79	0.54	0.54	0.52	0.45
skewness(y)	−0.11	−0.17	−0.38	−0.47	−0.97	−1.05
Kurtosis(y)	3.26	3.63	3.12	3.81	4.75	5.75
Sum of squared residuals	0.0076	0.0112	0.0016	0.0028	0.002	0.0003
Akaike criterion	−7.35	−6.96	−8.89	−8.34	−9.34	−11.21

Note: The normalized income level is denoted Y, and its logarithm as y. The income transition model is described in Table VI. The goodness of fit tests are performed on the 24 income groups as displayed in Figure 1.

tertiary education, but did not fit as well the distribution for those with tertiary education. The Akaike criteria of goodness of fit for the jointly estimated model was not as good as the generalized β-distribution in both cases because of the greater number of parameters in our income model. However, this is no great cause for concern as our model is a model primarily of the dynamic income processes rather than the eventual steady state. The analysis does however draw attention to the need to model entries into and exits from the labour market at the beginning and end of an individual's working life as this is the most probable reason why our conditional model overpredicts the 'fatness' of the high-income tails of the steady-state distribution.

7. CONCLUSIONS

One of the major macroeconomic changes over the last twenty years has been the huge increase in income inequality. The most common measure of income inequality in the literature is the Gini coefficient. In the UK, the Gini coefficient increased by about 8 points from around 26 in 1977 to 33.7 in 1991 (Goodman and Webb, 1994). To find levels of income inequality comparable to those found in 1991, one is forced to look as far back as the early part of the twentieth century. Atkinson (1996) summarizes the two popular explanations for this increasing income inequality. The first is that increasing competition from Third World trade reduced the wages of the unskilled relative to the skilled; and the second is that recent technological innovations have reduced the demand and therefore the wages of the unskilled. Both these explanations are concerned with the level of or return to different types of employment.

While our model does not address these questions directly, it does provide a framework for evaluating the effects of such changes on income dynamics and the distribution of income. In this discussion, we specialize to the linear version of our model **(L)**, where the comparative statics are transparent. Recall, from Corollary 1 of Section 3 that the variance of log-incomes in the steady state is

$$v_{*L} = \tilde{\sigma}^2 + \frac{\theta}{1-\theta}\sigma^2 + \frac{\theta}{(1-\theta)^2}\tilde{\mu}^2$$

An increase in earnings inequality can be attributed to changes in the recurrent income distribution $\tilde{\mu}, \tilde{\sigma}^2$; in mobility, θ; or in the variance of income growth σ^2 for different types of workers, as m_{*L} decreases with θ in this case.

Increased flexibility in labour markets, possibly in response to international competition, is likely to decrease θ. This is actually equalizing in the long run, as v_* increases with θ. A once-for-all decrease in θ may increase the dispersion of incomes on the transition path, but is eventually equalizing. At the very least, this model suggests that it is too early to deduce that increased flexibility has increased the long run dispersions of incomes.

An explanation based on increasing wage differentials can be pinned down to a decrease in the expected fall-back income, $\tilde{\mu}$, of the unskilled (defined as the mean of their respective recurrent income distribution). Such an explanation would imply that within-group variances have actually *decreased* (noting that $v_*(y)$ increases with $\tilde{\mu}$). Alternatively, an increase in expected fall-back income, $\tilde{\mu}$, of the skilled implies no change in the unskilled earnings distribution, and an increase in both mean and variance of the skilled earnings distribution. The welfare implications of the two changes are substantially different. The skilled are strictly better off if their fall-back income increases.

Finally, an increase in σ^2—the variance of earnings growth—will certainly increase the spread of the long-run distribution. Its welfare effects depend critically on the extent to which individuals are risk-averse. While this explanation is not directly suggested by any of the prevailing hypotheses, it is compatible with a dynamic version of the skill-differential hypothesis (i.e. an increase in expected fall-back income of the skilled).

Welfare comparisons are usually made *ex ante*, while earnings distributions are observed *ex post*. The arguments here suggest that some increases in observed inequality may be responses to changes which are *ex ante* desirable, and others manifestly not. A decomposition of inequality into static and dynamic components is essential in evaluating the quantitative importance of these factors. Our model provides a framework for this decomposition.

Unfortunately, panel data do not go back far enough in time to estimate shifts in all parameters ϕ. However, specialized versions of the model—especially (**L**)—can be estimated from time-series of cross sections, via the invariant distribution F_{*L}. The shape, and moments of this distribution are known (e.g. Theorem 2 and Corollary 1), and its parameters can be estimated from cross-sectional distributions. It is also possible to extend our analysis to measure the impact of parameter changes on the transitional dynamics of the earnings distribution.

We have estimated our model for only two population sub-groups, those with and without tertiary education. However, even this distinction yields a substantial insight into the way in which education affects earnings. We expect that the development of this type of model and its application to various subgroups of the population is likely to yield valuable insights about how their different characteristics affect people's earnings.

ACKNOWLEDGEMENTS

ESRC under grant no R000237688. The data used in this paper were made available through The Data Archive and were originally collected by the ESRC Research Centre on Micro-Social Change at the University of Essex. Neither the original collectors of the data nor the Archive bear any responsibility for the analyses or interpretations presented here. We would like to thank two anonymous referees and, in particular, an editor of this journal for their helpful comments.

REFERENCES

Abowd J, Card D. 1989. On the covariance structure of earnings and hour changes. *Econometrica* **57**: 411–445.

Atkinson A. 1970. On the measurement of inequality. *Journal of Economic Theory* **2**: 244–263.

Atkinson A. 1996. Seeking to explain the distribution of income. In *New Inequalities*, Hills J (ed.). Cambridge University Press: Cambridge; 19–48.

Atkinson A, Bourguignon F, Morrisson C. 1992. *Empirical Studies of Earnings Mobility*. Harwood Academic Publishers: Chur.

Atkinson A, Rainwater L, Smeeding T. 1995. Income distribution in OECD countries. Social Policy Studies 18, OECD: Paris.

Champernowne D. 1953. A model for income distribution. *Economic Journal* **53**: 318–351.

Dickens R. 1997. Caught in a trap? wage mobility in Great Britain: 1975–94. Discussion Paper 365, CEPR, London School of Economics.

Fisk P. 1961. The graduation of income distribution. *Econometrica* **29**: 171–185.

Frydman H. 1984. Maximum likelihood estimation in the mover-stayer model. *Journal of the American Statistical Association* **79**(387): 632–638.

Goodman A, Webb S. 1994. For richer for poorer: The changing distribution of income in the United Kingdom, 1961–91. Commentary 42, Institute of Fiscal Studies.

Goodman L. 1961. Statistical methods for the mover-stayer model. *Journal of the American Statistical Association* **56**: 841–868.

Gottschalk P, Moffitt R. 1994. The growth of earnings instability in the US labour marker. *Brookings Papers on Economic Activity* **2**: 217–273.

Hart P. 1976. The dynamics of earnings. *Economic Journal* **86**: 551–565.

Lombard F. 1987. Rank tests for changepoint problems. *Biometrika* **74**: 615–624.

Majumder A, Chakravarty S. 1990. The distribution of income: Development of a new model and its application to U.S. data. *Journal of Applied Econometrics* **5**(2): 189–197.

McDonald J, Mantrala A. 1995. The distribution of income: Revisited. *Journal of Applied Econometrics* **10**(2): 201–205.

Sampson M. 1990. A Markov chain model for unskilled workers and the highly mobile. *Journal of the American Statistical Association* **85**(409): 177–180.

Stokey NL, Lucas REJ. 1989. *Recursive Methods in Economic Dynamics*. Harvard University Press: Cambridge, MA.

