Review of Economic Studies (2009) **01**, 1–33 0034-6527/09/00000001$02.00

Modelling income processes with lots of heterogeneity

M. BROWNING
DEPARTMENT OF ECONOMICS, OXFORD UNIVERSITY
M. EJRNÆS
DEPARTMENT OF ECONOMICS, UNIVERSITY OF COPENHAGEN
J. ALVAREZ
BANK OF SPAIN, MADRID
First version received October 2007; final version accepted September 2009 (Eds.)

We model earnings processes allowing for lots of heterogeneity across agents. We also introduce an extension to the linear *ARMA* model that allows that the initial convergence to the long run may be different from that implied by the conventional *ARMA* model. This is particularly important for unit root tests which are actually tests of a composite of two independent hypotheses. We fit to a variety of statistics including most of those considered by previous investigators. We use a sample drawn from the PSID, and focus on white males with a high school degree. Despite this observable homogeneity we find more latent heterogeneity than previous investigators. We show that allowance for heterogeneity makes substantial differences to estimates of model parameters and to outcomes of interest. Additionally we find strong evidence against the hypothesis that any worker has a unit root. JEL codes: J30 ,C23

Keywords:*earnings, heterogeneity, ARMA, initial conditions, unit root*

1. INTRODUCTION.

Estimates of the earnings process of individuals and households are required for a number of purposes. These include: testing between different models of the determinants of income distribution (see Neal and Rosen (2000) and Rubinstein and Weiss (2006)); determining the earnings risk faced by individuals and households (see Carroll and Samwick (1997)); modelling the incidence and persistence of low income spells (see Atkinson, Bourguignon and Morrisson (1992)); modelling the time series variation in the earnings distribution (see Moffitt and Gottschalk (2002)); modelling labour supply (see Abowd and Card (1989)); the calibration of consumption and saving models and dynamic GE models (see Browning, Heckman and Hansen (2000)); modelling anticipated earnings growth for use in consumption Euler equations (see Browning and Lusardi (1996)); as an input to schooling choices (see Cunha, Heckman and Navarro (2005)); and predicting future earnings paths given individual information (see Chamberlain and Hirano (1999)). We shall return to a discussion of some of these issues but for now this will suffice to motivate the widespread interest in estimating earnings processes.

Various dynamic models for individual workers have been assumed in the earnings dynamics literature; see, for example, Hause (1977), Lillard and Willis (1978), Lillard and Weiss (1979), MaCurdy (1982), Abowd and Card (1989), Gottschalk and Moffitt (1994), Baker (1997), Chamberlain and Hirano (1997), Geweke and Keane (2000), Haider (2001), Ulrick (2000), Meghir and Pistaferri (2004), Guvenen (2007) and Guvenen and Smith (2008). As the latter two papers and Baker (1997) emphasise, there are two major strands in the earnings process literature. In the first strand (that takes off from Lillard and Weiss (1979)) researchers allow for heterogeneity in trends and usually find that shocks do not have a permanent component. In the second strand (starting with MaCurdy (1982) and Abowd and Card (1989)), there is no allowance for heterogeneous trends and researchers find that shocks are persistent, usually opting for everyone having a unit root. The most important conclusion that we draw from our reading of this literature is that, whatever the process chosen, only limited allowance is made for heterogeneity. As we shall show below, conventional processes in either of these strands are unable to account simultaneously for a wide range of features of the observed data. To capture all of the features we see in the data we shall show that we need a more general process than, say, an $ARMA\,(1,1)$ and we need to allow for lots of heterogeneity.

Other investigators have, of course, also allowed for heterogeneity but in a somewhat limited fashion. In particular, it has always been assumed that everyone has the same process, but with, for instance, different means, trends and/or variances. We consider whether different workers have different processes - some with a unit root, some with a stationary $AR(1)$ and others with an $MA(1)$ process, for example. Specifically, we allow for heterogeneity in the starting value, the variance of shocks, MA and AR parameters, deterministic trends and the deterministic speed of convergence from the initial value to the long run process. We also allow for different measurement error variances across agents.

Correlated heterogeneity in a dynamic context gives rise to an 'initial conditions' problem (see, for example, Hsiao (1986), chapter 4 or Arellano (2003), chapter 6). We adopt a generalisation of the approach of allowing that the unobserved heterogeneity is a parametric random function of the starting value of the process[1] (see, for example, Chamberlain (1980), Anderson and Hsiao (1982), Blundell and Smith (1991), Arellano and Carrasco (2003) and Wooldridge (2005)). For example, for a stationary AR(1) process with different means for each worker we might allow that the mean is a function of the starting values plus a random disturbance term. The advantages and disadvantages of this approach to dealing with unobserved heterogeneity are discussed in section 2.3 below. The main corollary of this approach is that we take a fully parametric approach to modelling the income process, conditional on a parametric starting value distribution.

There are three broad approaches to estimating the parameters of a fully parametric model in our context. One option is to first conduct a time series analysis on each person and then to use this to generate a model of unobserved heterogeneity using parametric distributions for the unknown parameters - a 'bottom-up' approach. The problem with following this strategy is that the individual estimates suffer from considerable small sample ('small T') and endogeneity biases. It might be possible to implement analytic or simulation based small sample corrections to the estimators properties (see, for example, Shaman and Stine (1988a and 1988b), Kiviet and Krämer (1992) or Kiviet and Phillips (1998)) but these corrections impose stronger assumptions on the distributional properties of the errors than those we would like to impose *a priori*.

A second alternative is to specify a general joint distribution of parameters and then, using either conditional maximum likelihood (CML) or GMM, to 'test down' to a more parsimonious model. The main problem with this approach is that we do not have any prior idea *at all* about the distribution of the parameters. Which parameters should be heterogenous and what (joint) distribution should we take for them? In the model we develop below we have eight parameters for each worker and we also allow for

1. Below we shall be careful to distinguish between the starting values of a process and the initial observation of a process (that is, the first value observed). For the moment, simply assume that we observe all processes from their starting point.

heterogeneous measurement error. Nowhere in the literature is there any indication of how to specify a general joint distribution for these parameters nor is there any current hope of identifying the joint (conditional) distribution nonparametrically.[2]

The third general approach, which we have followed, is to conduct an explicit exploratory analysis of a series of models starting with a restricted model that has been widely used in the literature and moving to more general model in a series of steps. At each stage the generalisation is chosen to deal with the worst empirical failing of the current model (a 'fire fighting' strategy). This procedure is not path independent (which is true of any exploratory specific to general analysis) but if we end up with a model that captures all of the different aspects of the data the literature has considered then it may be considered satisfactory. In practice, this eventually yields a general model which we then represent as a nonlinear factor model. To save on space, in the exposition below we do not give details of our exploration procedure (the exact details for an earlier version can be found in Alvarez, Browning and Ejrnæs (2002)) but rather start from the most general model.

This exploratory approach requires the fitting of a relatively large number of more and more complicated fully parametric models. To fit these models we use Simulated Minimum Distance (SMD) (also known as 'indirect inference'); see Smith (1993), Gourieroux and Monfort (1996) and Hall and Rust (2003). Since SMD is now a well established technique we provide only an outline of the specific SMD estimation procedure we use; see section 3.3 and appendix A4. To implement SMD we fit to a set of statistics or auxiliary parameters (ap's). We choose a set that provide a rich description of the data and some supplementary ap's that capture most of the variation in the data that previous investigators have used. The auxiliary parameters are then calculated for the data. The SMD procedure finds a set of parameter values such that the parametric model produces close to the same values for the auxiliary parameters. To evaluate whether the model (with a given set of parameters) produces the same auxiliary parameters, simulated data are generated. Then the auxiliary parameters are calculated for the simulated data. We then estimate the parameters of the model such that the weighted distance between the auxiliary parameters for the data and the simulated data is minimized.

The sample used in this study is drawn from the PSID sample from 1968 to 1993. We use exactly the sample used in Meghir and Pistaferri (2004); this is to ensure that the conclusions we draw, some of which are radically different from previous investigators, are not due to having a different sample. In order to obtain a more homogenous sample we stratify the sample and focus on white, male, high school graduates aged 25 to 55.

There are four principal new results arising from the empirical analysis. First, we have to extend the conventional ARMA model to take account of the fact workers close to the start of their earnings process have a different convergence to the 'long run' than implied by an $ARMA(1,1)$ model. Second, we find strong evidence of much more heterogeneity than previous researchers have allowed for. For example, we find that the variances of the shocks vary considerably across workers. Our third finding is that once we incorporate such extensive heterogeneity, the preferred model has everyone having an ARMA process with a deterministic trend and an autoregressive parameter below unity; *there is no sign of anyone having a unit root.* The latter finding is in accord with contributions such as Baker (1997), Haider (2001) and Guvenen (2007). Finally, we find

2. Showing nonparametric identification (or non-identification) for the joint distribution of several parameters for each worker is rarely possible and our case is no different. Indeed, even for models that allow for a good deal less heterogeneity than ours, showing nonparametric identification is problematic. In this paper we do not consider nonparametric identification.

that we need to allow for measurement error with variances that are heterogeneous, with some workers giving consistently accurate reports and others who persistently report with a lot of noise. We also show that allowing for a lot of heterogeneity can make a big difference for substantive questions of interest.

In the next section we present a general model for earnings processes and discuss how we allow for observed and unobserved heterogeneity, measurement error and the fact that not all workers are observed from the start of their earnings process. In section 3 we present details of the sample selection, a description of the data and the auxiliary parameters we use in estimation and testing for goodness of fit. We also present an outline of the SMD procedure used. Section 4 presents our results, including a discussion of four outcomes of substantive interest. The conclusions are given in section 5; these are largely a listing of our main findings and our beliefs about where to go next.

2. A GENERAL MODEL FOR EARNINGS.

2.1. *An extended ARMA model.*

In this section we present the model for a single worker that underpins our empirical analysis. The exposition of our model and heterogeneity structure given below might suggest that we started with a general parametric model and then used a general to specific procedure to 'test down'. In fact, the converse is true. We started with a very restricted model and then added extra features (more heterogeneity, particular functional forms, the extensions of the basic ARMA model given below and allowance for heterogeneous measurement error variances) to capture different facets of the data. For example, we tried several different distributions for the short run variance term and found that the lognormal worked well (in terms of fitting the data) whereas other simple distributions such as the exponential did not and more complicated forms, such as the generalised Gamma, did not do significantly better than the lognormal. As another example, we tried various support conditions for the distribution of the AR and MA parameters but the variant given below (which takes 'natural bounds') was preferred. This specific to general specification search lead to a preliminary preferred model with considerably fewer parameters than in the most general model below. We then generalised the preliminary preferred model to arrive at what we call our general model. The first version of this paper (available as Alvarez *et al* (2002)) put this specific to general 'fire fighting' specification search at the heart of the paper. In this version we emphasise the substantive implications of our results rather than how we arrived at them.

Our general process for an individual worker is given by:

$$\begin{aligned} y_t = &\left[\delta\left(\left(1-\omega^t\right)-\beta\left(1-\omega^{t-1}\right)\right)+\alpha\beta\right]+\beta y_{t-1} \\ &+\alpha\left(1-\beta\right)t+\left(\varepsilon_t+\theta\varepsilon_{t-1}\right) \end{aligned} \tag{2.1}$$

where t is age (minus 25). An alternative way to write (2.1) is as:

$$y_t = \delta\left(1-\omega^t\right)+\alpha t+\beta^t y_0+\sum_{s=0}^{t-1}\beta^s\left(\varepsilon_{t-s}+\theta\varepsilon_{t-s-1}\right) \tag{2.2}$$

This is an extended ARMA model for which α is a drift parameter; δ is a long run mean, net of the trend (which we shall usually call the 'long run mean' in all that follows); β is the AR parameter; θ is the MA parameter and ω is a novel parameter that is discussed below.[3] In all that follows we make a distinction between the parameters of the process for a single worker, *the model parameters*, and the parameters of the distributions for heterogeneity, *the heterogeneity distribution parameters* (these are presented in subsection 2.3). The shock ε_t is the stochastic component with mean zero (in the next subsection more details on the stochastic component are provided). This specification represents a significant generalisation of the conventional ARMA scheme which corresponds to the restriction $\omega = \beta$:

$$\begin{aligned} y_t &= [\delta(1-\beta)+\alpha\beta] + \beta y_{t-1} + \alpha(1-\beta)t + (\varepsilon_t + \theta\varepsilon_{t-1}) \\ &= \delta\left(1-\beta^t\right) + \alpha t + \beta^t y_0 + \sum_{s=0}^{t-1}\beta^s\left(\varepsilon_{t-s} + \theta\varepsilon_{t-s-1}\right) \end{aligned} \tag{2.3}$$

To motivate this generalisation that introduces an additional model parameter, observe that if we set $\beta = 1$ in (2.2) then the effect of the initial value and any subsequent shock never decays; this is the *essential* implication of the unit root restriction. In this case (2.1) is:

$$\Delta y_t = \left[\delta\left(\omega^{t-1}(1-\omega)\right) + \alpha\right] + (\varepsilon_t + \theta\varepsilon_{t-1}) \tag{2.4}$$

Since $\omega \le 1$, in the long run this gives a unit root with drift:

$$\Delta y_t \simeq \alpha + (\varepsilon_t + \theta\varepsilon_{t-1}) \text{ for } t \text{ large} \tag{2.5}$$

so that the long run process is independent of ω. For the process in its first few periods, however, setting $\omega = \beta$ is restrictive. To see this note that with $\beta = 1$ we have:

$$E(\Delta y_1) = \delta(1-\omega) + \alpha \tag{2.6}$$

Imposing $\omega = 1$ implies that the expected change at the beginning of the process is equal to the expected change for the process in the long run (in this case, the drift α). Thus the unit root hypothesis, $\beta = 1$, in the usual $ARMA$ model ((2.3)) is seen to be a composite of two independent hypotheses: persistence of the effects of shocks and the first differenced process behaving initially as it does in the long run. This composite nature of the unit root hypothesis has been recognized in the time series literature (see, for example, Müller and Elliott (2003)), but there the main concern is how this affects unit root tests.[4] The economic implications are not considered a significant restriction since it is usually assumed that the process has been running for a long time when we first observe it. For the earnings case we observe workers close to the start of the process and we have to take account of this. Thus our extension of the usual model is needed to break the link between two hypotheses that are logically distinct. This is potentially

3. More flexible forms of the deterministic component can be adopted (using, for example, splines) but that adds substantially to the number of parameters in the presence of heterogenous parameters. Our approach is relatively parsimonious, easily interpreted and fits the data well.

4. In a time series context and using a Bayesian framework, Jarocinsky and Marcet (2009) illustrate the importance of different assumptions on the initial conditions for estimation of autoregressive models. They derive a model with a prior on the initial growth rates that replaces the assumption on initial conditions. This approach has some disadvantages in relation to ours. Firstly, we have to determine the number of initial periods on which the prior is established. Second, it is in general complicated to translate the prior on the growth rates to a prior on the unobservables. Finally, to incorporate individual heterogeneity in the prior gives rise to a much less parsimonious scheme than ours.

important since it avoids rejecting the important component of the unit root hypothesis, $\beta = 1$ in equation (2.1), simply because the initial behaviour does not conform to (2.6).[5]

In our empirical analysis we consider three classes of models which can all be seen as special cases of the extended ARMA model. In the first class are *stable models* ((2.1) with $\beta < 1$) in which the effects of shocks to the income process ultimately die out.[6] Within this class the deterministic component of earnings is given by:

$$E\left(y_t \mid y_0\right) = \delta\left(1 - \omega^t\right) + \alpha t + \beta^t y_0 \tag{2.7}$$

so that, net of the trend, the *ex ante* expected value is a linear combination of the starting value, y_0, and the long run mean, δ, with weights for the latter that tend to unity with age and weights for the starting value that tend to zero. Combinations of $\beta \lesseqgtr \omega$ allow for a flexible adjustment to the long run process without restricting the latter.[7] Thus the introduction of the ω parameter in (2.1) also represents a significant weakening for the stable model; as we shall see, the data strongly support such an extension.

In the second class of models we consider *unit root processes* ((2.1) with $\beta = 1$) which are characterised by permanent shocks. Although tests for a unit root against a trend-stationary process often lack power in a time series context, the recent panel data literature makes it clear we can have powerful tests for a unit root; see Im, Pesaran and Shin (2003). Within the unit root class we also consider the 'simple unit root' model which is a further restricted version of the unit root model. That is, (2.1) with $\beta = \omega = 1$ and $\delta = \alpha = 0$. This simple unit root model has been widely used in the literature. One reason for this is that it neatly capture the distinction between permanent and transitory shocks (see Appendix subsection Appendix A.1). The other reason is that it is believed that it provides a good fit to the data; prominent recent examples include Meghir and Pistaferri (2004), Blundell *et al* (2008) and De Santis (2007). As we shall see below the widespread belief that everyone has a unit root is misplaced; in fact, we find that no one has a unit root.

The third class of models we consider are *mixture models* in which we allow that the earnings of some workers have a unit root and that of others is a stable process.[8] Details of how we do this are given in subsection 2.3.

2.2. *The stochastic component.*

Turning to the stochastic component of the process, we make the following assumptions on the error terms in (2.1):

$$\begin{gathered} E\left(\varepsilon_t\right) = 0 \text{ and } E\left(\varepsilon_t \varepsilon_s\right) = 0 \text{ for } s \neq t \\ E_{t-1}\left(\left(\varepsilon_t\right)^2\right) = \nu_t + \frac{\exp(\varphi)}{1 + \exp(\varphi)}\left(\varepsilon_{t-1}\right)^2 \end{gathered} \tag{2.8}$$

5. In the empirical section we shall present evidence that this novel allowance for initial convergence captures what we see in the data better than, for example, allowing for an $ARMA(2,1)$ process. Moreover introducing a second order AR parameter increases considerably the number of parameters since then we have to model the two starting values.

6. We make a distinction between stable models (those with an AR parameter of less than unity) and stationary models which are stable and have further restrictions on the initial conditions and trends.

7. When considering different combinations of $(\beta, \omega, \delta, y_0)$ it is important to note that we shall be modelling deviations from age means so that for some workers earnings (relative to the mean) can fall from the starting value and then rise again. This would not be the case if we restrict $\omega = \beta$.

8. We also considered models in which some workers had an explosive root ($\beta > 1$). This never improved the fit and is not discussed further.

where $E_{t-1}(.)$ denotes the expectations conditioned on $(\varepsilon_0, y_0, ..., y_{t-1})$. The third condition allows for a time varying 'base' variance (the t subscript on ν) and an ARCH component. The time component is to allow for a common time effect for the variance, see Moffitt and Gottschalk (2002). We found a quadratic time trend to be adequate. To do this we assume that the variance in period p for a worker h (ν_t in equation (2.8) with an h subscript) evolves according to:

$$\nu_t = \nu_0 e^{\left(\kappa_1 p + \kappa_2 p^2\right)} \tag{2.9}$$

where ν_0 is an individual baseline variance and $p = 0$ in the first year (1968).[9] The ARCH component is a simpler specification than used in Meghir and Pistaferri (2004) but, as we shall see below, it suffices when we allow for more heterogeneity than they allow for. Furthermore, we assume that ε_t conditional on ε_{t-1} is normally distributed with $\varepsilon_0 \equiv 0$.[10]

To complete our specification for an individual worker, we have to allow for measurement error. We shall follow most other researchers in assuming that the measurement error is additive and serially uncorrelated:

$$\begin{aligned} y_t^{obs} &= y_t + u_t \\ E\left(u_t\right) &= 0,\ E\left(u_t y_t\right) = 0, \\ E\left(u_t\right)^2 &= \lambda^2 \\ E\left(u_t u_s\right) &= 0, s \neq t \end{aligned} \tag{2.10}$$

We further assume that the measurement errors are normally distributed.

The specification (2.1), (2.8) and (2.10) gives eight *model parameters* per worker:

$$\text{Model parameters: } \{\nu_0, \theta, \alpha, \beta, \delta, \omega, \varphi, \lambda\} \tag{2.11}$$

(plus the common time parameters κ_1 and κ_2 in equation (2.9)). This specification includes almost all models suggested in the literature except that some researchers allow for an $MA(2)$ component.[11] As we shall see below, our preferred specification captures all of the higher order auto-correlations in the data without recourse to an $MA(2)$ term.

2.3. *Allowing for heterogeneity in the model parameters.*

In this subsection we show how we incorporate heterogeneity. A crucial feature of the model is that we allow for correlated heterogeneity. To do this we shall adopt a framework within which we first model the starting value for worker h (which, as discussed in the next subsection, is when the process starts and not necessarily the initial observation), y_{h0}, and then condition further heterogeneity on these values; a discussion of the genesis and advantages and disadvantages of this approach are given after the formalities. The equation for the starting value is:

$$y_{h0} = \tau_1 + \tau_2 z_h + \exp\left(\tau_3 + \tau_4 z_h\right) \eta_{h0} \tag{2.12}$$

9. Previous investigators use a quadratic multiplicative factor to capture the time effects. We prefer the exponential form to ensure that the variance ν_t is always positive.

10. Geweke and Keane (2000) need a mixture of two normals to fit their model. However, they assume that variances are homogeneous whereas we allow variances to vary across agents.

11. The most conspicuous class not covered are copula models; see Bonhomme and Robin (2006). We shall return to this in the concluding section.

where η_{h0} is a standard Normal random variable and z_h is the year of birth of worker h (to allow for the fact that we first observe workers in different years).[12] Note that we allow that both the mean and the variance of the starting values can vary by birth cohort.

In the model developed in the two subsections above we had eight model parameters for each worker (see (2.11)). In all that follows we shall impose that the ARCH parameter φ is the same for everyone but we shall allow that the other seven parameters can be heterogeneous.[13] The parameterisation for the joint distribution for the six model parameters (ν_0, θ, α, β, δ, ω) conditions on the initial value and uses a nonlinear triangular factor structure with 6 latent factors, $(\eta_1, ...\eta_6)$. We take these six factors to be mutually independent standard Normals and independent of η_0 in (2.12). Denoting the inverse logit function by $\ell(z) = e^z/(1+e^z)$, the functional forms we adopt are given by:

$$\begin{aligned}
\nu_{h0} &= \exp\left(\phi_{11} + \phi_{12} y_{h0} + \psi_{11}\eta_{h1}\right) \\
\theta_h &= \ell\left(\phi_{21} + \phi_{22} y_{h0} + \psi_{21}\eta_{h1} + \psi_{22}\eta_{h2}\right) - 0.5 \\
\alpha_h &= \phi_{31} + \phi_{32} y_{h0} + \sum\nolimits_{i=1}^{3} \psi_{3i}\eta_{hi} \\
\beta_h &= \ell\left(\phi_{41} + \phi_{42} y_{h0} + \sum\nolimits_{i=1}^{4} \psi_{4i}\eta_{hi}\right) \\
\delta_h &= \phi_{51} + \phi_{52} y_{h0} + \sum\nolimits_{i=1}^{5} \psi_{5i}\eta_{hi} \\
\omega_h &= \ell\left(\phi_{61} + \phi_{62} y_{h0} + \sum\nolimits_{i=1}^{6} \psi_{6i}\eta_{hi}\right)
\end{aligned} \tag{2.13}$$

so that we restrict $\nu_{h0} > 0$, $\theta_h \in (-0.5, 0.5)$, $\beta_h \in (0,1)$ and $\omega_h \in (0,1)$.[14]

Thus our most general heterogeneity structure allows for seven latent factors (including that for the starting value, η_{h0}, in equation (2.12)); in the empirical analysis below we find that we actually need far fewer than seven latent factors. Even though we have a model that is nonlinear in the factors the system is invariant to the triangular ordering of the factors, because all factors are normally distributed and the nonlinear functions we use preserve the linear index of the factors. To see this, the reduced form of the model parameters $m_h = (\nu_{h0}, \theta_h, \alpha_h, \beta_h, \delta_h, \omega_h)'$ is given by $m_h = g(\Phi Z_h + \Psi\eta_h)$ where $g(.)$ is a (6×1) vector of nonlinear functions $g_j(.)$, $Z_h = (1, y_{h0})'$, Φ is a 6×2 matrix of the ϕ parameters, Ψ is a 6×6 lower triangular matrix (containing the ψ parameters) and $\eta_h \sim N(\mathbf{0}, I_6)$. Thus $\xi_h = \Psi\eta_h \sim N(\mathbf{0}, \Psi\Psi')$. The covariance matrix is $\Psi\Psi'$ and can be represented by the matrix $\tilde{\Psi}\tilde{\Psi}'$ where $\tilde{\Psi}$ is the Cholesky matrix of $\Psi\Psi'$. Since $\tilde{\Psi}$ is upper triangular this is equivalent to writing the model with ω_h depending only on the initial value and one factor, δ_h depending on two factors and so on. The can be done for any re-ordering of the six model parameters.

The unit root model restricts $\beta_h = 1$ for all h (but allows that ω_h may be heterogeneous and less than unity). For the mixture model we take a censored

12. Since we stratify on gender, race and education, birth year is the only remaining observable, heterogeneous and time invariant variable. We also experimented with a mixture of two Normals for the initial distribution and did not find the generalisation significant.

13. We take the ARCH parameter to be homogenous because we do not have a source of variation that would allow us to identify heterogeneity. We present further details in section 3.2.

14. We also allowed the end points of supports for the AR and MA parameters (β and θ respectively) to be estimated. This gave only a small improvement in fit; consequently we prefer to stay with the 'natural' bounds.

parameterisation for the AR parameter:

$$\beta_h = \min\left\{(1+B)\,\ell\left(\phi_{41} + \phi_{42}y_{h0} + \sum\nolimits_{i=1}^{4}\psi_{4i}\eta_{hi}\right), 1\right\} \text{ with } B > 0 \tag{2.14}$$

That is, we take a support of $(0, 1+B)$ for the unrestricted distribution and then censor it at unity.

The final model parameter we have to consider is the measurement error variance, λ. Although most previous researchers assume that this is the same across units ($\lambda_h = \lambda$ for all h) we allow that it may vary across workers. If we make the homogeneity assumption and also assume that the lower bound support of the error variance υ in (2.8) is zero, then the measurement error variance is formally identified. However, the homogeneity assumption for measurement error variances is problematic since there are some workers in our sample who have very little variation in observed log earnings around their idiosyncratic trend. This implicitly gives an upper bound on the common measurement error variance that is very low. To overcome this, we allow that the measurement error variance is heterogeneous but uncorrelated with the other model parameters. To account for the heterogeneity in measurement error variances, we assume that the distribution of the measurement error standard deviations in (2.10) is lognormal:

$$\lambda_h = \exp\left(\varsigma_0 + \varsigma_1\eta_{h\lambda}\right) \tag{2.15}$$

where η_λ is a standard Normal variable. Given the parametric assumptions made above, the parameters $(\varsigma_0, \varsigma_1)$ are identified.[15]

Our approach to modelling the heterogeneity in the model parameters is a generalisation of the approach adopted by Chamberlain (1980), Blundell and Smith (1991), An and Liu (2000), Wooldridge (2005) and Arellano and Carrasco (2003). This approach has several advantages. First, we can establish consistency of our estimator as the number of cross-section units increases, holding the number of time periods constant. This avoids the 'incidental parameters' problem; see Arellano and Hahn (2006) for a discussion of the problems that this causes in estimation in nonlinear panel data models. Second, this approach can accommodate stationary models with the initial conditions given by the process as a special case, but it is not restricted to this. This is particularly useful if the model is, in fact, non-stationary since then the initial values do not have a distribution that is readily related to the process. A third (mundane but extremely important) advantage of this way of incorporating heterogeneity is that it is easy to implement. This was important in our context since we undertook a good deal of exploratory analysis. A fourth advantage, which is particularly emphasised by Wooldridge (2005), is that this procedure allows us to generate quantitative predictions for mean (or quantile) outcomes if something in the underlying process changes. To calculate these from the estimates of the individual earnings processes requires more than consistent estimates of the common parameters of the processes, it also requires a explicit specification of the heterogeneity. If we know the functional relationship between heterogeneity and the starting values then we can calculate the required outcomes, given that we have the starting values. An additional advantage accrues in our case since we model parametrically the marginal distribution of the starting values (as in (2.12) above). In this case we can report *all* of the information needed for anyone to simulate using the estimated process. The main disadvantage of the parametric approach, as compared

15. The 'parametric' qualification is critical here; the nonparametric identification of the measurement error distribution is an open question.

with a semi-parametric approach (which would also give consistency as the number of cross-section units becomes large) is precisely that we have to make some parametric assumptions. The discipline here is that the final model has to fit a wide range of different statistics.

2.4. *Starting values and initial observations.*

The panel data literature has emphasized the importance of modelling initial conditions, especially for panels in which the time dimension is small (see, for example, Arellano (2003)). In this subsection we consider the sampling complication that arises because we do not observe all workers from the start of their process, so that the initial observation in our data is not the starting value y_0. In our PSID sample below about half of the sample are first observed at age 25 but the other half are first observed at a later age. The distinction between the starting value and the initial observation is not always explicitly considered in the earnings process literature (a notable exception is Geweke and Keane (2000)) but it is critical when we have heterogeneous model parameters. To illustrate this point, consider an earnings process that starts at age 25 for all workers (so that y_{h0} corresponds to log earnings at age 25). Suppose that the process for subsequent values is given by a random walk with heterogenous drifts:

$$\Delta y_{ht} = \alpha_h + \varepsilon_{ht} \text{ with } \varepsilon_{ht} \sim N(0, \sigma_\varepsilon^2) \text{ and } \alpha_h \sim N(0, \sigma_\alpha^2)$$

where α_h is independent of y_{h0}. Now allow that some of the sample are observed from when they are 25 years old whilst the rest are observed from when they are 30 years old. The initial observation for a worker h who is observed from when he is 30 is:

$$y_{h5} = y_{h0} + 5\alpha_h + \sum_{t=1}^{5} \varepsilon_{ht}.$$

If we calculate the covariance between the initial observation (y_{h0} for some and y_{h5} for others) and the idiosyncratic drift, α_h, we shall have a non-zero value since for those who are first observed at age 30 we have $cov\,(y_{h5}, \alpha_h) = 5\sigma_\alpha^2 \neq 0$. Thus the sampling and the ignoring of the distinction between the starting value and the initial observation would lead us to erroneously conclude that the drifts are correlated with the starting value. The assumptions made in (2.12) and (2.13) accommodate the distinction between those who are observed from the start of the process (at age 25 for us) and those who have a later initial observation. The assumptions we have made above in (2.13) and (2.12) implicitly assume that the selection into one group or the other is independent of the model parameter distributions, once we condition on date of birth.

3. THE DATA.

3.1. *Sample selection.*

In this study, we use the PSID data for the 26 years from 1968 to 1993.[16] The sample drawn is exactly the same as in Meghir and Pistaferri (2004) (henceforth MP); we select male workers aged between 25 to 57 who are in the sample for at least nine years (for a detailed description see MP). We take age 25 to be the starting age for our process. The process, of course, starts at an earlier age for most high school graduates. The decision to restrict attention to the 'mature' part of the life-cycle was driven by a desire to take a 'standard' data set (for example, Geweke and Keane (2000) who also take 25 as the starting age). As we shall see below, we challenge some inferences that are widely accepted in the literature; by taking a data set that has been used by other authors, we can be sure that our differing conclusions are not because we have different data. It will be clear that if interest centers on, for example, the evolution of post-schooling earnings for high school graduates, then an earlier starting date would be appropriate; the methods described below can easily be extended to that case.

The MP sample consists of $2,069$ individuals, with $31,631$ observations. The earnings variable includes all after tax income from labour, deflated to the year 1992. For individuals in this sample the relevant variables we observe are education, race, age and birth cohort. We deal with some of the observable heterogeneity by stratifying on education and working with the high school sample. Furthermore, we only consider white workers. This gives a sample size of 749 with workers being observed between 9 and 26 years, which gives in total $11,503$ observations. We run a first round regression of log earnings on year dummies and age dummies.[17] In all that follows we work with the residuals from this regression which we shall term log earnings for convenience.

In figure 1 we present two sets of sample paths. The top panel gives the paths from age 25 to 36 for 10 workers who are in the middle of the earning distribution at age 25. The most important feature of this figure is that even for workers who have an almost identical starting value the realisations are diverse. For example, some paths are very volatile whereas others are quite smooth. Additionally the values at age 36 vary a good deal across the sample. Determining whether these differences are due to the random realisations of the earnings process or to a deterministic or stochastic trend is one of the primary purposes of the analysis of earnings processes. The lower panel in figure 1 shows the paths for workers with the 8 lowest and the 8 highest starting values (amongst those observed from age 25); here we have not conditioned on being in the sample at age 36. Once again there is a good deal of variation between those who have similar starting values. Additionally we see that there is a great deal of persistence although most of those who start low catch up somewhat and those who start high tend to decline. One important feature of those who start very low is that most of the 'catching up' seems

16. The value in the survey year actually refers to the previous year (that is, the 1968 value is for the year 1967). We follow the convention in the literature which is to refer to the survey year as the observed year. We cannot take the data beyond 1993 because of changes in the variable definitions. From $1968-1993$ the income variable includes income from farm and unincorporated businesses, while after 1993 these two items are not included. Furthermore, from 1997 the information on income are only collected every second year. This is why most other investigators stop at 1993. Note that we do have a substantial sample of mature workers in our sample, up to age 55. Some of these have been followed for all 26 years of the panel.

17. The first round regression differs from MP since we use a more flexible form in age and race, but we do not include regional dummies or a dummy for residence in a SMSA. We have also tried using the exact same first round regression and the results for the income process are almost unchanged.

[hpb]

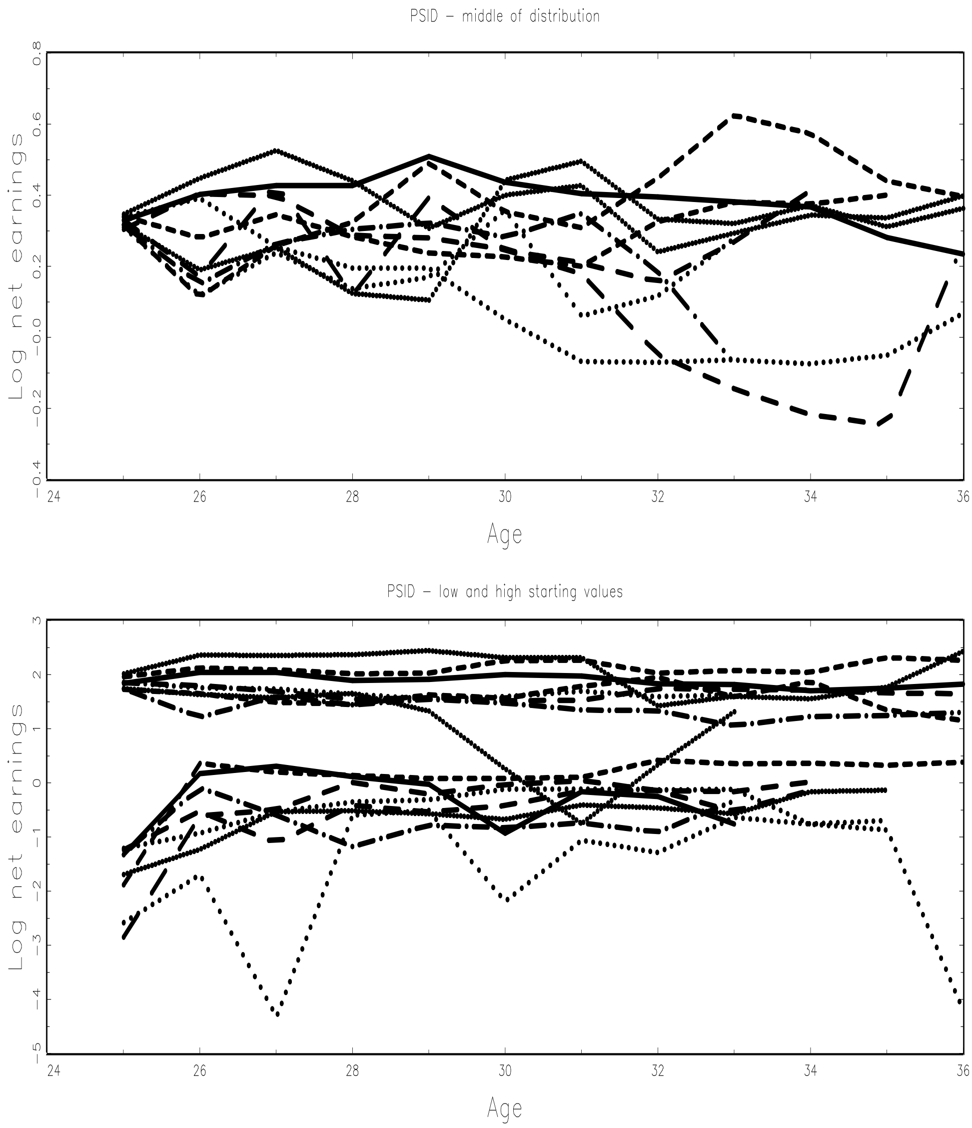

FIGURE 1
Sample paths for age 25-36

to be concentrated in the first few years but they display more volatility than those who start high. The most important conclusion we draw from these figures is that there seems to be a great deal of heterogeneity in the processes driving the realisations and this heterogeneity is dependent on the starting value.

3.2. *The choice of auxiliary parameters.*

In this subsection we present a full data description for our sample. This data description assumes particular importance for us since we use it in estimation and in judging the goodness of fit of our estimated processes. As discussed in the introduction, we shall use Simulated Minimum Distance (SMD) to estimate our various models; details are given below. SMD requires the specification of a set of auxiliary parameters (statistics such as means and covariances, transition probabilities and autocorrelations) which are matched between the observed data and simulated data generated under a particular model specification and a set of parameter values.

In all we use 45 auxiliary parameters (ap's) for our sample; exact details of the construction of the ap's are given in the Appendix Appendix A.2. The choice of statistics we consider is motivated by two considerations. The motivation for the first set of 35 auxiliary parameters is that they provide a detailed description of the distribution of current log earnings conditional on lagged earnings. Specifically, for each worker we run a regression of current log earnings on a constant, lagged log earnings and a trend. We record 5 parameters from each regression: the three coefficient parameter estimates, the log of the residual variance and the first order auto-correlation of the OLS residuals. To illustrate our motivation, consider the OLS slope parameter. For any worker, this is certainly not an unbiased estimator of the 'true' slope parameter but it is closely related to it; this is used in the estimation procedure. More complicated procedures, which includes small sample corrections, would give a closer correspondence to parameters but these are contentious and not widely used. Finally, the OLS route is transparent and quick (as opposed to, for example, Kalman filter based ML estimation of an $ARMA(1,1)$ process for each real and simulated worker); speed is important for a simulation based method such as SMD.[18] We use the mean, variance and covariances of these 5 parameters over the workers and the covariances with the values of the initial observations to create 25 ap's. Following the discussion in subsection 2.4, we also calculate 9 additional statistics to allow for the fact that for about half our sample, the initial observation is not at age 25 (see the Appendix subsection Appendix A.2 for details). Finally we calculate the variance within the bottom quartile of the errors to facilitate estimation of the measurement error variance (see subsection 2.3 below). In all we have 35 auxiliary parameters in this first set. We reiterate that these estimates do not have any specific interpretation and are used simply to capture the important features of the joint distribution of the observed conditional earnings paths.

Whereas the first set of ap's is chosen to facilitate estimation, our second set of ap's is motivated by a very different consideration; these are statistics that are of substantive interest. We wish to make sure that the final model we end up with can account for *all* the results currently in the literature. Since different investigators fit to different statistics, this requires considering a wide number of (correlated) statistics. For example, MaCurdy (1982) and Abowd and Card (1989) model the auto-covariance structure of first differences of log earnings whereas others base their analyses on mobility measures such as short run and long run transition matrices between different quintiles of the earnings distribution. We also use short run and medium run transition measures from the bottom quantile; see Geweke and Keane (2000). The other two major features we wish to capture are the time series properties of the cross-section variance of earnings (see Moffitt and Gottschalk(2002)) and the conditional heteroskedasticity that Meghir

18. We would need at least 540,000 MLE calculations per iteration of the optimisation program for the number of replications we use.

and Pistaferri (2004) identify. In all, we use a wide selection of sets of statistics giving 10 auxiliary parameters of substantive interest; details are given in appendix Appendix A.2 and a discussion is given here.

In the first column of the Table in appendix $A3$ we present the sample values for the full set of ap's (in some cases the values have been rescaled (as shown) to make reading easier). The two values ($vtrend$ and $vtrsq$) give the coefficients from a regression of the cross-section variances on trend and trend squared. These are both highly significant. Taking into account that the trend is measured in decades and that we have multiplied the parameter estimates by 10, the parameter estimates imply that the cross-section variance first increases from 0.09 in 1968 to 0.24 in 1985 and then decreases to 0.20 in 1993. This is qualitatively similar to the result of Moffit and Gottschalk (2002) (see also figure 2(b) in Meghir and Pistaferri (2004)). It is also consistent with the widespread finding that the inequality of earnings increased through our sample period (see, for example, Buchinsky and Hunt (1999)) but note that finding is for the population as a whole and not following the same group through time.

The three statistics $(dvar, dauto1, dauto2)$ are the variance and first two autocorrelations for first differenced log earnings; these correspond to the statistics used in Abowd and Card (1989). The variance is 0.078 which represents quite high volatility for growth. The first two auto-correlations are -0.26 and -0.06 respectively; these autocorrelations are qualitatively similar to Abowd and Card (1989) (see, for example, their PSID sample of males from 1969-1979 with the SEO sub-sample excluded, see their Table V). The ranges (over years) of the Abowd and Card statistics are: $dvar \in [0.09, 0.20]$, $dauto1 \in [-0.54, -0.10]$ and $dauto2 \in [-0.15, -0.005]$. Thus our data (which are more homogeneous than those of Abowd and Card)) shows a good deal less variance in growth but similar autocorrelations.

The next three rows $(arch1, arch2, arch3)$ give statistics on the conditional heteroskedasticity; as can be seen, these are all 'significant'. The ARCH ap's are based on moment conditions used in Meghir and Pistaferri (2004), who estimate a homogenous ARCH effect for each worker. We also investigated alternative (and simpler) ap's for ARCH such as the autocorrelation of $(\Delta y_{ht})^2$, and we find that our preferred model also fit these well (see Appendix Appendix A.2 for details). However, none of the suggested ARCH statistics in the literature can be used to detect heterogeneity in the ARCH parameter and therefore we refrain from include heterogeneity in the ARCH parameter. The final two auxiliary parameters $(p(t, t+1), p(t, t+10))$ show the short run and long run persistence of low earnings (here defined as being in the lowest quintile). We see that about three quarters of workers in the bottom quintile in any year are also in there in the subsequent year but only one half are still in the bottom quintile ten years out.

3.3. *Estimation procedure.*

We use Simulated Minimum Distance (SMD) to estimate the model. Appendix Appendix A.4 gives the details of the estimation procedure. There are several advantages to using SMD rather than CML or GMM techniques. The main advantage is that it is very easy to use since we need to conduct only informal prior analysis of the relationship between the model and the data. This simplicity is particularly important in exploratory analysis in which we examine a number of quite different models in order to capture the heterogeneity in the processes. Although it is possible to derive a likelihood function for some of models we consider, it would be very arduous. It would also be disheartening since we typically discard any model quite quickly (since simple models do not fit the data). A second and closely related advantage is that SMD can be used even when the likelihood function is very difficult (or even impossible) to formulate. For example, in the models below we wish to make allowance for considerable correlated heterogeneity, heterogeneous measurement error, an unbalanced panel and for ARCH effects. A third advantage is that we can explicitly account for the sampling scheme; in our case we can readily allow that for some workers we do not observe them from the starting date of their process. A fourth advantage is that we can fit to the statistics of the data that are of direct substantive interest. For example, for earnings processes we often interested in the dynamics of low earnings spells so statistics that capture this are natural choices to include in our set of auxiliary parameters. A final advantage is that when a simple model fits badly the SMD procedure often suggests a very natural dimension in which to generalise the model. Of course, there are also drawbacks. The first of these is that we need to specify a set of auxiliary parameters to fit to, which has a certain *ad hoc* quality.[19] Second, the procedure is inefficient relative to maximum likelihood (that is, it will not generally attain the CR lower bound unless the ap's are particularly well chosen).

As discussed at the beginning of subsection 2.1 we started with a very restricted model (the simple unit root model) and built up to a general model by adding new features to capture the main shortfalls in goodness of fit. Having done this, we then generalised that model to the wider encompassing model that is given in section 2. For this encompassing model there are many 'insignificant' heterogeneity distribution parameters (2.13). To drop these we adopt a stepwise procedure. The first step is to take the encompassing variant and drop each parameter, one at a time. The parameter that gives the smallest rise in the criterion is then excluded from the model and we continue to the next step. This repeats the procedure and leads to excluding another parameter. The stepwise dropping of parameters continues until the increase in the criterion from dropping a parameter is 'large' (details are given in the next subsection). For the stable model, for example, we exclude 17 heterogeneity parameters in this way.[20]

4. RESULTS.

4.1. *The fit of different models.*

Table 1 gives the goodness of fit (gf) statistics for the important variants of the encompassing model. We present results for the three broad classes of models: stable, mixture and unit root. For the stable model we present the results for the most general model (labelled *SG*) and the preferred restricted variant (labelled *SP*) (the exact form of the preferred stable model is given at the beginning of the next subsection). Additionally, we present results for the preferred model with $\omega = \beta$ (the conventional *AR* model with full heterogeneity in parameters), labelled *SN*. For the mixture model (labelled *MP*) we present results for the preferred variant with a value of $B = 0.2$; see (2.14). This value of B is set such that a reasonable fraction of the workers (in this case, 22%) have a unit root. This is for illustrative purposes only; the optimal value of B is very close to zero and only a very small fraction of workers have a unit root at this optimal value. For the unit root models we present the general version (labelled *UG*) and the simple unit root model ($\beta = \omega = 1$ and $\delta = \alpha = 0$), labelled *US*.

19. Exactly the same can be said of most choices of moments to fit for GMM.
20. This procedure is exceedingly time consuming and it is not feasible to investigate the properties of the model selection procedure by, for example, bootstrapping.

[htbp]

Model:	*Stable*			*Mixture*	*Unit root*	
	SG	*SP*	*SN*	*MP*	*UG*	*US*
# parameters	42	23	21	23	34	14
Degrees of freedom	3	22	24	22	11	31
χ^2 *statistic*	17.3	21.1	33.4	22.0	89.3	144.2
Probability (%)	0.06	51.4	9.7	44.7	0.00	0.00
G = general variant; P = preferred variant;						
N=no omega (preferred); S = simple unit root model						

TABLE 1

Goodness of fit for different models

Referring to Table 1 we see clear and strong statistical evidence in favour of the stable model with no unit roots for anyone. In particular, according to the comparison shown in the Table in appendix A.3, the main dimensions in which the general stable model is superior to the general unit root models are the distribution of the variance of the transitory shocks (mean and variance), the residual autocorrelation (mean), the distribution of earnings at age 25 (variance), the autocorrelations of the first differences and one of the ARCH parameters. A further dimension in which the fit is much better is for the common time trends for the variances. The mixture model with $B = 0.2$ fits considerably worse than its stable counterpart. The general unit root model[21] is strongly rejected against the general stable model (a $\chi^2(17)$ statistic of 72.0) and the simple unit root model is, in its turn, strongly rejected against the general unit root model (a $\chi^2(20)$ of 54.9).[22]

For the preferred stable model, the two additional parameters for the ω distribution are highly significant (comparing SN with SP gives a $\chi^2(2)$ statistic of 12.3). We have also conducted a test for whether an $ARMA(2,1)$ would do as well as our novel ω scheme. If there is a significant second order AR term then it would show up for older workers, for whom our ω-effect would be very close to zero. To test for this, we constructed ap's that captures the first and second order autocorrelations for workers when they are aged 35 or above (selecting on having at least 8 observations above this age). This includes about half of the original sample. The moment $\chi^2(2)$ test statistic is 1.32 (with insignificant values for each component). Thus the current scheme correctly predicts the first and second order auto-correlations for older workers without any need for a second order process.

In the next subsection we present a detailed account of the preferred stable model. In particular, we present all of the information that a reader needs to generate individual earnings processes from age 25 to 55 for our population of white male high school graduates who were born between 1921 and 1960.

21. Recall that this is the weak form of the unit root hypothesis: $\beta = 1$ (that is, the shocks cumulate). This is the only restriction from the stable model since we do allow for heterogeneous variances, drifts and ω's. (see the discussion after (2.1)).

22. Our 'simple' unit root model is actually more general than most used in the literature since it allows for heterogeneous variances and MA parameters.

4.2. *Parameter estimates.*

4.2.1. Distribution parameters.. The preferred stable model is given by:

$$\begin{aligned} y_{h0} &= \exp(\tau_3)\,\eta_{h0} \\ \nu_{h0} &= \exp(\phi_{11} + \phi_{12}y_{h0} + \psi_{11}\eta_{h1}) \\ \theta_h &= \ell(\phi_{21} + \psi_{22}\eta_{h2}) - 0.5 \\ \alpha_h &= \phi_{31} + \phi_{32}y_{h0} + \psi_{31}\eta_{h1} \\ \beta_h &= \ell(\phi_{41} + \psi_{41}\eta_{h1} + \psi_{42}\eta_{h2}) \\ \delta_h &= \phi_{51} + \phi_{52}y_{h0} + \psi_{51}\eta_{h1} + \psi_{52}\eta_{h2} \\ \omega_h &= \ell(\phi_{61} + \psi_{62}\eta_{h2}) \end{aligned} \tag{4.16}$$

In Table 2 we present the heterogeneity distribution parameter estimates. The $\chi^2(1)$ statistics in the final column are the quasi-LR values for excluding the corresponding parameter. One feature to note is that two of the distribution parameters for the novel model parameter, ω, we have introduced, ϕ_{61} and ψ_{62}, are highly significant. Note also that the heterogeneity term in the measurement error standard deviation (ς_1) is highly significant. The estimates of the measurement error parameters imply a mean of 9.2% for the measurement error distribution; this value is comparable to the value found by MP assuming a homogeneous measurement error variance.

One important aspect of our results is that we did not find any significant cohort effects for the initial distribution variance (that is, $\tau_4 = 0$); this is mainly because we have allowed for an exogenous time trend in variances. The variance (ν_0), the trend term (α) and the long run mean term (δ) all depend on the starting value (that is, $\phi_{i2} \neq 0$ for these model parameters). Although we have 6 heterogeneous model parameters we only need two latent factors for them, η_1 and η_2 (that is, $\psi_{ij} = 0$ for $j > 2$). Thus we have a relatively simple structure with three latent factors (including the starting value as a latent factor) and relatively parsimonious dependence of the heterogeneity in model parameters (as compared with the general model (2.13)). To complete the specification, the common age pattern for log earnings (which we took out in a preliminary regression) is very well described by the quadratic:

$$\Delta_a = 8.83 + 0.56a - 0.057a^2 \tag{4.17}$$

where a is age in decades. Given this, the equation for the starting distribution (2.12), the first order process (2.1) and the parameter estimates in Table 2 we can simulate series for log earnings. Taking the exponential of the resulting series gives earnings in thousands of 1992 dollars.

4.2.2. Model parameters.. Although the distribution parameters are of some interest, we are more interested in the model parameters. In Table 3 we present some summary statistics for these (and for the starting value) and the correlations between them. The first three rows present statistics of the marginal distribution of each model parameter. The most important feature to note is that most of the model parameters are highly dispersed; this justifies our approach of allowing for lots of heterogeneity. We shall discuss the variance of the short run shocks, υ, in detail in the next subsection. The MA parameters, θ, are centred above zero and only 33% of the population having a negative value. This contrasts with most previous estimates which have a negative (homogeneous) MA parameter. This 'change of sign' is attributable to the allowance for heterogeneous measurement error. The AR parameters, β, are widely dispersed with a median value of

[htbp]

Distribution Parameter	Corresponding model parameter	Estimated value	χ^2 (1) for exclusion
τ_3	y_0	−0.922	264*
ϕ_{11}	ν_0 (variance)	−4.684	1000+
ϕ_{12}	ν_0	−0.856	10.8
ϕ_{21}	θ (MA coefficient)	0.297	5.3
ϕ_{31}	α (trend)	0.045	114
ϕ_{32}	α	−0.038	14.6
ϕ_{41}	β (AR coefficient)	1.737	157
ϕ_{51}	δ (long run mean)	−0.319	3.3
ϕ_{52}	δ	0.912	5.1
ϕ_{61}	ω (initial adjustment)	3.510	404
ψ_{11}	ν_0	1.481	152
ψ_{22}	θ	−0.674	13.2
ψ_{31}	α	0.043	35.8
ψ_{41}	β	−0.732	28.5
ψ_{42}	β	1.435	35.3
ψ_{51}	δ	−0.931	27.7
ψ_{52}	δ	−0.487	15.2
ψ_{62}	ω	−0.772	10.3
φ	φ (ARCH)	−0.428	9.5*
ς_0	λ (measurement error)	0.684	75.0
ς_1		−2.693	1000+
κ_1 κ_2	Time trend in variance	0.052 −0.006	11.4 22.0
* value for setting parameter to −100			

TABLE 2

Parameter estimates for stable model

0.85. A significant fraction of the population have persistent shocks; for example, 39.4% of the population have an AR parameter above 0.9 and the ninth decile is 0.98. The median trend (the α parameter) is 4.4% per decade and some of the population have very high growth; for example, 10% of the population have a trend of above 10% per decade. The median long run (net of trend) mean, δ, is -0.323 which is below the mean of the starting distribution. This does not imply that earnings are falling: we have taken out the common nonlinear age effect in the initial regression (see (4.17)) and, as we have seen, the idiosyncratic trends are generally positive. Finally, the initial adjustment terms, ω, are mostly close to unity implying that the initial adjustment to the 'long run' process is quite slow (see equation (2.2)). From the Table it will be quite clear that the distribution of the persistence of shocks (as measured by the AR parameter) is quite different to the distribution of the decay of the effect of the initial value. This indicates that we do need the extra parameter, ω, in (2.1). Moreover, omitting this term biases us toward finding AR parameters closer to unity, particularly for samples that contain younger workers. This would lead to an under-rejection of the conventional unit root hypothesis.

Turning to the correlations between model parameters, we see that there are several large (absolute) values. This is to be expected given that we have only two latent factors and many model parameters that are correlated with the starting value. There is a high correlation between the long run mean (δ) and the trend (α). This is close to -1 indicating that a high positive trend is associated with a low long run mean, net of the trend. The overall impact, also taking into account the correlations with the starting values are difficult to visualise; we shall return to this in the next subsection. There is also a strong negative correlation between the AR and MA parameters. The strong positive correlation between the trend and the variance indicates that those with a steeper trajectory experience much higher volatility.

4.3. *Outcomes of interest.*

The results presented above have implications for all uses of earnings processes but here we concentrate attention on just four outcomes of interest. These are the distribution of the short run variance of earnings; short run and long run mobility out of low earnings; the level and dispersion of lifetime earnings and the evolution of the cross-section inequality in earnings with age. The first of these is an important input for saving and consumption simulation models. Mobility out of low earnings is of intrinsic interest. The third outcome is increasingly recognised as being an important element in school and career choices. The final outcome is of interest for theories of human capital that emphasise the early trade-off between wages and human capital accumulation.[23] For each outcome we compare the predictions from the preferred stable model, the simple unit root model and the general unit root model.

Turning first to short run variability, this is given by[24]:

$$\text{std of shock for worker } h = \sqrt{(1+\exp(\varphi))\, v_{0h}} \tag{4.18}$$

The top panel of Table 4 compares the distributions of this for the preferred variants of our three models. As can be seen, there are marked differences across models. In particular, the simple unit root (which imposes homogeneity on the variance of the shocks) shows

23. See Rubinstein and Weiss (2005), section 4.2, which discusses the significance of the shape of the cross-section variance against age curve for distinguishing between human capital models and alternatives such as search and learning.

24. We use the value for 1968; that is, with the trend set to zero.

	y_0	$\sqrt{\nu_0}$	θ	α	β	δ	ω
Marginal distributions							
1st decile	−0.509	0.037	−0.139	−0.012	0.403	−1.730	0.927
Median	0	0.096	0.076	0.044	0.851	−0.323	0.971
9th decile	0.509	0.247	0.265	0.100	0.980	1.090	0.989
[tbp] Correlations							
corr(y_0)	1	−0.183	−0.007	−0.348	0.002	0.318	−0.004
corr($\sqrt{\nu_0}$)	–	1	0.006	0.846	−0.379	−0.757	0.005
corr(θ)	–	–	1	−0.078	−0.808	0.434	0.901
corr(α)	–	–	–	1	−0.394	−0.897	0.009
corr(β)	–	–	–	–	1	−0.012	−0.607
corr(δ)	–	–	–	–	–	1	0.386

TABLE 3

The parameter distribution

much lower volatility. For the preferred stable model, the short run variability is very dispersed and skewed; this is in line with the qualitative results on subjective perceptions given in Dominitz and Manski (1997). Although the median is relatively low, over 10% of the population have a short run standard deviation of over 0.29. These estimates imply that in any period the chance of a 20% fall in earnings is 1% at the median and 20.4% for the top decile. Thus the significance of the precautionary motive for saving is highly skewed and very strong for a small proportion of the population. As already mentioned, the short run variance is strongly positively correlated with the trend, α, and strongly negatively correlated with the long run mean, δ; the overall impact of this on lifetime outcomes will be examined below.

The second outcome of interest we examine is mobility out of the bottom earnings quintile. The lower panel of Table 4 reports transition probabilities of movements at the bottom of the distribution. The first two measures relate to the probability of staying in the bottom quintile after one and ten years conditional on being there at age 25. The final measure gives the probability of being in the bottom quintile at age 50 given that the worker was in the bottom quintile at age 35. Once again, the three models give different predictions with the unit root generally giving less mobility. An important motivation in Geweke and Keane (2000) for the introduction of a mixture of Normals for the shocks was the overestimation of the persistence of low income. As can be seen, allowing for lots of heterogeneity achieves the same outcome. One interesting feature of these results is that the transition probabilities are dependent on age; the probability of moving out of the bottom quintile is much higher for those in the bottom quintile at age 25 rather than those initially aged 35. Indeed having relatively low earnings seems to be very persistent

[tbp]

	Preferred stable model	General unit root model	Simple unit root model
Short run variability			
Standard deviation of shocks (allowing for $ARCH$)			
First decile	0.043	0.041	0.024
Median	0.113	0.105	0.077
Ninth decile	0.298	0.272	0.243
Mobility			
Transitions from bottom quintile to bottom quintile (probability (%)):			
age 25 → 26	70.5	73.5	74.5
age 25 → 35	37.6	37.1	50.1
age 35 → 50	53.3	54.4	62.6

TABLE 4

The parameter distribution

after age 35 with only a 47% chance (for the preferred stable model) of being out 15 years later.

To examine the implications of our models for lifetime income we simulate the paths of earnings from age 25 to 55 for synthetic workers who were born in 1945. We first simulate $1,000$ values for the starting values and calculate 33 equally spaced percentiles between the first decile and the ninth decile (that is, the 10%, 12.5%, ... 87.5% and 90% percentiles of the starting values). For each of these percentiles we simulate $100,000$ earnings paths, without measurement error (giving $3.3 * 10^6$ paths in all). We then take these paths and add in the age dummy coefficient values that we took out in the original regression. Finally we discount each of the paths back to age 25 using a 3% real rate and divide by 31 to give annual values in 1992 thousands of dollars. We then calculate the median and interquartile range of the earnings paths for each of the 33 percentile starting values. The three curves for the same models as in the previous paragraph are shown in figure 2. The ranges of median annualised lifetime earnings (from the bottom decile to the top decile) are reflected in the horizontal lengths of the curves (the full range is given by the x-axis). In each case, the value corresponding to the first (respectively, ninth) decile of the starting distribution is the lowest (respectively, highest) median value shown. It can be seen that the three curves are quite different. First, the support of the medians are quite different, with the simple unit root model showing a much larger dispersion. The simple unit root model is almost linear and strictly decreasing. The general unit root model displays much more variance than either of the other two and a lower range of median values. The preferred stable model is non-monotone with higher interquartile ranges at the top and bottom of the median distribution and a minimum at close to the median starting values. This reflects what we saw in the raw data in figure 1; in particular, the higher variance seen for those with extreme starting values. If the curves

[hpb]

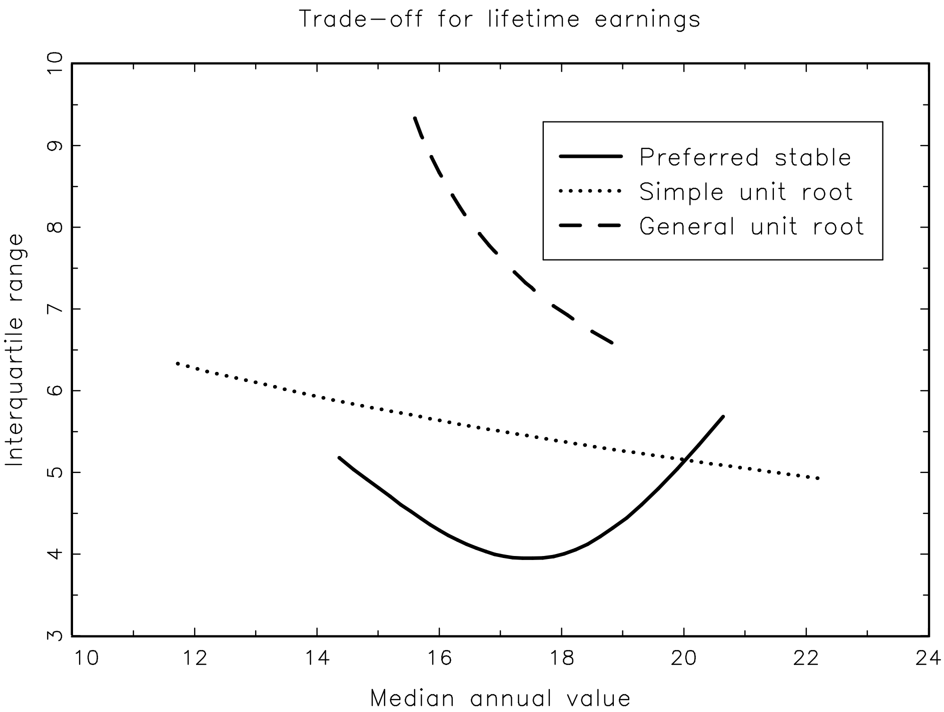

Figure 2
The trade-off in lifetime income between median and IQR.

in figure 2 are taken as inputs to models of schooling choices (as in Cunha *et al* (2005)), the results are likely to be very sensitive to the model chosen.

The final outcome we consider is for the cross-section variance of log earnings ('inequality') over age; see figure 3. These figures are based on $500,000$ simulated paths using the model starting values and estimated parameters but without introducing measurement error. Rubinstein and Weiss (2006) argue that different models of wage growth will lead to different developments in the cross-section variance. See, in particular discussion of the significance of the shape of the cross-section variance against age curve for distinguishing between human capital models and alternatives such as search and learning. For all models we have considered the cross-section variance is increasing with age and for all it is not too different from linear. The stable model has a much lower trend than the other models. Again this shows that for some outcomes the predictions of the types of models are very different.

5. CONCLUSIONS.

We have considered the evolution of earnings over age and introduced three novelties. First, we allow for extensive heterogeneity in all model parameters (except the ARCH

[hpb]

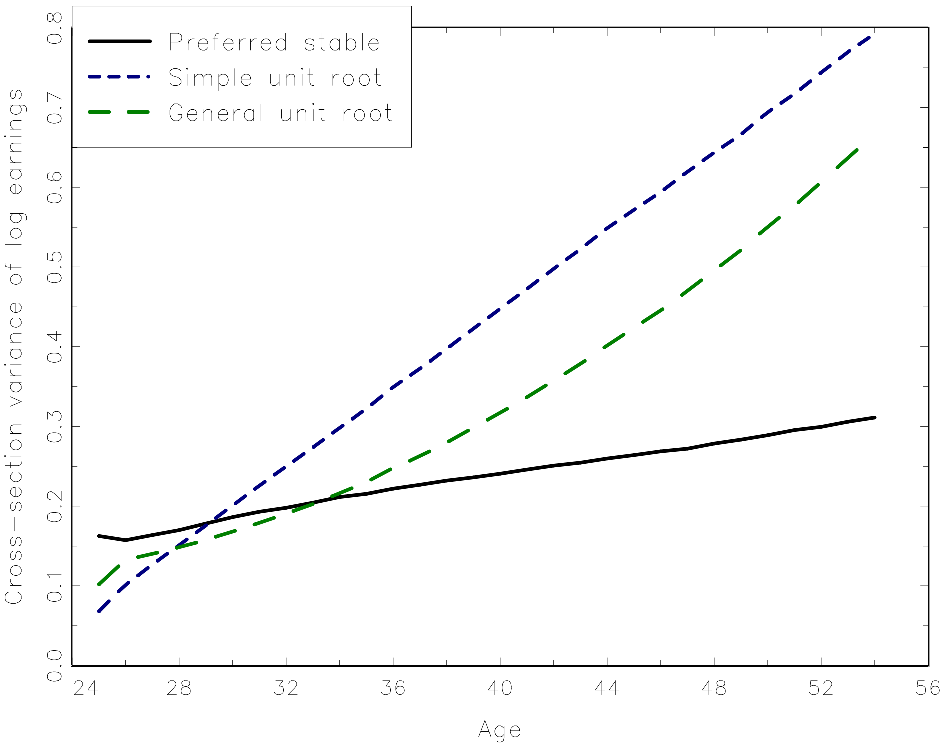

FIGURE 3
The evolution of inequality over the life-cycle.

term). We do this by introducing a nonlinear fully parametric multi-factor model. To estimate the model we employ a simulated minimum distance estimator. Second, we have introduced an extension of the conventional ARMA model to break the usual unit root hypothesis into its two constituent components. For the stable model this generalisation severs the link between the initial convergence to the 'long run' process and the AR parameter. Third, we allow for heterogeneous measurement error. Applying our methodology to a standard PSID sample we have a number of findings.

- We find much more heterogeneity in earnings processes than previous investigators have allowed for. Not allowing for lots of heterogeneity means that we miss important features of the data. Concentrating only on certain features of the data (for example, transition probabilities or the auto-covariance matrix of first differenced log earnings) gives a misleading picture of the need for heterogeneity.[25]

25. One implication of our finding is that transition probabilities vary across the earnings distribution. An alternative approach to capturing this is the copula approach of Bonhomme and Robin (2006). The relative merits of the two approaches in terms of fitting the conditional earnings process have not been systematically explored, but our approach has the advantages that the estimated structure is

- Unit root models are decisively rejected. Even a model with a mixture of a unit root and a heterogeneous AR parameter does no better, statistically, than the model with everyone having an AR parameter below unity. The model with everyone having an AR parameter below unity (the stable model) does best. This conclusion is in stark contrast to many recent papers in the literature which assume that everyone has a unit root. This divergence can be attributed to two factors. First, the extra degree of freedom that we allow for in our extended ARMA framework (captured by the ω parameter) means that the initial convergence of the process can be quite slow (implying values of ω close to unity) without simultaneously imposing that the AR parameter has to be close to unity. Second, our allowance for extensive heterogeneity. Finally, we should note that no previous investigators have used such a demanding range of goodness of fit measures so that most tests for a unit root may lack power
- The stable process is generally quite persistent with a median AR parameter of 0.85 and 39% of the population having an AR parameter above 0.9.
- There is significant and heterogeneous measurement error. The heterogeneity is statistically significant and has an important impact on the stochastic component of the process with lower variances for the shocks and a generally positive MA parameter for the model that allows for heterogeneous variances for the measurement error.
- The strongest heterogeneity is in the variances of short run of shocks; this distribution is very skewed This implies that some of the population experience much more earnings variability than the median worker.
- The MA term is heterogeneous and mostly positive. This contrasts with the usual finding of a negative (homogeneous) MA term.
- We follow the recent literature and allow for an *ad hoc* time varying aggregate increase in the variance of shocks. It is still an open question as to whether this is a genuine period effect or reflects a combined age and cohort effect.

Our results represent a fairly radical break with many aspects of past modelling of earnings processes and we see a great deal of work to be done. Amongst the issues on the agenda are the following. It would be desirable to estimate age-cohort-period effects simultaneously with the earnings process, rather than using the two step procedure that takes out the mean effects in a first regression. Many of the papers on earnings trajectories also consider the links to unemployment and hours fluctuations and with time varying covariates such as family composition and marital status. It would be desirable to use our approach with lots of heterogeneity for these associated processes. The factor approach we have adopted is well suited for this. Similar remarks apply to modelling jointly income and consumption processes. Finally, we have made strong functional form assumptions to estimate the conditional distribution of earnings. It would be desirable to relax these assumptions. Ideally we should find what distributions are identified nonparametrically and use a semiparametric estimation scheme. But note that we have found a scheme that gives a very good fit to the data and there are limits on what would be gained by loosening the specification.

very familiar (an extended ARMA) and the results can be used directly in simulations.

APPENDIX A. APPENDIX.

Appendix A.1. *The simple unit root model as an error component model*

In many derivations, log income is assumed to be the sum of a 'permanent' component, p_{ht}, and a transitory component, e_{ht}:

$$y_{ht} = p_{ht} + e_{ht} \tag{A19}$$

The permanent component is modelled as a random walk:

$$p_{ht} = p_{ht-1} + \zeta_{ht} \qquad \zeta_{ht} \sim iid(0, \sigma^2_\zeta)$$

By assumption the transitory component has low persistence and is modelled as a $MA(q)$-model. The simple unit root model as we model it here is equivalent to such a model with $q = 0$, so that e_{ht} is serially uncorrelated. Furthermore, we assume that $e_{ht} \sim iid(0, \sigma^2_e)$. The log income process (A19) can then be formulated as:

$$\Delta y_{ht} = e_{ht} - e_{ht-1} + \zeta_{ht}$$

This model can be seen as a special case of the simple unit root model:

$$\begin{aligned} \Delta y_{ht} &= \varepsilon_{ht} + \theta\varepsilon_{h,t-1} \text{ with } E(\varepsilon_{ht}) = 0 \text{ and } E(\varepsilon_{ht}\varepsilon_{hs}) = 0 \text{ for } t \neq s \\ V(\varepsilon_{ht}) &= \sigma^2_\varepsilon \end{aligned}$$

where the following equations are satisfied:

$$\begin{aligned} \sigma^2_e &= -\theta\sigma^2_\varepsilon \\ 2\sigma^2_e + \sigma^2_\zeta &= (1+\theta^2)\sigma^2_\varepsilon \end{aligned}$$

Note that this requires $\theta < 0$. Additionally, if we allow that the error variances in the permanent-transitory model (σ^2_e and σ^2_ζ) are heterogeneous, then the parameters σ^2_ε and θ will also be heterogeneous.

Appendix A.2. *The construction of the auxiliary parameters.*

The first set of auxiliary parameters are based on OLS regressions of current log earnings on lagged log earnings and a trend for each of our H workers individually:

$$y_{ht} = \beta_{0h} + \beta_{1h} y_{ht-1} + \beta_{2h} t + u_{ht},\ t = 2, ..., T_h$$

where T_h denotes the number of periods for which we observe h. We term the three parameter estimates (IN,SL,TR) for intercept, slope and trend.[26] For each worker we also calculate the log of the residual variance (LV) and the first order auto-correlation of the OLS residuals (AU). We also take the initial observation, if this is recorded at age 25 (that is, if the initial observation is also the starting value, see subsection 2.4), with a missing value otherwise; we denote this $Y1$ This gives a $H \times 6$ matrix of values and estimates: ($Y1$,IN,SL,LV,AU,TR). We first calculate the mean and standard deviation of each of these (denoted in our results by $M(.)$ and $S(.)$ respectively). For example, the mean of the log variances is:

$$M(LV) = \frac{1}{H}\sum_{h=1}^{H} \ln\left(\frac{1}{T_h - 4}\sum_{t=2}^{T_h}(\hat{u}_{ht})^2\right)$$

The next 15 statistics are the *correlations* between ($Y1$,IN,SL,LV,AU,TR). The latter allow us to identify dependencies between heterogeneity distributions and with the initial observation; see subsection 2.3. We denote the correlations between, say, $Y1$ and SL by $C(Y1SL)$.

The next six ap's are designed to allow for the fact that we do not observe everyone from age 25. To do this we compute the correlation between ($Y1$,IN,SL,LV,AU,TR) and the year of birth (denoted YO). The first of these, $C(Y1YO)$ allows us to identify cohort effects (if any) in the level of the initial value. To pick up cohort effects in the dispersion of the initial value we include an auxiliary parameter defined by:

$$C(YSYO) = \sum\nolimits_{h \in H(25)} (y_{h1} - \bar{y}_1)^2 \left(b_h - \bar{b}\right) \tag{A20}$$

26. The trend coefficients are multiplied by 10 to make the results easier to read.

where $H\,(25)$ is the set of workers who are first observed at age 25 and b_h is the year of birth for worker h. The final OLS based ap is for the identification of the variance of the measurement error; for this we take the mean of the lowest quintile of the pooled residual variances.

In total, this gives 35 $(= 12 + 15 + 6 + 1 + 1)$ ap's for the data description.

The other 10 auxiliary parameters we calculate are to capture particular statistics of interest in other papers in the literature. The first set of these captures the change in the dispersion of the distribution of earnings over time.[27] The interest here lies in the time series trend in inequality emphasised by Gottschalk and Moffitt (1994). Specifically, we calculate the cross-section unconditional variance in each year and then regress these time series of statistics on a trend (measured in decades) and trend squared and record the coefficient value on the trend and trend squared. These are denoted $vtrend$ and $vtrsq$ respectively.

The next three auxiliary parameters are based on the time series of differenced data; these statistics are included for comparability with MaCurdy (1982) and Abowd and Card (1989) and authors who follow them in basing their estimates on the auto-covariance matrix of first differenced log earnings. We take first differences for each worker $\Delta y_{ht} = y_{ht} - y_{h(t-1)}$ and then record the mean across the sample of the variance and the first two auto-correlations of these first differences. We denote these $dvar$, $dauto1$ and $dauto2$.

The next set of auxiliary parameters are three ARCH statistics that are chosen to capture exactly the departures from homoskedasticity that Meghir and Pistaferri (2004) identify. Specifically we construct deviations from their cross-section means of the following three statistics: $(\Delta y_{ht})^2$, $(\Delta y_{ht}\Delta y_{ht-1})$ and $(\Delta y_{ht}\,(y_{ht+1} - y_{ht-2}))$. Then we record the first order auto-correlations for each of these as $arch1$, $arch2$ and $arch3$. Alternative ap's such as the autocorrelation of $(\Delta y_{ht})^2$ could also have been used. We test the fit of the first, second and third autocorrelation of $(\Delta y_{ht})^2$ for our preferred specification. We find that the $\chi^2(3)$ statistics based on the differences between the data and simulated model is 4.83. Thus the preferred model fits these ap's.

Finally we include two mobility measures. Since the usual concern is with the duration of low income spells we concentrate on that. The statistics are the mean over years of the proportions of those in the bottom quintile in the one year and in the next year and the proportion of workers who are in the bottom quintile one year and ten years later. These two measures pick up the short run and long run persistence of low earnings. These three ap's are denoted $p(t, t + 1)$ and $p(t, t + 10)$ respectively.

Counting up we see that we have a total of 45 statistics: 35 OLS based statistics (including the ap's to pick up cohort effects and the ap to pick up low residual variances), 2 trend coefficients, 3 means of first differenced statistics, 3 ARCH statistics and 2 mobility measures.

Appendix A.3. *The values of the auxiliary parameters.*

The Table in this appendix presents the data and simulated value for the 45 auxiliary parameters and for three classes of models: the simple unit root model given ($\beta = \omega = 1$ and $\delta = \alpha = 0$ and (ν_0, θ) heterogeneous); the general unit root model ($\beta = \omega = 1$ and $(\nu_0, \theta, \delta, \alpha)$ heterogeneous) and the preferred stable model (all parameters heterogeneous and the model selection procedure imposed). The definitions of the ap's can be found in appendix $A2$.

27. Recall that in the first round we regress on time dummies so that the mean of the residuals in each year is zero. Thus we do not have to consider changes in the mean over time.

[p]

Auxiliary parameter fits				
Ap Name	Data value	Simple unit root	General unit root	Preferred stable
M(Y1)	0.000 (0.024)	0.013 [0.347]	0.001 [0.014]	0.000 [0.012]
M(IN)	0.003 (0.018)	0.020 [0.624]	0.011 [0.294]	0.011 [0.290]
M(SL)	0.192 (0.013)	0.150 [2.123]	0.154 [1.917]	0.189 [0.167]
M(LV)	−0.203 (0.002)	−0.213 [2.791]	−0.210 [1.923]	−0.205 [0.603]
M(AU)	−0.018 (0.006)	−0.034 [1.832]	−0.036 [2.099]	−0.024 [0.698]
M(TR)	−0.015 (0.014)	−0.007 [0.377]	−0.010 [0.198]	−0.021 [0.307]
S(Y1)	0.473 (0.034)	0.312 [3.186]	0.365 [2.122]	0.416 [1.123]
S(IN)	0.502 (0.026)	0.561 [1.479]	0.541 [0.981]	0.489 [0.319]
S(SL)	0.356 (0.011)	0.368 [0.773]	0.366 [0.619]	0.358 [0.131]
S(LV)	0.066 (0.002)	0.060 [1.984]	0.059 [2.482]	0.064 [0.564]
S(AU)	0.161 (0.005)	0.153 [0.887]	0.145 [1.851]	0.163 [0.269]
S(TR)	0.397 (0.032)	0.405 [0.181]	0.442 [0.949]	0.389 [0.153]
C(Y1IN)	3.386 (0.500)	3.112 [0.367]	4.068 [0.910]	2.954 [0.577]
C(Y1SL)	0.249 (0.363)	−0.376 [1.149]	−0.196 [0.818]	−0.187 [0.801]
C(Y1LV)	−1.564 (0.430)	−2.257 [1.074]	−1.875 [0.483]	−1.173 [0.606]
Values in (.) are standard deviations				
Values in [.] are absolute t-values				

Auxiliary parameter fits (continued)

Ap Name	Data value	Simple unit root	General unit root	Preferred stable
C(Y1AU)	0.262 (0.473)	0.102 [0.225]	0.170 [0.130]	0.054 [0.294]
C(Y1TR)	−1.414 (0.567)	0.232 [1.934]	−2.401 [1.159]	−1.409 [0.006]
C(INSL)	0.269 (0.421)	−0.309 [0.916]	−0.201 [0.745]	−0.020 [0.457]
C(INLV)	−1.114 (0.536)	−1.166 [0.064]	−1.297 [0.228]	−0.556 [0.695]
C(INAU)	−0.378 (0.380)	−0.389 [0.021]	−0.234 [0.252]	−0.052 [0.571]
C(INTR)	−6.329 (0.325)	−5.278 [2.155]	−5.990 [0.696]	−6.198 [0.268]
C(SLLV)	0.457 (0.387)	0.783 [0.561]	−0.250 [1.217]	1.083 [1.077]
C(SLAU)	1.062 (0.413)	2.047 [1.590]	2.113 [1.697]	2.120 [1.708]
C(SLTR)	−0.470 (0.472)	−0.055 [0.585]	0.414 [1.248]	−0.074 [0.558]
C(LVAU)	0.884 (0.384)	0.768 [0.201]	0.248 [1.105]	1.684 [1.389]
C(LVTR)	−1.266 (0.644)	−0.099 [1.209]	−0.871 [0.409]	−1.648 [0.395]
C(AUTR)	0.960 (0.463)	0.397 [0.811]	0.462 [0.717]	0.154 [1.161]
C(Y1YO)	−0.134 (0.304)	0.246 [0.832]	−0.022 [0.244]	−0.017 [0.256]
C(INYO)	0.062 (0.423)	−0.158 [0.346]	−0.300 [0.569]	−0.023 [0.134]
C(SLYO)	0.005 (0.359)	−0.866 [1.618]	−0.592 [1.110]	−0.087 [0.170]

Values in (.) are standard deviations

Values in [.] are absolute t-values

Auxiliary parameter fits (continued)

Ap Name	Data value	Simple unit root	General unit root	Preferred stable
C(LVYO)	0.617 (0.364)	−0.480 [2.009]	0.538 [0.146]	0.656 [0.070]
C(AUYO)	−0.461 (0.376)	−0.667 [0.366]	−0.304 [0.279]	−0.599 [0.245]
C(TRYO)	−0.004 (0.379)	0.222 [0.396]	−0.009 [0.010]	−0.249 [0.431]
C(YSYO)	0.018 (0.003)	0.007 [2.340]	0.010 [1.737]	0.012 [1.392]
MLOWV	−0.289 (0.003)	−0.296 [1.501]	−0.291 [0.400]	−0.293 [0.854]
VTREND	1.781 (0.337)	0.674 [2.189]	0.497 [2.539]	1.186 [1.176]
VTRSQ	−0.532 (0.145)	0.050 [2.684]	0.094 [2.885]	−0.277 [1.174]
DVAR	0.782 (0.080)	0.578 [1.704]	0.588 [1.612]	0.692 [0.745]
DAUTO1	−2.585 (0.175)	−2.942 [1.361]	−3.214 [2.400]	−2.295 [1.107]
DAUTO2	−0.555 (0.152)	−0.030 [2.300]	0.004 [2.453]	−0.812 [1.127]
ARCH1	0.256 (0.074)	0.113 [1.301]	0.162 [0.857]	0.143 [1.029]
ARCH2	−0.097 (0.052)	0.125 [2.838]	0.123 [2.812]	−0.070 [0.353]
ARCH3	0.064 (0.085)	0.207 [1.126]	0.198 [1.052]	0.248 [1.449]
p(T,T1)	0.752 (0.014)	0.761 [0.388]	0.750 [0.147]	0.742 [0.530]
p(T,T10)	0.532 (0.035)	0.568 [0.630]	0.481 [1.039]	0.521 [0.280]

Values in (.) are standard deviations

Values in [.] are absolute t-values

Appendix A.4. *The SMD procedure used*

We present here the step by step procedure for the SMD. The procedure is illustrated for the following simple unit root model without idiosyncratic drifts or our ω generalisation and with time invariant variances for each unit:

$$\Delta y_{ht} = \varepsilon_{ht} + \theta_h \varepsilon_{ht-1} + \Delta u_{ht},\ t = 1, 2...T \tag{A21}$$

where u_{ht} is measurement error. The unit specific *model parameters* are the initial value (y_{h0}); the baseline variance of the shocks (ε_{ht}); the MA parameter (θ) and the measurement error (u_{ht}) standard deviation: $(y_{h0}, \nu_{h0}, \theta_h, \lambda_h)$. All four are taken to be heterogenous. The *distribution parameters* of this model are

$$\Theta = \{\tau_1, \tau_2, \tau_3, \tau_4, \phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}, \psi_{11}, \psi_{21,} \psi_{22}, \varphi, \varsigma_0, \varsigma_1\}.$$

where φ is a common $ARCH$ parameter.

1 From the data we construct a vector of the year of birth, $\mathbf{z} = \{z_1, z_2, ..., z_H\}$ and a (balanced panel) $H \times 33$ matrix of log earnings[28] from age 25 to age 57, Y, where y_{it} is log earnings for unit i at age t; with missing values if i is not observed at age $t + 25$. Define a corresponding matrix D which has $d_{ht} = 1$ if the earnings of worker h at age $t + 25$ is observed and zero otherwise.

2 Define a mapping (procedure) from any Y and $\mathbf{z}$ that returns a vector of auxiliary parameters (ap's), $\Lambda(Y, \mathbf{z})$.

3 Calculate the ap's on the basis of the data Λ^{Data}. Calculate the weighting matrix, Ω, as the covariance matrix of the ap's, using a bootstrap procedure on the original data.

4 Select a number of even number of replications R. For each observed unit we shall construct R simulated individuals. The even number is required since we take antithetic draws for each model parameter for each simulated unit (see the next step). In our estimation we take $R = 2$.

5 Generate three sets of random draws that are kept fixed across all iterations of the estimation procedure. Let W^1 (for the four model parameter $(y_{h0}, \nu_{h0}, \theta_h, \lambda_h)$ draws) be a $\left(\frac{H*R}{2} \times 4\right)$ matrix, where each element is a random draw from the standard normal distribution. Then stack this matrix and its negative to give a full (antithetic) $H * R \times 4$ matrix $W = \begin{bmatrix} W^1 \\ -W^1 \end{bmatrix}$ with typical element w_{ij}. Let E (the shocks) be a $H * R \times 36$ matrix, where each element, e_{it}, is a random draw from the standard normal distribution. (Even though we only use 32 shocks per unit in the simulations, we construct 36 and discard the first four to minimise any starting effects in simulating with an MA and $ARCH$ process). Let M (the measurement errors) be an $H * R \times 33$ matrix, where each element m_{ij} is a random draw from the standard normal distribution. All the draws are mutually independent. Also construct a vector $\mathbf{z}^R$ by stacking $\mathbf{z}$ R times.

6 Generate the simulated data $\mathbf{y}_h = \{y^s_{h0}, y^s_{h1}, ..., y^s_{h32}\}$ for $H * R$ simulated workers (see equation (A21)) and for a given set of parameters Θ.

(a) Generate the starting values for $H * R$ simulated workers, where the actual year of birth for each corresponding worker is used:

$$y^s_{h0} = \tau_1 + \tau_2 z^R_h + \left(\tau_3 + \tau_4 z^R_h\right) w_{h1}$$

(b) Generate the $H * R$ model parameters v_h, θ_h and λ_h

$$\begin{aligned}
\nu_{h0} &= \exp\left(\phi_{11} + \phi_{12} y^s_{h0} + \psi_{11} w_{h2}\right) \\
\theta_h &= \frac{\exp(\phi_{21} + \phi_{22} y^s_{h0} + \psi_{21} w_{h2} + \psi_{22} w_{h3})}{1 + \exp(\phi_{21} + \phi_{22} y^s_{h0} + \psi_{21} w_{h2} + \psi_{22} w_{h3})} - 0.5 \\
\lambda_h &= \exp(\varsigma_0 + \varsigma_1 w_{h4})
\end{aligned}$$

28. Strictly, the residuals from a first round regression.

(c) Update variances and shocks recursively:

$$\sigma_{h1}^2 = v_{h0}$$
$$\varepsilon_{h1} = \sqrt{\sigma_{h1}^2} e_{h1}$$
$$\sigma_{ht}^2 = v_{h0} + \left(\frac{e^{\varphi}}{1+e^{\varphi}}\right)(\varepsilon_{ht-1})^2 \quad t = 2, ...36$$
$$\varepsilon_{ht} = \sqrt{\sigma_{ht}^2} e_{ht} \quad t = 2, ...36$$

(d) Drop the first three values of $\boldsymbol{\varepsilon}_h$ to take out starting effects from the MA and $ARCH$. Re-label ε_{ht} from $t = 0$ to $t = 32$.

(e) Define simulated log earnings recursively:

$$y_{ht}^s = y_{ht-1}^s + \varepsilon_{ht} + \theta_h \varepsilon_{ht-1} \quad t = 1, ...32$$

(f) Generate measurement error $u_{ht} = \lambda_h m_{ht+1} \quad t = 0, ...32$

(g) Add measurement error:

$$y_{ht}^{obs,s} = y_{ht}^s + u_{ht}$$

7 To mimic the sampling frame the simulated earnings are included in the simulated data set if $d_{ht} = 1$. That is, the simulated data matrix $Y^s = \left[y_{ht}^{obs,s}\right]$ is element by element multiplied by the sample D matrix defined in step 1, stacked R times.

8 Calculate the ap's on the basis of simulated data set, Λ^s.

9 Calculate the weighted distance between the ap's based on the data and simulations $(\Lambda^{data} - \Lambda^s)' \left(\widehat{\Omega} * \frac{1+R}{R}\right)^{-1} (\Lambda^{data} - \Lambda^s)$ where R is the number of replications of the data.

10 Embed steps 6 to 9 in a minimisation routine and find the parameters that minimise the criterion.

Acknowledgement

We are grateful to two referees, the editor, Manuel Arellano, Bo Honoré, Hide Ichimura and John Rust for very helpful comments. We also thank participants in the 2000 UK Econometrics Study Group and 2001 ESEM, Lausanne for comments. This research was supported by the Danish Social Science Research Foundation (SSF) and the Danish National Research Foundation through its grant to CAM

REFERENCES

Abowd, J.M. and D. Card (1989), "On the Covariance Structure of Earnings and Hours Changes", *Econometrica*, 57, 411-445.

Alvarez, J., Browning, M. and M. Ejrnæs (2002), "Modelling Income Processes with lots of heterogeneity", CAM working paper 01.

An, M.Y. and M. Liu (2000), "Using indirect inference to solve the initial conditions problem", *Review of Economics and Statistics*, 82, 656-667.

Anderson, T.W. and C. Hsiao (1982), "Formulation and Estimation of Dynamic Models Using Panel Data, *Journal of Econometrics*, 18,67-82.

Arellano, M. (2003), *Panel Data Econometrics*, Oxford University Press

Arellano, M. and Carrasco, R. (2003), "Binary Choice Panel Data Models with Predetermined Variables", *Journal of Econometrics*, 115, 125-157.

Arellano, M. and J. Hahn (2006), "Understanding Bias in Nonlinear Panel Models: Some Recent Developments", Invited Lecture, Econometric Society World Congress, London, August 2005.

Atkinson, A.B., Bourguignon, F. and C. Morrisson (1992), *Empirical Studies of Earnings Mobility.* Harwood Publishers.

Baker, M. (1997), "Growth-Rate Heterogeneity and the Covariance Structure of Life-Cycle Earnings", *Journal of Labor Economics*, 15(2), 338-75.

Blundell, R., L. Pistaferri I. and Preston (2008), "Consumption Inequality and Partial Insurance." *American Economic Review*, 98(5), 1887-921.

Blundell, R. and R. Smith (1991), "Initial Conditions and Moment Restrictions in Dynamic Panel Data Models", *Annales d'Économie et de Statistique*, 20/21, 109-123.

Bonhomme, S. and J.M. Robin (2006), "Modeling individual earnings trajectories using copulas", in *Structural Models of Wage and Employment Dynamics*, Henning Bunzel, Bent J. Christensen, George R. Neumann, Jean-Marc Robin (eds), Emerald Group Publishing.

Browning, M., L. Hansen and J. Heckman (2000) "Micro data and general equilibrium models", chapter 7 in *Handbook of Macroeconomics*, volume 1A, edited by John Taylor and Michael Woodford, Elsevier Science, Amsterdam.
Browning, M. and A. Lusardi (1996), "Household Saving: Micro Theories and Micro Facts", *Journal of Economic Literature*, 34(4), 1474-1855.
Buchinsky, M. and J. Hunt (1999), "Wage Mobility in the United States", *The Review of Economics and Statistics*, 81(3), 351-368 .
Carroll, C. and A. Samwick (1997), "The Nature of Precautionary Wealth", *Journal of Monetary Economics*, 40, 41-71.
Chamberlain, G. (1980), "Analysis of Covariance with Qualitative Data", *Review of Economic Studies*, 47, 225.238.
Chamberlain, G. and K. Hirano (1999), "Predictive Distributions Based on Longitudinal Earnings Data", *Annales d'Économie et de Statistique*, 55/56, 211-42.
Cunha, F., J. Heckman and S. Navarro (2005), "Separating Uncertainty from Heterogeneity in Life Cycle Earnings.", *Oxford Economic Papers-New Series*, 57(2), pp. 191-261.
De Santis, Massimiliano (2007), "Individual Consumption Risk and the Welfare Cost of Business Cycles.", *American Economic Review*, 97(4), 1488-506.
Dominitz, J. and C. Manski (1997), "Using Expectations Data to Study Subjective Income Expectations", *Journal of the American Statistical Association*, 92, 855-867.
Geweke, J. and M. Keane (2000), "An Empirical Analysis of Income Dynamics Among Men in the PSID: 1968-1989", *Journal of Econometrics*, 96, 293-356. .
Gottschalk, P. and R. Moffitt (1994), "The Growth of Earnings Instability in the U.S. Labor Market", *Brookings Papers on Economic Activity*, 2, 217-72.
Gourieroux, C. and A. Monfort, (1996) *Simulation-Based Econometric Methods*, CORE Lectures, Oxford University Press.
Gourieroux, C., Monfort, A. and E. Renault (1993), "Indirect Inference", *Journal of Applied Econometrics*, 8, 85-118.
Guvenen, F. (2007), "Learning Your Earning: Are Labor Income Shocks Really Very Persistent?", *American Economic Review*, 97(3), pp. 687-712.
Guvenen, Fatih and Anthony Smith (2008), "Inferring Labor Income Risk from Economic Choices: An Indirect Inference Approach", manuscript, 2008.
Haider, S. J. (2001). "Earnings instability and earnings inequality of males in the United States: 1967-1991." *Journal of Labor Economics* 19(4): 799-836.
Hall, G. and J. Rust (2003), "Econometric Methods for Endogenously Sampled Time Series: The Case of Commodity Price Speculation in the Steel Market". Unpublished Manuscript, University of Yale.
Hause, J.C. (1977), "The Covariance Structure of Earnings and On-the-job Training Hypothesis", *Annals of Economic and Social Measurement*, 6, 335-365.
Hsiao, C. (1986), *Analysis of Panel Data*, Cambridge University Press.
Im, Kyung So, Hashem Pesaran and Yongcheol Shin (2003), "Testing for Unit Roots in Heterogeneous Panels" *Journal of Econometrics*, 115(1), pp. 53-74.
Jarocinsky M. and A. Marcet (2009) "Autoregression and Priors About Initial Growth Rates", Unpublished Manuscript
Kiviet, J.F. and W. Krämer (1992), "Bias of s^2 in the Linear Regression Model with Correlated Errors", *The Review of Economic and Statistics*, 362-365.
Kiviet, J.F. and G.D. Phillips (1998), "Degrees of Freedom Adjustment for Disturbance Variance Estimators in Dynamic Regression Models", *Econometrics Journal*, 1, 44-70.
Lillard, Lee and Y. Weiss (1979), "Components of Variation in Panel Earnings Data: American Scientists 1960-1970", *Econometrica*, 47, 437-454.
Lillard, Lee and R. Willis (1978), "Dynamic Aspects of Earning Mobility" *Econometrica*, 46(5), 985-1012.
MaCurdy, T.E. (1982), "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis", *Journal of Econometrics*, 18, 83-114.
Meghir, C. and L. Pistaferri (2004), "Income Variance Dynamics and Heterogeneity", *Econometrica*, 72, 1, 1-32.
Moffitt, R. A. and P. Gottschalk (2002). "Trends in the transitory variance of earnings in the United States." *Economic Journal*, 112(478): C68-C73.
Müller, U.K. and G. Elliott (2003), "Tests for unit roots and the initial condition", *Econometrica*, 71, 1269-1286.
Neal, D. and S. Rosen (2000), "Theories of the distribution of earnings", chapter 7 in *Handbook of Income Distribution*, volume 1, edited by A. B. Atkinson and F. Bourguignon, Elsevier Science, Amsterdam.

Rubinstein, Y. and Y. Weiss (2006), "Post-school earnings: Search versus human capital", *Handbook of the Economics of Education*, E. Hanushek and F. Welch (eds), Amsterdam: North-Holland.

Shaman, P. and R.A. Stine (1988a), "The Bias of Autoregressive Coefficient Estimators", *Journal of the American Statistical Association*, 83, 842-848.

Shaman, P. and R.A. Stine (1988b), "A Fixed Point Characterization for Bias of Autoregressive Estimators", *The Annals of Statistics*, 17, 1275-1284.

Smith, A.A. (1993), "Estimating nonlinear time-series models using simulated vector autoregressions", *Journal of Applied Econometrics*, 8, S63-S84.

Ulrick, S. (2000), "Using Semiparametric Methods in an Analysis of Earnings Mobility", manuscript, Department of Economics, University of Iowa.

Wooldridge, J. (2005), "Simple Solutions to the Initial Conditions Problem for Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity," *Journal of Applied Econometrics* 20, 39-54