ISSN 1471-0498

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

MODELLING INCOME PROCESSES WITH LOTS OF HETEROGENEITY

Martin Browning, Mette Ejrnaes and Javier Alvarez

Number 285

October 2006

Manor Road Building, Oxford OX1 3UQ

Modelling income processes with lots of heterogeneity.*

Martin Browning
Department of Economics
University of Copenhagen
Martin.Browning@econ.ku.dk

Mette Ejrnæs
Department of Economics
University of Copenhagen
Mette.Ejrnes@econ.ku.dk

Javier Alvarez
Bank of Spain

June 2006
(First draft: June 2002)

Abstract

All empirical models of earnings processes in the literature assume a good deal of homogeneity. In contrast to this we model earnings processes allowing for lots of heterogeneity between 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 our models 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 much greater latent heterogeneity than previous investigators.

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

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 tests, ARMA.

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)); 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 Gottschalk (1997)); 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)) and predicting future earnings paths given individual information (Chamberlain and Hirano (1997)). We shall return to a discussion of these issues but for now this will suffice to motivate our interest in estimating earnings processes.

Various dynamic models for individual workers have been assumed in the earnings dynamics literature. The most important feature that emerges from our reading of this 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 (1997), Ulrick (2000) and Meghir and Pistaferri (2004)) is that, whatever the process chosen, only limited allowance is made for heterogeneity. As we shall show below, conventional processes with limited heterogeneity are unable to account for the observed facts. There are two broad reactions to this: we could allow for more complicated processes (for example, by introducing nonlinearities or by allowing for ARCH effects as in Meghir and Pistaferri (2004)) or we could allow for more heterogeneity. It is the latter line that we follow below.[1] Other

[1] Having said this, we also find that we need to allow for different convergence rates at the

investigators have also followed this path but in a limited fashion. In particular, it is always assumed that everyone has the same process, but with, for instance, different means and/or variances[2]. We are sceptical that everyone has the same process with much the same parameters. Rather it may be that 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. Given this, our approach is to allow for a good deal more heterogeneity than previous investigators. To do this we have to allow for correlated heterogeneity which, 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[3] (see, for example, Chamberlain (1980), Anderson and Hsiao (1982), Blundell and Smith (1991), Wooldridge (2000) and Arellano and Carrasco (2003)). 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 starting values.

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

start of the process than are allowed for with ARMA processes. This requires an extension of the latter so that we also allow for a 'more complicated process'.

[2]This is also true of the 'large T, large N' panel data literature on unit roots which always tests between everyone having a unit and no one having a unit root, see Baltagi and Kao (2000) for a survey.

[3]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.

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 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.[4] The third general approach, which we follow here, 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 models. To fit these models we use Simulated Minimum Distance (SMD) (also known as 'indirect inference'); see 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 A3. SMD is used to estimate the parameters of a fully parameterized model. To implement SMD we choose a set of statistics; these statistics we will call 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

[4] 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.

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. The SMD 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, 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 by education and race. In this paper the focus is on white males 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 that we observe some workers close to the start of their earnings process and the initial convergence to the 'long run' is different to the process implied by the standard ARMA model. Second, we find strong evidence of much more heterogeneity than previous researchers have allowed for. 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. This contrasts sharply with the widespread belief that all workers have a unit root with no idiosyncratic drift. Finally, we find 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.

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 AR parameter but the variant given below 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:

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

where t is age (minus 25), $|\beta| \leq 1$ and $|\omega| \leq 1$. An alternative way to write (2.1) is as:

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

This is an extended ARMA model with drift, where α is the drift parameter, δ is the long run mean, net of the trend (which we shall usually call the 'long run mean' in all that follows), β the AR parameter, θ the MA parameter and ω is a novel parameter that is discussed below. 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 $\omega = \beta$:

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

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 $\omega \leq 1$, 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. 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 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).

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.[5] Within this class the deterministic component of earnings is given by:

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

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.[6] 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. Within this class we also consider the consensus model which is a further restricted version of the unit root model ((2.1) with $\beta = \omega = 1$ and $\delta = \alpha = 0$). The consensus 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 A.1). The other reason is that it is believed that it provides a good fit to the data; as we shall see below the latter belief is misplaced. The third class of models contains *mixture models* in which we allow that the earnings of some workers has a unit root and that of others is a stable process. Details of how we do this are given in the subsection after the next.

[5]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.

[6]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 $\omega = \beta$.

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):

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

where $E_{t-1}(.)$ denotes the expectations conditioned on $(\varepsilon_0, y_0, ..., y_{t-1})$. The third condition allows for an ARCH component. This 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 $\varepsilon_0 \equiv 0$.

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:

$$\begin{aligned} y_t^{obs} &= y_t + u_t \\ E(u_t) &= 0, E(u_t y_t) = 0, \\ E(u_t)^2 &= \lambda^2 \\ E(u_t u_s) &= 0, s \neq t \end{aligned} \tag{2.9}$$

We further assume that the measurement errors are normally distributed.

The specification (2.1), (2.8) and (2.9) give eight *model parameters* per worker[7]:

$$\text{Model parameters: } \{\nu, \theta, \alpha, \beta, \delta, \omega, \varphi, \lambda\} \tag{2.10}$$

This specification includes almost all models suggested in the literature except that some researchers allow for an $MA(2)$ component.[8] 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

[7] For the mixture model the specification includes an extra parameter.

[8] The most conspicuous class not covered are copula models; see Bonhomme and Robin (2004). We shall return to this in the concluding section.

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:

$$y_{h0} = \tau_1 + \tau_2 z_h + (\tau_3 + \tau_4 z_h)\,\eta_{h0} \tag{2.11}$$

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).[9] 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 (2.10)). 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. The parameterisation for the joint distribution for the six model parameters (ν, 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 (2.11). Denoting the inverse logit function by $\ell(z) = e^z/(1+e^z)$, the functional forms we adopt are given by:

$$\begin{aligned}
\nu_h &= \exp(\phi_{11} + \phi_{12} y_{h0} + \psi_{11}\eta_{h1}) \\
\theta_h &= \ell(\phi_{21} + \phi_{22} y_{h0} + \psi_{21}\eta_{h1} + \psi_{22}\eta_{h2}) - 0.5 \\
\alpha_h &= \phi_{31} + \phi_{32} y_{h0} + \sum_{i=1}^{3} \psi_{3i}\eta_{hi} \\
\beta_h &= \ell\left(\phi_{41} + \phi_{42} y_{h0} + \sum_{i=1}^{4} \psi_{4i}\eta_{hi}\right) \\
\delta_h &= \phi_{51} + \phi_{52} y_{h0} + \sum_{i=1}^{5} \psi_{5i}\eta_{hi} \\
\omega_h &= \ell\left(\phi_{61} + \phi_{62} y_{h0} + \sum_{i=1}^{6} \psi_{6i}\eta_{hi}\right)
\end{aligned} \tag{2.12}$$

so that we restrict $\nu_h > 0$, $\theta_h \in (-0.5, 0.5)$, $\beta_h \in (0,1)$ and $\omega_h \in (0,1)$. Thus our heterogeneity structure allows for seven latent factors (including that for the

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

starting value, η_{h0}, in equation (2.11)); in the empirical analysis below we find that we actually need far fewer than seven latent factors.

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 have two alternative options. The first is to have one segment of the population who have a unit root with the others having a heterogeneous AR parameter $\beta_h < 1$ with the distribution given in (2.12). If the probability of having a unit root, π, is a constant then this nests the stable model with one extra parameter. The alternative is to assume that the stable AR parameter is common for those who do not have a unit root ($\beta = \beta_0 < 1$) and that the probability of having a unit root is heterogeneous and given by $\pi_h = \beta_h$ where the latter is given by (2.12).[10] This is not nested within the stable model but it does nest the pure unit root model. For reasons that will become clear below, we concentrate on the second alternative.

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 (2.9) is lognormal:

$$\lambda_h = \exp\left(\varsigma_0 + \varsigma_1 \eta_{h\lambda}\right) \tag{2.13}$$

where η_λ is a standard Normal variable. Given the parametric assumptions made above, the parameters $(\varsigma_0, \varsigma_1)$ are identified.[11]

Our approach to modelling the heterogeneity in the model parameters is a generalisation of the approach adopted by Chamberlain (1980), Blundell and Smith

[10] Since it requires one further heterogeneous model parameter, we have not tried to model the overall nesting model with a heterogeneous stable AR parameter and a heterogeneous probability.

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

(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 (2.11) 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 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 (1997)) 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}. Suppose now 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 (2.11) and (2.12) 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 (2.12) and (2.11) 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. The sample drawn is exactly the same as in Meghir and Pistaferri (2004) (henceforth MP); we select male workers aged between 25 to 55 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. 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. For individuals in this sample the relevant variables we observe are race, education, age and birth cohort. We deal with some of the observable heterogeneity by stratifying on education and race and working with the white, high school sample. This gives a sample size of $1,104$ with workers being observed between 9 and 26 years. We then run a first round regression of log earnings on year dummies and age dummies. In all that follows we work with the residuals from this regression which we shall term log earnings for convenience.

In figure 3.1 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 drift or trend is one of the primary purposes of the analysis of earnings processes. The lower panel in figure 3.1 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 to be concentrated in the first few years. 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.

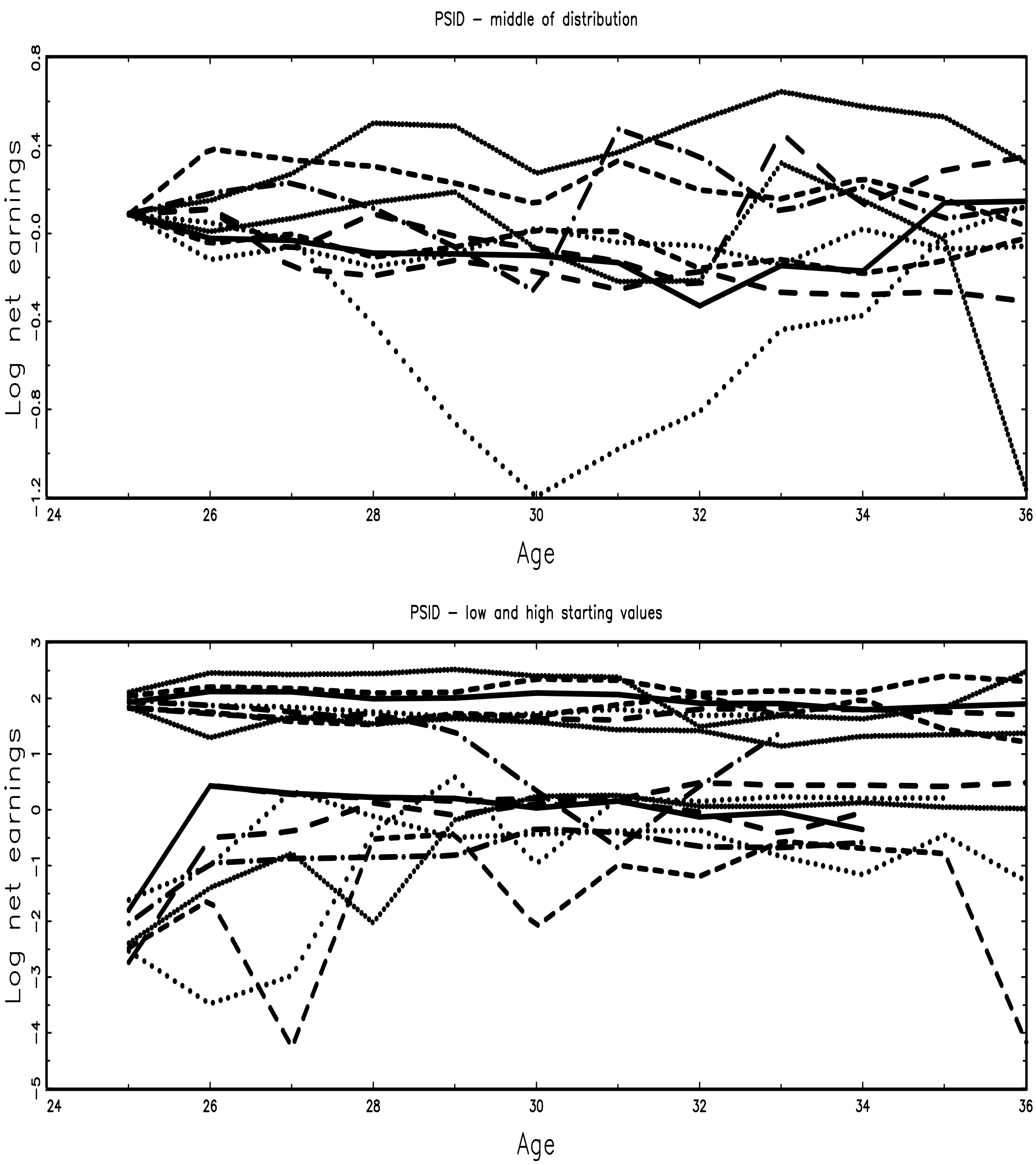

Figure 3.1: Sample paths for age 25-44

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 A.2. The choice of statistics we consider is motivated by two considerations. The motivation for the first set of auxiliary parameters we consider 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. As we shall see, having a close link between some of the statistics and the parameters is useful in many ways. More complicated procedures which include small sample corrections which give a closer correspondence to parameters could be implemented 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. 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. We also calculate 9 additional statistics to allow for the fact that for many in our sample the initial observation is not at age 25 (see the Appendix subsection A.2 for details and subsection 2.4 for the rationale). Finally we calculate the variance within the bottom quartile 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 specific interpretation and are 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 Geweke and Keane (1997) (and others) base their analyses on mobility measures such as short run and long run transition matrices between different quintiles of the earnings distribution. The other two major features we wish to capture are the time series properties of the cross-section variance of earnings (see Gottschalk and Moffitt (1995)) and the conditional heteroskedasticity that Meghir 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 A.2 and a discussion is given here.

In the first column of Table 3.1 we present the sample values for the second set of 10 ap's (for the sake of space we do not report the values for the first set of 35 auxiliary parameters since these are of little intrinsic interest); in some cases the values have been rescaled (as shown) to make reading easier. The first 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; the significance of the quadratic term argues against an underlying unit root process (which implies a linear trend for variances for the population). 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 increases nonlinearly over time, from an initial value of 0.09 to 0.214 after 10 years but only 0.233 after 20 years. This is qualitatively similar to the result of Gottschalk and Moffit (1994). It is also consistent with the widespread finding that the inequality of earnings has been increasing through our sample period (see, for example, Buchinsky and Hunt (1997)) but note that finding is for the population as a whole and not following the same group through time. The finding that the variance is increasing over time is not necessarily evidence that uncertainty or 'risk' is increasing: it may simply reflect the fact that *some* workers have a unit root with a consequent 'fanning out' of variance over time.

The next 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

Auxiliary parameters for data and preferred models.					
	Data	Parametric models			
Auxiliary Parameter	Value	Consensus	Unit root	Stable	Mixture
$vtrend$	1.78	-0.17	0.14	0.64	0.80
($\times 10$)	(0.33)	[4.0]	[3.3]	[2.3]	[2.0]
$vtrsq$	-0.53	0.22	0.13	-0.09	-0.12
($\times 10$)	(0.14)	[3.6]	[3.1]	[2.1]	[1.9]
$dvar$	0.78	0.47	0.53	0.64	0.66
($\times 10$)	(0.08)	[2.6]	[2.1]	[1.2]	[1.0]
$dauto1$	-2.59	-3.21	-3.06	-2.33	-2.33
($\times 10$)	(0.17)	[2.4]	[1.9]	[1.0]	[1.0]
$dauto2$	-0.56	0.06	0.16	-0.76	-0.55
($\times 10$)	(0.15)	[2.7]	[3.1]	[0.92]	[0.1]
$arch1$	0.26	0.09	0.06	0.15	-0.01
	(0.07)	[1.6]	[1.8]	[1.0]	[2.4]
$arch2$	-0.10	0.25	0.20	-0.06	0.02
	(0.05)	[4.6]	[3.9]	[0.4]	[1.6]
$arch3$	0.06	0.17	0.18	0.26	0.19
	(0.08)	[0.9]	[1.0]	[1.6]	[1.0]
$p(t, t+1)$	0.75	0.76	0.76	0.74	0.72
	(0.01)	[0.5]	[0.3]	[0.6]	[1.5]
$p(t, t+10)$	0.53	0.60	0.51	0.53	0.48
	(0.03)	[1.2]	[0.5]	[0.2]	[1.1]
Goodness of fit statistic	-	187.1	95.4	29.4	38.4
Degrees of freedom	-	31	25	20	18
Note: the first 35 auxiliary parameters are not shown.					
(.) = standard deviation for the data value					
$[t]$ = absolute t-values for the difference between the simulated and data values					

Table 3.1: Auxiliary parameters for data and preferred models.

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 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. Simulated minimum distance.

As shown in the last section, we consider a number of different models for the earning process. All the models are fully parametric and the most complicated has 41 parameters. To conduct an exploratory analysis of a series of more and more complicated models we need an estimation strategy which is fast and can handle a large set of parameters. Furthermore, we will evaluate the models by goodness of fit based on our auxiliary parameter (see the discussion of auxiliary parameter in section 3.2). Minimising the goodness of fit statistic is itself an estimator; in fact it is effectively Simulated Minimum Distance (SMD). This was first introduced in Lee and Ingram (1991) and Smith (1993) in a time series context. It was also used in Duffie and Singleton (1993) in an asset pricing model using time series data and Hall and Rust (2003) (who suggest the term SMD) who employ it in a time series model with sample based observations. It is closely related to other simulation methods such as the Method of Simulated Moments (see Stern (1997)); indirect inference (see Gourieroux, Monfort and Renault (1993)) and Efficient Method of Moments (see Gallant and Tauchen (1996)). Ultimately all of these methods take their inspiration from Lerman and Manski (1981) and the important finding that we can consistently estimate using simulation methods with a fixed number of simulations, (see McFadden (1989) and Pakes and Pollard (1989)). SMD proceeds in a number of steps. First we calculate some 'well chosen' statistics of the data; these are the 45'sample auxiliary parameters' discussed in the last subsection. Next we take a parametric model for the data generating process and simulate for particular parameter values. Then we calculate the value of the auxiliary parameters for the simulated data. If the model is well specified (in a well defined sense) and we have the 'correct' value for the model parameters then these simulated auxiliary parameters have the same (unknown) probability limit as the sample auxiliary parameters. The SMD estimator of the model parameters is then the value of the model parameters that minimises the weighted distance between the sample auxiliary parameters and the simulated auxiliary parameters. We present a fuller description of the SMD procedure we use (which includes the bootstrap for the weighting matrix and antithetic sampling) in subsection A.3.

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 and for ARCH effects. Likelihood functions for this are not easily derived; we present an explicit discussion in Appendix subsection A.4. A third advantage is that we can explicitly account for the sampling scheme; in our case we 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.[12] 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).

[12]Exactly the same can be said of most choices of moments to fit for GMM.

4. Results.

4.1. Choosing a preferred model.

We first present the goodness of fit (gf) for the three broad classes of models: stable, unit root and mixture. For each we present the results for the most general model and for our preferred restricted variant. Additionally, we present results for the consensus unit root model ($\beta = \omega = 1$ and $\delta = \alpha = 0$). Table 4.1 gives the gf statistics for each variant and Table 3.1 presents the detailed fits for the ap's of substantive interest for the preferred variants. For each of the general cases there are many redundant heterogeneity distribution parameters (2.12); for example, for the stable model we can exclude 15 heterogeneity parameters and the gf statistic only rises by 4.4. As we shall see in the next subsection, all of the remaining parameters in the preferred models are 'significant'. Referring to Table 4.1 we see clear and strong statistical evidence in favour of the stable model. The general unit root model[13] is strongly rejected against the general stable model (a $\chi^2(8)$ statistic of 67.7) and the consensus model is, in its turn, strongly rejected against the general unit root model (a $\chi^2(18)$ of 94.4). The fit of the mixture model is better than the pure unit root model (which it nests) but is considerably worse than for the stable model.[14] One notable feature of the preferred stable model is that we fit the second order auto-correlation term ($dauto2$) without recourse to an $MA(2)$ component.

The preferred stable model fits reasonably well with a $\chi^2(20)$ gf statistic of 29.4. As can be seen in the detailed results in Table 3.1, the ap's the preferred stable model conspicuously fails to fit are those capturing the trend and trend squared for variances over calendar time.[15] We conclude that these common non-linear time effects for variances need to be modelled separately from the general heterogeneous processes given by (2.1) and (2.12); see Gottschalk and Moffitt

[13] Recall that this is the weak form of the unit root hypothesis: $\beta = 1$ (that is, the shocks cumulate) but not $\omega = 1$ (see the discussion after (2.1)). Hence we drop 8 heterogeneity parameters (see (2.12)).

[14] For completeness, we record that the mixture model that has a fixed probability of a unit root and a heterogenous AR parameter (so that it has one more parameter than the general stable model, see the discussion after (2.12)) has a gf statistic of 24.1 so that it does not improve significantly on the stable model. Moreover the model has only a small proportion of the population with a unit root.

[15] The $\chi^2(18)$ gf statistic for the preferred stable model without fitting the two time trend ap's is 21.8. The qualitative implications of the estimates for this model are very similar to those reported below for the model.

Model:	Stable		Unit root			Mixture	
	G	P	G	P	C	G	P
# parameters	40	25	32	20	14	41	27
Degrees of freedom	5	20	13	25	31	4	18
χ^2 statistic	25.0	29.4	92.7	95.4	187.1	35.8	38.4
Probability (%)	0.01	8.1	0.00	0.00	0.00	0.00	0.35
G = general variant; P = preferred variant; C = consensus model							

Table 4.1: Goodness of fit for different models

(1994). 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.

4.2. Parameter estimates.

4.2.1. Distribution parameters.

In Table 4.2 we present the heterogeneity distribution parameter estimates for the preferred stable model; see (2.1) for the exact role of each parameter. The $\chi^2(1)$ statistics in the final column are the quasi-LR values for setting the corresponding parameter to zero. 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.

The variance (ν), the trend term (α), the long run mean term (δ) and the initial adjustment term (ω) 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 factor η_0 for the starting value) and relatively parsimonious dependence of the heterogeneity in model parameters (as compared with the general model (2.12)). 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:

$$\Delta_a = 8.83 + 0.56a - 0.057a^2 \tag{4.1}$$

where a is age in decades. Given this, the equation for the starting distribution (2.11), the first order process (2.1) and the parameter estimates in Table 4.2 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 4.3 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 largely positive, with a median value of 0.232, but 33% of the population have a negative value. This contrast 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 also widely dispersed with a median value of 0.793 but 27.4% of the population having an AR parameter above 0.9. The median trend (the α parameter) is 7.7% per decade and some of the population have very high growth; for example, 15% of the population having a trend of above 20% per decade. The median long run (net of trend) mean, δ, is 0.43 below the mean of the starting distribution but this does not imply that earnings are falling: we have taken out the common nonlinear age effect in the initial regression (see (4.1)) 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 in (2.1).

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

Distribution Parameter	Corresponding model parameter	Estimated value	$\chi^2(1)$
τ_1	y_0 (starting value)	0.484	7.4
τ_2	y_0	-0.914	6.3
τ_3	y_0	-4.896	108
τ_4	y_0	7.519	103
ϕ_{11}	ν (variance)	-5.069	1000+
ϕ_{12}	ν	-0.867	9.7
ϕ_{21}	θ (MA term)	1.003	2.2
ϕ_{31}	α (trend)	0.085	13.5
ϕ_{32}	α	-0.126	5.1
ϕ_{41}	β (AR term)	1.343	30.5
ϕ_{51}	δ (long run mean)	-0.479	1.4
ϕ_{52}	δ	2.649	5.3
ϕ_{61}	ω (initial adjustment)	3.827	115
ϕ_{62}	ω	-0.457	2.2
ψ_{11}	ν	1.601	33.6
ψ_{22}	θ	-2.308	13.0
ψ_{31}	α	0.109	24.8
ψ_{41}	β	-0.644	18.1
ψ_{42}	β	1.231	24.7
ψ_{51}	δ	-1.687	14.6
ψ_{52}	δ	-0.686	18.7
ψ_{62}	ω	-1.035	32.2
φ	$ARCH$	-0.621	6.4*
ς_0	λ (std of	0.640	18.7
ς_1	measurement error)	-2.586	1000+
* value for setting parameter to $-\infty$ (no ARCH)			

Table 4.2: Parameter estimates for stable model

	y_0	$\sqrt{\nu}$	θ	α	β	δ	ω
Marginal distributions							
1st decile	-0.211	0.028	-0.376	-0.066	0.394	-2.770	0.921
Median	0.072	0.077	0.232	0.077	0.793	-0.318	0.978
9th decile	0.358	0.221	0.482	0.222	0.958	2.160	0.994
Correlations							
corr(y_0)	1	-0.088	0.004	-0.238	-0.008	0.297	-0.079
corr($\sqrt{\nu}$)	–	1	-0.010	0.838	-0.391	-0.770	0.000
corr(θ)	–	–	1	-0.012	-0.754	0.349	0.805
corr(α)	–	–	–	1	-0.421	-0.930	0.008
corr(β)	–	–	–	–	1	0.079	-0.557
corr(δ)	–	–	–	–	–	1	0.279

Table 4.3: The parameter distribution

latent factors and many model parameters that are correlated with the starting value. The highest correlation is 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.

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.[16] For each outcome we compare

[16]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

	Stable model	Unit root model	Mixture model
Short run variability			
Standard deviation of shocks (allowing for $ARCH$)			
First decile	0.032	0.032	0.031
Median	0.089	0.088	0.093
Ninth decile	0.250	0.242	0.279
Mobility			
Transitions from bottom quintile to bottom quintile (probability (%)):			
age 25 $\rightarrow$ 26	59.9	68.6	61.9
age 25 $\rightarrow$ 35	28.7	29.6	31.6
age 35 $\rightarrow$ 50	56.8	62.6	56.4

Table 4.4: The parameter distribution

the predictions from the preferred stable model, the preferred unit root model and the preferred mixture model.

Turning first to short run variability, when we account for ARCH this is given by:

$$\text{std of shock for worker } h = \sqrt{(1+\varphi^2)\, v_h} \tag{4.2}$$

The top panel of Table 4.4 compares the distributions of this for the preferred variants of our three models. As can be seen, this is very similar across models. The more important point is that 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 0.24. 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. 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 4.4 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

and alternatives such as search and learning.

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. In this case the three models give somewhat different predictions with the unit root generally giving more mobility, but the qualitative implications are similar. The most interesting feature here is that 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 after age 35 with only a 43.2% 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, add in the age dummy coefficient values that we took out in the original regression and discount back to age 25 using a 3% real rate. The three curves are shown in figure (4.1). The ranges of median annualised lifetime earnings (from the bottom decile to the top decile) are reflected in the horizontal lengths of the curves. It can be seen that the three curves are quite different. The heterogenous unit root model is almost linear and strictly increasing, as we would expect for a unit root. The other two models display much less variance; this is what we expect since all shocks are transitory. The stable and mixture models are non-monotone with higher interquartile ranges at the top and bottom of the median distribution. This is a reflection of what we saw in the raw data in figures 3.1. Finally, the support of the three distributions are quite different, with the unit root model showing a much larger dispersion. If these curves are taken as inputs to models of schooling choices, then the results are likely to be very sensitive to the model chosen.

The final outcome we consider is for the cross-section standard deviation of log earnings over age; see figure (4.2). In Rubinstein and Weiss (2005) it is argued that different models of wage growth will lead to different developments in the cross-section variance.[17] For all models the latter is increasing with age and for the unit root and mixture models it is close to linear. The stable model has that inequality increases quite quickly at the start of the process but then levels off to a much lower level than the other two models. Again it shows that for some outcomes the predictions of the types of models are very different.

[17] This picture is comparable to Figure 6a in Rubinstein and Weiss (2005)

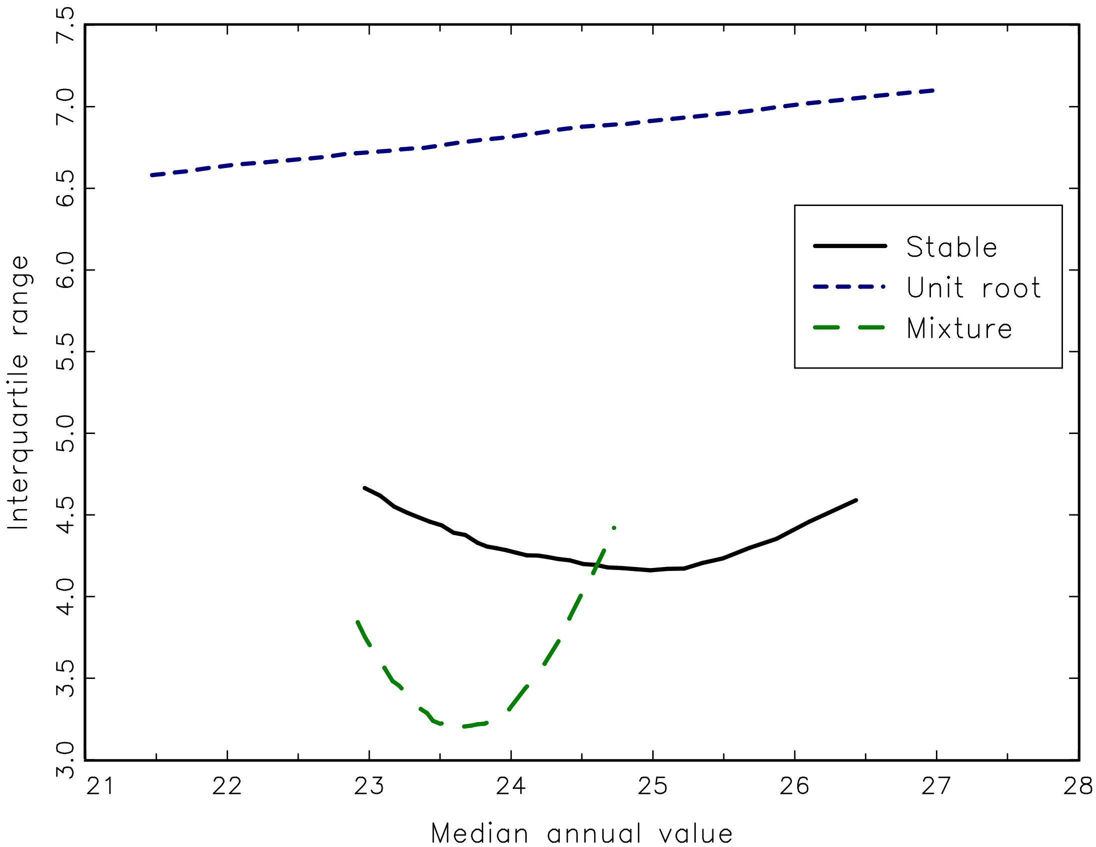

Figure 4.1: The trade-off in lifetime income between median and IQR.

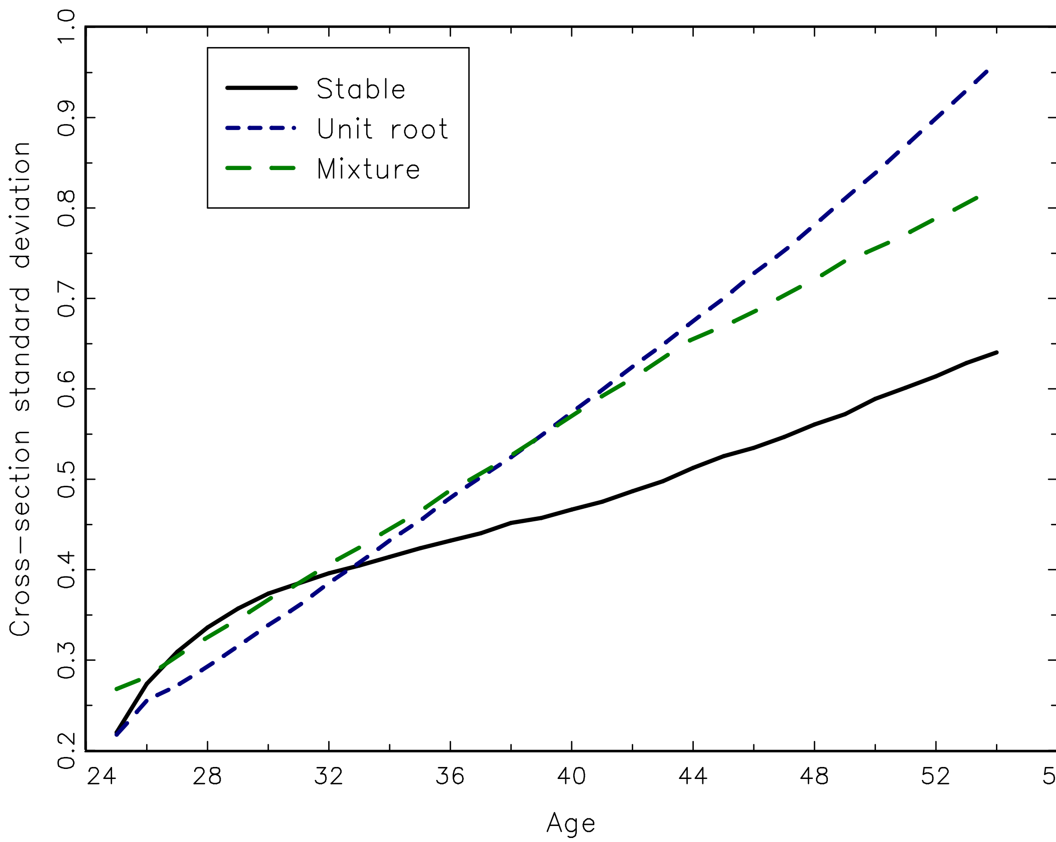

Figure 4.2: The evolution of inequality over the life-cycle.

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 term). We do this by introducing a nonlinear multi-factor model. To deal with the resulting curse of dimensionality 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. Fro 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. 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 (2004). 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 very familiar (an extended ARMA) and the results can be used directly in simulations.

- Unit root models are decisively rejected. The model with a mixture of a unit root and a homogeneous AR parameter does better statistically than the unit root model. The model with everyone having an AR parameter below unity (the stable model) does best. This conclusion is in stark contrast to most papers in the literature which find evidence in favour of everyone having 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 maintained unit root lack power

- The stable process is generally quite persistent with a median AR parameter of 0.79 and 27.4% 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 face 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 MA term. The difference arises because we allow for a heterogeneous measurement error.

- The preferred model fails to capture the nonlinear trend in variances over calendar time. Despite using a very flexible model with a great deal of heterogeneity it looks like we need to introduce some common secular variation, as in Gottschalk and Moffitt (1994).

These results represent a fairly radical break with 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. This is particularly important if we are to allow common time trends in variances.

- We have only allowed for first order schemes (only one lag of earnings and an $MA(1)$). Although some of our auxiliary parameters would have picked up evidence of any need for longer lags, it is possible that some of the heterogeneity we find is spurious and is induced by ignoring dependence on higher order lags.

- The earlier literature on earnings also considered the links to unemployment and hours fluctuations and with time varying covariates such as family and marital status. It would be desirable to use our approach with lots of heterogeneity for these ancillary processes. The factor approach we have adopted is well suited for this.

- 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.

A. Appendix.

A.1. The consensus 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}:

$$y_{ht} = p_{ht} + e_{ht} \tag{A.1}$$

The permanent component is modelled as a random walk:

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

By assumption the transitory component has low persistence and is modelled as a $MA(q)$-model. The consensus 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 $e_{ht} \sim iid(0, \sigma_e^2)$. The log income process (A.1) 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 consensus model:

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

where the following equations are satisfied:

$$\begin{aligned} -\sigma_e^2 &= \theta\sigma_\varepsilon^2 \\ 2\sigma_e^2 + \sigma_\zeta^2 &= (1+\theta^2)\sigma_\varepsilon^2 \end{aligned}$$

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

A.2. The construction of the auxiliary parameters.

The first set of auxiliary parameters are based on the OLS regression 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. For each worker we also calculate the log of the residual variance (lv) and the first order auto-correlation of the OLS residuals (au). This gives a $H \times 5$ matrix of estimates for (in, sl, tr, lv, au). We then calculate 26 statistics based on these OLS regressions. The first ten of these are the mean and variance of the the estimated parameters; for example, the mean of the log variances is given by:

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

The next 15 statistics are the covariances across the five parameter estimates and its covariances with the initial observation $y_{h\tau}$. 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 quarter of the residual variances.

The other 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.[18] The interest here lies in the time series trend in inequality emphasised by Gottschalk and Moffitt (1995). 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 $dauto2$.

[18] 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.

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 $arch3$.

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 36 statistics: 26 OLS based statistics (including the ap to pick up low residual variances), 2 trend coefficients, 3 means of first differenced statistics, 3 ARCH statistics and 2 mobility measures. To be sure there is an element of arbitrariness in this choice but it does have the virtues of being fast to compute and it captures most of the concerns of previous investigators.

Moreover, we also have to estimate the parameters of the distribution of the starting values, including allowance for dependence on the year of birth of the respondent. To construct ap's for this we have to make a clear distinction between the starting value (the value at age 25) and the initial observed value, see subsection 2.4. Only the latter is observed for all workers so that we have to base our ap's on these observations. We take as our ap's the mean and variance of the initial observations and the covariance between the year of birth and the initial observation $(y_{h\tau})$ and the five OLS parameters (in, sl, tr, lv, au). Finally we include an ap to capture the dependence of the variance of the starting values on year of birth. Specifically we compute the mean of the product of the initial value squared and the trend (both as deviations from the sample mean):

$$c(y_{h\tau}, z_h) = \frac{1}{H}\sum\left((y_{h\tau} - \bar{y}_\tau)^2\right)(z_h - \bar{z}) \tag{A.2}$$

where $y_{h\tau}$ is the initial observation for worker h and z_h is the year of birth. In all, this gives 9 extra ap's, for a total of 45 ap's.

A.3. The SMD procedure used

The SMD estimator

We present here the step by step procedure for the SMD. The procedure is illustrated for the following simple model:

$$\Delta y_{ht} = \varepsilon_{ht} + \theta_h \varepsilon_{ht-1} \tag{A.3}$$

The *distribution parameters* of this model (including the starting values) are

$$\Theta = \{\tau_{1,}\tau_2, \tau_3, \tau_4, \phi_{11}, \phi_{12}, \phi_{21}, \phi_{22}, \psi_{11}, \psi_{12,}\psi_{22}, \varphi, \varsigma_0, \varsigma_1\}.$$

1. Define a set of auxiliary parameters (ap), Λ (a vector)

2. 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.

3. From the data we construct a vector of the year of birth, $z = \{z_1, z_2, ..., z_H\}$ and a $H \times 31$ matrix D which has $d_{ht} = 1$ if the earnings of worker h at age $t + 25$ is observed and zero otherwise.

4. Generate three sets of random draws. Let W be a $H \times 4$ matrix, where each element w_{ij} is a random draw from the standard normal distribution. Let E be a $H \times 30$ matrix, where each element e_{ij} is a random draw from the standard normal distribution. Let M be $H \times 31$ matrix, where each element m_{ij} is a random draw from the standard normal distribution. All the draws are mutually independent.

5. Generate the simulated data $y_h^s = \{y_{h0}^s, y_{h1}^s, ..., y_{h30}^s\}$ for H artificial workers (see equation (A.3)) and for a given set of parameters Θ.

    1. Generate the starting value for H artificial workers, where the actual year of birth for each worker, z_h is used:

    $$y_{h0}^s = \tau_1 + \tau_2 z_h + (\tau_3 + \tau_4 z_h)\, w_{h1}.$$

    2. Generate the model parameters v_h, θ_h and λ_h

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

3. Generate simulated earnings for H artificial workers recursively:

$$\begin{aligned}
\sigma_{h1}^2 &= v_h \\
y_{h1}^s &= y_{h0}^s + \sqrt{\sigma_{h1}^2} e_{h1} \\
\sigma_{ht}^2 &= \frac{e^{\varphi}}{1+e^{\varphi}} \sigma_{ht-1}^2 e_{ht-1}^2 + v_h \quad t = 2, ...30 \\
y_{ht}^s &= y_{ht-1}^s + \sqrt{\sigma_{ht}^2} e_{ht} + \theta \sqrt{\sigma_{ht-1}^2} e_{h,t-1} \quad t = 2, ...30
\end{aligned}$$

4. Generate measurement error u_{ht}

$$u_{ht} = \lambda_h m_{ht} \quad t = 0, ...30$$

   Add measurement error:

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

6. To mimic the sampling frame the simulated earnings are included in the simulated data set if $d_{ht} = 1$.

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

8. A second replication of ap's based on simulated data set is constructed by repeating step 5 to 7 but with using $-W$, $-E$ and $-M$ instead. This is an antithetic procedure which gives an increase in the precision of the estimator that is equivalent to more than doubling the replications.

9. Calculate the weighted distance between the ap's based on the data and simulations $(\Lambda^{data} - \Lambda^s)'\widehat{\Omega}^{-1}(\Lambda^{data} - \Lambda^s)$.

10. Minimize the distance by repeating step 5 to 10 with a new set of parameters.

A.4. Likelihood function of the preferred model

In this section, we derive the likelihood function for the preferred stable model without measurement errors[19]. The model is given by:

$$\begin{aligned}
y_{ht} &= \left[\delta_h \left(\left(1-\omega_h^t\right) - \beta_h \left(1-\omega_h^{t-1}\right)\right) + \alpha_h \beta_h\right] + \beta_h y_{ht-1} + \quad (A.4) \\
& \alpha_h \left(1-\beta_h\right) t + \left(\epsilon_{ht} + \theta_h \epsilon_{ht-1}\right), \quad (A.5) \\
h &= 1, ..., H; t = 1, ..., T. \quad (A.6)
\end{aligned}$$

[19] The introduction of measurement errors increases the complexity of the algebra but it does not change the main conclusions of the analysis.

The individual model parameters and the starting value can be written as

$$v_h = e^{\phi_{11}+\phi_{12}y_{h0}+\psi_{11}\eta_{h1}} \quad (A.7)$$

$$\theta_h = \frac{e^{\phi_{21}+\psi_{22}\eta_{h2}}}{1+e^{\phi_{21}+\psi_{22}\eta_{h2}}} - 0.5 \quad (A.8)$$

$$\alpha_h = \phi_{31} + \phi_{32}y_{h0} + \psi_{31}\eta_{h1} \quad (A.9)$$

$$\beta_h = \frac{e^{\phi_{41}+\psi_{41}\eta_{h1}+\psi_{42}\eta_{h2}}}{1+e^{\phi_{41}+\psi_{41}\eta_{h1}+\psi_{42}\eta_{h2}}} \quad (A.10)$$

$$\delta_h = \phi_{51} + \phi_{52}y_{h0} + \psi_{51}\eta_{h1} + \psi_{52}\eta_{h2}$$

$$\omega_h = \frac{e^{\phi_{61}+\phi_{62}y_{h0}+\psi_{62}\eta_{h2}}}{1+e^{\phi_{61}+\phi_{62}y_{h0}+\psi_{62}\eta_{h2}}} \quad (A.11)$$

$$y_{h0} = \tau_1 + \tau_2 z_h + (\tau_3 + \tau_4 z_h)\eta_{h0} \quad (A.12)$$

where the latent factors $\eta_{h0}, \eta_{h1}$and η_{h2} are standard normally distributed: $\eta_{h0}\tilde{}N(0,1)$, $\eta_{h1}\tilde{}N(0,1)$, $\eta_{h2}\tilde{}N(0,1)$ and mutually independent and independent of a time invariant observable variable, z_h. Let Γ_h denotes the vector of model parameters $\Gamma_h = \{v_h, \theta_h, \alpha_h, \beta_h, \delta_h, \omega_h\}$.

For the stochastic term, ϵ_{ht} we allow for ARCH effects. The first and second conditional moments of ϵ_{ht} are:

$$E(\epsilon_{ht}|y_{ht-1}, y_{ht-2}, .., y_{h0}, \Gamma_h) = 0 \quad (A.13)$$

$$E(\epsilon_{ht}^2|y_{ht-1}, y_{ht-2}, .., y_{h0}, \Gamma_h) = \frac{exp(\varphi)}{1+exp(\varphi)}\epsilon_{ht-1}^2 + v_h \quad for\ t > 1 \quad (A.14)$$

$$E(\epsilon_{h1}^2|y_{h0}, \Gamma_h) = v_h \quad (A.15)$$

In order to derive the density function for the model we also make the standard assumption that $\epsilon_{h0} \equiv 0$. Furthermore, we assume that the conditional distribution of ϵ_{ht} given $y_{ht-1}, y_{ht-2}, .., y_{h0}, \Gamma_h$ is normal.

In this model we have 23 heterogeneity distribution parameters which we denote by Θ :

$$\Theta = \left\{ \begin{array}{c} \{\tau_1, \tau_2, \tau_3, \tau_4\}, \{\phi_{11,}\phi_{12}, \phi_{21}, \phi_{31}, \phi_{32}, \phi_{41}, \phi_{51}, \phi_{52}, \phi_{61,}\phi_{62}\}, \\ \{\psi_{11}, \psi_{22}, \psi_{31}, \psi_{41}, \psi_{42}, \psi_{51}, \psi_{52}, \psi_{62}\}, \varphi \end{array} \right\}.$$

The density function

When deriving the density function we distinguish between the case where the starting value is observed and where the starting value is unobserved.

Case 1: The starting value is observed. The density function is derived in three steps. First we derive the joint density for a worker conditioned on the model parameters and the starting value, or as we shall show this is equivalent to condition on the set of heterogeneity distribution parameters, the observed time invariant variable, z_h and the three latent factors. In the next step we derive the unconditional density for one worker. Finally, we derive the joint likelihood function for the whole sample.

The joint density for a worker h with starting value y_{h0} conditional on $\epsilon_{h0} = 0$ and Γ_h is given by:

$$f(y_{h1}, ..., y_{hT}|\epsilon_{h0} = 0, y_{h0}, \Gamma_h) = \prod_{t=1}^{T} f(y_{ht}|y_{ht-1}, \epsilon_{h0} = 0, y_{h0}, \Gamma_h). \tag{A.16}$$

By using the assumption of normality of ϵ_{ht} given $y_{ht-1}, y_{ht-2}, .., y_{h0}, \Gamma_h$ and denoting $\mu_{ht} = \delta_h \left((1 - \omega_h^t) - \beta_h \left(1 - \omega_h^{t-1}\right)\right) + \alpha_h \beta_h + \alpha_h (1 - \beta_h) t$ we have

$$f(y_{ht}|y_{ht-1}, \epsilon_{h0} = 0, y_{h0}, \Gamma_h) = \varphi \left(\frac{y_{ht} - (\mu_{ht} + \beta_h y_{ht-1}) - \theta_h \epsilon_{ht-1}}{\sqrt{var(\epsilon_{ht}|y_{ht-1}, y_{ht-2}, .., y_{h0}, \Gamma_h)}} \right), \tag{A.17}$$

where φ is the density function of a standard normal and $var(\epsilon_{ht}|y_{ht-1}, y_{ht-2}, .., y_{h0}, \Gamma_h)$ is defined above. The disturbances are obtained using that

$$\begin{aligned} \epsilon_{h1} &= y_{h1} - (\mu_{h1} + \beta_h y_{h0}) \\ \epsilon_{h2} &= y_{h2} - (\mu_{h2} + \beta_h y_{h1}) - \theta_h \epsilon_{h1} \\ \epsilon_{h3} &= y_{h3} - (\mu_{h3} + \beta_h y_{h2}) - \theta_h \epsilon_{h2} \\ & \quad ..... \end{aligned} \tag{A.18}$$

From equation (A.7)-(A.12) it is seen that the model parameters are completely determined by the heterogeneity distribution parameter Θ, the time invariant variable, z_h and the three latent factors $\eta_{h0}, \eta_{h1}$and η_{h2}. This means that instead of conditioning on the model parameters and starting value we can condition on Θ, z_h and η_{h0}, η_{h1} and η_{h2} :

$$f(y_{h1}, ..., y_{hT}|\epsilon_{h0} = 0, \Theta, z_h, \eta_{h0}, \eta_{h1}, \eta_{h2}) = f(y_{h1}, ..., y_{hT}|\epsilon_{h0} = 0, y_{h0}, \Gamma_h) \tag{A.19}$$

where the model parameters and starting value is given by equation (A.7)-(A.12).

In the second step we derive the "unconditional" density for worker h. By using that the three latent factors are standard normal distributed and mutually independent we get

$$\begin{aligned} f(y_{h1},....,y_{hT}|\epsilon_{h0} &= 0,\Theta,z_h) = \qquad (A.20) \\ \int\int\int f(y_{h1},....,y_{hT}|\epsilon_{h0} &= 0,\Theta,z_h,\eta_{h0},\eta_{h1},\eta_{h2})\varphi(\eta_{h0})\varphi(\eta_{h1})\varphi(\eta_{h2})d\eta_{h0}d\eta_{h1}d\eta_{h2}, \end{aligned}$$

where

$$\begin{aligned} f(y_{h1},....,y_{hT}|\epsilon_{h0} &= 0,\Theta,z_h,\eta_{h0},\eta_{h1},\eta_{h2}) \qquad (A.21) \\ &= \prod_{t=1}^{T} f(y_{ht}|y_{ht-1},\epsilon_{h0}=0,z_h,\Theta,\eta_{h0},\eta_{h1},\eta_{h2}) \end{aligned}$$

Finally, to derive the likelihood function for the whole sample we use independence across workers

$$L(\Theta|z_1,...,z_H) = \prod_{h=1}^{H} f(y_{h1},....,y_{hT}|\epsilon_{h0}=0,\Theta,z_h). \qquad (A.22)$$

Case 2: The starting value is unobserved. In this case we do not observe the starting value but instead we have the initial observation y_{hs}. The joint density of the observed sample for a worker h conditional on $\epsilon_{h0}=0$, y_{h0} and Γ_h is given by

$$\begin{aligned} f(y_{hs},....,y_{hT}|\epsilon_{h0} &= 0,y_{h0},\Gamma_h) = \\ \prod_{t=s+1}^{T} f(y_{ht}|y_{ht-1},\epsilon_{h0} &= 0,y_{h0},\Gamma_h) f(y_{hs}|\epsilon_{h0}=0,y_{h0},\Gamma_h) \qquad (A.23) \end{aligned}$$

where $f(y_{ht}|y_{ht-1},\epsilon_{h0}=0,y_{h0},\Gamma_h)$ is defined as in the case 1 with disturbances defined by:

$$\begin{aligned} \epsilon_{hs+1} &= y_{hs+1} - (\mu_{hs+1} + \beta_h y_{hs}) - \theta_h \epsilon_{hs} \\ \epsilon_{hs+2} &= y_{hs+2} - (\mu_{hs+2} + \beta_h y_{hs+2}) - \theta_h \epsilon_{hs+1} \\ &\;..... \end{aligned}$$

Note that unless we assume $\epsilon_{hs}=0$ the disturbances are not observed in this case. This assumption is difficult to verify in this case, because it implies that there is no shock the period the worker enters the sample.

Moreover, with respect to the distribution of y_{hs} given $\epsilon_{h0} = 0, y_{h0}, \Gamma_h$ for $s > 1$, we can write recursively the process so that we have

$$y_{hs} = \delta_h \left(1 - \omega_h^s\right) + \alpha_h s + \beta_h^s y_{h0} + \epsilon_{hs} + (\beta_h + \theta_h)(\epsilon_{hs-1} + \beta_h \epsilon_{hs-2} + ... + \beta_h^{s-2}\epsilon_{h1}) \tag{A.24}$$

Given this, the conditional moments of y_{hs} given $\epsilon_{h0} = 0, y_{h0}$ and Γ_h, can be shown to be:

$$\begin{aligned} E(y_{hs}|\Gamma_h, \epsilon_{h0} &= 0, y_{h0}) = \delta_h \left(1 - \omega_h^s\right) + \alpha_h s + \beta_h^s y_{h0} & \text{(A.25)} \\ V(y_{hs}|\Gamma_h, \epsilon_{h0} &= 0, y_{h0}) = v_h[\rho_s + (\beta_h + \theta_h)^2(\rho_{s-1} + \beta_h^2 \rho_{s-2} + ... + \beta_h^{2s-4})] \end{aligned}$$

where $\rho_s = \sum_{t=0}^{s-1} \left(\frac{e^{\varphi}}{1+e^{\varphi}}\right)^t$. The distribution of y_{hs} given $\epsilon_{h0} = 0, y_{h0}, \Gamma_h$ for $s > 1$ depends on the distribution of the disturbances ϵ_{ht} given y_{h0}, Γ_h (for $t < s$). This distribution is determined from the ARCH assumption and it is not normal. To illustrate this point consider for example the case $s = 2$, the second and fourth conditional moments of y_{h2} are given by:

$$\begin{aligned} V(y_{h2}|\Gamma_h, \epsilon_{h0} &= 0, y_{h0}) = v_h[1 + \frac{e^{\varphi}}{1+e^{\varphi}} + (\beta_h + \theta_h)^2] & \text{(A.26)} \\ E[((y_{h2} - E(y_{h2}|\Gamma_h, \epsilon_{h0} &= 0, y_{h0}))^4|\Gamma_h, \epsilon_{h0} = 0, y_{h0}] \\ &= E[(\epsilon_{h2} + (\beta_h + \theta_h)\epsilon_{h1})^4|\Gamma_h, \epsilon_{h0} = 0, y_{h0}] & \text{(A.27)} \\ &= 3V[(y_{h2}|\Gamma_h, \epsilon_{h0} = 0, y_{h0})]^2 - 6v_h^2 \left(\frac{e^{\varphi}}{1+e^{\varphi}}\right)^2 \\ &\neq 3V[(y_{h2}|\Gamma_h, \epsilon_{h0} = 0, y_{h0})]^2 & \text{(A.28)} \end{aligned}$$

Therefore, the ARCH effects induces non normality on the conditional distribution. Similarly the marginal distribution of a process with ARCH errors will have a non normal distribution (Engle (1982)).

An alternative strategy is to derive the joint density of the vector $(y_{hs}, y_{hs+1}, .., y_{hT})$. Instead of relying on the conditional distribution on y_{h0}, we define the joint distribution of the vector of observations for individuals with initial observation y_{hs} :

$$\begin{aligned} & f\ (y_{hs}, y_{hs+1}, .., y_{hT}; \Theta) \\ = & \int\int\int f\ (y_{hs}, y_{hs+1}, .., y_{hT}|z_h, \eta_{h0}, \eta_{h1}, \eta_{h2}; \Theta)\varphi(\eta_{h0})\varphi(\eta_{h1})\varphi(\eta_{h2})d\eta_{h0}d\eta_{h1}d\eta_{h2} \end{aligned}$$

In order to obtain the density function $f\ (y_{hs}, y_{hs+1}, .., y_{hT}|z_h, \eta_{h0}, \eta_{h1}, \eta_{h2}; \Theta)$, note that we can write the process iteratively and that in this case we do not condition on $\epsilon_{h0} = 0$:

$$\begin{aligned} y_{ht} &= \delta_h \left(1 - \omega_h^t\right) + \alpha_h t + \beta_h^t y_{h0} + \epsilon_{ht} \\ &\quad + (\beta_h + \theta_h)(\epsilon_{ht-1} + \beta_h \epsilon_{ht-2} + ... + \beta_h^{t-2}\epsilon_{h1}) + \beta_h^{t-1}\theta_h \epsilon_{h0} \\ &= \delta_h \left(1 - \omega_h^t\right) + \alpha_h t + \beta_h^t(\tau_1 + \tau_2 z_h + (\tau_3 + \tau_4 z_h)\eta_{h0}) + \epsilon_{ht} \\ &\quad + (\beta_h + \theta_h)(\epsilon_{ht-1} + \beta_h \epsilon_{ht-2} + ... + \beta_h^{t-2}\epsilon_{h1}) + \beta_h^{t-1}\theta_h \epsilon_{h0} \end{aligned}$$

Note that the joint distribution of $y_{hs}, y_{hs+1}, .., y_{hT}$ given $z_h, \eta_{h0}, \eta_{h1}, \eta_{h2}$ is obtained from the joint conditional distribution of the disturbances ϵ_{ht} which is not a normal distribution. The main conclusion is that there is no closed form for the joint distribution and we have to rely on simulation methods or non parametric density estimation (see Diebold and Schuermann (1996)). This is the case even when $\theta_h = 0$.

As shown above, in the case where the starting value is not observed and in the presence of ARCH effects there exists to our knowledge no closed solution to the likelihood function. Estimating the parameters using a likelihood approach therefore have to rely on simulated maximum likelihood, which we will leave for future work.

References

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

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

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

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

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

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

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

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

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

[10] Baltagi, B.H. and C. Kao (2000), "Nonstationary Panels, Cointegration in Panels and Dynamic Panels: A Survey", CPR working paper No 16.

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

[12] Bonhomme, S. and J.M. Robin (2004), "Modeling individual earnings trajectories using copulas", mimeo, Paris 1.

[13] 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.

[14] Browning, M. and A. Lusardi (1996), "Household Saving: Micro Theories and Micro Facts", *Journal of Economic Literature*, 34(4), 1474-1855.

[15] Buchinsky, M. and J. Hunt (1997), "Wage Mobility in the United States", Working paper, NBER, 5455.

[16] Carroll, C. and A. Samwick (1997), "The Nature of Precautionary Wealth", *Journal of Monetary Economics*, 40, 41-71.

[17] Chamberlain, G. (1980), "Analysis of Covariance with Qualitative Data", *Review of Economic Studies*, 47, 225.238.

[18] Chamberlain, G. and K. Hirano (1997), "Predictive Distributions Based on Longitudinal Earnings Data", manuscript, Department of Economics, Harvard University.

[19] Diebold, F.X. and T. Shuermann (1996), "Exact maximum likelihood estimation of observation-driven econometric models", NBER technical working paper 194.

[20] Dominitz, J. and C. Manski (1997), "Using Expectations Data to Study Subjective Income Expectations", *Journal of the American Statistical Association*, 92, 855-867.

[21] Duffie, D. and K.J. Singleton (1993), "Simulated Moments Estimation of Markov Models of Asset Prices", *Econometrica*, 61, 929-952.

[22] Engle, R.F. (1982), "Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation", *Econometrica*, 50, 987-1007.

[23] Gallant, A.R. and G. Tauchen (1996), "Which Moments to Match?", *Econometric Theory*, 12, 651-681.

[24] Geweke, J. and M. Keane (1997), "An Empirical Analysis of Income Dynamics Among Men in the PSID: 1968-1989", *Journal of Econometrics*, forthcoming.

[25] Gottschalk, P. (1997), "Inequality, Income Growth, and Mobility: The Basic Facts", *Journal of Economic Perspectives*, 11(2), 21-40.

[26] 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.

[27] Gottschalk, P. and R. Moffitt (1995), "Trends in the Autocovariance Structure of Earnings in the US: 1969-1987", manuscript, Department of Economics, Brown University.

[28] Gourieroux, C. and A. Monfort, (1996) *Simulation-Based Econometric Methods*, CORE Lectures, Oxford University Press.

[29] Gourieroux, C., Monfort, A. and E. Renault (1993), "Indirect Inference", *Journal of Applied Econometrics*, 8, 85-118.

[30] 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.

[31] Hause, J.C. (1977), "The Covariance Structure of Earnings and On-the-job Training Hypothesis", *Annals of Economic and Social Measurement*, 6, 335-365.

[32] Hsiao, C. (1986), *Analysis of Panel Data*, Cambridge University Press.

[33] 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.

[34] 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.

[35] Lee, B.S. and B.F. Ingram (1991), "Simulation Estimation of Time-Series Models", *Journal of Econometrics*, 47, 197-205.

[36] Lerman S. and C. Manski (1981), "On the use of simulated frequencies to approximate choice probabilities", in *Structural Analysis of Discrete Data with Econometric Applications*, C. Manski and D. McFadden (eds), Cambridge: MIT Press.

[37] Lillard, L.A. and Y. Weiss (1979), "Components of Variation in Panel Earnings Data: American Scientists 1960-1970", *Econometrica*, 47, 437-454.

[38] Lillard, Lee A. and R. Willis (1978), "Dynamic Aspects of Earning Mobility" *Econometrica*, 46(5), 985-1012.

[39] 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.

[40] McFadden, D. (1989), "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration", *Econometrica*, 57, 995-1026.

[41] Meghir, C. and L. Pistaferri (2004), "Income Variance Dynamics and Heterogeneity", *Econometrica*, 72, 1, 1-32.

[42] Müller, U.K. and G. Elliott (2003), "Tests for unit roots and the initial condition", *Econometrica*, 71, 1269-1286.

[43] 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.

[44] Pakes, A. and D. Pollard (1989), "Optimization Estimators", *Econometrica*, 57, 1027-1057.

[45] Rubinstein and Weiss (2005), "Post Schooling Growth: Investment, Search and Learning" Manuscript.

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

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

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

[49] Stern, S. (1997), "Simulation Based Estimation", *Journal of Economic Literature*, 35, 2006-2039.

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

[51] 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,