Modelling income processes with lots of heterogeneity.*

Javier Alvarez
Alicante University

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

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

November 2001

Abstract

All empirical models of earnings processes in the literature assume a good deal of homogeneity. For example, all authors assume either that everyone has a unit root process or that everyone has a stationary process. In contrast to this we model earnings processes allowing for lots of heterogeneity between agents. To do this we have to formulate a series of increasingly complex processes which make maximum likelihood or GMM procedures very onerous. To avoid this we use a simulated minimum distance (SMD) estimation procedure. This is the first time that such an estimator has been applied to dynamic panel data models. We fit our models to a variety of statistics including most of those considered by previous

*We are grateful to Manuel Arellano, Bo Honoré, Hide Ichimura, John Rust and participants in the 2000 UK Econometrics Study Group and 2001 ESEM, Lausanne for very helpful comments. We are grateful to Mette Lien Larsen for research assistance. This research was supported by the Danish Social Science Research Foundation (SSF).

investigators (for example, trends in the cross-section variance and transition probabilities from low income states). The principal sample we use is of a group of Danish male workers followed for 16 years. The sample we draw is very homogeneous in terms of observables such as education, age, experience, marital status and all have full year, full time employment during the period considered. Despite this observable homogeneity we find much greater latent heterogeneity than previous investigators. Applying our methods to a more heterogeneous sample drawn from the PSID we find a completely different but still very heterogeneous process is needed. This suggests that not only do processes vary a lot within groups they also vary between different samples so that detailed modelling is required in each instance. We show that allowance for heterogeneity makes substantial differences to inferences of interest. For example, we find that workers appear to trade off mean for variance in their choice of earnings process. Such a conclusion would be ruled out by a model that did not allow for correlated heterogeneity.

1. Introduction.

Estimates of the earnings process facing 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.

In Table A1 in Appendix A we present a summary of a number of the significant contributions to the earnings dynamics literature. Two important features emerge from this Table. First, whatever the process chosen, only limited allowance is made for heterogeneity. A second, related feature is that some investigators assume that everyone has a stationary process and others that everyone has a unit root but no one allows that some agents may have a stationary process and others a unit root. For a number of reasons it may be that the distinction between unit root and stationary processes is not too important if we impose that all workers are in one or the other category. First, on the estimation side, if the auto-regressive (AR) parameter in the stationary model is close to unity then with a short time series (our panel covers 16 years) it is difficult to distinguish between the two processes unless we impose enough structure so that we can also exploit the cross-section variation. Second, in most applications we take the earning process to have a finite horizon (35 years, say) so that the impacts of shocks never completely disappear even for stationary process. Finally, for many uses of our estimates, it is the path of future *discounted* earnings flows that is important. If we take a discount rate of, say, three percent then much of the distinction between the 'permanent' effects of shocks in a unit root process and in a stationary process are lost. Suppose, however, that some agents have a unit root and others have a stationary process with a relatively small AR parameter. In a model in which we impose that everyone has the same process this will likely lead to a conclusion that the common process is stationary with a high AR parameter (and, potentially, the appearance of a spurious moving average in the errors since we are mixing heterogeneous AR processes). It will also lead to considerable bias in the estimation of the mean long run effects of an earnings shock.

As we shall see below, simple 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 (see Meghir and Pistaferri (2001)) or we could allow for more heterogeneity. It is the latter line that we follow below. Other 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[1]. We are sceptical that everyone has the same process with much the same parameters. Rather it may be that different agents have different processes - some with a unit root, some with a stationary AR(1) and others with an MA(1) process, for example. Mixing from different populations with different processes is known to lead to more complicated processes if the pooled data is treated as homogeneous. Given this, our approach is to allow for a good deal more heterogeneity than previous investigators. We shall show that models that only allow for heterogeneity in the levels

[1] 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 recent survey.

perform very poorly. We then show that there is some improvement if we allow for uncorrelated heterogeneity ('random coefficients' or 'random effects') in certain dimensions but this still does not take us very far towards a satisfactory fit to the data. Thus we have to allow for correlated heterogeneity which gives rise to the usual 'initial conditions' problem (see, for example, Hsiao (1986), chapter 4). We adopt the approach of allowing that the unobserved heterogeneity is a parametric random function of the initial conditions (see, for example, Chamberlain (1980), Anderson and Hsiao (1982), Blundell and Smith (1990) and Wooldridge (2000)). For example, for a stationary AR(1) process with different means for each agent we might allow that the mean is a function of the initial values plus a random disturbance term. Parametric modelling of correlated heterogeneity has a number of advantages. First, it 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 second advantage of this way of incorporating heterogeneity is that it is easy to implement. This is important in our context since we shall be undertaking a good deal of exploratory analysis. A third advantage is that we can establish consistency of our estimator as the number of cross-section units increases, holding the number of time periods constant. That is, this avoids the 'incidental parameters' problem even when we cannot difference away the latter. A final and important advantage, which is particularly emphasised by Wooldridge (2000), is that this procedure allows us to generate quantitative predictions consequent on a change in the underlying process. For example, suppose the government introduces a new policy that reduces the short run cross-section variance of earnings (for example, by increasing income tax progressivity). Examples of outcomes that we are interested in are the consequent changes in the distribution of short run and long run risk facing an 'average' household and the persistence of poverty. 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.[2] If we know the functional relationship between heterogeneity and initial conditions then we can calculate the required outcomes. 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 parametric

[2] Of course, there may also be general equilibrium effects from such a change. We are here implicitly making whatever ancillary assumptions are needed on what else is being held constant.

assumptions. The discipline here is that the final model has to fit a wide range of different statistics.

There are three broad approaches to the econometric analysis in this context. One option is to first conduct an analysis of time series 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 and endogeneity biases. It might be possible to implement analytic or simulation based small sample corrections to the estimators properties (see, for example, Shaman and Stine (1988a and 1988b), Kiviet and Krämer (1992) or Kiviet and Phillips (1998)) but these corrections impose stronger assumptions on the distributional properties of the errors than those we would like to impose *a priori*. A second alternative is to specify a general joint distribution of parameters and then, using either conditional maximum likelihood (CML) or GMM, to 'test down' to a more parsimonious model. The main problem with this approach is that we do not have any prior idea *at all* about the distribution of the parameters. Which parameters should be heterogenous and what (joint) distribution should we take for them? Nowhere in the literature is there any indication of how to find a starting general distribution. The third general approach, which we follow here, is to conduct an explicit exploratory analysis of a series of models starting with (restricted) consensus models 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, of course, path independent (which is generally true of any exploratory analysis) but if, as we do, we end up with a model that captures all of the different aspects of the data the literature has considered then it may be satisfactory. Of course, there may be other general models that do as well and then identification would require the use of additional information.

This exploratory approach requires the fitting of a relatively large number of more and more complicated models. To fit these models we use what has come to be known as Simulated Minimum Distance (SMD). This was first introduced in Lee and Ingram (1990) in a time series context. It is also used in Duffie and Singleton (1993) in an asset pricing model using time series data and Hall and Rust (1999) (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)). As far as we aware this is the first application of SMD to panel data. SMD proceeds in a number of steps. First we calculate some 'well chosen' statistics of the data (these are known as sample auxiliary parameters). If we do not know the data generating process then these sample auxiliary parameters have an unknown probability limit (as the number of cross-section units becomes large). Next we take a parametric model for the data generating process and simulate for a particular parameter value. Then we calculate the value of the auxiliary parameters for the simulated data. If the model is well chosen in a certain 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.

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 subject to the identifying conditions discussed below which require that the auxiliary parameters chosen be 'relevant'. This 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 each model we consider, it would be very arduous. It would also be disheartening since we typically discard any model quite quickly. 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. Likelihood functions for this are not easily derived. A third 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 fourth advantage is that when a simpler model fits badly the SMD procedure often suggests a very natural dimension in which to generalise the model. Finally, SMD does preclude the use of more conventional techniques. Once a particular model has been selected that fits the data well in all the dimensions considered we can adopt, say, a conditional ML scheme for the final estimation. 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. Second, the procedure is inefficient relative to maximum likelihood (that is, it will not generally attain the CR lower bound). Our feeling is that efficiency is of less importance when we are in a state of ignorance and that it is better to have an inefficient estimator of a model that fits well rather efficient estimates of a poorly fitting model.

The estimation method described above is applied to the study of two samples. The first is a sample of male workers drawn from Danish administrative data from 1981 to 1996. These data have a number of advantages which will be discussed below, but for now we simply note that although we draw a *very* homogeneous and balanced panel of workers, the number of cross-sections unit is large (2119) and the time series dimension is also relatively long ($T+1 = 16$). Moreover, these data are drawn from administrative and tax records and are likely to be less susceptible to measurement error than survey information. We also present comparison results on a much less homogeneous sample drawn from the U.S. PSID. This is a sample of 792 workers observed for 16 years over the period 1968 to 1998.

The main result of the empirical analysis is that even for the very homogenous Danish sample, there is much more heterogeneity than previous researchers have allowed for. We also find that we need a different process for the Danish sample than for the PSID sample. For the Danish sample we found a unit root for everyone, albeit with very heterogeneous parameters. For the PSID sample, we need to allow for a mixture of agents with a unit root and some with a stationary process. We then go on to show that the additional heterogeneity we find has a substantial impact on outcomes of substantive interest.

In the next section we discuss the formalities of the SMD estimator. In section 3 we discuss the choice of models and which auxiliary parameters to choose for the SMD estimation step. We also present results from a Monte Carlo study that uses simulated data that replicates the main features of the data used in the empirical section. In section 4 we discuss the data we use. In section 5 we present results. Section 6 presents an analysis of whether allowing for heterogeneity makes much difference for some selected areas of substantive interest.

2. Simulated minimum distance (SMD).

In this section, we present a simulated minimum distance estimator (SMD) for the parameters of a model of the individual income process. To motivate the utility of the method, suppose we have a sample $(y_1, ..., y_H)'$ where $y_h = (y_{h0}, y_{h1}, ..., y_{hT})$ for $(T+1)$ periods on each of H agents and we specify a model for the individual income given by a first order Markov process:

$$\{f_0(y_{ht}|y_{h,t-1};\theta)\}_{t=1...T,h=1...H} \text{ with } y_{h0} \text{ given} \tag{2.1}$$

where θ is a k-vector of parameters in a compact space Θ. The model is assumed to be correctly specified in the sense that there exists a value of the parameters $\theta = \theta_0$ such that $f_0(y_{ht}|y_{h,t-1};\theta_0)$ is the true generating process. Given this, we could apply maximum likelihood (ML) procedures or method of moment (MOM) procedures to generate an estimator for θ_0. For example, suppose that the conditional distribution given in (2.1) implies a set of q ($\geq k$) moment conditions on any realisation for one unit, y, which takes the form $G(y, \theta_0) = 0$. For the sample described above, we take the empirical analogue of this, $G_H(\theta) = H^{-1}\sum_{h=1}^{H} G(y_h, \theta)$ and define the MOM estimator:

$$\hat{\theta}_{MOM} = \arg\min_{\theta} ||G_H(\theta)|| \tag{2.2}$$

for some distance function $||.||$. If $q = k$ then we define the MOM estimator implicitly by $G_H\left(\hat{\theta}_{MOM}\right) = 0$. Sometimes the evaluation of $G_H(\theta)$ is very onerous and we must have recourse to methods that rely on simulation. In this case we replace $G_H(\theta)$ by an unbiased simulator $\tilde{G}_H(\theta)$, where the simulator is for the model given in (2.1). This gives the Method of Simulated Moments (MSM) estimator:

$$\hat{\theta}_{MSM} = \arg\min_{\theta} ||\tilde{G}_H(\theta)|| \tag{2.3}$$

See Pakes and Pollard (1989) and McFadden (1989) for the original analyses and Stern (1997) for a survey of the subsequent literature.

The MOM and MSM methods rely on being able specify $f_0(.)$ and to derive implied moments $G(.)$. Sometimes this is very difficult or the resulting moment conditions are intractable. In this case, it may be possible to replace the true model with an auxiliary model that is 'close' to the true model and to work with that. This is the motivation for the closely related analyses of Lee and Ingram (1991), Duffie and Singleton (1993), Gallant and Tauchen (1996) (GT), Gourieroux, Monfort and Renault (1993) (GMR) and Hall and Rust (1999) (HR). These papers all present variants of what HR term *Simulated Minimum Distance* (SMD). Formally, MSM is also within the class of SMD estimators since we can always take the 'true' model as the auxiliary model. To elucidate how SMD generalises MSM, we give two leading examples.[3] Consider first a time series model

[3] These examples are illustrative only. For both of them we would usually use maximum likelihood or GMM.

in which we wish to estimate the parameters of an $MA(1)$ process. MSM takes the $MA(1)$ and finds associated moment conditions and then uses a simulator for these moment conditions. In contrast, an SMD estimator would take an auxiliary model such as an $AR(2)$ process and then match estimates of this model from the data and from a simulator generated under the $MA(1)$ assumption. As a second example, consider a multinomial Probit model. In this case maximum likelihood requires the evaluation of a high dimensional integral and we need to fall back on simulation. This requires an unbiased simulator for the choice probabilities under the Normal assumption. Given such a simulator, we would construct an MSM estimator based on the moments implied by the multinomial model (see, for example, Stern (1997), equation (2.27)). An alternative SMD estimator could be based on, for example, linear probability (OLS) estimates for each choice (except the last one). Thus we would estimate the parameters of these equations for the data and then choose model parameters to give the same parameters when we estimate on simulated data[4], where the simulations are based on the multinomial Probit model.

In our empirical work we also use SMD but our motivation for using it is somewhat different from the papers mentioned above. We start from the position in which we have almost no idea of the form of the true generating process. For example, even if we assume that everyone has the same finite parameter process we might allow that the parameters of this process vary across workers. The question that then arises is how we should allow for (possibly correlated) heterogeneity; the possibilities are literally without limit and the current literature gives almost no guide since it allows so little heterogeneity. Given this we have to conduct an exploratory analysis which involves formulating and estimating a series of increasingly general models until we find a satisfactory model. These models often involve integration across the multiple dimensions of the heterogeneity. In this case, it is prohibitively time and energy consuming to generate a series of ML models only to discard them almost immediately when they fail to 'fit' the data. Equally, we shall treat concerns regarding efficiency as being of second order importance in the exploratory phase, relative to fitting well many different aspects of the data. We suggest using a more efficient estimation procedure once a preferred model has been found. We now describe the main steps in the estimation method.

[4] We require equality since we have as many estimated parameters as model parameters (that is, $q = k$).

2.1. Sample auxiliary parameters.

The first step in the SMD procedure is to specify a set of q *sample auxiliary parameters* (to use the terminology of GMR). These are simply statistics of the data, denoted by the q-vector $\hat{\gamma}(y_1, ..., y_H)$. For the GMR Indirect Inference procedure the auxiliary parameters are maximisers of a given data dependent criterion which constitutes an approximation to the true DGP. In the GT approach the auxiliary parameters are data dependent functions that maximise a 'score generator' (quasi-likelihood function) which nests the true model. This typically involves taking a flexible (instrumental) model with a large number of parameters that provides a good approximation to any distribution thought possible (this is captured in an embedding assumption discussed below). In HR the auxiliary parameters are simply 'well-chosen' statistics of the data; this is the broad approach we adopt.[5] The central idea for SMD, as emphasised by HR, is very close to calibration. Given the model and choice of parameters we can generate model auxiliary parameters. We then choose the parameters of the model to give a best fit to the sample auxiliary parameters. The important conceptual point is that even if the probability limits of the auxiliary parameters under the true DGP are unknown, it is still possible to estimate the parameters of the true model under a set of assumptions. In the following we will state the assumptions needed.

In our panel data application the auxiliary parameters take a 'fixed T' cross-section mean form:

$$\hat{\gamma}(y_1, ..., y_H) = \frac{1}{H}\sum_{h=1}^{H} m(y_h), \tag{2.4}$$

where $m(.)$ is a vector valued function of the $T+1$ time series realisations for a given agent y_h. The first assumption concerns independence across cross-section units.[6]

Assumption A1

The vectors of observations y_h are independent across cross-section units

Moreover we also assume that the auxiliary parameters have an expectation and a covariance matrix under the true distribution.

Assumption A2

The auxiliary parameters, $\hat{\gamma}$, have expectations and a covariance matrix under the true distribution defined by:

[5] To reconcile this with the GMR approach take the data dependent criterion function to be $Q_{H,T} = -(\hat{\gamma} - \tau)'(\hat{\gamma} - \tau)$ and maximise this with respect to τ.

[6] In the empirical application we will relax this assumption and allow for time effects.

$$\gamma_0(\theta_0) = E_0(\hat{\gamma}(y_1,..,y_H)) \quad (2.5)$$
$$V_0(\theta_0) = Var_0(\hat{\gamma}(y_1,..,y_H)) \quad (2.6)$$

where E_0 and Var_0 denote the expectation and the covariance matrix with respect to the true distribution $f_0(;\theta_0)$.

We term $\gamma_0(\theta_0)$ the *population auxiliary parameter.* Note that assumptions *A1* and *A2* imply that:

$$p\lim_{H\to\infty} \hat{\gamma}(y_1,...,y_H) = \gamma_0(\theta_0) \quad (2.7)$$

so that $\hat{\gamma}(y_1,...,y_H)$ is a consistent estimator for $\gamma_0(\theta_0)$.

2.2. Simulated auxiliary parameters.

The second step of the estimation method is to simulate from the model $\{f_0(y_{ht}|y_{h,t-1};\theta)\}$ with a given set of parameters θ and on the basis of the simulations to calculate the auxiliary parameters. We assume that the model is so that it can be simulated, conditional on a set of initial values and on a given value of the parameters θ. The simulated paths depend on an $(HT \times 1)$ vector of draws from the error distribution specified by the model. These draws are kept fixed during the estimation process to stabilise the iterative estimation procedure and in order to satisfy the equicontinuity conditions necessary to establish asymptotic normality. Let $(y_1^s,...,y_H^s)'$ be a set of H simulated paths conditional on the starting values in the observed data (so that $y_{h0}^s = y_{h0}$ for all h and s). Replicating the procedure S times, we define the *simulated auxiliary parameters* by:

$$\gamma^{SH}(\theta; y_{10},...,y_{H0}) = \frac{1}{S}\sum_{s=1}^{S}\hat{\gamma}(y_1^s,...,y_H^s) \quad (2.8)$$

The notation emphasises that these values depend the initial values in the data but we shall usually simply write $\gamma^{SH}(\theta)$. We make the following assumptions on the simulated auxiliary parameters:

Assumption A3

$\gamma^{SH}(\theta)$ is a continuous function of θ

$\gamma^{SH}(\theta)$ is differentiable at $\theta_0$with a derivative matrix of full rank

We also assume that the model is chosen so that the simulated values converge uniformly (as H becomes large) in θ to a deterministic function $\gamma^{\infty}(\theta)$ that we term *predicted auxiliary parameters*:

Assumption A4

$$p\lim_{H\to\infty}\gamma^{SH}(\theta)=\gamma^{\infty}(\theta)\,,\ \textit{uniformly in }\theta$$

Finally, the limit function $\gamma^{\infty}(\theta)$ verifies the following assumption:

Assumption A5

$$\gamma^{\infty}(\theta)\ \textit{is one to one and}\ \bigtriangledown\gamma^{\infty}(\theta_0)\ \textit{is of full-column rank}$$

The last assumption we need is to establish a link between that the population auxiliary parameter and the predicted auxiliary parameter.

Assumption A6:

The model is correctly specified the sense that there exist a θ_0 *such that the simulated paths,* $y^s(\theta_0)$*, have same distributions as the observed data* y*.*

Assumption A6 is similar to assumption 4 in HR and assumption A5 in GMR. Assumption A6 implies that

$$\begin{aligned}\gamma_0(\theta_0) &= \gamma^{\infty}(\theta_0)\\ V_0(\theta_0) &= Var_0(m(y_h^s)).\end{aligned}$$

If first expression is violated we will say that the model is *misspecified.* Notice that the concept misspecified is defined relative to the choice of auxiliary parameters.

A weaker condition than Assumption 6 allows for an approximate model. In this case the value θ_0 minimises the distance between the simulated and the true generating process[7]:

$$\theta_0=\arg\min_{\theta\in\Theta}\|\gamma_M^{\infty}(\theta)-p\lim_{H\to\infty}\hat{\gamma}(y_1,...,y_H)\| \tag{2.9}$$

This approach has the advantage that under weak assumptions a true value for the model parameters always exists. We prefer not to go this far since the use of approximate models in the final use for the model (for example, calibrating a GE model) is problematic. Additionally inference with approximate models is not completely clear-cut.

[7] This is equivalent to what is proposed in Hall and Rust (1999), equation (95).

2.3. The estimator.

The third step in the estimation procedure is to define the estimator on the basis sample auxiliary parameter and the simulated auxiliary parameters. The central point of SMD is that we can replace the (unobserved) population auxiliary parameters and the (analytical intractable) predicted auxiliary parameters with the sample and simulated auxiliary parameters, respectively, and thereby obtain a consistent estimator for θ_0. To do this, we first specify a $q \times q$ data dependent, symmetric and positive definite matrix A_H with the property that $p\lim A_H = A_\infty$, a non-stochastic positive definite matrix. We define the *simulated minimum distance* (SMD) estimator by:

$$\hat{\theta}_{SMD} = \arg\min_{\theta\in\Theta} \left(\gamma^{SH}(\theta) - \hat{\gamma}(y_1, ..., y_H)\right)' A_H \left(\gamma^{SH}(\theta) - \hat{\gamma}(y_1, ..., y_H)\right) \quad (2.10)$$

Given continuity of $\gamma^{SH}(\theta)$ and compactness of Θ, the SMD estimator always exists. The estimate is locally unique if $\nabla_\theta \gamma^{SH}\left(\hat{\theta}_{SMD}\right)$ has full rank. In general the value and finite sample properties of $\hat{\theta}_{SMD}$ will depend on the choice of the weighting matrix.

Below we choose to work always with the just identified case for which we have as many parameters in our model as auxiliary parameters ($k = q$). Then the choice of weighting matrix is irrelevant and we set:

$$\gamma^{SH}\left(\hat{\theta}_{SMD}\right) = \hat{\gamma}(y_1, ..., y_H) \quad (2.11)$$

If there is no solution for this equation then we say that the model is *misspecified* (relative to the choice of auxiliary parameters).[8]

2.4. Inference.

We now establish the asymptotic properties of the SMD estimator defined in (11) under assumptions *A1-A6.* With respect to the role of our assumptions, note that the assumptions *A3-A4* are standard in order to establish consistency of the estimator. Pakes and Pollard established consistency of its simulation estimator in a more general context without assuming continuity in the criterion

[8] The same can be found when estimating non-linear models using moment conditions.

function[9]. The uniform convergence assumption *A4* is in the line of condition (iii) of Theorem 3.1 in Pakes and Pollard or assumption 6 in HR. As we have seen above, assumption *A5* is identification condition for θ_0 and it is similar in essence to assumption A4 in GMR or assumption 5 in HR.

In practice, in our empirical work, we only need the distribution of the difference $\left(\gamma^{SH}(\theta) - \widehat{\gamma}(y_1, ..., y_H)\right)$ for inference, but we shall also provided the distribution of the estimator $\widehat{\theta}_{SMD}$ to conform with standard presentations. To establish consistency, note that given assumptions *A1-A6* $p\lim_{H\to\infty} \widehat{\theta}_{SMD} = \theta_0$. To see this, note that given assumptions *A1-A6* , the criterion in (2.10) converges in probability uniformly to a limit function:

$$\left(\gamma^{\infty}(\theta) - \gamma_0(\theta_0)\right)' A_{\infty} \left(\gamma^{\infty}(\theta) - \gamma_0(\theta_0)\right), \tag{2.12}$$

and the limit function in this expression attains a unique global minimum at $\theta = \theta_0$, from which consistency follows.[10]

To derive the variance of

$$\left(\gamma^{SH}(\theta) - \hat{\gamma}(y_1, ..., y_H)\right)$$

we use the following decomposition:

$$\gamma^{SH}(\theta) - \hat{\gamma} = \left(\gamma^{SH}(\theta) - \gamma^{\infty}(\theta)\right) + \left(\gamma^{\infty}(\theta) - \gamma_0\right) + \left(\gamma_0 - \hat{\gamma}\right)$$

If the model is correctly specified (assumption 6) the middle term disappears when evaluating at $\theta = \theta_0$, we can write:

$$\sqrt{H}\left(\gamma^{SH}(\theta_0) - \hat{\gamma}\right) = \sqrt{H}\left(\gamma^{SH}(\theta_0) - \gamma^{\infty}(\theta_0)\right) + \sqrt{H}\left(\gamma_0 - \hat{\gamma}\right) \tag{2.13}$$

For the first term on the right hand side we have (using the definition given in (2.8)):

$$\begin{aligned} \sqrt{H}\left(\gamma^{SH}(\theta_0) - \gamma^{\infty}(\theta_0)\right) &= \sqrt{H}\left(\frac{1}{S}\Sigma_{s=1}^{S}\widehat{\gamma}(y_1^s, ..., y_H^s) - \gamma^{\infty}(\theta_0)\right) \\ &= \sqrt{H}\left(\frac{1}{SH}\Sigma_{s=1}^{S}\Sigma_{h=1}^{H} m(y_h^s) - \gamma^{\infty}(\theta_0)\right) \end{aligned} \tag{2.14}$$

[9] Notice that our SMD estimator may be defined in terms of their formulation (see example 4.1 in Pakes and Pollard) as the estimator which minimizes the random function:

$$||\tilde{G}_H(\theta)|| = ||A_H^{1/2}(\gamma_M^{SH}(\theta) - \hat{\gamma})||$$

[10] Consistency may be also obtained by application of Theorem 3.1 in Pakes and Pollard. It can be shown that assumptions *A1-A5* imply the necessary conditions for consistency in the theorem.

which is a sum of random variables which are independent across simulations and across cross section units. By the CLT, for fixed S we have:

$$\sqrt{H}\left(\gamma^{SH}\left(\theta_0\right)-\gamma^{\infty}\left(\theta_0\right)\right)\xrightarrow{d}N\left(0,\frac{1}{S}V_0\right) \tag{2.15}$$

If we allow the number of replications to become large then the simulated auxiliary parameters converge to the (non-stochastic) predicted auxiliary parameters. For the second term, we have, by standard CLT derivations, that:

$$\sqrt{H}\left(\gamma_0-\hat{\gamma}\right)\xrightarrow{d}N\left(0,V_0\right) \tag{2.16}$$

Since the simulations and the data follow the same process and they are independent this gives:

$$\sqrt{H}\left(\gamma^{SH}\left(\theta_0\right)-\hat{\gamma}\right)\xrightarrow{d}N\left(0,\frac{S+1}{S}V_0\right) \tag{2.17}$$

Thus the variance for estimation with only one replication of the model ($S=1$) is twice that of estimation that uses the analytical solution $\gamma^{\infty}\left(\theta_0\right)$. A test for the over-identifying restrictions implied by the model is based on the statistic:

$$\begin{aligned} GF &= \frac{HS}{S+1}\left(\gamma^{SH}\left(\widehat{\theta}_{SMD}\right)-\hat{\gamma}\left(y_1,...,y_H\right)\right)'\widehat{V}_0^{-1} \\ &\quad \left(\gamma^{SH}\left(\widehat{\theta}_{SMD}\right)-\hat{\gamma}\left(y_1,...,y_H\right)\right), \end{aligned} \tag{2.18}$$

where GF is asymptotically distributed as a χ^2 distribution with $q-k$ degrees of freedom when the model is well specified and the embedding assumption holds.

2.5. The asymptotic distribution of SMD estimator

We now turn to consider the asymptotic distribution of $\widehat{\theta}_{SMD}$. From the first order condition and a Taylor expansion of $\gamma_M^S\left(\hat{\theta}_{SMD}\right)$ about $\theta=\theta_0$, we can show that:

$$\sqrt{H}\left(\hat{\theta}_{SMD}-\theta_0\right)\xrightarrow{d}N\left[0,\frac{S+1}{S}G_1^{-1}G_2G_1^{-1}\right], \tag{2.19}$$

where

$$\begin{aligned} G_1 &= (p\lim_{H\to\infty}[\nabla\gamma^{SH}\left(\theta_0\right)])'A_{\infty}(p\lim_{H\to\infty}[\nabla\gamma^{SH}\left(\theta_0\right)]) \\ G_2 &= (p\lim_{H\to\infty}[\nabla\gamma^{SH}\left(\theta_0\right)])'A_{\infty}V_0A_{\infty}(p\lim_{H\to\infty}[\nabla\gamma^{SH}\left(\theta_0\right)]) \end{aligned} \tag{2.20}$$

Therefore, the optimal weight matrix is given by $A_H = V_0^{-1}$. In this case, the asymptotic variance of the optimal SMD estimator which uses a consistent estimator of V_0^{-1} as weight matrix is $\left(\frac{S+1}{S}\right) G^{-1}$, where

$$G = (p \lim_{H\to\infty} [\nabla \gamma_M^{SH} (\theta_0)])' V_0^{-1} (p \lim_{H\to\infty} [\nabla \gamma_M^{SH} (\theta_0)]). \tag{2.21}$$

As we stated above, we do not use this for inference, but rather changes in the value the over-identifying restriction value.

2.6. Choosing the number of auxiliary parameters.

The number of potential auxiliary parameters, q, usually exceeds the number of model parameters, k. This suggests two alternative estimation strategies. One is to use all the auxiliary parameters in fitting and to use (2.18) to test for the validity of the over-identifying restrictions. An alternative is the just identified (JI) procedure in which we choose k of the auxiliary parameters and fit on these. Then we can use the $q - k$ auxiliary parameters not used in fitting to generate goodness of fit test statistics similar to equation (2.18). The expression for the goodness of fit test is given by (2.18) replacing $\hat{\gamma}(y_1, ..., y_H)$ with the vector of statistics not used in fitting and $\gamma_M^{SH}\left(\widehat{\theta}_{SMD}\right)$ by the vector of simulated statistics using the estimated parameter values from the JI procedure.

There are a number of advantages to the JI procedure. The first is that the JI procedure focuses on the chosen statistics and allows us to examine directly how well we fit the other statistics. As we shall see, this usually gives clear indications of how to generalise the model. A second and important practical advantage of the JI procedure is that we can be sure that the iterative estimation procedure has converged since then the criterion should be zero if the model is correctly specified. As we shall see below, all but the simplest models have many local minima and it is a major practical matter to be able to ensure that we have a global minimum for the criterion. A third advantage of the JI procedure is that the parameter estimates are independent of the chosen weighting matrix (but note that the use of the JI procedure is simply a choice of weighting matrix for all the statistics that gives zero weight to some of them). The disadvantages of the JI procedure are that the parameter estimates are, of course, sensitive to which statistics we choose to fit to. We present some Monte Carlo evidence on the two approaches below; this investigation suggests strongly that the JI procedure also has other advantages.

3. An earnings process.

3.1. The model of earnings.

We shall follow previous investigators and adopt a two step procedure (see Appendix A). First we 'control' for age, experience, education and macro effects by selecting the sample on education and being in a particular year of birth group and then by regressing log earnings on age and experience variables and time dummies (details are given below). In the second step we model the residuals as a univariate process. Although this two step procedure is not necessarily coherent with general heterogeneity schemes we adopt it to minimise the differences between this study and previous studies so as to concentrate on the effect of allowing for unobservable heterogeneity. In the second stage we model the residuals from the first round regression as an ARMA(1,1):

$$y_{ht} = \alpha_h + \beta_h y_{h,t-1} + \varepsilon_{ht} + \theta_h \varepsilon_{h,t-1}, \;\; \varepsilon_{ht} \sim iid(0, \sigma_h^2) \tag{3.1}$$

where, to save, notation, y_{ht} denotes the residual for agent h in period t (which we shall refer to as earnings below).[11] Note that we do not restrict any of the parameters to be the same for different individuals. In particular, we do not restrict β_h so that some agents could have a unit root (or even an explosive process) and others a stationary process.

In our empirical work we shall consider three classes of models. First we restrict attention to models in which everyone has a unit root. We begin with the simplest such model:

$$\Delta y_{ht} = \varepsilon_{ht} + \theta \varepsilon_{h,t-1}, \;\; \varepsilon_{ht} \sim iid(0, \sigma^2) \tag{3.2}$$

in which everyone has the same (driftless) process with the same $MA(1)$ parameter and error variance. We then consider a sequence of increasingly general unit root models that culminate in a model that allows for considerable correlated heterogeneity in the parameters of individual processes. As we shall see this most general unit root model fails to fit the data in significant directions. We then consider a series of stationary models, beginning with the simplest in which we have:

$$y_{ht} = \alpha_h + \beta y_{h,t-1} + \varepsilon_{ht} + \theta \varepsilon_{h,t-1}, \;\; \varepsilon_{ht} \sim iid(0, \sigma^2) \tag{3.3}$$

[11] Some investigators use a model which has an MA(2) structure for first differences so that these models are not captured by our general model. In practice, we found no evidence of significant second order auto-correlation in our sample (see the dsicussion of the data below) so we simply consider only the first order scheme.

and the individual intercepts are a parametric random function of the initial values. Once again we consider a series of increasingly general models all of which maintain that the auto-regressive parameter β is the same for everyone and once again we conclude that we cannot fit the data with this class of model. Finally we consider models that allow for heterogeneity in the auto-regressive parameter with the possibility that a proper subset of agents have a unit root.

3.2. Which statistics to fit?

Below we shall present and fit 22 statistics derived from a panel of univariate earnings processes. The choice of statistics we consider is motivated by three (closely related) considerations. First, 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. Clearly these two sets of measures are closely related but there is no hope of finding a general analytical relationship that would allow us to fit to one set and then derive the implications for the other. Thus we use both sets of statistics.

A second motivation for some of the statistics we consider is that they can be readily related (informally) to the parameters of our models; that is they have an informal structural interpretation. Specifically, we run a simple OLS regression of current (log) earnings on lagged earnings for each agent (for the 16 years we observe each). We record the parameter estimates of the intercept and slope, the residual variance and the first order auto-correlation of the OLS residuals. To illustrate out motivation, consider the OLS slope parameter. This is certainly not an unbiased estimator of the 'true' slope parameter (see Kendall (1954)) 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 agent); speed is important for a simulation based method such as SMD.

A third motivation derives from a consideration of the uses for estimates of earnings processes (see the brief discussion at the start of the paper and section 6). Typically, a different use leads to emphasis on different statistics.

We now present a detailed description of the set of 22 auxiliary parameters that we use in the SMD estimator. As discussed above, the first set comes from OLS regressions for each agent. Specifically, we first run an OLS regression of y_{ht} on a constant and y_{ht-1} for $t = 3, 4..., 16$ (here and below we do not use the first observation since we use this in the simulations). We then take means and variances and covariances over the sample of the OLS intercept and slope parameters, the log of the residual variance[12] and the residual first auto-correlation. This gives us 14 statistics (four means, four variances and six covariances). The next statistic captures the change in the dispersion of the distribution of earnings over time.[13] We shall be considering a sample that is very homogeneous in terms of education, age and marital status so that the interest here lies in the time series trend in 'within group' inequality emphasised by Gottschalk and Moffit (1995). Specifically, we calculate the cross-section unconditional variance in each year and then regress these 15 time series statistics (recall again that we do not use the first period information) on a trend and record the coefficient value on the trend and the variance of the errors from this regression. As we shall see this trend is significantly positive. The source of this increase in the unconditional variance is of of particular interest. If the underlying processes (or some of them) have a unit root then the cross-section unconditional variance will increase over time even if we take the individual conditional variances to be constant. It is thus of considerable interest to test whether the increase in the unconditional variance is solely to some agents having a unit root or whether we also have some increase in the conditional variance which is the short earnings risk that agents face. The second statistic, the variance about the (deterministic) trend will be useful in determining the source of the heterogeneity in the individual variances; specifically , we can use it to distinguish whether macro shocks have a differential impact on the income processes.

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). We take first differences for each agent and then record

[12]The log of the variances is taken since this is closer to being Normally distributed than the variances.

[13]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 mean across the sample of the variance and the first three auto-correlations of the first differences. Finally we include three mobility measures. Since the usual concern is with the duration of low income spells we take the proportions of those in the bottom quintile in the second year of our data who are in the bottom quintile in the third year and in the final year and the proportion of agents who are in the bottom quintile for more than half of the observed periods. Counting up we see that we have a total of 22 statistics: 14 OLS based statistics, 2 trend coefficients, 3 means of first differenced statistics and 3 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, capturing most of the concerns of previous investigators and being close to (or perhaps even including) sufficient statistics for likelihood functions for all of our models.

3.3. Monte Carlo simulations.

To illustrate the use of our estimator and also to resolve some of the outstanding issues discussed above we present here a simple Monte Carlo study. We first simulate some synthetic data using a known process. We then estimate the parameters of the data many times using the SMD estimator. The variables in the study are the number of simulations used in the estimation (denoted S above) and whether to use the JI or the OI option.

We take the simple unit root model with an $MA(1)$ error:

$$y_{it} = \alpha + y_{i,t-1} + \varepsilon_{it} + \theta\varepsilon_{i,t-1} \tag{3.4}$$

The parameters for the simulation are chosen to be similar to the parameter estimates from our data (presented below) for this simple model: $\alpha = 0$, $\theta = -0.1$ and $\sigma_\varepsilon = 0.05$. To generate the synthetic data (which we do only once) we use a sample size the same as in our data ($H = 2119$) with the first two initial values for each simulated household being set equal to those in the data (recall that the estimation procedure only uses observations 2 to T so that the first observation for each household is irrelevant).

For each Monte Carlo simulation of the SMD estimation we choose the number of replications of the data $S = 1, 2$ or 5 and for each simulation we estimate using both the JI procedure and the OI procedure. For the JI procedure we fit to the mean of the OLS intercepts, the variance of the OLS residuals and the first order auto-correlations. Table 1 presents the results. We present statistics for the three parameter estimates and for the χ^2 goodness of fit test for the JI procedure and the χ^2 over-identification test statistics for the OI case. Since the two test statistics have degrees of freedom equal to 19 for the JI and OI cases, we would expect that the means and variances are centred on 19 and 38.

The principal features of the Monte Carlo results are:

1. All estimators look unbiased (the mean parameter estimates are very close to true values).

2. The OI estimator sometimes converges to the 'wrong' parameter estimates (it finds a local minimum rather than a global one). This results in high standard deviations for the parameter estimates (relative to the JI standard deviations) and very high values for the χ^2 test parameter. As can be seen, the means and variances for the latter are very high whereas the medians look reasonable.

3. The JI parameter estimates are relatively precise (recall that all the values are multiplied by 100) with the higher S estimates having less dispersion, almost exactly as predicted by the theory.

4. The means for the $\chi^2(19)$ goodness of fit statistics for the JI estimates are a little low, so that we would tend to under-reject relative to the nominal size but the bias is not dramatic.

5. The JI, $S = 1$ $\chi^2(19)$ estimates are better than for the higher S cases, in the sense that the mean and variance are closer to their theoretical counterparts.

The lessons we draw from these simulations are: the SMD model does a good job of estimating the parameters of a simple model; the OI procedure is unstable and sometimes converges to a local minimum; using one simulation of the data ($S = 1$) is better than using more simulations both because it is faster and also because it yields better test statistics in our simulations. In our empirical work below we consequently use the JI procedure with S set equal to unity.

4. The data.

The Danish earnings data come from the administrative data collected and collated by Statistics Denmark. This is based on information collected by a number of different administrative authorities. From the central register containing information on the *entire* population in Denmark over a 16 year period, a ten per cent sample is drawn. The most notable features of the Danish data are: a large and representative sample; no attrition (except for natural attrition due to death or immigration); low measurement error; real measures of experience (not age minus schooling) and the possibility to observe many other correlates (for example, residential and marital status, labor force status, health etc.) although only limited number of these used here.

The data cover the period 1981-1996 and contain annual information. We select a sample of male skilled blue collar workers who are some time from their training period. Specifically: men aged between 30 and 39 in 1981 and who have basic schooling (to age 16) and formal vocational training (for example, plumbers and electricians). To avoid problems with unemployment (which requires a separate study of its own), we condition on being continuously employed in a full-time job (with no self-employment) for the whole period. We select men who were continuously married or cohabiting with the same spouse in all 16 years. Finally we have eliminated individuals with an unreasonably low income in any year (individuals with an annual income below $30,000$ Danish Crowns in 1980). After all these selection criteria, we end up with a balanced panel of 2119 workers. Compared to the earlier studies of income dynamics this is a very homogeneous sample. The motivation for taking such a homogeneous sample is that if we find significant heterogeneity for this group then *a fortiori* it is likely for more heterogeneous groups considered in literature.

In the all following analyses we are using annual labour gross income deflated by the consumer price index. The earnings variable is constructed from the tax register and are therefore believed to be very reliable. To be consistent with most previous studies we first regress log earnings on year dummies and a set of time invariant covariates (year of birth and experience) and work with residuals from these regressions. We present a fuller description of the data and some summary statistics in Appendix D. In figure 4.1 we present five earnings paths (strictly, residuals) drawn at random from our sample. As can seen, the levels and the paths differ quite radically across individuals. Figures such as these (and those presented in Chamberlain and Hirano (1997)) underpin our feeling that we need more heterogeneity than is conventionally allowed for in previous studies.

The sample auxiliary parameters are presented in Column 1 of Table 2 (with bootstrap based standard errors below). The OLS statistics are given in rows 1 to 14. The mean of the OLS intercept is close to zero (0.005) with a value for the standard deviation equal to 0.128. The mean of the OLS autocorrelation. is also close to zero and negative (-0.0179) with a standard deviation of 0.213. On

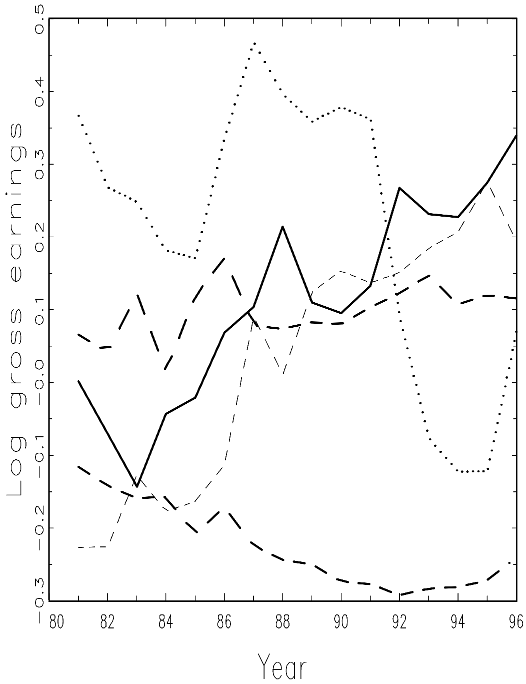

Figure 4.1: Five earnings paths.

the other hand, the distribution of the OLS slope parameter has a mean value of 0.6 with a standard deviation of 0.307. With respect to the covariances among the OLS parameters, we observe that there is a positive covariance between the intercept and the OLS log variance and negative covariances between the OLS slope and the OLS log variance and between the OLS intercept and the OLS slope.

To capture changes in the distribution of earnings over time we calculate the trend in variances and the variance about this trend, see rows 15 and 16. The variance of the first period levels of logs is 0.048 so that the trend coefficient of 0.003 represents a doubling in the variance over the 16 years of our data which is a pronounced increase in dispersion over time.[14] This is qualitatively similar to the result of Gottschalk and Moffit (1994) who also find an increase in the within group variance over time. It is also consistent with the very widespread finding that the inequality of earnings has been increasing through our sample period (see, for example, Buchinsky and Hunt (1997)) but note that 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 or all of the agents have a unit root with a consequent 'fanning out' of variance over time. It is an important goal of our research to see whether we can determine what is generating this increase in dispersion.

For the three statistics derived from first differences (rows 17-19) we see that the variance is 0.07 which represents quite a high variance for growth. The first two auto-correlations are -0.19 and -0.02 respectively which suggests that at most we have an MA(1) process in first differences. These statistics 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) except that they have somewhat more second order auto-correlation. The ranges (over years) of the Abowd and Card statistics are: $\text{var}(\Delta \ln y) \in [0.09, 0.20]$, $\text{auto1}(\Delta \ln y) \in [-0.54, -0.10]$ and $\text{auto2}(\Delta \ln y) \in [-0.15, -0.005]$. Thus our data shows a good deal less variance in growth, which may be partly attributable to the smaller measurement error we have and also to the fact that we draw a much more homogeneous sample.

The final three rows (20-22) give statistics on the incidence of low earnings. We can see that while 87.3% of workers in the bottom quintile in the second period

[14] The change in the variance is not monotonic but is highly correlated with the business cycle; although this is an important issue, we leave this aspect for future research.

are also in there in the third period, this percentage decreases to a 59% when we consider the final period. Finally, the incidence of situations of 'permanent' low earnings as measured by the percentage of workers in the bottom quintile for at least 8 of 15 years is approximately 20%.

5. Results.

5.1. Unit root models

The first model we consider (model 1) is extremely parsimonious and is used mainly to set the agenda. We assume that everyone has a unit root and that first differences of log earnings follow a driftless $MA(1)$ process with Normally distributed errors with the same parameters for everyone; see Table 3 for an exact statement of this and other forms we consider, as well as the choice of ap's and the parameter estimates for all the models we consider. We do not include a drift term (a constant) since the data are residuals from a regression on year dummies (amongst other things) and so the mean in each year is zero. This model could readily be estimated using a variety of conventional GMM, minimum distance or maximum likelihood techniques but we choose to use SMD for comparability with later, more complicated models. As discussed above, we shall use the JI procedure so we need to choose two of our auxiliary parameters for fitting. Since the parameters of the model are the variance and the $MA(1)$ parameter, we fit this model to the OLS analogues: the OLS mean log variance and first order autocorrelation. For this model only we also consider an alternative choice of ap's, so we designate the estimates model 1a.

Not surprisingly this very restricted model fails to fit the other auxiliary parameters in a number of dimensions. Referring to Table 2, column 2 we see first that the goodness of fit for model 1a rejects decisively (a χ^2 (20) statistic of 856). In figure 5.1 we present the distributions for the four OLS estimates for the data and for the simulated model. As can be seen, the model fits three of the distributions surprisingly well for such a simple specification. Note, however, that we use the same initial values in our simulations as in the data so that the good fit for the OLS intercept is to be expected. As can be seen, the model dramatically underestimates the (log of the) variance of the OLS errors. Referring to Table 3, we see that for the individual ap's the most important failures (in order of the listing in Table 3) are:

1. we overestimate the mean OLS slope parameter (*m(sl)*);

2. we underestimate the variance of the OLS intercepts (*v(in)*);

3. we dramatically underestimate the variance of the error log variance (*v(lv)*);

4. we underestimate the variance of the OLS auto-correlations ((*v(au)*).

5. there is a significant positive correlation between the OLS intercept and variance in the data (*c(in, lv)*) and a negative one in the simulated data;

6. we underestimate the trend in the cross-section variance (*vtrend*) and the variance about this trend (*vtrsd*);

7. we underestimate the variance in first differences ((*var, D)*);

8. the first order auto-correlation for the first differences ((*ac1, D)*) is smaller, is absolute value, for the simulated data than for the sample;

9. we underestimate the persistence of the low income state ((*p(2,3)* and *p(half)*).

The first issue we address is whether it might not be better to fit on the first differenced data (as in MaCurdy and Abowd and Card). Thus we also fit model 1 to the first two differences ap's ((*var, D)* and *(ac1, D)*) (see model 1b in Table 3). As we would expect, given the fit of model 1a to 'D, var', the estimated $MA(1)$ parameter is larger (in absolute value) than for model 1a. Apart from the obvious differences that arise from taking different ap's to fit, we see that model 1b fits significantly better for the OLS mean slope, the variance of the OLS intercepts and the trend of the cross-section variance. It is somewhat worse for the measures of the persistence of low income states. Overall, Model 1b fits worse in terms of the goodness of fit test statistic so that we shall continue with fitting to OLS parameters below. The latter strategy also allows us to choose ap's that are close to being 'structural'. One possible route to reconciling the results for models 1a and 1b may be to allow that the error term has a common ('macro' element) and an idiosyncratic component and that the effect of the former differs across workers. On the other hand, it may be that the difference is due solely to the use of a very restrictive model, so we postpone further consideration of this feature of our results until later.

One obvious but important point to emerge from the comparisons of models 1a and 1b is that fitting to different ap's may change the fit to ap's that are not used in fitting in ways that are very difficult to predict analytically. This justifies the 'fire-fighting' approach to the exploratory analysis we adopt in which we address

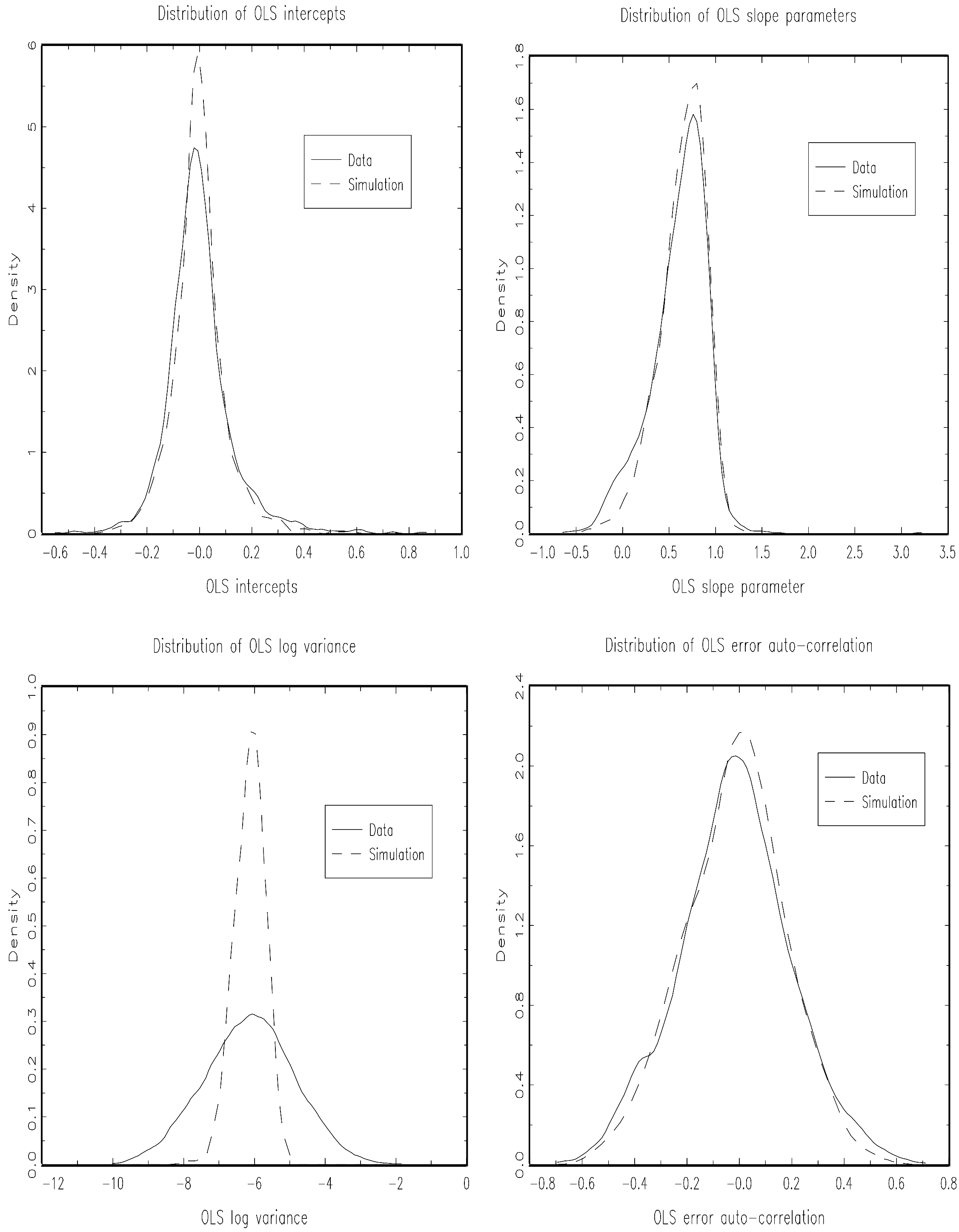

Figure 5.1: Model 1 fits for OLS parameters.

the most serious problems first and then see how the fix affects the fit to other ap's.

The most serious misfit (in terms of the t-value) of models 1a and 1b is to the dramatic under-prediction of the dispersion of the error log variance. This has been observed by other authors (Geweke and Keane (1997), Chamberlain and Hirano (1997) and Ulrick (2000)) and has been addressed in one of two ways. Either investigators take a different (fat tailed) distribution which is assumed common to everyone or they allow that agents have Normal errors with heterogenous variances. In model 2 we allow for the former by modelling the error variance as a mixture of zero mean Normals with a given mixing probability of 0.8.[15]. Specifically we assume that in any year and for any worker the standard deviation can be one of two values[16] (see Table 2). We fit this three parameter model to the same ap's as for model 1a and to the variance of the OLS log variances ('v(lv)'). One feature of the estimates of model 2 is that the $MA(1)$ parameter is much closer to zero than for the normal model; this is consistent with the view that an incorrect model may lead to spurious dynamics. Note as well that the mean standard deviation (0.042) is a little less than the value for model 1a. Referring to Table 3 we see the overall goodness of fit test has improved considerably but it still rejects decisively. This reflects that some fits have improved (the most obvious example being the variance of the log variances!) but most have not.

The alternative to allowing for a common fat tailed distribution is to assume that all agents have Normally distributed errors but with persistently different variances. Dominitz and Manski (1997) present direct evidence that there is a good deal of heterogeneity in subjective perceptions of short run household income risk. Model 3 captures this by assuming that the individual specific variances are constant over time and are distributed according to a particular distribution. We tried one parameter families such as the exponential (as in Chamberlain and Hirano (1997)) and the two parameter Beta, but these failed to fit the data in the sense that we could not find parameter values that drove the criterion to zero. We ended up using a two parameter lognormal distribution.[17] Specifically, we first draw a value for the variance of the error distribution for each simulated agent

[15] We could also have estimated the mixing parameter but there is no 'natural' auxiliary parameter to identify it. Put another way, a three parameter model for the variances is over-parameterised relative to our choice of auxiliary parameters.

[16] Recall that our sample excludes any worker who has any unemployment in the 16 years of the sample so this does *not* reflect intermittent large shocks due to unemployment.

[17] This distribution also has the important advantage that we can allow for correlated heterogeneity (see later models) in a simple way.

from a log Normal distribution with given mean and standard deviation (denoted by η_1 and η_2 respectively) and then keep this fixed over time for the simulated agent. Referring to the estimates for model 3 in Table 2 we note that the estimated mean variance is a little above that for the homogeneous model (model 1a) and the $MA(1)$ parameter is broadly similar. The variance of the lognormal distribution is highly significant in the sense that the goodness of fit statistic falls from 856 for the model with $\eta_2 = 0$ (that is, model 1a) to 410. We shall not discuss the fits for this model in detail but we do note that although the correlation between the OLS slope and log variances is much reduced, it is still of the 'wrong' sign. Figure 5.2 presents the distributions of the OLS error variances for the data and for models 2 and 3; it will be seen that model 3 seems to better capture all of the features of the data distribution. Given this and the fact that the heterogeneous variances model also fits rather better the other ap's than the fat-tailed alternative (if the difference between $\chi^2(19)$ values of 489 and 410 means anything) and since we have a pre-disposition toward modelling persistent heterogeneity, we shall only pursue the persistent heterogeneity model in the subsequent analysis.

The next model addresses explicitly that no model yet considered captures the correlations between the OLS error variances and the OLS intercepts and slopes. Thus we consider model 4 which allows for correlated heterogeneity. To do this, the only correlation between model parameters and data that we can allow for in our models is with the initial values (see the discussion in section 1). Specifically, we take model 3 and allow that the mean of the lognormal distribution for the variances is a deterministic function of the initial values (see Table 3 for details). Thus we have four parameters (θ, η_1, η_2 and λ) with model 3 if $\lambda = 0$ and model 1 if $\lambda = \eta_2 = 0$. Given the way we model the correlated heterogeneity the obvious additional auxiliary parameter to take to fit this model is the correlation between the OLS intercepts and log variances. Table 2 presents the results; the reduction is the goodness of fit parameter suggests that this new parameter is highly significant. The new model now fits reasonably well on a number of dimensions but still does not capture the variation in the OLS intercept and autoregressive parameter. To accommodate these, models 5 and 6 add (uncorrelated) heterogeneity in the drift terms and the $MA(1)$ parameter. Adding the latter achieves a considerable improvement in many dimensions. For example, model 6 captures much of the negative correlation between the OLS slope and variance parameters even though we do not model this explicitly nor use it in the fitting.

One concern that we have is how much bias is introduced by using just identified models rather than fitting to all the ap's simultaneously. To investigate

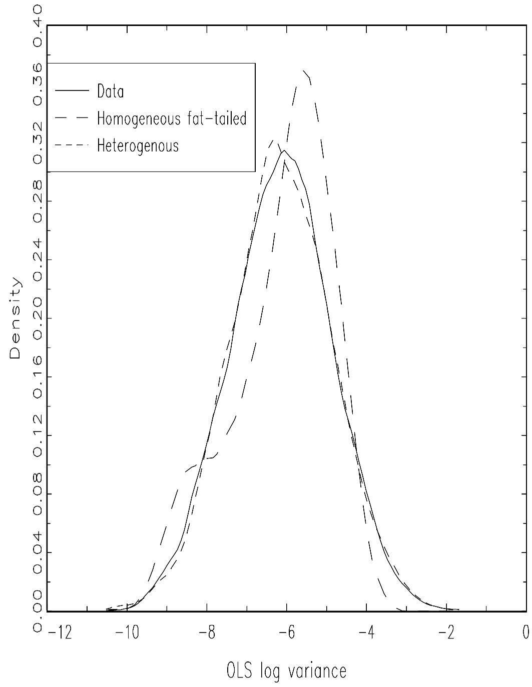

Figure 5.2: Distributions of OLS log variance

this, we also estimate model 6 using all 22 ap's. As we have emphasised above, we prefer the JI procedure since then we can be sure of convergence. For the OI estimation we use the JI estimates as starting values; we could not find other starting values that gave a lower criterion but this does not guarantee that we have reached a global minimum. The $\chi^2(16)$ statistic for the OI estimation procedure is 58.4. This is, of course, necessarily lower than the goodness of fit statistic of 60.5 for the JI estimates, but not much so. Consistent with this, the parameter estimates are very close to those for the JI estimates. We conclude that it does not make much difference whether we use the JI or the OI procedure for the final model. This does not rule out that using the OI procedure throughout our exploratory analysis might not lead us to a different preferred model.

Model 6 is the most general 'pure' unit root model we shall present. The goodness of fit statistic is still poor by conventional standards (a $\chi^2(16)$ of 60.5) but it is a considerable improvement on more restricted models. To reinforce this, in Table 4 we present formal tests for some restrictions on model 6. To do this we take the six ap's used to fit model 6 and in turn set each of four of the model parameters to zero using the same ap's to fit. This gives a $\chi^2(1)$ test statistic for each restriction. As can seen, all of the simpler variants of model 6 are decisively rejected against the more general model. In particular, the evidence is very strong for there being correlated heterogeneity between the initial value and the variance of the error terms; we shall return to the implications of this below.

Table 4. Tests for the unit root model

Restriction	Parameter restriction	$\chi^2(1)$
Zero mean, heterogeneous MA(1) parameter	$\theta_1 = 0$	26.0
Homogeneous MA(1) parameter	$\theta_2 = 0$	15.1
Uncorrelated heterogeneity in variance	$\lambda = 0$	145
No drift for anyone.	$\phi = 0$	24.5

The most general unit root model that we choose to fit (model 6) fits most of the features of the data quite well. In particular, the mean and the variance of the OLS slope parameters are well approximated even though the model imposes that everyone has a unit root. On the other hand, the fit for some of the other ap's is not so good. For example, we overestimate the trend in the variance ('vtrend') and underestimate the long run persistence of being in the lowest quintile ('p(2,16)'). As regards the former, the time path of the cross-section variance for the data is not linear, as we would expect from a model in which everyone has a unit root, but rather concave which suggests that some agents have a stationary process. Thus it seems worth exploring how well stationary models can fit the data.

5.2. Stationary models.

The first stationary model we consider (model 7) is an $ARMA(1,1)$ with the same AR parameter for everyone; that is, model 6 with a non-unit autoregressive parameter (see Table 3 for an exact statement).[18] Notice that the intercept is a linear function of the initial condition (with a coefficient restricted to be equal to $(1-\beta)$) plus a random error. The new auxiliary parameter used to fit β is the mean of the OLS slope. The estimated AR parameter is 0.89 and estimates for the other parameters are similar to that obtained for model 6. One noteworthy feature of model 7 is that it fits the negative correlation between the OLS slope and variance parameters very well (see Table 5) even though no explicit allowance is made for such a correlation. Given these estimates we can test between everyone having a unit root and everyone having a stationary process with the same AR parameter. Our SMD procedure gives two alternative tests for this.[19] The first SMD based test is a Wald test based on the t-value for the difference between the data and simulated mean OLS slope parameter, 'm(sl)', for the restricted model, 6; this is 1.63. An alternative test - which corresponds to a likelihood ratio test - follows the same procedure as for the tests of restricted versions of model 6. That is, we estimate our preferred unit root model (model 6) with the extra ap used in model 7 with β restricted to be equal to unity. The $\chi^2(1)$ statistic for the goodness of fit is 0.66. Thus both tests suggest that the pure unit root model would not be rejected against a stationary model with a homogeneous AR parameter. However, the underestimate for the variance of the OLS slope parameters in model 7 (a predicted value of 7.95 as against 9.44 for the data, with a t-value of -2.32) suggests that it may be worth investigating whether we should allow for a heterogeneous AR parameter. For the moment we shall maintain stationarity for everyone with a positive AR parameter, so that the values are constrained to lie within the open interval $(0,1)$; estimates with a support of

[18] We of course also tried the simpler stationary models corresponding to models 1 to 5; they all fit much worse than our starting point here and are not reported for the sake of expositional parsimony.

[19] In future work we shall investigate the properties of these procedures against more conventional tests for 'fixed T' contexts, such as those proposed in Harris and Tzavalis (1999). Other tests we would like to consider are likelihood ratio tests of models 6 and 7 such as those proposed by Vuong (1989).

$(-1, 1)$ gave very similar results to those reported here. To do this, in model 8 we model the AR parameters as being distributed as a logisitic distribution (see Table 3 for details). The parameters reported in Table 3 imply that the distribution of AR parameters is highly skewed; the mean and median are 0.92 and 0.95 respectively with a first percentile and first decile of 0.61 and 0.82 respectively. The $\chi^2(2)$ statistic for the test of the heterogeneous stationary model against a pure unit root model is 2.18. Consequently we conclude that the unit root model is to be (statistically) preferred to the heterogeneous stationary model. On the other hand, the parameter values of the latter and the pile-up of the distribution close to unity (see figure 5.1, panel B) suggest that a mixture model in which some agents have a unit root and others a stationary process might do better than either of the 'pure' models; it is to this that we now turn.

5.3. Mixed models.

The mixture model we consider is a mixture of the unit root model 6:

$$\begin{aligned}
\Delta y_{ht} &= \alpha_h + \varepsilon_{ht} + \theta_h \varepsilon_{h,t-1}, \\
&\quad \text{with } \alpha_h \tilde{} N\left(0, \sigma_\alpha^2\right), \\
&\quad \varepsilon_{ht} \tilde{} N\left(0, \sigma_h^2\right), \sigma_h^2 \tilde{} lN(exp(\lambda^u y_{h1}) * \eta_1^u, \eta_2^u) \\
&\quad \text{and } \theta_h \tilde{} N(\theta_1^u, \theta_2^u)
\end{aligned} \tag{5.1}$$

and the homogeneous stationary model 7:

$$\begin{aligned}
y_{ht} &= \alpha_h + \beta y_{ht} + \varepsilon_{ht} + \theta_h \varepsilon_{h,t-1}, \\
\alpha_h &= (1-\beta) y_{h1} + \upsilon_h, \upsilon_h \tilde{} N\left(0, \sigma_v^2\right), \\
&\quad \varepsilon_{ht} \tilde{} N\left(0, \sigma_h^2\right), \sigma_h^2 \tilde{} lN(exp(\lambda^s y_{h1}) * \eta_1^s, \eta_2^s) \\
&\quad \text{and } \theta_h \tilde{} N(\theta_1^s, \theta_2^s)
\end{aligned} \tag{5.2}$$

with a mixing parameter of π for the probability of being a stationary model.

Models 6 and 7 have 13 parameters between them. We do not have enough ap's to estimate these and the mixing parameter for the wholly unrestricted model, so that we need to impose some restrictions on the variation of parameters across the two models. To choose these restrictions, we first compare the parameter estimates for the pure models given in Table 3; for convenience, the comparisons are given in Table 6.

Table 6: comparison of parameter estimates.		
	Model 6 (unit root)	Model 7 (stationary)
β	1	0.89
$\sigma_\alpha, \sigma_\upsilon$	0.007	0.022
η_1	0.059	0.058
η_2	0.029	0.029
λ	1.51	1.45
θ_1	-0.148	-0.116
θ_2	0.283	0.237

As can be seen, the two estimates for the 'intercept' parameters (σ_α and σ_υ) differ radically. This simply reflects the very different role that the intercept plays in stationary and unit root models. For the former, heterogeneity in the intercept represents heterogeneity about the theoretically predicted intercept $(1-\beta)\, y_{h1}$ whereas for unit root models it represents heterogeneity in growth rates (drifts). As we would expect there is considerably more variation in the former than in the latter. On the other hand the other values are all quite similar (the MA parameter (θ_1 and θ_2) estimates are relatively imprecise so that the differences are not important statistically) so that it seems acceptable to restrict them to be equal across the two mixtures ($\eta_1^u = \eta_1^s, \eta_2^u = \eta_2^s$ etc.). This gives 9 parameters: $(\sigma_\alpha, \sigma_\upsilon, \eta_1, \eta_2, \lambda, \theta_1, \theta_2, \beta, \pi)$. Even this is too many, so that we impose that there is no heterogeneity in the drift for the unit root; that is, with σ_α we set to zero. To model the mixture we take draws from a unit uniform distribution and assign all simulated agents with a value above a certain cut-off (π, 'the probability of being stationary') to having a unit root with no drift. Otherwise a simulated agent is assigned to having a stationary process. As is usual with simulated discrete models we have to introduce some smoothness in the assignment to make the otpimising algorithms work; we do this by replacing the 0/1 assignment by a Normal cdf with a very small variance so that agents who have uniform draws that are distant from the cut-off point π are assigned zero or unity but a small number of agents close to the cut-off point are assigned a value between zero and unity. In practice, we can take a 'tight' bound and only one percent of our simulated agents have a probability of being stationary that is between 1% and 99%. The parameter estimates are given in Table 2 and the fits are given in Table 5. We estimate that 30% of our agents have a stationary process with an AR parameter of 0.93. Given these estimates we can also test the two preferred 'pure' models, 6 and 7. The latter is given by restricting π to unity; the $\chi^2(1)$ statistic is 0.26 so that we would certainly not reject the homogeneous stationary model against the mixture model. To test for the pure unit root model with heterogeneous drifts, note that if we impose $\pi = 0$ and $\beta = 1$ then we have model 6; the $\chi^2(2)$ statistic for this restriction is 2.17 so that we would not reject the unit root with heterogeneous drifts.

5.4. Some implications of the estimates.

We can now pull together the implications of the foregoing analysis. First, as we have seen, the unit root model (model 6) and the stationary model (model 7) cannot be rejected against the mixture model (model 9). We also found that the unit root model is not rejected against the stationary model. Thus we can conclude that the preferred model for this data is the one that imputes a unit root to everyone. Note that our version of the 'pure' unit root case allows for much more heterogeneity than any of the unit root models given in Table $A1$. We now briefly present two implications of this extra heterogeneity.

In figure 5.3 we present the distribution of variances implied by model 6 parameter estimates. We also add a vertical line at the (homogeneous but fat tailed) variance given for model 2. The important feature of the figure is that the mean of the distribution is close to the homogeneous estimate so that any inferences that use linear functions of the variance are unlikely to be much affected by allowing for heterogeneity. On the other hand, many uses of the earnings process use non-linear functions of the variance. A pre-eminent example is the literature on the importance of the pre-cautionary motive for saving (see, for example, Carroll and Samwick (1997)). Impatient agents with a precautionary motive and *given labour supply* have a saving function which is increasing and convex in earnings risk. Thus agents who have a variance of, say, twice the mode (that is, about 0.1) will have a much higher level of precautionary saving than those at the mode; at the very top of the distribution a variance of 0.2 is very high indeed (one third of the time earnings will either rise or fall by over 35%) and agents would need substantial buffer stocks to smooth consumption in the face of this variation. Conversely, very many agents will have almost no precautionary motive since their earnings variance is small. Thus the heterogeneity in earnings will translate into an even more dispersed buffer stock savings distribution. Note, however, that this assumes fixed labour supply; if some of the variation in earnings is due to variations in labour supply (perhaps in response to variations in wages) which is not to be accounted risk.[20]

[20] Although we have conditioned on being in full time employment for the whole 16 year period,

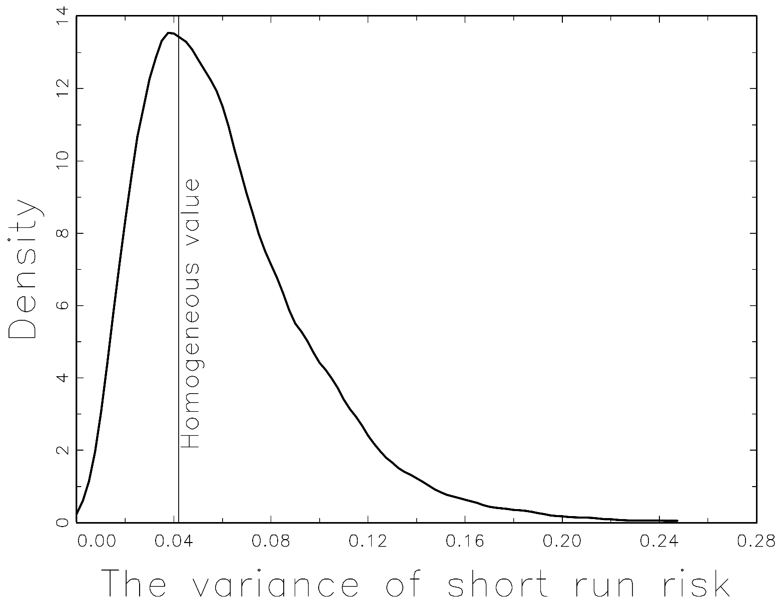

Figure 5.3: The distribution of variances

The other implication of the heterogeneity we have found that we wish to highlight is given in figure 5.4 which presents the relationship between the initial value and the variance. As can be seen, there is a positive trade-off between initial value and short run risk (the error variance) implied by the parameter estimates from model 6. The main curve is the deterministic function given by $\hat{\eta}_1 \exp\left(\hat{\lambda} y_{h1}\right)$ and the confidence intervals are given by the usual manipulations of the log Normal.[21] As can be seen the relationship is positive, which is consistent with a model in which agents trade off initial value against short run risk. This is a potentially important finding that can only be allowed for in a model which has correlated heterogeneity in the variances.

These two examples illustrate that some important inferences are sensitive to assumptions regarding heterogeneity. We consider now whether the process found on the Danish sample - everyone with a unit root and an MA(1) process with all parameters varying - can be considered 'generic' in the sense that we can use this for other samples. To do this, we consider a very different sample of workers drawn for the PSID.

5.5. Evidence from the PSID.

We have chosen to use a Danish sample since we can construct a very homogeneous sample but still have a relatively large data set. The question arises: is there

workers may have a lot of overtime. Note as well that this is more prevalent for the sample - those with a training as plumbers, electricians etc. - we have chosen.

[21] Formally the mean of the individual variances is:

$$\widehat{\eta_{1h}} = \hat{\eta}_1 \exp\left(\hat{\lambda} y_{h1}\right)$$

and the (common) standard deviation is $\sqrt{\widehat{\eta_2}}$. Define:

$$\begin{aligned} a_h &= \frac{1}{2} \ln \left(\frac{\widehat{\eta}_{1h}^4}{\left(\widehat{\eta}_{1h}^2 + \hat{\eta}_2\right)} \right) \\ b_h &= \sqrt{\ln \left(\frac{\hat{\eta}_2 + \widehat{\eta}_{1h}^2}{\widehat{\eta}_{1h}^2} \right)} \end{aligned}$$

which gives confidence bands of:

$$\exp\left(a_h \pm 1.96 b_h\right)$$

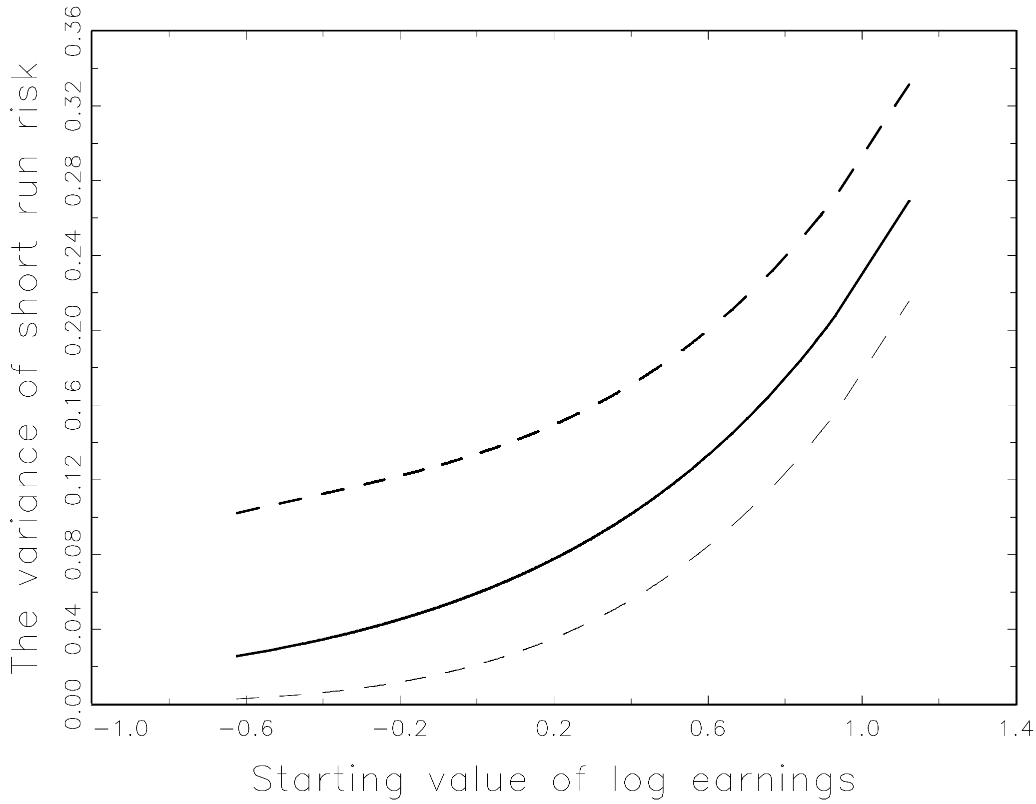

Figure 5.4: The trade-off between levels and risk.

something unusual about Danish workers that makes their earnings processes more heterogeneous than we would see elsewhere? To be sure, the Danish labour market environment is very different to that of, say, the U.S. or the U.K. (for example, high levels of unionisation; almost no employment protection laws; very generous unemployment insurance provisions etc.) and these differences might lead to more heterogeneous earnings processes. In this sub-section we provide a brief account of the application of our methods to a sample drawn from the PSID. The sample consists of 792 married men who are each followed for at least 16 years. Exact details of the sample can be found in the Appendix $D.2$. The main differences from the Danish sample are that we do not rule out unemployment and we have a more diverse educational background; a stringent selection as for the Danish data would leave us with a very small sample. In the first column of Table 7 we present the sample auxiliary parameter values. There are some noticeable differences between these and the corresponding values for the Danish data. First, for the OLS fits, the mean of the OLS slopes is much lower in the PSID data and the mean of the error variances is much higher. The former suggests that pure unit root models might not do as well as models with some stationary agents. The higher mean variance is probably caused by two factors: the sample selection which does not rule out unemployment spells for the PSID sample and the higher measurement error in the PSID data. The higher (in absolute value) mean auto-correlation and lower variance for the OLS auto-correlation are also consistent with the latter suggestion since first differencing noise leads to a negative $MA(1)$ error component and an attenuation of the variance of the auto-correlation parameter estimates.

In terms of the substantive ap's (those not based on the OLS regressions) we see that the PSID data has a much higher trend in the variance ('inequality' is increasing more over time); this may be attributable to the fact that in the PSID we have a more diverse sample in terms of education so that the trend over time contains a between group component as well as the within group component seen in the Danish data. The other striking feature of the comparison of substantive ap's is that the variance of the first differenced data is much higher in the PSID; this is consistent with the finding for the OLS error variances discussed above. All of this suggests that we should not automatically expect that the same class of model that was preferred for the Danish data (model 6 in which everyone has a unit root but parameters are very heterogeneous) will work well here.

The first model we consider for the PSID is the closest to the Abowd and Card and MaCurdy specifications (we shall refer to it as ACM) which they fitted on the PSID; this is model 2, which assumes that everyone has an $MA(1)$ in first differences with the same $MA(1)$ parameter and a non-Normal error distribution.[22] We fit to the mean of the OLS log variances ('m(lv)') and the variance and first order autocorrelation. of the first differences ('var, D' and 'ac1, D'); the latter two are chosen so as to be as close as possible to the original ACM procedure. The results are presented in Table 7. The goodness of fit statistic is very high so that the model does not fit very well the other ap's. In particular, the ACM model over-predicts the trend in variance very substantially. In terms of the OLS ap's, the ACM model substantially over-predicts the variance of the OLS intercepts and the error variances. It is important to note that none of the diagnostics used by Abowd and Card and MaCurdy would pick up these empirical failures of the model. Given that this pure unit root model does not do very well, it might be hoped that the more general one found on the Danish data (model 6) would do much better. In fact, the converse is true. Using the same ap's to fit as for the Danish data, model 6 does not even converge; the lowest value we can find (after an extensive search for starting values) is about 11. As can be seen from Table 7 the problem is with the variance of the OLS intercepts: we always over-predict which suggests that the model introduces too much variation into the drift terms. This is confirmed by the fact that the drift variance parameter is driven to zero in the estimation.

Clearly, then, we cannot simply export results on the preferred class of processes from one sample to another. We shall spare the reader the specification search made for the PSID sample used here, but the model that we found that fitted best was an eight parameter stationary $ARMA(1,1)$ model with heterogeneous AR parameters:

$$\begin{aligned} y_{ht} &= \alpha_h + \beta_h y_{h,t-1} + \varepsilon_{ht} + \theta_h \varepsilon_{h,t-1} \\ \alpha_h &= \gamma y_{h1} \\ & \quad \beta_h \tilde{} \exp\left(\beta_1 + \beta_2 \gamma_h\right) / \left(1 + \exp\left(\beta_1 + \beta_2 \gamma_h\right)\right) \\ & \quad \sigma_h^2 \tilde{} lN(\eta_1 \exp(\lambda y_{h1}), \eta_2) \\ & \quad \theta_h \tilde{} N(\theta_1, (\theta_2)^2) \end{aligned} \tag{5.3}$$

We refer to this as model 10 since it is not nested within any of the previous models (although it is close to model 8 with a different process for the intercept α_h). In model 10 we do not force the intercept to have a mean of $(1-\beta_h)\,y_{h1}$

[22] The minimum distance techniques used by Abowd and Card and MaCurdy did not require a specification for the error distribution and hence do not restrict the errors to be Normally distributed.

as for the Danish data; this weakening is necessary to accommodate the over-prediction of the OLS intercept variance. The choice of ap's (in fact the same as for model 8 for the Danish data) and the fit of this model are given in Table 7. The most satisfying aspect of this model is that the goodness of fit for the model - a $\chi^2\,(14)$ of 10.4 - indicates that this model does fit adequately all of the ap's we have chosen to model. The parameter value estimates are:

$$\begin{aligned} \hat{\gamma} &= 0.025,\, \hat{\beta}_1 = 2.42,\, \hat{\beta}_2 = 1.48,\, \hat{\lambda} = -0.22, \\ \hat{\eta}_1 &= 0.23,\, \hat{\eta}_2 = 0.13,\, \hat{\theta}_1 = -0.22,\, \hat{\theta}_2 = 0.17 \end{aligned} \tag{5.4}$$

The estimates for the AR parameters give a distribution that is skewed, the mean and median are 0.85 and 0.92 respectively, but still with a substantial number of agents with a low parameter: the first decile value is 0.63. Thus for our PSID sample it seems that we need to assume that everyone has a stationary process with heterogeneous AR parameters.

6. Conclusions.

In this paper we have considered the estimation on panel data of a univariate earnings process allowing for lots of heterogeneity. To do this we have adapted the SMD estimator to the dynamic panel, fixed T framework and we have shown how to conduct inference. The SMD approach is justified on the grounds that, *a priori*, we have no idea what form heterogeneity should take so that we need to start with simple models and build up to more general ones. A number of conclusions emerge for the modelling of earnings processes:

- For a very homogenous sample of Danish workers, we find significantly more heterogeneity than any previous investigator has allowed for.
- For our preferred model for the Danish data, we are unable to achieve a really satisfactory fit to the statistics not used in the fitting. This suggests to us that we need to relax the two step procedure that first regresses earnings on time dummies and individual characteristics. This two step procedure essentially assumes that all heterogeneity correlated with these factors can be captured through heterogeneity in the levels regression intercept. Thus we need to integrate the modelling of the wage equation (the levels equation) with the dynamic process.

- For the Danish sample we found that we can account reasonably well for the time series facts without assuming changes in parameters over time (except for those implicit in the first round regression), such as increasing mean risk.
- For the Danish sample a process that imputes a unit root to everyone does best.
- The specific conclusion on the Danish sample is not 'exportable' since an investigation on a PSID sample leads to a completely different process: everyone has a stationary process with heterogenous $AR(1)$ parameters. This process gives a good fit to all of the dynamic features of the PSID.
- The allowance for heterogeneity is critical for many substantive conclusions. We have chosen to highlight two: the impact on the distribution of precautionary saving and the relationship between the starting values and risk, but almost all uses of the process estimates will be seriously affected.

The foregoing analysis suggests a number for directions for future work. First it will be fruitful to check how well SMD based inferential procedures for fixed T univariate panel data do against other tests for, for example, unit roots versus stationary processes. Related to this, we plan to compare the SMD fits to ML fits for our preferred models. Second, it would also be of interest to develop formal tests between models such ours which have a good deal of individual 'fixed' heterogeneity and models that allow for time varying macro parameters. Third, the methods here could be applied to models of earnings processes that are not based on the conventional two step regression procedure. For example, for stationary models with heterogeneous $AR(1)$ parameters we could allow that this heterogeneity is correlated with age. Finally, the methods suggested here could be applied to multivariate processes for, for example, unemployment and earnings or for joint husband and wife earnings. The latter would allow us to consider, for example, whether earnings risks are correlated within the household.

To sum up: our broad conclusion is that there is much more heterogeneity in earnings processes both within groups and between groups than has previously been allowed for. This will make future modelling more difficult since we cannot simply pull down an earnings process off the shelf in any given context. Our results also indicate that not accounting for heterogeneity may lead to serious bias in substantive inferences.

A. Previous studies.

TABLE A1: Previous studies.

Author (year)	Data, sample size and period	Unit root or stationary
Hause (1977)	Swedish Low-income Commision study, 135 individuals 1951-1966 (7 periods)	unit root in levels
Lillard and Willis (1978)	PSID, 1144 individuals 1967-1973 (7 year)	non unit root in levels
Lillard and Weiss (1979)	National Science Foundation's Reg., 4330 obs 1960-1970 (every second year)	non unit root in levels
MaCurdy (1982)	PSID, 513 individuals 1967-1976 (10 years)	unit root in levels non unit root in first differences
Abowd and Card (1989)	PSID, 1448 1969-1979 (11 years) NLS, 1316 1966-1975 (6 years) Seattle and Denver Income Maintenance Experiment, 560 individuals 1971-1972 (8 quarters)	unit root in levels non-unit roots in first differences
Gottschalk and Moffitt (1994)	PSID, 2730 individuals 1970-1987	mainly non unit root one case with unit root
Baker (1997)	PSID, 534 individuals 1967-1986, (20 years)	non unit root (few specifications with unit root)
Chamberlain and Hirano (1997)	PSID, 813 individuals 1967-1991 (10 years)	non unit root in levels (but very close to unit root)
Geweke and Keane (1997)	PSID, 4766 individuals 1967-1989	non unit root in levels
Ulrick (2000)	PSID, 4766 individuals 1967-1989 (same data as G&K)	non unit root in levels

TABLE A1: Previous studies (continued).			
Author (year)	Process[1]	Heterogeneity[2]	Other features
Hause (1977)	MA(1), MA(2) or AR(1) with a unit root	intercept and slope of time trend	
Lillard and Willis (1978)	AR(1) (common factor model)	intercept	balanced sample
Lillard and Weiss (1979)	AR(1) (common factor model)	intercept and slope of time trend	unbalanced sample
MaCurdy (1982)	ARMA(0,2), ARMA(0,3) ARMA(0,1), ARMA(1,1)	intercept and slope of time trend	balanced sample
Abowd and Card (1989)	MA(2)	intercept and slope of time trend	
Gottschalk and Moffitt (1994)	ARMA(1,1)	intercept and slope of age	unbalanced sample
Baker (1997)	ARMA(1,2) (common factor model)	intercept and slope of age	balanced sample
Chamberlain and Hirano (1997)	AR(1) (common factor model)	intercept, variance and autoregressive parameter	balanced sample
Geweke and Keane (1997)	AR(1)	intercept	unbalanced sample
Ulrick (2000)	AR(1)	intercept	unbalanced sample

1. For unit roots the process is for the first difference.

2. For unit roots, heterogeneous intercepts are allowed for implicitly.

B. Comments on Table A.1.

Table A.1 summarises the features of ten leading empirical studies of earnings dynamics. Although many of the studies use the PSID data, there are considerable differences in how the sample is selected (this aspect is not included in the table). All studies except Lillard and Weiss (1979) focus entirely on males. Some of the studies restrict the sample to homogenous groups (e.g. narrowed age and cohort groups, scientists or husbands continuously married to the same wife). As regards sample size, all studies using a balanced data set contain less than 1500 individuals (the smallest sample consists of 135 individuals). The unbalanced samples are somewhat larger. The sample periods of the studies cover a 40 year period from 1951-1991, where the maximum period of a balanced data set is 20 years (Baker (1997)) and 22 for an unbalanced (Geweke and Keane (1997) and Ulrick (2000)).

In the third column we list whether the process contains a unit root. When the process does not contain a unit root process we do not necessarily mean that the process is assumed stationary. Most of the studies assuming no unit root process allow for some kind of non-stationarity. Some allow that the distribution of the initial observation might differ from the remaining process and others allow that the variances vary across time. About half of the studies assume a unit root, and most of these deal with it by working with earnings growth instead of the level of earnings (MaCurdy (1982), Abowd, Card (1989) and Baker (1997)).

When modelling the dynamic process all studies except Geweke and Keane (1997) and Ulrick (2000) use a two-step procedure. In the first step earnings or earnings growth are regressed on a number of individual characteristics (e.g. age) and time dummies, normally using OLS-estimation. Then subsequent analysis is based on the residuals from the first regression. In many of the studies the errors are assumed to follow an AR(1) process, which is equivalent to assuming a common factor model of the earnings or earnings growth. The parameters of the dynamic process are estimated in the second step. In Geweke and Keane (1997) and Ulrick (2000) the impact of the individual characteristics are estimated together with the dynamic process.

Column five reports the level of individual heterogeneity in the model. All of the models allow for individual heterogeneity in the level of the earnings. In some of the models a time trend is allowed to vary across individuals (given an individual specific effect in the levels the time trend is equivalent to an age trend). Furthermore one study (Chamberlain and Hirano (1997)) uses a model with heterogeneity in the variance of the error terms.

C. Tables.

TABLE 1: Monte Carlo results							
		$S = 1$		$S = 2$		$S = 5$	
	True	JI	OI	JI	OI	JI	OI
α	0	0.013	0.081	0.016	0.074	0.013	0.10
(sd)		(0.06)	(0.37)	(0.04)	(0.55)	(0.02)	(0.45)
σ_ε	5	5.00	5.01	5.00	5.01	5.00	5.03
(sd)		(0.02)	(0.16)	(0.02)	(0.24)	(0.01)	(0.19)
θ	-10	-10.1	-10.0	-10.1	-10.9	-10.1	-10.1
(sd)		(1.3)	(0.69)	(0.92)	(0.61)	(0.53)	(0.63)
mean(χ^2)		16.7	144	14.5	113	13.3	298
median(χ^2)		15.9	15.1	14.0	13.6	13.0	13.5
var(χ^2)		34.6	$2.8 * 10^6$	20.6	$2.1 * 10^6$	9.0	$8.0 * 10^6$
d.f.		19	19	19	19	19	19
# replications	-	3194		2104		622	
Means and standard deviations for parameter values multiplied by 100.							

TABLE 2: Results for unit root models								
Auxiliary Parameter	Data Value	Simulated models						
		1a	1b	2	3	4	5	6
m(in)	0.05	−0.07	0.34	0.15	0.32	0.34	0.08	0.55
	(0.28)	[−0.32]	[0.73]	[0.26]	[0.69]	[0.74]	[0.09]	[1.26]
m(sl)	60.04	64.36	57.59	67.92	62.39	62.42	58.18	61.58
	(0.67)	[4.58]	[−2.59]	[8.33]	[2.49]	[2.52]	[−1.96]	[1.63]
m(lv)	−613.99	−613.99	−538.41	−613.99	−613.99	−613.99	−613.99	−613.99
	(2.75)	[.]	[19.45]	[.]	[.]	[.]	[.]	[.]
m(au)	−1.79	−1.79	−3.93	−1.79	−1.79	−1.79	−1.79	−1.79
	(0.47)	[.]	[−3.24]	[.]	[.]	[.]	[.]	[.]
v(in)	1.65	1.06	1.62	1.26	1.42	1.38	1.65	1.65
	(0.10)	[−4.23]	[−0.22]	[−2.76]	[−1.68]	[−1.94]	[.]	[.]
v(sl)	9.44	6.48	7.89	8.10	7.09	7.10	7.80	9.95
	(0.45)	[−4.62]	[−2.42]	[−2.09]	[−3.68]	[−3.66]	[−2.56]	[0.80]
v(lv)	160.82	19.53	19.02	160.82	160.82	160.82	160.82	160.82
	(4.78)	[−20.91]	[−20.99]	[.]	[.]	[.]	[.]	[.]
v(au)	4.56	3.79	3.26	3.70	3.59	3.60	3.74	4.56
	(0.14)	[−3.79]	[−6.45]	[−4.27]	[−4.78]	[−4.76]	[−4.06]	[.]
c(in,sl)	−5.02	−1.75	−2.97	0.73	−4.08	−3.59	0.31	−1.75
	(3.29)	[0.70]	[0.44]	[1.24]	[0.20]	[0.31]	[1.15]	[0.70]
c(in,lv)	33.72	−0.82	−4.49	−2.34	−3.40	33.72	33.72	33.72
	(1.99)	[−12.24]	[−13.54]	[−12.78]	[−13.16]	[.]	[.]	[.]
c(in,au)	4.64	2.57	1.22	−1.96	1.92	0.63	−3.77	0.31
	(1.88)	[−0.78]	[−1.29]	[−2.49]	[−1.03]	[−1.51]	[−3.17]	[−1.63]
c(sl,lv)	−16.27	23.88	21.76	14.58	7.10	5.18	12.15	−9.29
	(2.42)	[11.72]	[11.11]	[9.01]	[6.83]	[6.27]	[8.30]	[2.04]
c(sl,au)	1.27	−8.82	−11.08	−19.98	−6.97	−7.11	−14.69	−1.90
	(1.85)	[−3.86]	[−4.73]	[−8.14]	[−3.16]	[−3.21]	[−6.11]	[−1.22]
c(lv,au)	2.65	1.58	−5.02	−8.99	−0.06	0.12	−7.49	8.35
	(2.12)	[−0.36]	[−2.56]	[−3.89]	[−0.91]	[−0.85]	[−3.39]	[1.90]

TABLE 2 (continued)								
Auxiliary Parameter	Data Value	Simulated models						
		1a	1b	2	3	4	5	6
vtrend	0.37	0.24	0.42	0.46	0.47	0.44	0.40	0.48
	(0.03)	[−3.48]	[1.12]	[2.21]	[2.47]	[1.68]	[0.75]	[2.77]
vtrsd	0.30	0.05	0.07	0.08	0.10	0.09	0.11	0.11
	(0.05)	[−3.30]	[−3.07]	[−2.97]	[−2.64]	[−2.79]	[−2.54]	[−2.58]
D, var	0.64	0.29	0.64	0.49	0.61	0.55	0.57	0.58
	(0.03)	[−9.20]	[.]	[−4.10]	[−0.91]	[−2.34]	[−1.93]	[−1.55]
D, ac1	−18.60	−9.85	−18.60	−4.92	−11.55	−11.56	−15.53	−14.07
	(1.26)	[4.93]	[.]	[7.70]	[3.97]	[3.96]	[1.72]	[2.55]
D, ac2	−1.29	0.11	−1.35	0.14	−1.34	−1.33	−1.41	−0.75
	(1.16)	[0.85]	[−0.04]	[0.86]	[−0.03]	[−0.03]	[−0.07]	[0.33]
p(2,3)	87.31	82.11	75.98	87.54	80.70	84.24	80.70	84.95
	(1.42)	[−2.59]	[−5.66]	[0.12]	[−3.30]	[−1.53]	[−3.30]	[−1.18]
p(2,16)	59.23	55.69	46.96	51.44	51.20	55.21	54.98	50.02
	(1.88)	[−1.33]	[−4.61]	[−2.92]	[−3.01]	[−1.51]	[−1.60]	[−3.46]
p(half)	19.87	18.88	18.17	18.92	18.74	18.97	19.02	19.02
	(0.31)	[−2.25]	[−3.86]	[−2.15]	[−2.57]	[−2.04]	[−1.93]	[−1.93]
GF test		855.89	1203.77	489.41	410.07	157.23	242.46	60.51
df		20.00	20.00	19.00	19.00	18.00	17.00	16.00

Notes. All values multiplied by 100.
(.) = standard deviation for the data value
[t] = t-values for the difference between the simulated and data values
[.] indicates that auxiliary parameter used for fitting
(in, sl, lv, au) = OLS intercept, slope, log variance and auto-correlation
m(.) = mean; v(.) = variance; c(.,.) = correlation
'vtrend' = trend of cross-section variance
'vtrsd' = standard deviation of errors from trend of cross-section variance
'D' = first differenced mean and first two auto-correlations
'p(2,t)' = probability of being in the bottom quintile in year t,
given in bottom quintile in year 2
'p(half)' = probability of being in bottom quintile for at least 8 of 15 years
'GF' is the chi-squared goodness of fit test statistic
'df' is the degrees of freedom for the goodness of fit test

TABLE 3: Models and estimates.	
Model	Process
1 (unit root)	$\Delta y_{ht} = \varepsilon_{ht} + \theta\varepsilon_{h,t-1}$ with $\varepsilon_{ht}\tilde{}N(0,\sigma^2)$
AP's for 1a	m(lv); m(au)
Estimates of 1a	$\hat{\theta} = -0.104$, $\hat{\sigma} = 0.053$
AP's for 1b	Variance and first order autocorrelation of differences
Estimates of 1b	$\hat{\theta} = -0.201$, $\hat{\sigma} = 0.079$
2 (unit root)	As model 1 with fat tails.
	$\varepsilon_{ht}\tilde{}N(0,\sigma_1^2)$ with probability 0.8
	and $\varepsilon_{ht}\tilde{}N(0,\sigma_2^2)$ with probability 0.2
AP's for 2	m(lv); m(au); v(lv)
Estimates of 2	$\hat{\theta} = -0.053$, $\hat{\sigma}_1 = 0.014$, $\hat{\sigma}_2 = 0.152$
3 (unit root)	As model 1 with uncorrelated heterogeneity in variances.
	$\varepsilon_{ht}\tilde{}N(0,\sigma_h^2)$ where $\sigma_h^2\tilde{}lN(\eta_1,\eta_2)$
AP's for 3	m(lv); m(au); v(lv)
Estimates of 3	$\hat{\theta} = -0.124$, $\hat{\eta}_1 = 0.065$, $\hat{\eta}_2 = 0.044$
4 (unit root)	As model 3 with correlated heterogeneity in variances.
	$\eta_{1h} = exp(\lambda y_{h1}) * \eta_1$
AP's for 4	m(lv); m(au); v(lv); c(in, lv)
Estimates of 4	$\hat{\theta} = -0.124$, $\hat{\eta}_1 = 0.062$, $\hat{\eta}_2 = 0.032$, $\hat{\lambda} = 1.31$
5 (unit root)	As model 4 with uncorrelated heterogeneity in drifts
	$\Delta y_{ht} = \alpha_h + \varepsilon_{ht} + \theta\varepsilon_{h,t-1}$, $\alpha_h\tilde{}N(0,(\phi)^2)$
AP's for 5	m(lv); m(au); v(lv); c(in, lv); v(in)
Estimates of 5	$\hat{\theta} = -0.167$, $\hat{\eta}_1 = 0.062$, $\hat{\eta}_2 = 0.034$, $\hat{\lambda} = 1.21$, $\hat{\phi} = 0.195$
6 (unit root)	As model 5 with heterogeneity in MA parameter.
	$\Delta y_{ht} = \alpha_h + \varepsilon_{ht} + \theta_h\varepsilon_{h,t-1}$, $\theta_h\tilde{}N(\theta_1,(\theta_2)^2)$
AP's for 6	m(lv); m(au); v(lv); c(in, lv); v(in); v(au)
Estimates of 6	$\hat{\theta}_1 = -0.148$, $\hat{\theta}_2 = 0.283$, $\hat{\eta}_1 = 0.059$, $\hat{\eta}_2 = 0.029$, $\hat{\lambda} = 1.51$, $\hat{\phi} = 0.007$
Estimates of 6 (Over-identified)	$\hat{\theta}_1 = -0.161$, $\hat{\theta}_2 = 0.276$, $\hat{\eta}_1 = 0.061$, $\hat{\eta}_2 = 0.031$, $\hat{\lambda} = 1.37$, $\hat{\phi} = 0.007$
Estimates of 6 ($S = 20$)	$\hat{\theta}_1 = -0.172$, $\hat{\theta}_2 = 0.300$, $\hat{\eta}_1 = 0.059$, $\hat{\eta}_2 = 0.030$, $\hat{\lambda} = 1.35$, $\hat{\phi} = 0.007$

TABLE 3 (contd.)	
Model	Process
7 (stationary)	Stationary process with homogeneous AR parameter.
	Model 6 with no unit root.
	$y_{ht} = \alpha_h + \beta y_{h,t-1} + \varepsilon_{ht} + \theta_h \varepsilon_{h,t-1}$
	with $\sigma_h^2 \tilde{} lN(\eta_1 \exp(\lambda y_{h1}), \eta_2)$, $\theta_h \tilde{} N(\theta_1, (\theta_2)^2)$
	and $\alpha_h = (1-\beta)\, y_{h1} + \upsilon_h$ with $\upsilon_h \tilde{} N(0, \phi)$
AP's for 7	m(lv); m(sl); m(au); v(in); v(lv); m(au); c(in,lv)
Estimates of 7	$\hat{\theta}_1 = -0.116$, $\hat{\theta}_2 = 0.237$, $\hat{\eta}_1 = 0.058$, $\hat{\eta}_2 = 0.029$,
	$\hat{\lambda} = 1.45$, $\hat{\phi} = 0.022$, $\hat{\beta} = 0.893$
8 (stationary)	As model 7 with heterogeneous AR parameters.
	$y_{ht} = \alpha_h + \beta_h y_{h,t-1} + \varepsilon_{ht} + \theta_h \varepsilon_{h,t-1}$
	$\beta_h \tilde{}\ \exp(\beta_1 + \beta_2 \gamma_h) \,/\, (1 + \exp(\beta_1 + \beta_2 \gamma_h))$ with $\gamma_h \sim N(0,1)$
AP's for 8	As model 7 plus v(sl)
Estimates of 8	$\hat{\theta}_1 = -0.133$, $\hat{\theta}_2 = 0.269$, $\hat{\eta}_1 = 0.058$, $\hat{\eta}_2 = 0.030$, $\hat{\lambda} = 1.40$, $\hat{\phi} = 0.016$, $\hat{\beta}_1 = 2.89$, $\hat{\beta}_2 = -1.042$
9 (mixed)	Mixture of models 6 and 7.
	With probability π: $y_{ht} = \alpha_h + \beta y_{h,t-1} + \varepsilon_{ht} + \theta_h \varepsilon_{h,t-1}$
	$\alpha_h = (1-\beta)\, y_{h1} + \upsilon_h$ and $\upsilon_h \tilde{} N(0, (\phi)^2)$
	with probability $(1-\pi)$: $\Delta y_{ht} = \varepsilon_{ht} + \theta_h \varepsilon_{h,t-1}$
AP's for 9	As model 8.
Estimates of 9	$\hat{\theta}_1 = -0.149$, $\hat{\theta}_2 = 0.278$, $\hat{\eta}_1 = 0.059$, $\hat{\eta}_2 = 0.030$ $\hat{\lambda} = 1.41$, $\hat{\phi} = 0.035$, $\hat{\beta} = 0.929$, $\hat{\pi} = 0.309$

Notes. (in, sl, lv, au) = OLS intercept, slope, log error variance, first order autocorrelation.

m(.) = mean; v(.) = variance; c(.,.) = correlation.

TABLE 5: Results for stationary and mixed models				
Auxiliary	Data	Stationary		Mixed
Parameter	Value	7	8	9
m(in)	0.05 [0.28]	0.52 [1.19]	0.47 [1.07]	0.39 [0.87]
m(sl)	60.04 [0.67]	60.04 [.]	60.04 [.]	60.04 [.]
m(lv)	−613.99 [2.75]	−613.99 [.]	−613.99 [.]	−613.99 [.]
m(au)	−1.79 [0.47]	−1.79 [.]	−1.79 [.]	−1.79 [.]
v(in)	1.65 [0.10]	1.65 [.]	1.65 [.]	1.65 [.]
v(sl)	9.44 [0.45]	7.95 [−2.32]	9.44 [.]	9.44 [.]
v(lv)	160.82 [4.78]	160.82 [.]	160.82 [.]	160.82 [.]
v(au)	4.56 [0.14]	4.56 [.]	4.56 [.]	4.56 [.]
c(in,sl)	−5.02 [3.29]	−6.23 [−0.26]	−2.46 [0.55]	−0.07 [1.06]
c(in,lv)	33.72 [1.99]	33.72 [.]	33.72 [.]	33.72 [.]
c(in,au)	4.64 [1.88]	3.41 [−0.47]	2.27 [−0.90]	−0.63 [−1.99]
c(sl,lv)	−16.27 [2.42]	−15.84 [0.13]	−13.10 [0.93]	−2.51 [4.02]
c(sl,au)	1.27 [1.85]	−2.31 [−1.37]	−4.42 [−2.18]	−1.66 [−1.12]
c(lv,au)	2.65 [2.12]	16.04 [4.47]	12.52 [3.30]	3.88 [0.41]

TABLE 5 **(continued)**				
Auxiliary	Data	Stationary		Mixed
Parameter	Value	7	8	9
vtrend	0.37 [0.03]	0.34 [−0.93]	0.39 [0.57]	0.55 [4.46]
vtrsd	0.30 [0.05]	0.23 [−0.93]	0.09 [−2.73]	0.07 [−3.06]
var, D	0.64 [0.03]	0.58 [−1.63]	0.58 [−1.49]	0.59 [−1.34]
ac1, D	−18.60 [1.26]	−14.77 [2.15]	−15.54 [1.72]	−13.16 [3.06]
ac2, D	−1.29 [1.16]	−3.49 [−1.34]	−2.67 [−0.84]	0.17 [0.89]
p(2,3)	87.31 [1.42]	82.35 [−2.48]	83.77 [−1.77]	83.06 [−2.12]
p(2,16)	59.23 [1.88]	49.79 [−3.54]	50.26 [−3.37]	46.01 [−4.96]
p(half)	19.87 [0.31]	18.92 [−2.15]	19.25 [−1.39]	18.88 [−2.25]
GF test df		66.10 15.00	56.22 14.00	92.22 14.00
Notes. See Table 3.				

TABLE 7: Results for PSID data				
Auxiliary Parameter	Data Value	Unit root 2	Unit root 6***	Stationary 10
m(in)	−0.99 [1.02]	−1.36 [−0.26]	0.43 [0.98]	−0.11 [0.61]
m(sl)	35.56 [1.15]	40.56 [3.08]	54.98 [11.95]	35.56 [.]
m(lv)	−344.81 [4.15]	−344.81 [.]	−348.15 [.]	−344.81 [.]
m(au)	−3.21 [0.57]	−8.24 [−6.29]	−3.66 [.]	−3.21 [.]
v(in)	8.26 [0.74]	19.26 [10.58]	9.87 [.]	8.26 [.]
v(sl)	10.38 [0.47]	15.00 [6.97]	8.74 [−2.47]	10.38 [.]
v(lv)	133.85 [6.01]	271.09 [16.15]	132.89 [.]	133.85 [.]
v(au)	2.54 [0.17]	3.57 [4.19]	3.24 [.]	2.54 [.]
c(in,sl)	6.36 [4.42]	6.47 [0.02]	5.42 [−0.15]	1.75 [−0.74]
c(in,lv)	−17.43 [3.99]	−5.29 [2.15]	−15.75 [.]	−17.43 [.]
c(in,au)	−5.76 [3.54]	−4.76 [0.20]	−5.06 [0.14]	−2.91 [0.57]
c(sl,lv)	−15.87 [3.46]	16.82 [6.68]	7.92 [4.86]	−6.23 [1.97]
c(sl,au)	−7.07 [3.87]	−29.63 [−4.12]	−15.19 [−1.48]	−9.21 [−0.39]
c(lv,au)	7.35 [3.61]	−20.12 [−5.39]	−7.27 [−2.87]	0.04 [−1.43]

TABLE 7 (continued)				
Auxiliary Parameter	Data Value	Unit root 2	Unit root 6***	Stationary 10
vtrend	0.88 [0.16]	3.32 [11.02]	4.37 [15.79]	0.79 [−0.44]
vtrsd	1.52 [0.39]	2.27 [1.34]	0.93 [−1.06]	1.08 [−0.79]
var, D	9.92 [0.58]	9.92 [.]	7.63 [−2.79]	9.45 [−0.57]
ac1, D	−35.67 [1.37]	−35.67 [.]	−23.06 [6.51]	−36.79 [−0.58]
ac2, D	−1.47 [1.64]	0.49 [0.84]	0.60 [0.89]	−0.23 [0.53]
p(2,3)	70.08 [3.06]	71.97 [0.44]	75.76 [1.31]	71.34 [0.29]
p(2,16)	45.45 [3.25]	47.98 [0.55]	53.66 [1.79]	43.56 [−0.41]
p(half)	16.41 [0.66]	18.43 [2.15]	18.43 [2.15]	17.05 [0.67]
GF test		680.82	431.21	10.39
df		19.00	16.00	14.00

*** Model 6 did not converge (see text for details).

D. Data descriptions

D.1. The Danish Data

In this paper we use two different data sets. The first data set comes from the Danish register data set, which contains information on the Danish population. Our main data set consists of 2119 men with a vocational training, who we follow for the period 1981-1996. We select so that all the sample have been full time employed in all 16 years and continuously married to or cohabiting with the same spouse during the sample period. The sample is restricted to men aged 30-39 in 1981. The final sample is a balanced sample where all individuals are exactly observed in 16 years.

The income variable is defined as gross annual earnings. This variable is defined on the basis of information from the tax authorities and is therefore very reliable. This means that measurement errors are not likely to be a major problem using the Danish data. Average real earnings for our sample increases about 25% over the data period. As in most of the previous studies of income dynamics we use a two-step procedure to when estimating the income process. In the first step log annuals earnings are regressed on individual characteristics and time dummies. In the sample of men we use age and age squared, experience and experience squared and time dummies as explanatory variables. The distribution of experience is presented in **Table D1**.

Table D1*:* **The Distribution of experience.**

Experience in 1981	Number in sample
Less than 6 years	7
6-8 years	27
9-11 year	234
12-14 years	600
15-16 years	529
17 years or more	722
Total	2119

D.2. The PSID data

The second data set is a subsample of the PSID. To make the data as comparable as possible we have here restricted the data such all the individuals are in the sample for exactly 16 years. The sample contains 792 men who are aged between 20-66 in the sample period[23]. They all report positive earnings and positive working hours in all 16 years. Furthermore they are continuously reported as the head of household and married or cohabiting. The data set is balanced in the sense that all individuals are observed

[23] In order to get a sample of a reasonal size we do not limit our selves to a narrow age group.

the same number of time periods, but they may be observed in different calendar years. The entire observation period is from 1968 to 1992, where 477 individuals are observed in 1968, all individuals are observed in the period 1977-1983 and only 43 in 1992. This data set is a good deal less homogenous compared to the Danish data sets in several aspects: more variations in working hours, wider age group and covers different calendar years. On the other hand the selection on being in the sample for at least 16 years with positive earnings in every year and being continuously married means that our sample is probably much more 'stable' than some of the other samples used in previous studies. The earnings variable is defined as total annual labour market income of the head. This variable is self-reported and corresponds to the year before the interview year.

To make this data set comparable to the Danish ones, we use the two-step procedure. The log earnings are regressed on age, age squared, educational dummies and year dummies. In the **Tables D2** and **D3** we present the distributions of age and educational attainment in 1977.

Table D2: The Distribution of age

Age in 1977	Number
20-24	25
25-29	118
30-34	145
35-39	133
40-44	106
45-49	113
50-54	101
55-59	47
60-64	4
Total	792

Table D3: The Distribution of educational attainment

Educational attainment in 1977	Number
can not read	6
0-5 grades	11
6-8 grades	62
9-11 grades	106
12 grade	376
college degree	231
Total	792

References

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

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

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

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

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

[6] 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.

[7] Browning, Martin, Lars Hansen and James 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.

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

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

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

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

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

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

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

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

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

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

[18] 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.

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

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

[21] Hall, G. and J. Rust (1999), "Econometric Methods for Endogenously Sampled Time Series: The Case of Commodity Price Speculation in the Steel Market". Unpublished Manuscript, University of Yale.

[22] Harris, R. and E. Tzavalis (1999), "Inference for Unit Roots in Dynamic Panels Where the Time Dimension is Fixed", *Journal of Econometrics*, 91, 201-226.

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

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

[25] Kendall, M.G. (1954), "Note on Bias in the Estimation of Autocorrelation", *Biometrika*, 41, 403-404.

[26] 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.

[27] 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.

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

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

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

[31] 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.

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

[33] Meghir, C. and L. Pistaferri (2001), "Income Variance Dynamics and Heterogeneity", Institute for Fiscal Studies, discussion paper 01/07.

[34] Neal, Derek and Sherwin 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.

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

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

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

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

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

[40] Vuong, Q.H. (1989), "Likelihood Ratio Tests for Model Selection and Non-Nested Hyptoheses", *Econometrica*, 57, 307-333.

[41] Wooldridge, J. (2000), "The Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity", manuscript, Department of Economics, Michigan State University.