Heterogeneity in the dynamics of labor earnings.*

Martin Browning† — Oxford University
Mette Ejrnæs‡ — University of Copenhagen

December 2012

Abstract

We survey the literature on individual earnings dynamics with a particular focus on allowing for pervasive heterogeneity across individuals. We structure the discussion around ARMA processes with nonlinear trends for each individual. We show that allowing for pervasive and co-dependent heterogeneity in individual parameters has a major impact on econometric modeling, estimation and substantive conclusions. We describe an econometric method that is suitable for models with pervasive heterogeneity. We develop a long list of statistics that describe any earnings panel in great detail and provide a demanding set of features of the data for fitting. This list encompasses most moments used in the literature as well as providing novel statistics based on individual regressions. Finally, we provide an empirical illustration using a long Danish panel. Based on this we provide some conclusions concerning earnings dynamics but emphasize that details will vary according to the sample.

Key words: ARMA, Simulated Minimum Distance, co-dependent heterogeneity, panel data.
JEL codes: J31, C23

*We thank an editor and Manuel Arellano, Richard Blundell, Fatih Guvenen, Bo Honoré, Dmytro Hryshko, Bent Nielsen, Anders Rahbek and the participants at the workshop on Earnings dynamics in London, March 2012 and the Danish microeconometric network meeting, June 2012 for useful comments.

†Corresponding Author: Department of Economics, Oxford University, Manor Road, Oxford OX1 3UQ, England. email: martin.browning@economics.ox.ac.uk.

‡Department of Economics, University of Copenhagen, Øster Farimagsgade 5,DK-1355 Copenhagen, Denmark. email: mette.ejrnes@econ.ku.dk.

Contents

1 Introduction.

During the last 35 years a large empirical literature on individual earnings dynamics has developed. Despite the number of papers, for many important issues no real consensus has emerged on how the evolution of individual earnings over the life cycle and the business cycle should be modelled. Different studies have found very different results and these differences impact on questions of interest. In this survey we take a rather narrow focus to concentrate on what seems to us the most important direction for future research: allowing for pervasive heterogeneity. Specifically, we consider models with conventional time series processes (an ARMA(1,2)) for each individual that can vary across individuals. For example, one worker might have a unit root process with a high short run variance whereas another has a stationary process with a low short run variance. We set up a model in which the earnings process of each worker is described by seven parameters. For the purposes of this survey we shall take the individual parameters to be invariant with age and period; this is largely to allow us to focus on the key aspect of heterogeneity. Given this perspective, the prime object of interest is the joint distribution of the individual parameters across the population, conditional on observable time-invariant factors such as schooling, race and cohort. In this view there is a particular emphasis on the correlation between parameters; we label this as *co-dependent heterogeneity*. For example, it may be that not only are the AR and short run variance parameters heterogeneous, but the two may be statistically dependent.

We have four broad reasons for our contention that it is important to allow for pervasive co-dependent heterogeneity. The principal reason is that there is empirical evidence that is impossible to reconcile with a model with limited heterogeneity. Below we will show that standard statistics of earnings processes, such as the cross sectional variance and autocorrelation functions indicate that heterogeneity is needed. Moreover, we will also provide novel statistics of the data which can be used to detect heterogeneity. Most prosaically, looking at individual time series provides a persuasive and simple indication that it is unlikely that the same process can describe earnings paths for all individuals. In Figure 1, we show earnings paths for 20 Danish men born in 1958 with vocational training; they are followed from close to the beginning of their labour market career in 1980 to 30 years ahead (the data will be described in detail later). We have randomly selected 10 men in the bottom decile and 10 men in the top decile of the entry earning distribution. There are two important features we take away from these figures. First, the earnings paths are very diverse, even when we condition on the initial value decile; specifically, there is a large variation in volatility and final values. Second, those from the bottom decile have a steeper trend in earnings as compared to the top decile which strongly suggests co-dependent heterogeneity in trends and the initial condition.

[Figure 1 about here]

The very different results found in the literature could also be seen as an indication of pervasive heterogeneity. The empirical results in the literature are based on different data sources, different samples and different conditioning variables. If all workers have the same process with very limited heterogeneity these differences should not matter much. However, if there is pervasive heterogeneity, these differences in samples will matter. In this context it is important for us to provide a first warning that the results from the Danish sample that we use to *illustrate* our methods will not necessarily have wide application. For example, we find evidence that no one in this sample has a unit root but we should not extrapolate from this and conclude that there are no samples in which some (or all) workers have a unit root.

The second motivation for allowing for pervasive heterogeneity derives from our view that models of the earnings process, even though atheoretic, should not rule out widely posited theory models. The classic example is the Becker-Mincer human capital model which predicts a negative correlation between entry earnings and subsequent earnings trends (for a given schooling level). More generally, Rubinstein and Weiss (2006) show that leading investment, search and human capital models predict that the previous level of earnings will affect future earnings growth. In a similar vein, matching or search models that have heterogeneity for firms and workers (for example, Postel-Vinay and Robin (2002)) are likely to yield complicated latent heterogeneity structures for individual earnings processes.[1] Eckstein and van den Berg (2007) present a decomposition of wage variation into components due to individual heterogeneity (candidate variables are ability and discount rates), firm heterogeneity (for example, capital structure and managerial productivity) and heterogeneity in search frictions (for

[1] Postel-Vinay and Turon (2010) provide estimates of a unit root model with limited heterogeneity for a structural search model with two sided heterogeneity. Unfortunately they do not consider moments of the data that would pick up heterogeneity in parameters.

example, arrival rates of job offers). As yet another example, Groes, Kircher and Manovskii (2012) present empirical analysis and a theory model of occupation switching that entails a good deal of heterogeneity. It is likely that any atheoretic process will require a good deal of latent heterogeneity to capture the predictions of such models.

Third, most interesting questions cannot be properly addressed unless heterogeneity is explicitly taken into account. An example is how expected life time earnings and earnings volatility vary across individuals. If individuals choose schooling and occupations so as to trade off life time earnings against volatility in earnings, we would expect to see a positive correlation between expected life time earnings and the variance of the earning shock.[2] Another example is the extent to which workers can self-insure against adverse earnings shocks. If individuals who have a high variance for shocks also experience highly persistent shocks, self-insurance is costly and social insurance is valuable. Conversely, if the persistence and variance of shocks are negatively correlated, the need for social insurance is weakened. To address these questions we need to allow not only that the persistence and variance of shocks are heterogeneous but that they are correlated. Given this perspective, analyses of the efficacy of social insurance programs (such as Huggett and Parra (2010)) need to be extended to allow that the value of social insurance may be highly idiosyncratic.

The final reason why it is important to consider heterogeneity is that it has

[2] Whether we consider the short run variance or the long run variance depends on the capital market structure the worker faces; see Cunha, Heckman and Navarro (2005).

a major impact on the validity of econometric modelling and estimation procedures. Below we shall illustrate how neglecting co-dependent heterogeneity may invalidate the inference on important parameters. For example, the commonly used approach of removing time effects by running an initial regression on time dummies will induce bias if there is heterogeneity in the persistence of a shock. Thus, we have to re-think econometric methods as soon as we introduce pervasive heterogeneity.

It will be clear that we believe that there is much more heterogeneity than is usually allowed for in the current context. If this is the case, then we are just at the beginning of modelling labour earnings processes and there is a still a great deal to be done on identification, incorporating observable heterogeneity and allowing for age/time varying parameters. Let the reader be warned that this survey has more loose ends than most surveys.

Our focus in this survey is on finding a comprehensive atheoretic *linear parametric process* that fits the data well and that can be approximately consistent with a wide range of theory models. Our emphasis on an atheoretic process should not be seen as saying that the theory is unimportant, but there already exist a number of fine surveys or reviews that focus on the interface between the theory and empirical processes; examples include Heckman, Lochner and Todd (2003), Rogerson, Shimer and Wright (2005), Rubinstein and Weiss (2006), Eckstein and van den Berg (2007) and Meghir and Pistaferri (2011). The restriction to linear models is because they are easier to handle and seem to fit the data well; but see Kniesner and Li (2002) for a nonlinear parametric model, Horowitz and Markatou (1996) for general semiparametric modelling and Bohomme and Robin (2010) for non-parametric modelling in this context. Moreover, we want a model which can describe the *level* of earnings as a number of interesting questions require knowledge on the level of earning as well as the growth of earning. In the end, we find that an $ARMA\,(1,2)$ model with a variety of supplementary features fulfills both requirements.

Section 2 presents candidate linear parametric processes and examines in detail the properties of the deterministic component, the stochastic component and the form of heterogeneity. The intention here is to derive a general process that encompasses all of the existing suggestions in the literature and allows for pervasive heterogeneity. The only time varying conditioning variable we allow for is labour market 'experience'. That is, we do not consider factors such as job loss due to plant closure or entry into or out of marriage. Although these events have considerable 'explanatory power' we keep to our task of finding a process that includes them implicitly as unobserved heterogeneity rather than explicitly as observed heterogeneity. This in sharp contrast to studies of earnings processes such as Topel and Ward (1992), Geweke and Keane (2000) and Altonji, Smith and Vidangos (2009) that do take account of observable and correlated factors. It is an open and interesting question as to whether unconditional models with pervasive heterogeneity can encompass such models. We introduce unobserved heterogeneity into the model using a latent factor structure. In this section we also provide a discussion of the challenges introducing heterogeneity imposes, especially when modelling initial conditions.

In section 3 we take a myriad of facts about earnings, based on a particular Danish sample. This serves to illustrate many of the points made in the previous section. The Danish data has three paramount advantages as compared to many other panel data sets. First, the series of earnings we can construct for any worker is very long; in our case, 1980 to 2009. The only attrition is due to death or emigration. The second advantage is that since the data are based on administrative information, we can follow everyone in the economy (including over one million male workers) and we can stratify on time invariant factors (such as cohort, education and race) quite finely and still have large samples. We focus on men born in 1958 in Denmark with vocational education. This allows us to control for permanent observable differences in a robust fashion. The third advantage is that the earnings reports are taken from tax returns which gives very low measurement error. Of course, our specific findings might be peculiar to the conditions for our particular cohort.[3] We argue that although the Danish labour market is unusual in some respects, it is not pathological. For example, in our data period, the Danish labor market is similar in terms of worker protection and turnover to that of the US; see Albæk and Sørensen (1998). Similarly, although the Danish tax and benefit system is more comprehensive than in most other countries, it is not wildly so. In the appendix, we show that the our Danish earnings data display all of the broad features of the dynamics seen in the US data.

[3] This cohort might be impacted by particular business cycles conditions, demand side shifts, policy reforms or institutional changes during their labour market career.

In section 4 we present estimation results from the empirical analysis on the Danish sample. We estimate the model presented in section 2 by using Simulated Minimum Distance ("indirect inference"). The justification for using SMD as a bias reduction estimation method in dynamic panel models is eloquently presented in Gouriéroux, Phillips and Yu (2010), albeit for a much simpler model than ours.[4] The parameters are chosen to fit a subset of the statistics from section 3 with the remaining subset being kept back for a goodness of fit test. The most important finding is that we find strong evidence for pervasive co-dependent heterogeneity. We find heterogeneity in both the trends, ARMA parameters and especially the variance across individuals. Models that restrict heterogeneity, for example, by assuming homogenous trends or a unit root model for everyone, are decisively rejected. Once again, we emphasise that although very similar general results have been found for the PSID (see Browning, Ejrnæs and Alvarez (2010)) these conclusions may not hold universally.

In section 5, we summarize what we believe are the main contributions of the survey. First, we believe there is accumulating evidence for much more pervasive heterogeneity than most of the previous literature considers. We provide a general ARIMA model and an empirical methodology that can serve as a general model that can be used to adapt the model to other data to hand. Related to this issue, we reiterate the widely held belief that if a parametric model is to

[4] As the 'bias reduction' terminology suggests, SMD estimates will not generally be consistent in N for fixed T. The best that can be hoped for is that the bias becomes negligible. In our empirical illustration we have a relatively long panel ($T = 30$) which should yield low fixed-T bias.

have any credibility, it needs to be tested against a wide range of goodness of fit (GF) test statistics. We suggest a specific set of 86 such statistics. These include most moments used in the literature as well as some novel individual regression based statistics.[5] This set provides a powerful test for any specific model of earnings processes (be it parametric or semi-parametric). Fitting to restricted subsets of this set (for example, just the autocovariances of first differences) is likely to lead to bias conclusions and lead to instability in general conclusions across studies using different subsets of the GF statistics. Finally, we show that allowing for pervasive heterogeneity can have a major impact on outcomes of interest.

2 Parametric earnings processes.

2.1 An ARMA(1,2) for each person.

We first address the specification of the deterministic component of the earnings process. Let y_{it} denote log earnings for person i at 'age' t; here age is in inverted commas as it is often 'potential experience' (current age minus age at completed schooling). We assume that the deterministic component of the earnings process has a *steady state mean* given by:

$$E(y_{it}) = \mu_i + \alpha_i(t_i - 1) + \tau_i(t-1)^2 \tag{1}$$

[5]The set we suggest in this survey includes some statistics designed to pick up time or age varying parameters. These could be further augmented by additional GF statistics. For example, statistics designed to pick up GARCH effects as in Meghir and Pistaferri (2004) and Browning *et al* (2010).

where μ_i is the *steady state initial value*. Allowing that all the parameters are heterogeneous, this form encompasses many theoretical models of the evolution of earnings; see Rubinstein and Weiss (2006) and Heckman, Lochner and Todd (2006).

For the stochastic component, we take an additive component to (1), denoted ε_{it}. An error components approach allows that this stochastic component can be decomposed into a autoregressive *persistent component*, p_{it}, and a *transitory shock*, u_{it}:

$$\begin{aligned} \varepsilon_{it} &= p_{it} + u_{it} \\ p_{it} &= \rho_i p_{i,t-1} + \eta_{it} \end{aligned} \tag{2}$$

where η_{it} is a *persistent shock*. This persistent-transitory (PT) model is a generalization of the very widely used permanent-transitory model (Friedman and Kuznets (1954), chapter 7) which imposes $\rho_i = 1$. Universally in the earnings literature, it is assumed that the shocks, u_{it} and η_{it}, are uncorrelated; this is a difficult assumption to maintain. For example, a loss of earnings due to lay-off and subsequent short spell of unemployment might lead to a temporary loss of earnings and a longer run impact on wages when subsequently re-employed; see, for example, Kletzer (1998). Ejrnæs and Browning (2012) discuss in detail the problems that arise if the zero correlation condition is not imposed.

Combining (1) and (2) we have:

$$\begin{aligned} y_{it} &= \left[\mu_i\left(1-\rho_i\right)+\rho_i\left(\alpha_i-\tau_i\right)\right]+\rho_i y_{i,t-1}+\left[\alpha_i\left(1-\rho_i\right)+2\rho_i\tau_i\right](t-1) \\ &\quad +\tau_i\left(1-\rho_i\right)(t-1)^2+u_{it}-\rho_i u_{i,t-1}+\eta_{i,t} \end{aligned} \tag{3}$$

If the persistent shock η_{it} is serially uncorrelated and the transitory shock is an $MA(1)$[6] then a general representation of this model is as an $ARMA(1,2)$ model with a quadratic trend given by:

$$\begin{aligned} y_{it} &= \left[\mu_i (1-\rho_i) + \rho_i (\alpha_i - \tau_i)\right] + \rho_i y_{i,t-1} + \left[\alpha_i (1-\rho_i) + 2\rho_i \tau_i\right](t-1) \\ &\quad + \tau_i (1-\rho_i)(t-1)^2 + \xi_{it} + \theta_{i1}\xi_{i,t-1} + \theta_{i2}\xi_{i,t-2} \end{aligned} \tag{4}$$

The $ARMA(1,2)$ model is a more general specification than the model given by (3) if the latter assumes zero correlation between the persistent and transitory shocks. Specifically, every persistent-transitory model of the form (3) can be written as an $ARMA(1,2)$ but not all values of the parameters $(\rho_i, \theta_{i1}, \theta_{i2})$ in (4) admit a PT representation *with zero correlation* between the transitory and permanent shocks; see Ejrnæs and Browning (2012). Thus the widespread choice of a permanent-transitory form ((2) with $\rho_i = 1$) as a basis for the stochastic component is doubly restrictive in that it requires restrictions for a PT representation and a value of unity for the persistence parameter, ρ.

If $\rho_i = 1$ the process has a unit root (UR) and can be written as:

$$\Delta y_{it} = (\alpha_i - \tau_i) + 2\tau_i (t-1) + \xi_{it} + \theta_{i1}\xi_{i,t-1} + \theta_{i2}\xi_{i,t-2} \tag{5}$$

The UR assumption is a composite of two hypotheses; this is reflected in the

[6]Some researchers allow the transitory shock to have a persistent AR component so that the effect of a transitory shock persists forever (albeit with a decaying influence). This seems to us to be a very odd assumption for a transitory process. Such an assumption gives, for example, an $ARMA(2,2)$ process with restrictions on the two AR parameters. This is neither nested within the following nor does it nest it.

fact that (5) has two fewer parameters (ρ_i and μ_i) than (4). The first restriction is that the impact of the initial value y_{i1} is permanent and the second restriction is that the effect of a subsequent shock is permanent.[7]

2.2 The properties of the shock.

The most common assumption for parametric models is:

$$\xi_{it} \sim N\left(0, \nu_i^2\right) \tag{6}$$

If we wish to generalize this, we have two basic issues to consider: the normality assumption and the time series properties of the moments of the shocks. Although we would always keep the zero mean assumption in (6) we might want to consider skewed distributions and fatter or thinner tails than the Normal. For example, Lillard and Willis (1978) observe that the shocks they estimate are fat tailed and slightly left skewed. Horowitz and Markatou (1996) use semiparametric estimation methods and conclude that the earnings error distributions are not Normally distributed.[8]

There are two broad avenues to weakening the normality assumption. The first is take a mixture of Normals, see Geweke and Keane (2000). Another, more

[7] Browning *et al* (2010) introduce an extra parameter to break the link between the AR parameter and the initial convergence. This allows that the persistence of a shock and the initial convergence are not governed by the same parameters. We shall not allow for this in the following.

[8] We do not separately allow for measurement errors so this is subsumed in the error term. If the measurement is serially uncorrelated then it appears as a part of the transitory shock in a PT representation.

parsimonious, approach is to allow for the shocks to take a three parameter, zero mean distribution directly. Here we can follow the finance literature; see Hansen, McDonald and Theodossiou (2007). In our context, a convenient distribution is the *translated hyperbolic sine* (ths) transformation:

$$\begin{aligned} \xi_{it} &= a_i + \nu_i \sinh\left\{c + dN(0,1)\right\} \\ \text{with } a_i &= \frac{\nu_i}{2}\left(e^{-c} - e^{c}\right) e^{\left(d^2/2\right)} \text{ and } \nu_i, d > 0 \end{aligned} \quad (7)$$

where $N(0,1)$ is a standard normal. The restriction on a_i is to give a zero mean for given $\{\nu_i, c, d\}$. The (homogeneous) parameters c and d control the skewness and kurtosis, respectively. The (heterogeneous) parameter ν_i determines the variance of the error distribution for a given i; specifically, the standard deviation of ξ_{it} is given by:

$$std\left(\xi_{it}\right) = \frac{\nu_i}{2}\sqrt{2 + \exp\left(d^2 - 2c\right) + \exp\left(d^2 + 2c\right)}\sqrt{\exp\left(d^2\right) - 1} \quad (8)$$

We use this *ths* form in our empirical illustration below.[9]

We might also allow that the parameters of the distribution of the shocks vary with age or calendar time; see Alvarez and Arellano (2004) for a discussion of the econometric problems that arise if we have time series heteroskedasticity. For example, we could allow that the variance of the shocks vary smoothly with potential experience:

$$\nu_{it} = \nu_i f\left(t; \beta_i\right)$$

where $f\left(t; \beta_i\right)$ is a positive function of potential experience that depends on the

[9] In our empirical illustration, tests for heterogeneity in the skewness and kurtosis parameters do not show any evidence that this is needed.

parameter vector β_i. Baker and Solon (2003) find evidence of an age effect in the variance.

As well as variations with experience, we might allow that there are common period effects. For example, Moffitt and Gottschalk (1994) find that earnings instability rose during the 1970's and 1980's in US; see Moffitt and Gottschalk (2012) for a recent review and discussion. Time varying variances could be motivated as labour supply effects, which in this case suggests that the variance should depend both on age and period crossed, since there is evidence suggesting that younger people are harder hit during a recession. We might also wish to allow for transitory dynamic sources of variation in the variance that can be captured by a GARCH structure (see Meghir and Windmeijer (1999), Meghir and Pistaferri (2004) and Browning *at al* (2010)).

As we discuss further in section 4, the whole issue of whether variances vary with age, period and/or previous shocks requires thorough re-examination once we allow for pervasive heterogeneity. This is particularly pressing if it is accepted that there is substantial cross-section variation in error variances, the ν_i's above, since composition effects could give what looks like parameter instability. In our empirical illustration below we ignore these issues; a full examination would be far beyond the scope of the current survey. This should not be read as our contending that such effects are unimportant. On the contrary, we believe that one of the most pressing issues in estimating with pervasive heterogeneity is to assess how much allowing for time or age varying homogeneous parameters would obviate the need for pervasive heterogeneity. Or whether we can separately identify the two, given the data to hand.

2.3 Initial conditions.

When considering individual income processes we have to pay attention to the initial value. In contrast to the time series literature that usually assumes that the process has been running for a long time, making assumptions on initial values redundant, individual income processes start when entering the labour market; see Topel and Ward (1992). We define the *initial value* as the value when the process *starts*, denoted y_{i1}; this is not necessarily the same as the *initial observation* for i.

For $t > 1$, the process (4) can be written:

$$y_{it} = \mu_i + \alpha_i(t-1) + \tau_i(t-1)^2 + \rho_i^{t-1}(y_{i1} - \mu_i) + \sum_{s=0}^{t-2} \rho_i^s (\xi_{i,t-s} + \theta_{1i}\xi_{i,t-s-1} + \theta_{2i}\xi_{i,t-s-2}) \tag{9}$$

where the value at t depends on $\xi_{i1}, ..., \xi_{it}$, as well as an additional error term, ξ_{i0}; all assumed mutually independent. Following Chamberlain (1980), Wooldridge (2005) advocates modelling the joint distribution of the latent parameters as dependent on $y_{\cdot 1}$. This simplifies identification and estimation but has two disadvantages from our perspective. First, we do not always observe the initial value (as opposed to the initial observation). Second, the need we see is for a fully parametric model if one wants to address important questions of how entry earnings are related to other features of the process, for example the variance or the trend. Given this, we need to model y_{i1}.

We model the initial value by:

$$y_{i1} = \mu_i + m_1 + c_0\xi_{i1} + c_{i1}\xi_{i0} + c_{i2}\xi_{i,-1}$$

where $\xi_{i,-1}$ is an additional term that is independent of all ξ_{it}.[10] If $|\rho_i| < 1$ we can impose covariance stationarity on the initial value y_{i1} by imposing restrictions on the parameters $\{m_1, c_0, c_{i1}, c_{i2}\}$. For *mean and covariance stationarity*, we require:

$$\begin{aligned} m_1 &= 0 \\ c_0 &= 1 \\ c_{i1} &= \rho_i + \theta_{i1} \\ c_{i2} &= \frac{\rho_i^2 + \rho_i\theta_{i1} + \theta_{i2}}{\sqrt{1-\rho_i^2}} \end{aligned} \quad (10)$$

(a proof is provided in the appendix A.2). If $|\rho_i| < 1$ but the restrictions (10) do not hold, we will call the model *stable.* Notice, that the conventional panel data 'unrestricted specification' for initial conditions (for example, Hsiao (1986), chapter 4 or Arellano (2003), section 6), assumes that y_{i1} is drawn from *one* common distribution. This does not admit a stationary model if the parameters $(\rho, \theta_1, \theta_2)$ are heterogeneous.[11]

Geweke and Keane (2000) and Browning *et al* (2010) also allow that the initial conditions depend on observable characteristics such as education, parental education, race and marital status. In Geweke and Keane (2000), the mean is a function of observables and in Browning *et al* (2010) both the mean and

[10] Below we actually take a generalisation of this; see equation (13).

[11] In appendix A.2, we derive the restrictions on c_0, c_{i1} and c_{i2} when we have a unit root.

dispersion of y_{i1} are functions of observables.

2.4 Heterogeneity in the parameters.

Unit roots and heterogeneous trends.

From the beginning of the empirical literature on earnings, there has been an unusual allowance for heterogeneity, not just in intercepts (as is conventional in the panel data literature), but also in trend parameters; a prime example is Lillard and Weiss (1979). One important issue is whether workers have different trends. As we have already discussed, the paths shown in Figure 1 provide strong visual evidence that trends are heterogeneous and that the trends are correlated with initial values. A logically distinct issue is whether individual processes have a unit root. Although logically distinct, tests for a unit root and heterogeneous trends (or drifts) are closely bound together. In the time series literature it is well established that for a single time series, including a trend in a regression when no trend is present in the data considerably lowers the power of a unit root test (see case 4, chapter 17 of Hamilton (1994)). Similar reasoning carries over to the panel data context; see Levin *et al* (2002) and Im *et al* (2003). Both of these papers use a Dickey-Fuller type statistic to test for unit roots. Critically, these analyses allow for 'lots of heterogeneity'; specifically both papers allow for heterogeneity in variances and the MA parameters.[12] Levin *et al* (2002) also

[12] Allowing for heterogeneous short run variances and MA parameters in the panel data context is important since the power of unit root tests in the time series context depends on the short run and long run variance. Imposing that everyone has the same variance when this is not the case biases against the hypothesis that everyone has a unit root.

allow for heterogeneity in trends while Im *et al* (2003) allow for heterogeneity in the AR parameter under the alternative (that is, they test for everyone having a unit root against an alternative in which some have a stationary process). Both of these papers develop test statistics based on individual regressions. For us, the important lesson we draw from these papers is that it is possible to craft powerful tests for a unit root and/or heterogeneous trends using individual regressions.

We can use the deterministic component of the conditional model (4) where, for convenience, we drop the quadratic trend (τ_i) and assume that the AR parameter is homogeneous to line up with the discussion in the previous literature:

$$y_{it} = (\mu_i(1-\rho) + \rho\alpha_i) + \rho y_{i,t-1} + \alpha_i (1-\rho)(t-1) \tag{11}$$

Table 1 presents the four possible cases. In all cases μ_i may be heterogenous and may be correlated with α_i when the latter is heterogeneous. Guvenen (2009) terms the homogeneous drift unit root model the 'restricted income profiles' model (RIP).[13] Conversely we may have a stationary model with heterogeneous trends, which Guvenen (2009) terms the 'heterogeneous income profile' model (HIP). The other two cases do not have associated terminology, even though they have been used in the literature.

[Table 1]

Guvenen (2009) and Hryshko (2012) present analysis that reinforces findings

[13] Guvenen (2009) actually characterizes a model as RIP if it has an AR parameter very close to unity and homogenous trends. For fixed-T panels, the distinction between a unit root and an AR parameter close to unity is blurred.

from the time series literature, but do not allow for as much heterogeneity as Levin *et al* (2002) and Im *et al* (2003). Guvenen (2009) shows that if the true model is a HIP model, but it is estimated as a RIP model it will lead to an upward bias in the (homogeneous) AR parameter (ρ). Specifically, Guvenen (2009), finds that the AR estimate is found to be very close to unity in a model with no heterogeneity in trends, while the unrestricted model has an AR estimate of 0.82. That is, a misspecified RIP model will bias the results towards a unit root model. Hryshko (2012) shows that if the true model is a RIP model but is it misspecified as a HIP model this biases the results towards finding heterogenous trends.

To distinguish between HIP and RIP most of the literature applies the MaCurdy (1982) test, which uses the autocorrelation function for first differences. The HIP model would imply positive higher order autocorrelations. Most of the papers that favour a RIP model use the absence of significant positive autocorrelations as evidence against the HIP. However, Guvenen (2009) shows that it may require up to the 12th order autocorrelation to see the difference between RIP and HIP. Furthermore, Guvenen (2009) also shows how different assumptions on initial conditions can affect these tests.

Rather than relying entirely on the implications for first differences (earnings growth), some researchers (notably, Lillard and Weiss (1979), Baker (1997) and Guvenen (2009)) focus on the implications for the level of earnings and, in particular, the evolution of the cross sectional variance with age. This is motivated by the observation that the HIP model can have a U-shaped path for variances with a minimum close to the initial period (see, for example, Rubinstein and Weiss (2006)) whereas a unit root with a homogeneous drift implies that it should be linearly increasing. In Table 2, we summarize the implications of different assumptions for the cross-section variance. As seen from the table, there are distinct differences between the two models and it would seem to be a straight forward task to distinguish on the basis of this simple set of statistics. Empirically, however, it turns out that one needs fairly long time series that start near the beginning of the process in order to distinguish the two types of models. Moreover, panels with attrition and replacement may be subject to composition effects which can give rise to spurious patterns.

[Table 2]

The literature has not reached a conclusive consensus of whether there is a unit root or not in earnings processes and the degree of heterogeneity in trends/drifts. Lillard and Weiss (1979), Hause (1980), Baker (1997), Haider (2001) and Guvenen (2009) find evidence in favour of a stable model with heterogenous trends that are correlated with the initial condition. Browning *et al* (2010) test the weaker hypothesis that some agents have a unit root and others a stable process; they reject the hypothesis that anyone has a unit root and find evidence of heterogeneous trends. Alvarez and Arellano (2004) reject a unit root, even with homogeneous trend parameters (the top left case in Table 1). MaCurdy (1982), working with first differences, finds no evidence of a unit root and recommends modelling earnings *growth* as an $ARMA\,(0,2)$ (with homogeneous parameters) which implies an $ARMA\,(1,2)$ for levels. MaCurdy finds that the AR parameter is very close to unity and rejects heterogeneity in the trends on the grounds that first differences do not display any high order (positive) auto-correlation.

On the other hand Abowd and Card (1987), (1989); Topel (1991); Topel and Ward (1992), Meghir and Pistaferri (2004) and Hryshko (2012) find evidence for a unit root without heterogeneous drift (the RIP model). Common to most of these papers is that they apply a MaCurdy style test. An exception is Hryshko (2012), who sets up a model that nests both HIP and RIP (his model contains all four cases in Table 1 as special cases). He finds the PSID data are well described by a RIP model and does not reject the restriction to homogenous trends.[14]. The papers that find support for the RIP model mainly use information on earnings growth for identification and do not use information on the levels of earnings. The underlying idea seems to be that using first differences removes heterogeneity in levels as well as initial condition problems. This is in contrast to the unit root panel data literature and to the papers finding support for the HIP model, where it is shown that levels contain important information on the dynamics.

Most of the earnings studies mentioned above do not allow for heterogeneity in the $ARMA$ parameters. Two exceptions to this are Wansbeek and Knaap (1999) and Browning *et al* (2010). Wansbeek and Knaap (1999) allow for a heterogeneous ρ parameter and a heterogeneous linear trend (α) and double

[14] This test and most of the tests in the literature are conditioned on particular dynamic structure of the transitory component; for example an $MA(2)$ process. Furthermore, none of the tests allow for the impact of initial conditions

differences to eliminate μ and α. Browning *et al* (2010) suggest an alternative approach which we now present in detail.

Incorporating co-dependent latent heterogeneity.

In this survey we consider the most general process from the previous two subsections; we postpone allowing for observable differences until subsequent subsections. Our model nests both RIP and HIP as special cases (as seen in Table 1)). In this model we have seven heterogeneous parameters to describe the earning process[15] and we also have the initial value which has to be modelled. We take as the object of interest the joint distribution of these parameters for a given schooling/cohort sub-population.

Our seven heterogenous *model parameters* are: ν_i for the standard deviation of the error term; μ_i, α_i and τ_i for the deterministic process and ρ_i, θ_{1i} and θ_{2i} for the dynamics of the process. To model the joint distribution of these model parameters we follow Browning *et al* (2010) and employ a triangular factor model with standard Normal factors denoted by η_{ik}, $k = 1, ..7$. The system is

[15] We also have two homogeneous model parameters, c and d, in the ths transformation (7).

given by :

$$
\begin{aligned}
\nu_i &= \exp\left(\phi_1 + \exp\left(\psi_{11}\right)\eta_{i1}\right) \\
\mu_i &= \phi_2 + \psi_{21}\eta_{i1} + \exp\left(\psi_{22}\right)\eta_{i2} \\
\alpha_i &= \phi_3 + \sum\nolimits_{k=1}^{2} \psi_{3k}\eta_{ik} + \exp\left(\psi_{33}\right)\eta_{i3} \\
\tau_i &= \phi_4 + \sum\nolimits_{k=1}^{3} \psi_{4k}\eta_{ik} + \exp\left(\psi_{44}\right)\eta_{i4} \\
\rho_i &= \ell\left(\phi_5 + \sum\nolimits_{k=1}^{4} \psi_{5k}\eta_{ik} + \exp\left(\psi_{55}\right)\eta_{i5}\right) \\
\theta_{1i} &= 2l(\phi_6 + \sum\nolimits_{i=1}^{5} \psi_{6k}\eta_{ik} + \exp\left(\psi_{66}\right)\eta_{i6}) - 1 \\
\theta_{2i} &= 2l(\phi_7 + \sum\nolimits_{i=1}^{6} \psi_{7k}\eta_{ik} + \exp\left(\psi_{77}\right)\eta_{i7}) - 1 \qquad (12)
\end{aligned}
$$

where the $\exp\left(\psi_{ii}\right)$ ensures that standard deviations are positive and $l(x) = \exp\left(x\right) / \left(1 + \exp\left(x\right)\right) \in (0, 1)$, so that $\rho_i \in (0, 1)$ and $\theta_{mi} \in (-1, 1)$ for $m = 1, 2$. We refer to the ϕ and ψ parameters as *distributional parameters,* since they determine the joint distribution of the model parameters. In estimation, some of the ψ_{jk}'s may be set to zero. Reduced factor models are given by setting $\psi_{jj} = -\infty$ and $\psi_{jk} = 0$ for $k > j$.

The initial value is specified as discussed in section 2.3, but with an extension to allow for dependence on the other model parameters:

$$
\begin{aligned}
y_{i1} &= \mu_i + m_1 + \tilde{c}_0\xi_{i1} + \tilde{c}_1\left(\rho_i + \theta_{i1}\right)\xi_{i0} + \tilde{c}_2\left(\frac{\rho_i^2 + \rho_i\theta_{i1} + \theta_{i2}}{\sqrt{1 - \rho_i^2}}\right)\xi_{i,-1} \\
&\quad + \psi_{81}\eta_{i1} + \psi_{83}\eta_{i3} + \psi_{84}\eta_{i4} \qquad (13)
\end{aligned}
$$

This is analogous to the scheme presented in Heckman (1981) in which the initial condition is modelled as a function of latent parameters; this is the antithesis of the scheme advocated in Wooldridge (2005). Following (10), stationarity implies:

$$\begin{aligned} m_1 &= 0,\ \tilde{c}_0 = \tilde{c}_1 = \tilde{c}_2 = 1 \\ \psi_{81} &= \psi_{83} = \psi_{84} = 0 \end{aligned} \tag{14}$$

To illustrate the flexibility this gives, consider the correlation between the initial value and subsequent earnings growth. From equation (9), (with, once again, the quadratic trend term set to zero) expected earnings growth between $t-1$ and t is given by

$$E\left(\Delta y_{it}|y_{i1}\right) = \alpha_i + \rho_i^{t-2}(\rho_i - 1)(y_{i1} - \mu_i) \quad t > 3 \tag{15}$$

For a stable model ($\rho_i \in (0,1)$) the second term on the right hand side gives a negative correlation which dies away exponentially. However, there may be a permanent negative relationship if, for example, $\alpha_i > 0$ and $\psi_{83} < 0$. Importantly, even in a unit root model ($\rho_i = 1$) we can have a correlation between the initial value and subsequent growth. This example also illustrates the very important point that when we have a unit root and pervasive heterogeneity (in this case, a heterogeneous drift term), first differencing does *not* remove the impact of the initial value.

Specifying the initial conditions correctly is important not only for the validity of particular econometric estimators but also for substantive uses of the estimates. For example, in the context of schooling choices, Cuhna *et al* (2005) make a distinction between what is known to the agent at the start of the earnings process ('heterogeneity') and what is yet to be revealed ('uncertainty').

Maximum uncertainty holds if, at the start of the process, agents know only their initial value and the joint distribution of all the other parameters conditional on this value and variables such as schooling level, race and cohort. The opposite case is one in which the parameters are known at the start of the process and only the subsequent shocks are unknown. As another example, when modelling consumption we might assume that agents know the parameters of their income process from the start or that they initially only know the joint distribution conditional on the initial value and learn about their own particular process as they age. Chosen consumption paths and the correlation of consumption changes with income shocks in early life depend critically on which case holds; see Guvenen (2007).

Incorporating observed heterogeneity.

Most previous studies take account of observed heterogeneity. Generally these studies accommodate variation in the parameters due to observable time invariant factors such as cohort, race and education by stratifying. However, a large number of studies, especially those that employ the permanent-transitory approach, 'take out' heterogeneity due to observables by running a first round regression. In a first round OLS regression (FRR) log earnings are regressed on a set of individual characteristics and time dummies and the residuals from this regression are used in the subsequent analyses. The set of individual characteristics that have been used includes: age (or experience); cohort; race; region of residence; demographics (such as marital status); health or disability and labour force status variables.

First round regression is an *ad hoc* way of modeling observed heterogeneity. Generally the motive and justification for the regression is not stated; a notable exception is MaCurdy (1982) which provides a justification for using residuals from a FRR in a model with latent heterogeneity in the intercept. MaCurdy makes the telling point that this regression should take account of possible dependencies between the heterogeneous intercept and the observable time varying sources of heterogeneity; later researchers seem to have neglected this point. From our perspective, which emphasizes that all parameters are likely to be heterogeneous, even a first round within estimation is problematic.

In a model such as that developed above, there are two alternative ways of including observed time invariant heterogeneity. The first alternative is to stratify on variables such as education and race that take on a limited number of values. For time invariant covariates that take on many ordered values, such as cohort, it is more practical to allow the model parameters to depend on observed heterogeneity as well as unobserved factors. For example, given a variable z_i we might modify the first two terms in (12) to:

$$\begin{aligned} \nu_i &= \exp\left(\phi_1 + \gamma_1 z_i + \exp\left(\psi_{11}\right)\eta_{i1}\right) \\ \mu_i &= \phi_2 + \gamma_2 z_i + \psi_{21}\eta_{i1} + \exp\left(\psi_{22}\right)\eta_{i2} \end{aligned}$$

Similarly, we could allow that the initial value, equation (13) depends on the observable variable.

It is less obvious how the time varying covariates should be included in the model, such that the reduced form model remains consistent with a wide range of structural models. An example is including a control for living in a city. In this case we can posit that this is a price effect with the price of human capital being higher in big cities. However, living in a big city might also be correlated with productivity. A conventional FRR would not wash out such effects; rather approaches such as Altonji, Smith and Vidangos (2009), which carefully model the exogeneity of observable events (such as a move to a city), would be needed.

First round regression with time dummies.

Related problems arise when introducing calender time effects. An approach used in very many studies is to include time dummies in the FRR to remove any macro shock and then work with the residuals from this regression. In this subsection we consider the impact of using a FRR on time dummies if we have pervasive heterogeneity. In a model with heterogenous model parameters, a first round regression would tend to induce bias in estimates of the distribution of the model parameters. To illustrate this point, consider a simple model with $\mu_i = \alpha_i = \tau_i = \theta_{1i} = \theta_{2i} = 0$ and common time effects:

$$y_{it} = \rho_i y_{it-1} + \omega_t + \xi_{it} \tag{16}$$

where ω_t is the macro shock. To simplify the calculations we assume that there are two types of individuals:

$$\text{Type 1} \quad : \quad y_{it} = 0.5 y_{it-1} + \omega_t + \xi_{it}$$

$$\text{Type 2} \quad : \quad y_{it} = y_{it-1} + \omega_t + \xi_{it}$$

Figure 2 illustrates the points by simulating the model above. In the simulations we take $var(\xi_{it}) = (0.75)^2$ and one macro shock with a value of -6 occurring in period 10. The upper panel shows the simulated income paths of the two types. For a type 1 person the impact of a shock decays quickly and the income level is quickly almost back to the level before the shock. For the type 2 person, all shocks are permanent, so the shock has a lasting impact. In the lower panel, we graph the residuals based on a first round regression of the same two persons. From this graph, we cannot see the macro shock at period 10. Instead we see that process of the residuals look very different. Especially the process for the type 1 person now looks like it is very persistent. This implies that working with the residuals of the first round regression makes the dynamics of the processes of the two individuals look more similar than using the original data. In appendix A.1 we formally show how the first round regression will upward bias the AR parameter of the persons with the lowest AR parameter.

[Figure 2]

3 The data and estimation.

3.1 The Danish data.

We use Danish administrative data to illustrate the estimation of the earnings model outlined in subsection 2.4. Our estimation sample is a narrowly defined sample, which is selected to make it as homogenous as possible. In this study, we model men born in 1958 who have vocational training. By focusing on a single birth cohort, we ensure that all individuals are exposed to the same general labour market conditions. The price of this is of course a very specific sample and that we are not able to separate time effects from age effects. We select a balanced sample consisting of Danish citizens who lived in Denmark in the period $1980 - 2009$ and who have not retired, worked as self-employed or been a full-time student within the sample period. Thus we observe all individuals in our sample from very close to the beginning of their labour market career. In this empirical illustration we wish to minimize the impact of unemployment and labor supply fluctuations so we select out individuals who have annual earnings below $50,000$ Danish Crowns (about $\$8,300$ in 2009 prices) in any year or who are employed for less than 80% of any year. We shall return to 'controlling' for employment effects in the conclusion. We have $2,122$ workers in our final sample; the details of the selection criteria are given in Table B.1 in the data appendix. Earnings are defined as gross annual earnings from the tax record reported by the employer and deflated by the consumer price index. In all the analyses we use log real gross earnings, which implies that changes in earnings reflect both changes in labour supply and wages. By the construction of the sample of men who work at least 80% of the time, we limit the impact of labour supply. [16].

Figure 3 shows the sample mean and variance of earnings as a function of potential experience. The mean is increasing in potential experience. As an aside we note this path of means would not be well fitted by a quadratic in age, a point that Murphy and Welch (1990) make forcibly. The cross sectional variance starts by falling, reaching a minimum at around 3 years of potential experience and then starts to increase.[17] Rubinstein and Weiss (2006) and Guvenen (2009) find the same fall in the variance in the beginning of the career, although both papers find that it occurs a bit later than we find. From 20 years of experience the cross sectional variance is almost constant.

[Figure 3 about here]

Table 3 displays the autocorrelations for earnings growth (not levels). The table shows that our sample exhibits the general pattern found in the literature. There is a large negative first order autocorrelation and then smaller but significantly negative correlations (up to lag 5). In two respects, however, our sample deviates from studies based on the PSID sample. First, we find a numerically lower first order autocorrelation compared to most other studies based on PSID.

[16] Our approach is not so different from the rest of the literature. Most of the literature takes earnings and 'controls' for labour supply effects in a first round regression. This gives something like full year equivalent earnings.

[17] Paths for other cohorts (not shown) indicate strongly that the initial fall is not a period effect.

The difference can be explained by the larger measurement error in the PSID. Second, we find some significant higher order autocorrelations (for example, order 12 and 15), but the size of the coefficients are at the same magnitude as found in, for example, Guvenen (2009). We attribute the difference to the fact that we have smaller measurement error and a large, more homogenous sample. The fact that the autocorrelations first are negative (for lag 1 to 5) then becomes positive (for lag 12) and later negative (for lag 15) is consistent with a model with heterogenous quadratic trends.[18]

[Table 3 about here]

To summarize, our sample is a highly selected sample, but it exhibits the same general features as the PSID. In the data appendix B, we provide a detailed comparison between our sample and the PSID in terms of the standard descriptive statistics. The conclusion based on the comparison is that the Danish sample looks very similar to the ('standard') sample we draw from the PSID in terms of autocorrelation pattern and size with the exception that we find significant higher order autocorrelation.

3.2 Estimation.

The literature on earnings processes has tended to follow panel data dynamic model estimation albeit without any reference to papers such as Levin *et al* (2002) or Im *et al* (2003). A fine overview of the estimation methods for models

[18] This follows from extending the calculations in Guvenen (2009) equation (8) to include a quadratic trend where α_i and τ_i (in our notation) are negatively correlated.

with limited heterogeneity can be found in MaCurdy (2007). One problem has been that many of the techniques suggested are highly model specific and small changes in the specification can lead to large changes in the choice of estimation procedure (see, for example, Guvenen (2009) or Hryshko (2012)). Moreover, many procedures for particular models treat parameters that are of interest to us as 'nuisance parameters'. In particular, most procedures are not designed to allow for pervasive co-dependent heterogeneity.

To estimate the parameters of our model, we suggest the use of Simulated Minimum Distance (SMD) ('indirect inference'). Gouriéroux, Phillips and Yu (2010) provide a persuasive defence for using SMD in the context of estimating a fully parametric dynamic panel. The principal advantages are: it is easy to use and hence it facilitates generalizing a given model; it automatically partially corrects for the bias induced by the presence of the lagged dependent variable and it can automatically consider any statistics that previous researchers have used in estimation. Note, however, that SMD does not yield consistent estimates if we have fixed T; the best we can hope for is that the bias is vanishingly small for relatively large T.

SMD requires the specification of a set of statistics which are known as *auxiliary parameters* (ap's). SMD proceeds by comparing ap's based on the sample with ap's based on simulated data from the model. The distribution parameters are determined by minimizing the weighted distance between the two sets of ap's. The ap's can be moments or functions of moments but could also be other statistics such as long or short run transitions. When choosing the ap's, the ap's should have a probability limit as the number of cross-section units becomes large (but this probability limit does not have to be known nor be anything of direct interest). Furthermore, it is critical that the ap's for the simulated data depend on all the distributional parameters.

Our aim is to estimate the joint distribution of all the model parameters. At present we seem to be far from having nonparametric identification results for structures such as a given parametric model for each worker with an unrestricted distribution for the model parameters, (but neither do we have any non-identification counter-examples). Given that our prime interest is the joint distribution of the model parameters, a more modest goal is to estimate features of the joint distribution of the model parameters with negligible bias. For example, if we are interested in knowing if the intercept and trends are correlated we need an estimate of this correlation.

Choosing auxiliary parameters.

The critical decision for SMD estimation is the choice of ap's. Our choice is dictated by several broad considerations. First, we need ap's that are 'bound' to crucial aspects of the model. For example, if we wish to estimate the dispersion of short run variances (the ν_i's above) then we need at least one statistic that varies with this. We also wish to include most moments that have been used in previous studies. This reflects our belief that choosing a subset of moments for fitting may lead to biased inferences. As an obvious example, if none of the ap's could indicate heterogeneity in the AR parameters, even if it present, then it is no surprise that we cannot reject the hypothesis that the persistence parameter is homogeneous. Finally, the ap's should be quick to calculate since they are embedded in an iterative estimation procedure.

Below we will describe the set of ap's we have chosen.[19] We have organized the ap's into five groups: cross sectional variance, autocorrelations, transitions, first period observations and statistics based on individual regressions. The first three groups contain the moments which are used in the existing literature. In estimation, we will use the last two groups for fitting the parameters and the first three groups for checking the goodness of fit. By using the 'conventional' moments for goodness of fit we test if our model is consistent with the moments normally considered. We recognize that there might be a high correlation between some of the ap's.

Cross section variances.

The first group of ap's relates to the cross sectional variance. As shown in Table 2, the pattern of cross section variance varies between a unit root model, a stationary model and a stable model. Also, all existing studies use information on the cross sectional variance for estimation and a number of studies are particularly concerned about how the cross sectional variance develops over age and calender time; see, for example, Moffitt and Gottschalk (2002). In our case (because we only look at one birth cohort) these effects coincide. To capture the pattern in the cross sectional variance in Figure 3, we regress the cross sectional

[19] Details are given in appendix C.1.

variance on a fourth order polynomial:

$$CSV_t = \kappa_0 + \kappa_1 t + \kappa_2 t^2 + \kappa_3 t^3 + \kappa_4 t^4 + w_t$$

and use the four estimates of the coefficients $\kappa_0, \kappa_1, \kappa_2$ and κ_3 as ap's. This makes 4 ap's relating to the cross sectional variance.

Autocorrelations.

Most of the previous literature relies on autocovariances/correlations of either y_{it} or Δy_{it} (see, for example, Lillard and Weiss (1979), Hause (1980), Abowd and Card (1989), MaCurdy (1982), Baker (1997), Moffitt and Gottschalk (2002), Guvenen (2009), Hryshko (2012)). Consequently our second group of ap's contains functions of autocorrelations. Since most of the studies mentioned run a first round regression on time dummies and use residuals from this regression we do this, giving residuals denoted $\tilde{y}_{it}$. The ap's are $corr(\tilde{y}_{it}, \tilde{y}_{it-1})$, $corr(\tilde{y}_{it}, \tilde{y}_{it-2})$ and $corr(\tilde{y}_{it}, \tilde{y}_{it-3})$ and $var(\Delta\tilde{y}_{it})$, $corr(\Delta\tilde{y}_{it}, \Delta\tilde{y}_{it-1})$ and $corr(\Delta\tilde{y}_{it}, \Delta\tilde{y}_{it-2})$. Meghir and Pistaferri (2004) identify and estimate the variance of the permanent shock using the correlation between long difference and short difference: $Cov(\tilde{y}_{it+2} - \tilde{y}_{it-3}, \tilde{y}_{it} - \tilde{y}_{it-1})$, so we also include this as an ap. This gives 7 ap's

Guvenen (2009) observes that the pattern of autocovariances over potential experience contains information on any heterogeneity in trends. Therefore, we also include ap's that describe how the autocorrelation function evolves over potential experience. We focus on the first and second autocovariances for different experience (see the data appendix, Figure B.4). We found that the following relationship gives a good description of the age evolution:

$$Cov(y_{it}, y_{it-k}) = \tau_{k0} + \tau_{k1}\frac{1}{t} + \text{residual for } k = 1, 2$$

The OLS estimates of the coefficients are taken as ap's. We run these regressions for both levels and first differences; giving 8 ap's.

Following MaCurdy (1982), we also calculate the partial autocorrelations for first differences. We include these for lags 1 to 4, 7, 10 and 12 (see the description of the partial autocorrelations in the data appendix). We find a significant positive partial autocorrelation for lag 12 while the lags of lower order are negative. Finally, we include measures of how the previous level of earnings is correlated with earnings growth. Rubinstein and Weiss (2006) show that this correlation can be used to distinguish between different theory models and that it varies over the life cycle. We calculate the correlation from a regression of Δy_{it+4} on y_{it} where we do separate regressions for three different bands of potential experience: $1-9$, $10-18$ and $19-26$ (see appendix B). The correlations are -0.034, -0.007 and -0.019 respectively; Rubinstein and Weiss (2006) also find negative correlations for US data.

In total, we have 25 ap's relating to autocorrelations.

Transitions.

Finally, studies on income mobility use transitions between quantiles in the income distribution; see, for example, Geweke and Keane (2000) and Atkinson *et al* (1992). To capture this aspect, we include two measures of mobility: short run persistence and long run persistence of low income. The short run persistence we define as the conditional probability of staying in the lowest quintile for two subsequent years $P(t, t+1)$. In the data we find that this transition probability is 0.75 suggesting that only one out of four immediately moves out of a low earnings state. The long run persistence is defined as the conditional probability of a workers being in the bottom quintile given that they were in the bottom quintile 10 years before, $P(t, t+10)$; this transition probability is 0.49 which indicates a high degree of long run persistence in having low earnings.

First period observation.

As discussed in section 2.3, it is important to model the first observation flexibly. To identify the parameters of the first observation we include 9 ap's describing the distribution of the first three observations. We calculate the mean, standard deviation, skewness and kurtosis of the first observation. Then we use the mean and standard deviation of the second observation as well as the correlations: $corr(y_{i1}, y_{i2})$, $corr(y_{i1}, y_{i3})$ and $corr(y_{i2}, y_{i3})$. These correlations allow us to distinguish between a stable and a stationary model.

Individual regression based (IRB) ap's.

The final set of ap's is based on individual regressions. We calculate estimates for each individual and use the estimates to calculate 'aggregated statistics' for the whole sample. This idea of individual regressions was first proposed by Baker (1997) although he only uses it for visual inspection and not for estimation. In

Browning *et al* (2010) it was used as ap's for estimation. As already discussed, using individual regressions has been applied in the literature on unit root tests in panel data, see Levin *et al* (2002) and Im *et al* (2003). Both papers run individual based regressions to construct the DF test statistics. Levin *et al.* (2002) show that it is possible to construct a powerful unit root test against an alternative where everyone has a stable process with heterogenous trends based on individual regressions with 25 time periods.[20]

We seek a parsimonious and fast method for generating ap's that relate to the joint distribution of the parameters. Equation (4) is nonlinear in parameters and involves an $MA\,(2)$ error term. It is not feasible to embed this in an iterative simulation based estimation scheme in which the auxiliary parameters have to be calculated very many times. Instead, we exploit the two stage structure that motivated (4) to derive a three step procedure for calculating the IRB ap's. The first step, which is analogous to (1) employs a regression of log earnings on a quadratic of experience for each agent:

$$y_{it} = b_{1i} + b_{2i}t + b_{3i}t^2 + r_{it} \tag{17}$$

In the second step, we take the estimated residuals from this regression:

$$\hat{r}_{it} = y_{it} - \left(\hat{b}_{1i} + \hat{b}_{2i}t + \hat{b}_{3i}t^2\right),\ t = 1, ..., T \tag{18}$$

and regress $\hat{r}_{it}$ on $\hat{r}_{i,t-1}$ and record the OLS parameter $\hat{b}_{4i}$. Finally we take the

[20] The use of individual regressions in itself requires a reasonably large value for T which reinforces the use of SMD as a bias reduction method.

estimated residuals from the second regression:

$$\hat{u}_{it} = \hat{r}_{it} - \hat{b}_{4i}\hat{r}_{it-1},\ t = 2, \ldots, T \tag{19}$$

and record the standard deviation, denoted $\hat{b}_{5i}$, and the first two autocorrelations of the $\hat{u}_{it}$'s, denoted $\hat{b}_{6i}$ and $\hat{b}_{7i}$ respectively. Finally we calculate the skewness and kurtosis of the $\hat{u}_{it}$'s, $\hat{b}_{8i}$ and $\hat{b}_{9i}$. Thus for each individual we have 9 individual estimates, which are estimated on 30 observations. To illustrate these individual estimates, we consider the estimate $\hat{b}_{4i}$ and the log of the estimate of the standard deviation $ln(\hat{b}_{5i})$. The AR estimates $\hat{b}_{4i}$ are distributed between -0.4 and 0.9, with an average around 0.35 and a standard deviation of 0.24. These are biased estimates of the individual AR parameters, due to small sample bias (see Kendall (1954)) and may also be affected by the MA component, so the size of the IRB estimates does not have any meaningful interpretation. However they do provide information on the true distribution of ρ. Similarly we can obtain information of the true distribution of the individual variances, ν_i, from the estimates, $ln\left(\hat{b}_{5i}\right)$. To obtain information on the codependent heterogeneity we calculate the correlation between the estimates; for example $corr(\hat{b}_{4i}, ln\left(\hat{b}_{5i}\right))$ (which has a value of 0.56 in our sample) provides information on the correlation between the AR parameters and the variances.

Based on these 9 individual estimates we calculate aggregated sample statistics. For $\hat{b}_{1i} - \hat{b}_{7i}$, we calculate means, standard deviations and correlations. This give 35 ap's. We also calculate the correlation between these 7 parameter and y_{i1}, which adds 7 more ap's. Finally we calculate the means and standard deviations of the $\hat{b}_{8i}$'s and $\hat{b}_{9i}$'s as ap's. In total we have 46 IRB ap's.

4 Estimation results.

4.1 The fit of different models.

For the estimation we use Simulated Minimum Distance (SMD) and the ap's described in section 3. We have in total 86 ap's. We use the 46 IRB ap's and the 9 ap's for the initial observation for estimation (that is, these ap's are matched to the data using a conventional weighting matrix). The remaining 31 ap's from the first three sets are kept back for goodness of fit tests.

In Table 4 we present results for four stable models and a unit root model. All of these models have fewer parameters than the 55 ap's used for estimation; thus, an OI test can be performed as well as the goodness of fit test. In appendix C.2 we present all the details of the estimates (see Table C.3 and C.4).

[Table 4 about here]

The first column of Table 4 gives fit statistics for our preferred model; this has five factors (the implications are given below). Tests for fewer factors and for stationarity both reject; the $\chi^2(7)$ quasi-LR test statistic for stationarity (see (14)) is 130.5. Although the the OI test statistic for the fitting to our 55 fitting ap's is marginal (a probability of 0.5%), all 55 ap's used in estimation are fitted quite well. The goodness of fit test is less favorable with the worst fit for the ap's relating to the cross sectional variance and the autocorrelation of third order. Our reading of the results suggests that the model cannot fully fit the individual trends for older ages; the model over-predicts the trend in variance and cannot capture the negative correlation between earnings growth and levels for older ages. This suggests that we need to allow age dependence for some of our model parameters (for example, the variance or the persistence).

The next four columns give the results for restricted versions of the preferred model. The first is a model with a homogeneous AR parameter. This has a significantly worse fit for both the ap's used in fitting and for the GF ap's. The parameter estimate for the AR parameter is 0.62. The third column presents the results for the model with no heterogeneity in the quadratic trend.[21] As can be seen, we strongly reject that the trend parameters are homogeneous. In this survey we have been at pains to emphasize that it is not enough to allow for heterogeneity in the model parameters, we must also allow them to be co-dependent. The fourth column of the Table makes this clear. In this variant, we have set all the loading factors in (12) for cross-effects to zero (that is, $\psi_{jk} = 0$ for all $j \neq k$). The fit for both sets of ap's is dramatically worse. For the fitting ap's, this is because there is no mechanism for fitting the correlations between the IRB estimates. For all three variants, the OI test is strongly rejected in models with more limited heterogeneity. Since the OI test is largely based on the IRB ap's, this suggests that these ap's have more power in discriminating between models with more or less heterogeneity.

The last column of Table 4 reports the results for the unit root model. We see that the unit root model performs considerably worse than the stable model

[21] Our estimate of the (heterogeneous) AR parameter distribution in this variant has them all bunched very close to unity. This re-produces the the result reported in Guvenen (2009) and Hryshko (2009) that closing down trend heterogeneity leads to an upwards bias in the AR parameter.

with a homogenous AR parameter. The unit root model does not fit ap's relating to the MA term and first observations very well even though we do allow for a flexible correlation for the first three observations (for details see appendix A.2).

Turning to the GF test we find that all the models have problems fitting the ap's for GF test. We find that all of the estimated stable models fit well the autocorrelations of earnings growth. This suggests these moments lack much power for discriminating between different models. Instead we find that the autocorrelations in level and transition probabilities are poorly fitted when homogeneity is imposed.

4.2 The implications of the preferred model.

We now turn to our main specification and discuss the results in more detail. While some of these results are very similar to those found for the PSID in Browning *et al* (2010)), there are also results where the sign and size of the parameters are particular to this sample. Estimation with pervasive heterogeneity on more samples is required before we can begin to draw strong general conclusions.

The magnitude of the heterogeneity found in the preferred model is displayed by the first, fifth and ninth deciles of the model parameters presented in the top panel of Table 5. The first feature is that the standard deviation of the earnings shock varies enormously across workers: the median is 0.068 and the 90 percent band is from 0.042 to 0.109. These differences will impact substantially on the cross-section distribution of risk.[22] Turning to the deterministic trend, the initial trends (the parameter α) are very dispersed with even the first decile having positive growth (albeit of only 4% per decade). We find a concave function for most of our sample with the median turning point at 41 years of experience.[23] There is considerable heterogeneity in the concavity with an the turning point ranging from 29 years to 50 years. There is considerable heterogeneity in the AR parameter; indicating a large variation in the persistence of the shocks. The median for the AR parameter is 0.63 with a 90 percent band of $[0.11, 0.96]$. This median is a bit lower than normally found in the literature but very close to the estimate in the model with a homogeneous AR parameter. Finally, the parameters of the translated hyperbolic sine imply that the skewness is 0.27 and the kurtosis is 4.93, suggesting that the distribution is a little skewed to right and with fatter tails than the normal.[24]

[Table 5 about here]

Turning to the co-dependence between parameters (the bottom panel of Table 5), we note that there are some very strong correlations. There is a very high negative correlation between the linear trend parameter, α, and the

[22] From equation (8), the implied standard deviation of the shock ξ_{it} has a median of 0.08 and a 90 percent band of $[0.05, 0.13]$.

[23] We have chosen to present the turning point rather than the coefficient on the quadratic trend τ, since it is easier to interpret.

[24] Lillard and Weiss (1979), Geweke and Keane (2000) and Guvenen *et al* (2012) find that the distribution of shocks are left skewed. The difference with our result is probably due to our sample selection, in which we have conditioned on workers having limited unemployment in any year.

turning point which implies that those who start with the highest growth peak earlier. There are also strong correlations between the variance and the trend parameters. Finally, the ARMA parameters $(\rho, \theta_1, \theta_2)$ all exhibit strong mutual dependence. Most of the correlations shown are of a 'significant' size which reemphasizes our concern for allowing for co-dependent heterogeneity.

Based on the estimated distribution of the model parameters (see Appendix Table C.5) we can present the implications of co-dependence for the set of illustrative issues mentioned in the introduction. For the first, we calculate the implied correlation between expected earnings growth in the first 10 years of the labour market career and the initial value. While this raw correlation is small there is a considerable correlation between the starting value and the trends in the beginning of the career working through the second term of equation (15). Taking this into account the total correlation is -0.7 which can be seen as support for the Becker-Mincer hypothesis on human capital formation. The second issue is whether the persistence of a shock and the variance are correlated; we find a positive correlation of 0.25 indicating that those who experience large variance shocks also have more persistent shocks. This implies that self-insurance will be relatively more costly for high variance workers. Finally we can examine if expected life time earnings are correlated with the short run variance. This correlation is found to be positive and large (0.55) which suggests that there is an initial trade-off between mean and variance.

5 Concluding remarks.

The contributions of this survey are fourfold. First, we derive some broad lessons on earnings dynamics based on the Danish sample. We emphasise again that our sample is a very particular sample (which is true of any sample drawn!) but it displays many of the same features seen for the sample from the PSID used in many previous studies. Our principal conclusion is that there is strong evidence for pervasive co-dependent heterogeneity. More specific conclusions we draw are that: there is a large cross-section variation in the short run variance of earnings; the $ARMA$ parameters display robust evidence of being heterogeneous; no one in this sample has a unit root; deterministic (quadratic) trends are heterogeneous; individual processes are stable but not stationary; the initial value y_1 is more dispersed than the steady state initial value, μ;
The specific conclusions may vary from sample to sample. This implies that for each sample an idiosyncratic analysis is required.

Our second contribution is to emphasise the usefulness of individual regression based (IRB) statistics which more effectively exploit the individual time series information. Without such information it is difficult to test between models with more or less heterogeneous parameters and/or more or less persistent processes. In the context of heterogeneity in trend parameters, Guvenen (2009) argues that long individual time series are needed but even in this case the suggested test based on autocorrelations does not seem to have much power against stable models. Furthermore, these tests only consider heterogeneity in the trends not in, for example, dynamics or variance. IRB statistics can be used to detect heterogeneity in all parameters and our empirical analyses suggest that these statistics are more powerful in rejecting limited heterogeneity. We also found, that transition probabilities (suggested in Geweke and Keane (2000)) are difficult to fit unless pervasive heterogeneity is included. Finally, the autocorrelation and partial autocorrelations in levels seem difficult to fit in models with limited heterogeneity.

Third, we reiterate the general point that if a parametric model is to have any credibility it must be checked against a wide range of discerning goodness of fit (GF) measures. Ideally, these GF tests should be tailored to plausible alternative hypotheses. We provide a candidate list of 86 statistics that any proposed model should fit. In the current context, we emphasize that if there are only weak tests for pervasive heterogeneity (for example, moments based on earnings growth), it is unlikely that researchers will find evidence of such heterogeneity.

Finally, we show that pervasive heterogeneity can have a major impact on econometric modelling and estimation procedures. Among the issues that arise when lots of heterogeneity is introduced are: the specification of initial conditions; the use of a first round regression to 'control' for the effect of observables and introducing time effects. Handling the initial condition problem, which in the context of an earnings process is important because the process actually starts at entry to the labour market, depends on the sort of heterogeneity in the model. We discuss how the usual estimation procedure of doing a first round regression is invalid unless the heterogeneity is present in a very limited way. Also, the commonly used approach of removing calender time effect by subtracting time averages invalidates conventional inference on the dynamic parameters, if there is heterogeneity in these parameters and we will not be able to recover even the average of the parameters of interest.

Future Issues.

1. Nonparametric identification for earnings models with pervasive heterogeneity.

2. Allowing for heterogeneous time effects with some individuals being more exposed to business cycle shocks than others.

3. How much of the latent heterogeneity would be absorbed by more explicit conditioning on observables? For example, quits and promotions and changes in marital status.

4. How good an approximation is the ARIMA model with pervasive heterogeneity for theory models with heterogeneity?

5. The error distribution and how all moments vary with time and experience.

6. Econometric methods for dynamic models with pervasive heterogeneity.

References

[1] Abowd, J. M. and D. Card. 1987. "Intertemporal Labor Supply and Long-Term Employment Contracts", *American Economic Review*, 77(1), pp. 50-68.

[2] Abowd, J. M. and D. Card. 1989. "On the Covariance Structure of Earnings and Hours Changes", *Econometrica*, 57(2), pp. 411-45.

[3] Albæk, K. and B. E. Sørensen. 1998. "Worker Flows and Job Flows in Danish Manufacturing, 1980-91", *Economic Journal*, 108(451), pp. 1750-71.

[4] Altonji, Joseph G.; Anthony Smith and Ivan Vidangos. 2009. "Modeling Earnings Dynamics", NBER WP 14743.

[5] Alvarez, Javier and Manuel Arellano. 2004. "Robust Likelihood Estimation of Dynamic Panel Data Models", mimeo, CEMFI, Madrid.

[6] Arellano, Manuel. 2003. *Panel Data Econometrics*. Oxford: Oxford University Press.

[7] Atkinson, Anthony; François Bourguignon and C. Morrisson. 1992. *Empirical Studies of Earnings Mobility*. Harwood Publishers.

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

[9] Baker, M. and G. Solon. 2003. "Earnings Dynamics and Inequality among Canadian Men, 1976-1992: Evidence from Longitudinal Income Tax Records", *Journal of Labor Economics*, 21(2), pp. 289-321.

[10] Bonhomme, S. and J. M. Robin. 2010. "Generalized Non-Parametric Deconvolution with an Application to Earnings Dynamics", *Review of Economic Studies*, 77(2), pp. 491-533.

[11] Browning, M.; M. Ejrnæs and J. Alvarez. 2010. "Modelling Income Processes with Lots of Heterogeneity", *Review of Economic Studies*, 77(4), pp. 1353-81.

[12] Chamberlain, G. (1980), "Analysis of Covariance with Qualitative Data", *Review of Economic Studies*, 47, 225-238

[13] Cunha, F.; J. Heckman and S. Navarro. 2005. "Separating Uncertainty from Heterogeneity in Life Cycle Earnings", *Oxford Economic Papers*, 57(2), pp. 191-261.

[14] Eckstein, Z. and G. J. van den Berg. 2007. "Empirical Labor Search: A Survey", *Journal of Econometrics*, 136(2), pp. 531-64.

[15] Ejrnæs, M. and M. Browning. 2012. "The Persistent-Transitory Representation for Earnings Processes", mimeo, Department of Economics, Oxford University.

[16] Friedman, Milton and Simon Kuznets. 1954. *Income from Independent Professional Practice New York*, National Bureau of Economic Research.

[17] Geweke, J. and M. Keane. 2000. "An Empirical Analysis of Earnings Dynamics among Men in the PSID: 1968-1989", *Journal of Econometrics*, 96(2), pp. 293-356.

[18] Gottschalk, P. and R. A. Moffitt 1994 "The growth of earnings instability in the U.S. Labor Market" *Brookings Papers on Economic Activity*, vol. 1994(2), pp. 217-272

[19] Gouriéroux, C.; P. C. B. Phillips and J. Yu. 2010. "Indirect Inference for Dynamic Panel Models", *Journal of Econometrics*, 157(1), pp. 68-77.

[20] Groes, Fane; Philipp Kircher and Iourii Manovskii. 2012. "The U-Shapes of Occupational Mobility", mimeo, Copenhagen Business School.

[21] Guvenen, F. 2007. "Learning Your Earning: Are Labor Income Shocks Really Very Persistent?", *American Economic Review*, 97(3), pp. 687-712.

[22] Guvenen, F., S. Ozkan and J. Song 2012. "The Nature of Countercyclical Income Risk", manuscript.

[23] Guvenen, F. 2009. "An Empirical Investigation of Labor Income Processes", *Review of Economic Dynamics*, 12(1), pp. 58-79.

[24] Haider, S. J. 2001. "Earnings Instability and Earnings Inequality of Males in the United States: 1967-1991", *Journal of Labor Economics*, 19(4), pp. 799-836.

[25] Hamilton, James. 1994. *Time Series Analysis*. Princeton: Princeton University Press.

[26] Hansen, Christian; James McDonald and Panayiotis Theodossiou. 2007 "Some Flexible Parametric Models for Partially Adaptive Estimators of Econometric Models", http://www.economics-ejournal.org/economics/journmalarticles/2007-7.

[27] Hause, J. C. 1980. "The Fine-Structure of Earnings and the on-the-Job Training Hypothesis", *Econometrica*, 48(4), pp. 1013-29.

[28] Heckman, J. J. 1981. "The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process and some Monte Carlo evidence". in *Structural Analysis of Discrete Data*, ed. D McFadden, C. Manski, pp. 179–95. Cambridge: MIT Press

[29] Heckman, James; Lance Lochner and Petra Todd. 2003. "Fifty Years of Mincer Earnings Regressions," NBER Working Paper 9732.

[30] Horowitz, J. L. and M. Markatou. 1996. "Semiparametric Estimation of Regression Models for Panel Data", *Review of Economic Studies*, 63(1), pp. 145-68.

[31] Hryshkoy, Dmytro 2012. "Rip to Hip: The Data Reject Heterogeneous Labor Income Profiles," mimeo, University of Alberta.

[32] Hsiao, Cheng. 1986. *Analysis of Panel Data*, Cambridge, UK: CUP.

[33] Huggett, M. and J. C. Parra. 2010. "How Well Does the Us Social Insurance System Provide Social Insurance?", *Journal of Political Economy*, 118(1), pp. 76-112.

[34] Im, K. S.; M. H. Pesaran and Y. Shin. 2003. "Testing for Unit Roots in Heterogeneous Panels", *Journal of Econometrics*, 115(1), pp. 53-74

[35] Kendall, M. G. 1954. "Note on bias in estimation of autocorrelation", *Biometrika*, 41(3-4), pp. 403-404.

[36] Kletzer, Lori G., 1998. "Job displacement". *Journal of Economic Perspectives*, 12 (1), pp. 115–136.

[37] Kniesner, Thomas J. and Qi Li. 2002. "Nonlinearity in Dynamic Adjustment: Semiparametric Estimation of Panel Labor Supply", *Empirical Economics*, 27(1), pp. 131-48.

[38] Levin, A.; C. F. Lin and C. S. J. Chu. 2002. "Unit Root Tests in Panel Data: Asymptotic and Finite-Sample Properties", *Journal of Econometrics*, 108(1), pp. 1-24.

[39] Lillard, L. A. and Y. Weiss. 1979. "Components of Variation in Panel Earnings Data - American Scientists 1960-70", *Econometrica*, 47(2), pp. 437-54.

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

[41] 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(1), pp. 83-114.

[42] MaCurdy, Thomas. 2007. "A Practitioner's Approach to Estimating Intertemporal Relationships Using Longitudinal Data: Lessons from Applica-

tions in Wage Dynamics", *Handbook of Econometrics*, ed. James Heckman and Edward Leamer, 4057-167. Elsevier.

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

[44] Meghir, Costas and Luigi Pistaferri. 2011. "Earnings, Consumption and Life Cycle Choices", *Handbook of Labor Economics*, ed. Orley Ashenfelter and David Card, 773-854. Elsevier.

[45] Meghir, Costas and Frank Windmeijer. 1999. "Moment Conditions for Dynamic Panel Data Models with Multiplicative Individual Effects in the Conditional Variance", *Annales d'Économie et de Statistique*, (55/56), pp. 317-30.

[46] Moffitt, R. A. and P. Gottschalk. 2002. "Trends in the Transitory Variance of Earnings in the United States", *Economic Journal*, 112(478), pp. C68-C73.

[47] Moffitt, R. A. and P. Gottschalk. 2012. "Trends in the Transitory Variance of Male Earnings, Methods and Evidence", *Journal of Human Resources*, 47(1), pp. 204-36.

[48] Murphy, K. M. and F. Welch. 1990. "Empirical Age-Earnings Profiles", *Journal of Labor Economics*, 8(2), pp. 202-29.

[49] Postel-Vinay, F. and J. M. Robin. 2002. "Equilibrium Wage Dispersion with Worker and Employer Heterogeneity", *Econometrica*, 70(6), pp. 2295-350.

[50] Postel-Vinay, F. and H. Turon. 2010. "On-the-Job Search, Productivity Shocks, and the Individual Earnings Process", *International Economic Review*, 51(3), pp. 599-629.

[51] Rogerson, R.; R. Shimer and R. Wright. 2005. "Search-Theoretic Models of the Labor Market: A Survey", *Journal of Economic Literature*, 43(4), pp. 959-88.

[52] Rubinstein, Yona and Yoram Weiss. 2006. "Post Schooling Wage Growth: Investment, Search and Learning", *Handbook of the Economics of Education*, ed. E. Hanushek and F. Welch, 1-67. Elsevier.

[53] Topel, R. 1991. "Specific Capital, Mobility, and Wages - Wages Rise with Job Seniority", *Journal of Political Economy*, 99(1), pp. 145-76.

[54] Topel, R. H. and M. P. Ward. 1992. "Job Mobility and the Careers of Young Men", *Quarterly Journal of Economics*, 107(2), pp. 439-79.

[55] Wansbeek, Tom J. and Thijs Knaap. 1999. "Estimating a Dynamic Panel Data Model with Heterogenous Trends", *Annales d'Économie et de Statistique*, (55/56), pp. 331-49.

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

Brief explanations for the annotated references.

- Abowd and Card (1989): “A seminal paper on earnings dynamics using the permanent-transitory model”
- Baker (1997): “An perceptive challenge to the then emerging consensus that everyone has a unit root”
- Browning *et al* (2010): ”Introduces and finds pervasive heterogeneity in the earnings process”
- Geweke and Keane (2000): ”The use of transition probabilities as a basis for estimation”
- Guvenen (2009): “An in-depth study of the implications of heterogeneity for auto-covariance structures”
- Hryshko (2012): “Shows the identification of the HIP and RIP models from auto-covariance structures”
- Lillard and Weiss (1979): “The first study that discusses implications of the Becker-Mincer for earnings processes”
- MaCurdy (2007): “The best overview of the estimation of earnings dynamics with limited heterogeneity”
- Moffitt and Gottschalk (2012): “The latest study on time varying variances in the earnings process”

- Rubinstein and Weiss (2006): "A thorough survey of the theory implications for empirical facts on earnings process"

Tables and Figures.

	Trend	
	Homogenous	Heterogenous
Stable	$\rho < 1$, $\alpha_i = \alpha$	$\rho < 1$ (HIP)
Unit root	$\rho = 1$, $\alpha_i = \alpha$ (RIP)	$\rho = 1$

Table 1: Hypotheses on stability and heterogeneity

	$\rho = 1$	$\rho < 1$	$\rho < 1$
Heterogeneity	*unit root*	*stable*	*stationary*
No heterogeneity $\alpha_i = \alpha$	increasing linear	monotone	constant
α_i (not correlated with intercept μ_i)	increasing	non-monotone	increasing
α_i (correlated with intercept μ_i)	U-shaped	non-monotone	U-shaped

Table 2: The cross sectional variance

Order	coeff	(p-value)	Order	coeff	(p-value)
1	-0.257	(0.00)	9	-0.005	(0.35)
2	-0.024	(0.00)	10	0.005	(0.36)
3	-0.022	(0.00)	11	-0.001	(0.30)
4	-0.020	(0.00)	12	0.017	(0.00)
5	-0.018	(0.00)	13	0.002	(0.76)
6	-0.000	(0.99)	14	-0.006	(0.33)
7	0.000	(0.98)	15	-0.012	(0.03)
8	-0.001	(0.20)			

Table 3: The autocorrelation for earnings growth in the samples

	Stable models				Unit root
	Preferred model	Homogeneous AR	Homogeneous trend	No co-dependence	
No parameters	41	34	28	20	29
df	14	21	27	35	26
OI test statistic	31.0	145.8	226.9	456.4	287.5
Goodness of fit (χ^2 (31))	90.7	205.3	164.5	226.4	301.7

Table 4: Specification tests

	y_1	ν	μ	$\alpha \times 10$	tp^*	ρ	θ_1	θ_2
Marginal distributions								
1st decile	12.21	0.042	12.38	0.04	29.0	0.11	-0.309	0.011
Median	12.43	0.068	12.45	0.32	41.2	0.63	0.021	0.100
9th decile	12.67	0.109	12.52	0.58	50.3	0.96	0.347	0.188
Correlations								
corr(y_1)	1	0.26	0.15	-0.04	-0.16	0.27	-0.31	-0.70
corr(ν)		1	0.13	0.53	-0.74	0.25	-0.32	-0.38
corr(μ)		–	1	-0.70	0.25	0.17	-0.21	-0.47
corr(α)		–	–	1	-0.84	0.22	-0.09	-0.03
corr(tp)		–	–	–	1	-0.52	0.27	0.44
corr(ρ)		–	–	–	–	1	-0.43	-0.23
corr(θ_1)		–	–	–	–	–	1	0.54

* tp is defined as the turning point of the quadratic trend function in years

Table 5: The parameter distribution

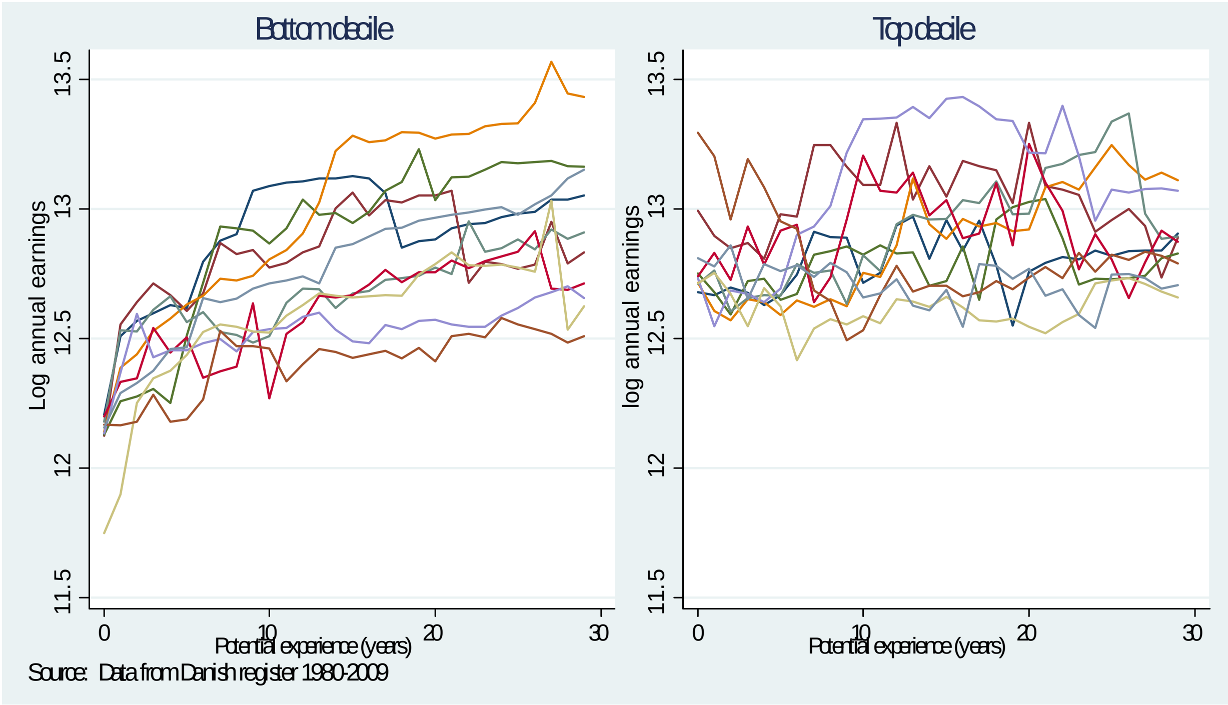


Figure 1: Individual earnings path

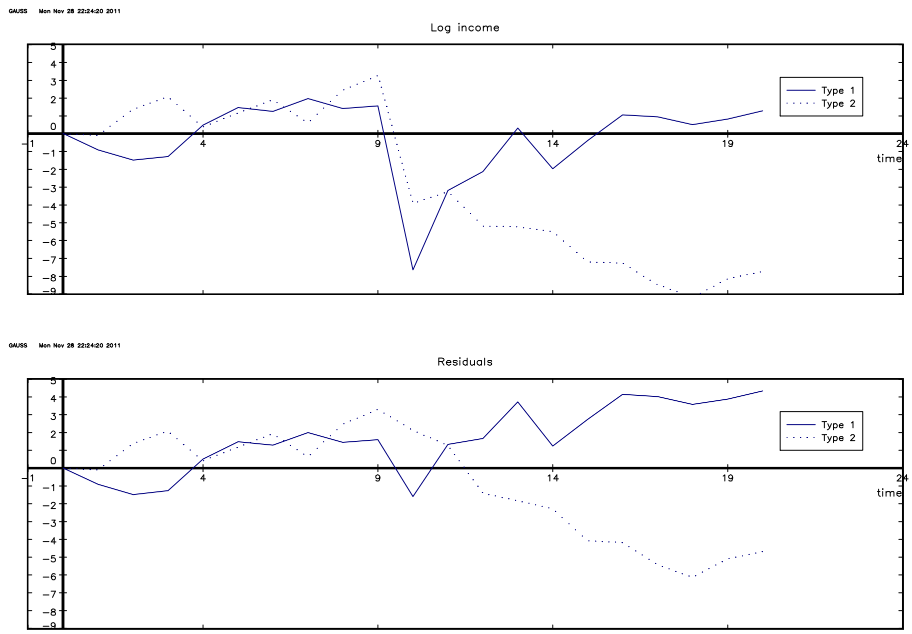


Figure 2: The impact of a first round regression

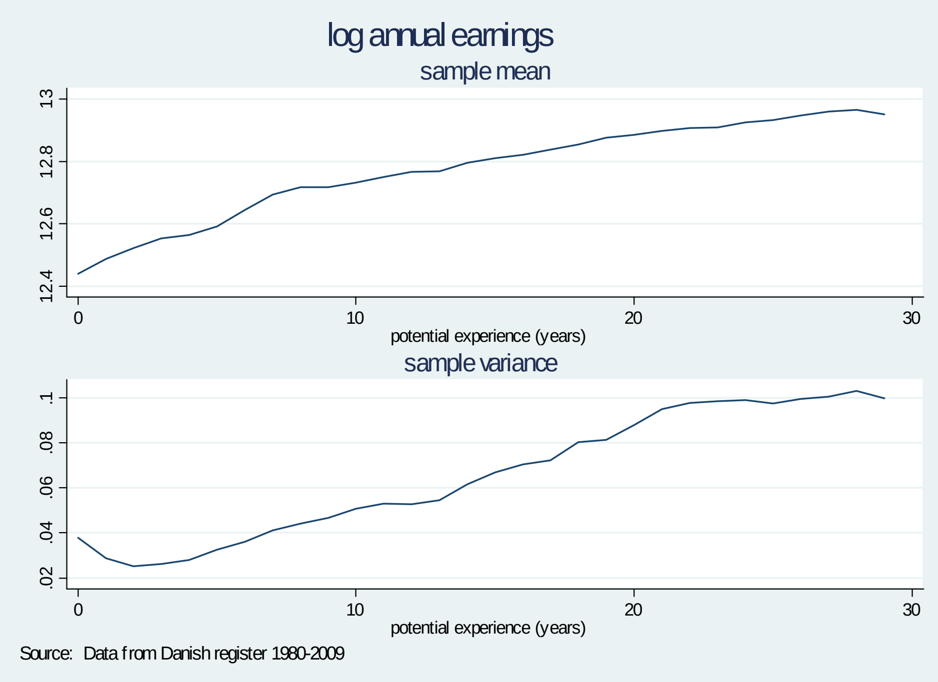


Figure 3: The sample mean and variance of log annual earnings

A Technical appendix (Online-only material).

A.1 Proofs for the implications of a first round regression.

The following model is considered:

$$\begin{aligned} \text{Type 1} &: \quad y_{it} = \rho_1 y_{it-1} + \omega_t + \xi_{it}, \rho_1 < 1 \\ \text{Type 2} &: \quad y_{it} = \rho_2 y_{it-1} + \omega_t + \xi_{it}, \rho_2 < 1, \end{aligned}$$

where we assume that the variance of the shocks are: $V(\omega_t) = \sigma_\omega^2$ and $V(\xi_{it}) = \sigma_\xi^2$. The initial value is given such that we have a stationary process with mean $E(y_{i1}) = 0$ and the variances are:

$$\begin{aligned} \text{Type 1:} \qquad V(y_{i1}) &= \frac{\sigma_\omega^2 + \sigma_\xi^2}{(1-\rho_1^2)} \\ \text{Type 1:} \qquad V(y_{i1}) &= \frac{\sigma_\omega^2 + \sigma_\xi^2}{(1-\rho_2^2)} \end{aligned}$$

In the first round regression, the common time specific mean is removed. The time specific mean is given by:

$$\bar{y}_t = \pi \bar{y}_t^1 + (1-\pi)\bar{y}_t^2.$$

where π is the fraction of type 1. We assume we have a balanced sample of N individuals. The residuals from a first round regression for an individual of type 1 and type 2 are given by:

$$\begin{aligned} \text{Type 1} &: \quad y_{it} - \bar{y}_t = \rho_1(y_{it-1} - \bar{y}_{t-1}) + (\xi_{it} - \bar{\xi}_t) + (1-\pi)(\rho_1 - \rho_2)\bar{y}_{t-1}^2 \\ \text{Type 2} &: \quad y_{it} - \bar{y}_t = \rho_2(y_{it-1} - \bar{y}_{t-1}) + (\xi_{it} - \bar{\xi}_t) + \pi(\rho_2 - \rho_1)\bar{y}_{t-1}^2 \end{aligned}$$

First we notice that the model is no longer an $AR(1)$ model. We can also show that the first order autocorrelation is no longer equal to the autoregressive parameters. The first order autocorrelation for the residuals is defined

$$\hat{\rho} = \frac{Cov((y_{it} - \bar{y}_t)(y_{it-1} - \bar{y}_{t-1}))}{Var(y_{it-1} - \bar{y}_{t-1})}.$$

We can calculate this for each of the two types. For type 1 it is given by.:

$$\begin{aligned}
\hat{\rho}_1 &= \rho_1 + (1-\pi)(\rho_1 - \rho_2)\frac{Cov((y_{it-1} - \bar{y}_{t-1}), \bar{y}^2_{t-1})}{Var(y_{it-1} - \bar{y}_{t-1})} \\
&= \rho_1 + (1-\pi)(\rho_1 - \rho_2)\frac{Cov(y_{it-1}, \bar{y}^2_{t-1}) - \pi Cov(\bar{y}^1_{t-1}, \bar{y}^2_{t-1}) - (1-\pi)Var(\bar{y}^2_{t-1})}{Var(y_{it-1} - \bar{y}_{t-1})} \\
&= \rho_1 + (1-\pi)(\rho_1 - \rho_2)\frac{\frac{\sigma^2_w}{1-\rho_1\rho_2} - \pi\frac{\sigma^2_w}{1-\rho_1\rho_2} - (1-\pi)\frac{\sigma^2_w + \sigma^2_\xi/((1-\pi)*N)}{1-\rho_2^2}}{Var(y_{it-1} - \bar{y}_{t-1})} \\
&= \rho_1 + (1-\pi)^2(\rho_1 - \rho_2)\frac{\frac{\sigma^2_w}{1-\rho_1\rho_2} - \frac{\sigma^2_w + \sigma^2_\xi/((1-\pi)*N)}{1-\rho_2^2}}{Var(y_{it-1} - \bar{y}_{t-1})}
\end{aligned}$$

where $Var(y_{it-1} - \bar{y}_{t-1})$ is the variance of the residuals for a type 1 person. If $\rho_2 > \rho_1$ the bias will always be positive. If $\rho_1 > \rho_2$ the bias can be both negative, positive and zero.[25] We can also calculate the asymptotic bias for ρ_2

$$\hat{\rho}_2 = \rho_2 + (\pi)^2(\rho_2 - \rho_1)\frac{\frac{\sigma^2_w}{1-\rho_1\rho_2} - \frac{\sigma^2_w + \sigma^2_\xi/(\pi*N)}{1-\rho_1^2}}{Var(y_{it-1} - \bar{y}_{t-1})}$$

where $Var(y_{it-1} - \bar{y}_{t-1})$ is the variance of the residuals for a type 2 person. If $\rho_1 > \rho_2$ the bias will always be positive This indicates that the first round regression will always induce an upward bias on the lowest AR parameter, and in many cases give an upward bias for both estimates.

A.2 Appendix: Initial condition.

Stable ARMA process.

In this subsection we describe how one can model initial conditions in a stable ARMA(1,2) model ($\rho_i < 1$). Abstracting from the deterministic part, the model

[25] The bias will be negative if $\sigma^2_\omega = 0$.

is given by:

$$
\begin{aligned}
y_{i1} &= c_0\xi_{i1} + c_{i1}\xi_{i0} + c_{i2}\xi_{i-1} \\
y_{it} &= \rho_i y_{it-1} + \xi_{it} + \theta_{i1}\xi_{it-1} + \theta_{i2}\xi_{it-2} \text{ for } t = 2, 3, .., T
\end{aligned}
$$

For a stationary process the auto-covariance functions are time invariant and are given by:

$$
\begin{aligned}
\gamma_{i0} &= \frac{1 + \theta_{i1}^2 + \theta_{i2}^2 + 2\rho_i^2\theta_{i2} + 2\rho_i\theta_{i1}(1 + \theta_{i2})}{1 - \rho_i^2} \\
\gamma_{i1} &= \rho_i\gamma_{i0} + \theta_{i1} + \theta_{i2}(\theta_{i2} + \rho_i\theta_{i1}) \\
\gamma_{i2} &= \rho_i\gamma_{i1} + \theta_{i2} \\
\gamma_{it} &= \rho_i\gamma_{it-1} \text{ for } t > 2
\end{aligned}
$$

We can determine the values of $\{c, c_{i1}, c_{i2}\}$ such that the process has constant variance and covariances from the three equations:

$$
\begin{aligned}
V(y_{i1}) &= V(y_{i2}) = \gamma_{i0} \\
cov(y_{i1}, y_{i2}) &= cov(y_{i2}, y_{i3}) = \gamma_{i1} \\
cov(y_{i1}, y_{i3}) &= cov(y_{i2}, y_{i4}) = \gamma_{i2}
\end{aligned}
$$

This gives:

$$
\begin{aligned}
c_0 &= 1 \\
c_{i1} &= \rho_i + \theta_{i1} \\
c_{i2} &= \frac{\rho_i^2 + \rho_i\theta_{i1} + \theta_{i2}}{\sqrt{1 - \rho^2}}
\end{aligned}
$$

Unit root model.

In this subsection we describe how one can model initial conditions in a unit root model with a MA(2) error. The idea is to model the first observation y_{i1}, which is normally not done because most studies are working with first differences. However, it is still an issue how to start a process which contains a MA component. When dealing with a unit root model different issues than in the stationary case arises. The model is given by (when abstracting from the deterministic part):

$$\begin{aligned} y_{i1} &= m_1 + c_{i0}\xi_{i1} + c_{i1}\xi_0 + c_2\xi_{i-1} \\ y_t &= y_{it-1} + \xi_{it} + \theta_{i1}\xi_{it-1} + \theta_{i2}\xi_{it-2} \text{ for } t = 2, 3, .., T, \end{aligned}$$

where the process depends on $\xi_{it}, ..., \xi_{i1}$ as well as two additional error terms: ξ_{i0} and $\xi_{i,-1}$, which are all independent. If the parameters c_{i0} and c_{i1} vary freely the variance and covariance will not necessarily evolve like a linear trend for the first three observations. If we want to impose a time invariant correlation it will imply restrictions on c_{i1} and c_{i2}. In a unit root model we have:

$$\begin{aligned} Cov(\Delta y_t, y_{t-1}) &= (\theta_{i1} + \theta_{i2}(1 + \theta_{i1}))\nu_i \;\text{ for } t = 2, 3, .., T \\ Cov(\Delta y_t, y_{t-2}) &= \theta_{i2}\nu_i \;\text{ for } t = 2, 3, .., T \end{aligned}$$

Similarly we can get for the first observation:

$$
\begin{aligned}
E(y_{i1}) &= m_1 \\
V(y_{i1}) &= (c_{i0}^2 + c_{i1}^2 + c_2^2)\nu_i \\
Cov(\Delta y_{i2}, y_{i1}) &= (c_{i0}\theta_{i1} + c_{i1}\theta_{i2})\nu_i \\
Cov(\Delta y_{i3}, y_{i1}) &= (c_{i0}\theta_{i2})\nu_i
\end{aligned}
$$

If we impose that $c_{i0} = 1$ and $c_{i1} = \theta_{i1} + 1$ we will have that the covariances $Cov(\Delta y_{it}, y_{it-1})$ and $Cov(\Delta y_{it}, y_{it-2})$ are constant over time and variances and covariances will evolve as linear functions. We also see that m_1 determines the mean of y_{it} and c_2 will be determined by the variance of y_{i1}.

B Data appendix (Online-only material).

B.1 The sample selection.

In this paper we use the Danish administrative data to illustrate the estimation of earnings processes. Our core sample is a narrowly defined sample, which is selected to make it as homogenous as possible. In this appendix we will discuss the implication of our sampling strategy. We focus on men born in 1958 with vocational training. We select those who are Danish citizens, lived in Denmark in the period 1980-2009, who have not retired, worked as self-employed or being a full-time student within the period. Furthermore, they should always have annual earnings above 50.000 Danish Crowns,[26] and being employed for at least 80% in any year (the impact of the selection criteria are shown in the Table (B.1)). Our core sample is a balanced sample. We also construct three

[26] measured in 2009 prices.

comparison samples to illustrate how sensitive our results are to the sample selection. The three samples are: a sample of older workers (born in 1952), an unbalanced sample and a sample where we allow for more unemployment. The sample selection for these samples are shown in table (B.1).

The sample we use is a highly selected sample but as we will show it is not pathological and it exhibits the same features as the samples based on PSID.

B.2 Standard descriptions of the data.

In this subsection we present descriptions of the data, which are considered to be important features of the data in studies of earnings dynamics. These descriptions include mean, variances and autocorrelations of earnings and especially earnings growth. We will also provide a comparison between the Danish data and the PSID. In Figure (B.1) and (B.2) we show the averages and standard deviation of the log earnings (deflated by the consumer price index) for the four samples. Figure (B.1) shows that the average log earnings are increasing over time with a largest increment at earlier ages while when the sample reaches 50 the earnings are almost constant. The figure also displays that "more unemployed sample" and "the unbalanced sample" have on average lower earnings. Looking at the graphs we see very small business cycle effects even for the more unemployed sample. The business cycle effects are on the other hand very visible when we consider the cross sectional standard deviation for the "more unemployed sample", where the spikes in the variance coincide with high level of unemployment (in 1981, 1993 and 2009)(see Figure (B.2)). For the core sample and the older sample where we condition on high degree of employment in all

Sample	Core	Older	Unbalanced	Unemp.
Cohort	1958	1952	1958	1958
Period	1980-2009	1980-2009	1980-2009	1980-2009
Balanced/Unbalanced	balanced	balanced	unbalanced	balanced
Ages	22-51	28-57	22-51	22-51
Men with vocational training at 35	15,550	17,992	15,550	15,550
- Only Danish citizens	15,019	17,691	15,019	15,019
- Living in DK in all 30 years	13,533	15,626	13,533	13,533
- Never retired	12,887	14,204	-	12,887
- Never self-employed	9,450	10,281	-	9,450
- Not full-time student in the period	7,335	9,627	-	7,335
- Always earnings above 50k DK[1]	5,685	7,644	-	5,685
- Fulltime (more than 80%) all years	2,122	3,793	-	-

[1] Measured in 2009 prices

Table B.1: Sample selection of core sample and alternative samples

years, we do not see the business cycle effects. We have chosen our core sample such that business cycles effects are almost not present and we can ignore this aspect when modelling the earnings process. This is not because we find it unimportant, but in this survey we want to focus on the dynamics of the process.

A comparison of the cross sectional variance between our sample and the PSID shows that the variance in our sample is in general lower (due to a more homogenous sample) but the development of the variance shows the same features as found in the PSID sample (see Guvenen (2009) and Rubinstein and Weiss (2006)). We find that the cross sectional variances start by falling and reaches its minimum around the age of 25. Rubinstein and Weiss (2006) and Guvenen (2009) find the same fall in the variance in the beginning of the career, although both papers find that it occurs a bit later than we find. Again we think the difference can be attributed to the fact that we work with a more

Figure B.1: The mean earnings for the four samples

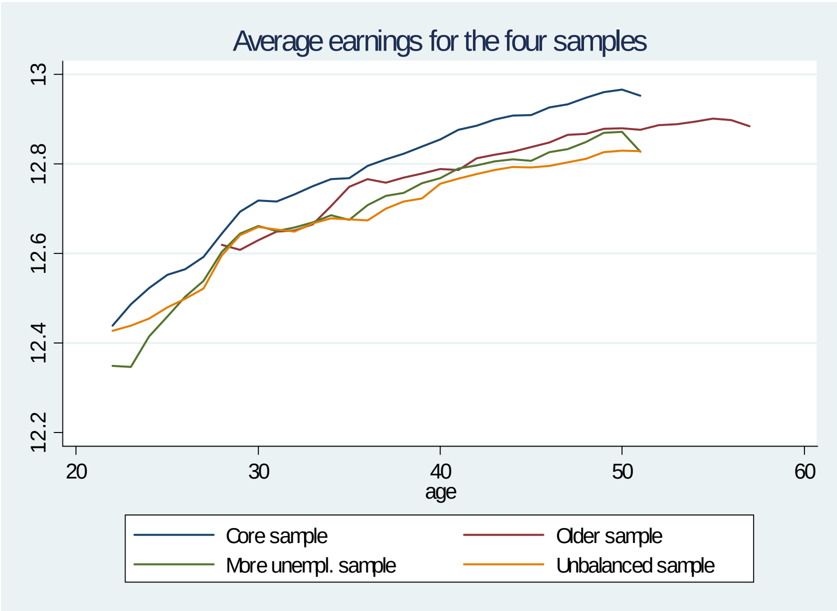
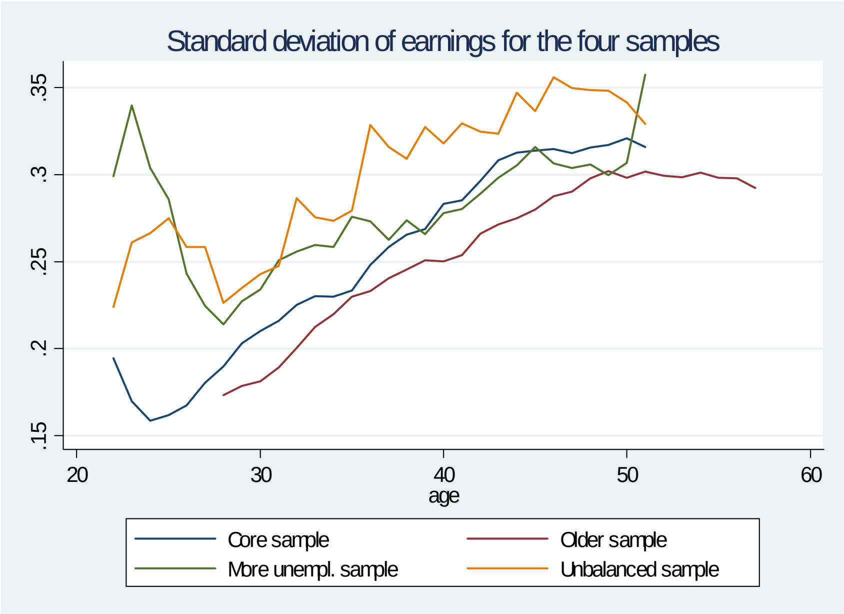


Figure B.2: The standard deviation of earnings for the four samples



homogenous sample in terms of education. We also find that the cross sectional variance flattens out after 25 years of (potential) experience around the age of 47. This is in line with what Rubinstein and Weiss (2006) finds but in contrast to Guvenen (2009) that finds that the cross sectional variance still increases after 30 years of experience.

To display the dynamics of the process we show the first 15 autocorrelation for the process in levels and first differences in Table (B.2). The table shows that the autocorrelation in levels is decreasing from around 0.93 for the first order autocorrelation to about 0.52 for 15. order autocorrelation. Compared to Baker (1997) the autocorrelation in our sample is higher. When we compare the autocorrelation function for the first differences, these estimates are significant up to lag 5. Compared to the other studies, we find higher order significant autocorrelations in earnings growth than most of the literature (Above and Card (1989) and Meghir and Pistaferri (2004) find that only 2. order autocorrelation in earnings growth,[27] Hryshko (2012) finds that 3. order autocorrelation and Guvenen (2009) finds up to 4. order autocorrelation is significant). We also find higher order (>10) correlations are significant. This is not found in the PSID (Guvenen 2009), but may be due to very small sample size. When we compare the size of the autocorrelation in first difference they are not bigger that in the PSID. For the first order autocorrelation we find a somewhat numerical lower estimate. This is consistent with the PSID containing more measurement error than the Danish data. Finally, we reproduce the graph from Guvenen (2009)

[27] Meghir and Pistaferri (2004) finds that 3. order is on the borderline depending on the explicit sample and test procedure.

Order	Core sample levels	1.diff	Older sample 1. diff	Unbalanced 1. diff	More unemployed 1. diff	PSID Guvenen[1] 1. diff	M&P[2] 1. diff
1	0.939*	-0.257*	-0.228*	-0.227*	-0.269*	-0.317*	-0.278*
2	0.909*	-0.024*	-0.040*	-0.038*	-0.068*	-0.026*	-0.072*
3	0.882*	-0.022*	-0.014*	-0.004	-0.017*	-0.019*	-0.022*
4	0.857*	-0.020*	-0.030*	-0.019*	-0.021*	-0.021*	0.006
5	0.833*	-0.017*	-0.012*	-0.009	-0.010*	-0.001	
6	0.811*	0.000	0.009*	0.001	-0.003	-0.002	
7	0.787*	0.000	-0.019*	-0.002	-0.002	0.004	
8	0.762*	-0.007	-0.004	-0.007	-0.009*	0.004	
9	0.736*	-0.005	-0.009	-0.004	-0.005	-0.009	
10	0.710*	0.005	-0.004	0.0214*	0.002	-0.014	
11	0.681*	-0.006	-0.005	-0.016*	-0.000	0.010	
12	0.648*	0.017*	0.005	0.017*	0.006	-0.008	
13	0.609*	0.002	-0.001	0.002	-0.003	0.005	
14	0.566*	-0.006	-0.009*	0.004	-0.010*	-0.019	
15	0.520*	-0.012*	-0.012*	-0.009	0.004	0.044	

[1]Guvenen (2009), see Table 3

[2]Meghir and Pistaferri (2004), calculated on the basis of table 1

Table B.2: The autocorrelation for the samples

Figure 6 that shows the autocovariance as a function of potential experience (see Figure (B.3)). Here we get a very similar picture to what Guvenen (2009) finds for the high school sample.

When making the comparison of the dynamics across the different Danish samples we see some differences. This indicates that the sample selection might also affect the results. Similar differences is found when different sample of PSID are compared (see the last two column of the table).

Rubinstein and Weiss (2006) point out that most theory models have clear predictions on the correlation between earnings growth and the prior level. Furthermore, the pattern of the correlation over the life cycle can be used to distinguish between the types of models. This pattern is documented for US data in

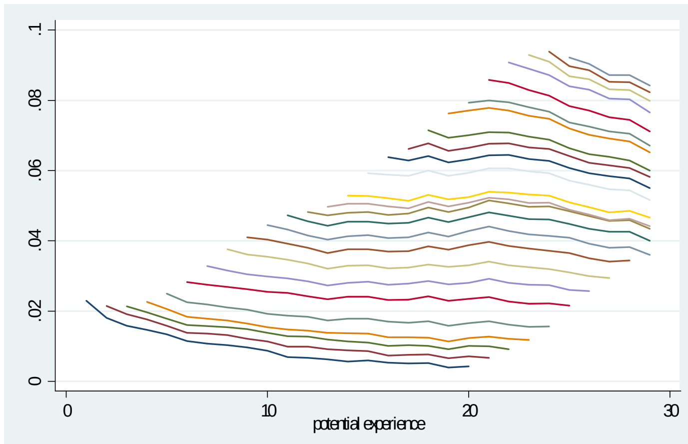


Figure B.3: Covariance structure of log earnings (similar to Guvenen (2009) figure 6)

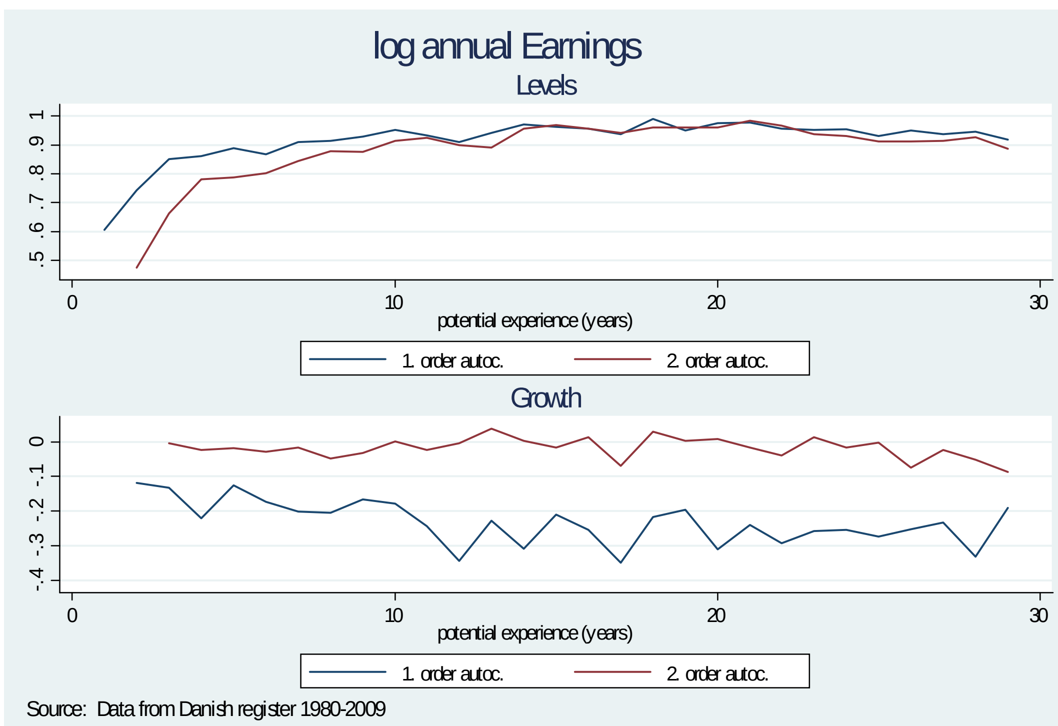


Figure B.4: The autocorrelation function

Rubinstein and Weiss (2006), Figure 8a-8e. Below the we show the regression coefficient of Δy_t on y_{t-4} for different levels of potential experience.[28] The regression coefficient and confidence interval is shown in Figure (B.5). The figure shows that the correlation is mainly negative and but becomes insignificant for potential experience in the range of 10-12,13-15 and 16-18 years of experience. The same pattern of the correlation displays an inverted U-shape. The pattern is in general not consistent with a RIP type of model. The pattern we find for the Danish data are very comparable to what is found for US data (although Rubinstein and Weiss (2006) do not estimate the correlation for experience above 18 years).

B.3 Heterogeneity in the data.

In this subsection we provide additional descriptions of the data, which highlight heterogeneity in the data. We illustrate co-dependent heterogeneity by considering how the process develops conditioned on the first observations; this complements Figure 1. In the Figure (B.6), we display the average earnings conditioned on the position in the distribution based on the mean values from 1980 to 1982. The graph show that those who start in the bottom of the distribution on average have a steeper trend in earnings than those who start in the middle or top of the distribution. This could indicate that the trend and the intercept might be negatively correlated or a stable process with a initial values far from the stationary conditions. We also see a tendency to convergence in the level.

[28] We use y_{t-4} to avoid that measurement error or the dynamics of transitory shocks can affect this correlation.

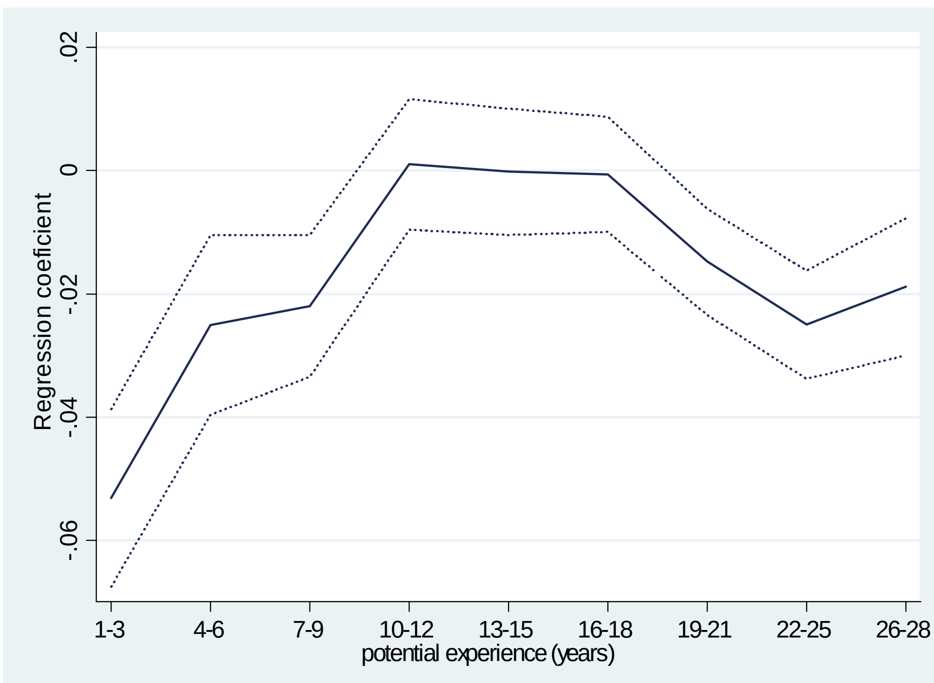
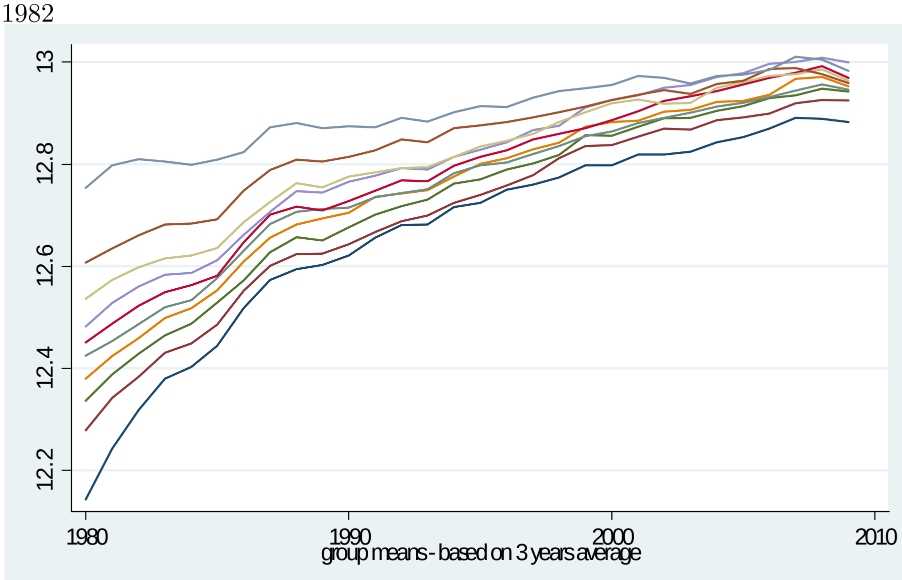


Figure B.5: Regression coefficient for Δy_{it} on y_{it-4}

Figure B.6: Earnings path conditioned on the position in the distribution 1980-1982



B.4 Alternative descriptions.

In this subsection we provide additional descriptions of the data. An alternative way of displays the dynamics of the process is by examining the partial autocorrelations. The partial autocorrelations are function of autocovariances and but as pointed out in MaCurdy (1982) the partial autocorrelation is very useful to detect the AR part of the process. In Figure (B.7) we display the partial autocorrelation for earnings growth together with the confidence interval. The figure shows that the coefficients are negative and significant up to lag 8 and lag 12 is again significant but positive. These features are not consistent with a standard unit root model with no heterogeneity.

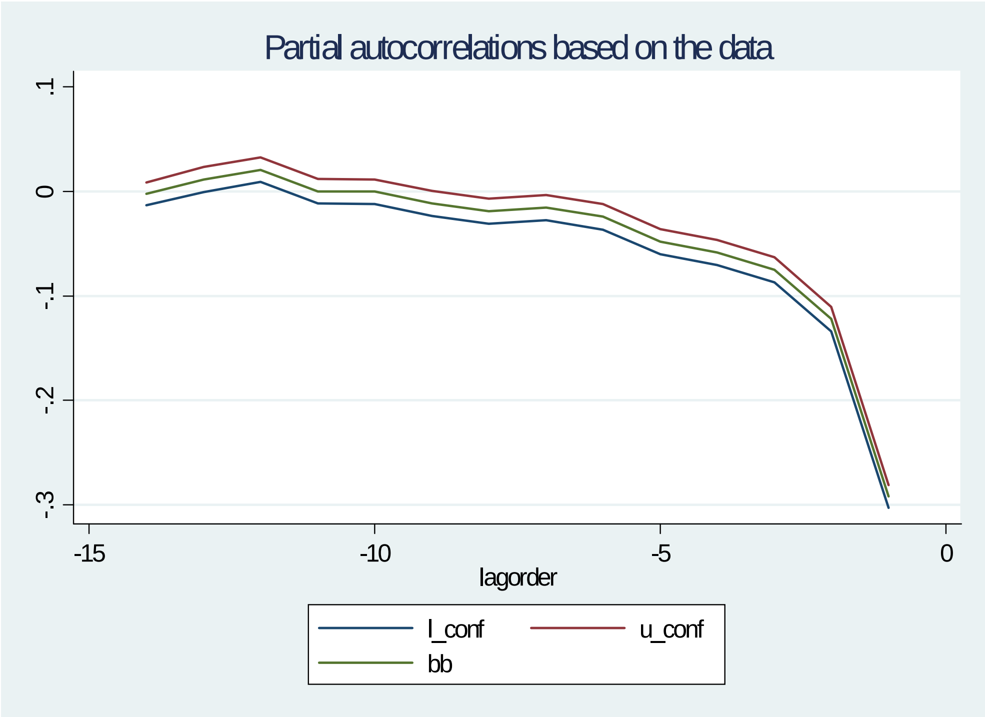


Figure B.7: Partiel autocorrelation function for earnings growth

C Estimation results appendix (Online-only material).

C.1 Simulation procedure: Stable process.

We simulate a stationary process with heterogeneous quadratic trends. Each unit has seven free parameters: ν_i, ρ_i, θ_{1i}, θ_{2i}, the mean μ_i and the trend parameters α_i and τ_i.

1. Simulate $\{\mu_i, \nu_i, \alpha_i, \tau_i, \rho_i, \theta_{1i}, \theta_{2i}\}_{i=1,..H}$ from the assumed joint distribution (see section 2.4). This includes support constraints such as $\nu_i \geq 0$, $\rho_i \in (-1, 1)$ and, perhaps, $\theta_{ik} \in (-1, 1)$ for $k = 1, 2$. It may also involve constraints on the trend parameters so that the turning points for the quadratic trend are 'reasonable'.

2. Draw $u_{it} \sim N(0, 1)$, $i = 1, 2, ..H$ and $t = -1, 0, 1, ..T$. Construct $\xi_{it} = \nu_i u_{it}$.[29]

3. When allowing for nonstationary initial conditions we could choose an unrestricted joint distribution for heterogeneous $\{m_1, c_0, c_{i1}, c_{i2}\}$. However this adds very many new distribution parameters. Instead we take scalars m_1, c_0, $\tilde{c}_1$ and $\tilde{c}_2$ and set:

$$\begin{aligned} c_{i1} &= \tilde{c}_1 (\rho_i + \theta_{i1}) \\ c_{i2} &= \tilde{c}_2 \left(\frac{(\rho_i)^2 + \rho_i \theta_{i1} + \theta_{i2}}{\sqrt{1 - (\rho_i)^2}} \right) \end{aligned}$$

   Then the test for the initial conditions having the stationary distribution is simply a test of $m_1 = 0$ and $c_0 = \tilde{c}_1 = \tilde{c}_2 = 1$.

[29] In practice we take the hyperbolic sine transformation to give some skewness and fat tails.

4. Construct initial conditions (see equation (13):

$$y_{i1} = \mu_i + m_1 + c_0\xi_{i1} + c_{i1}\xi_{i0} + c_{i2}\xi_{i,-1} + \psi_{81}\eta_{i1} + \psi_{83}\eta_{i3} + \psi_{84}\eta_{i4}$$

   Note that the simulated initial values have two sources of variation: one due to the heterogeneity in the means and the other due to the stochastic variation.

5. Construct the subsequent values recursively for $t = 2, ..T$:

$$\begin{aligned} y_{it} &= [(1-\rho_i)\,\mu_i + \rho_i(\alpha_i - \tau_i)] + \rho_i y_{i,t-1} \\ &\quad + [\alpha_i\,(1-\rho_i) + 2\rho_i\tau_i]\,(t-1) + \tau_i\,(1-\rho_i)\,(t-1)^2 \\ &\quad + \xi_{it} + \theta_{i1}\xi_{h,t-1} + \theta_{i2}\xi_{i,t-1} \end{aligned}$$

C.2 Estimation results.

ap name	Data	Preferred	Hom. AR	Hom. trend	No co-dep.	Unit root
M(Y1)	12.439	12.440	12.450	12.452	12.454	12.461
	(0.004)	[-0.195]	[-1.831]	[-2.054]	[-2.411]	[-3.583]
M(IN)	12.462	12.457	12.461	12.479	12.471	12.461
	(0.004)	[0.796]	[0.086]	[-2.950]	[-1.560]	[0.103]
M(AR)	0.349	0.347	0.352	0.350	0.365	0.321
	(0.005)	[0.220]	[-0.373]	[-0.141]	[-1.959]	[3.566]
M(TR)	0.311	0.320	0.323	0.281	0.270	0.309
	(0.007)	[-0.870]	[-1.192]	[3.061]	[4.165]	[0.230]
M(QU)	-0.049	-0.051	-0.053	-0.042	-0.039	-0.046
	(0.002)	[0.688]	[1.523]	[-2.445]	[-3.401]	[-1.004]
M(LV)	-0.266	-0.265	-0.264	-0.265	-0.264	-0.267
	(0.001)	[-0.539]	[-1.159]	[-0.412]	[-1.124]	[0.965]
M(AU1)	0.089	0.087	0.077	0.065	0.068	0.058
	(0.003)	[0.525]	[2.634]	[5.412]	[4.740]	[6.879]
M(AU2)	-0.011	-0.013	-0.007	-0.012	-0.002	-0.034
	(0.004)	[0.397]	[-0.526]	[0.285]	[-1.470]	[3.710]
M(SKW)	0.226	0.212	0.193	0.217	0.240	0.200
	(0.018)	[0.504]	[1.204]	[0.319]	[-0.517]	[0.935]
M(KUR)	0.948	0.973	1.026	1.013	0.949	0.981
	(0.041)	[-0.398]	[-1.267]	[-1.048]	[-0.007]	[-0.535]
S(Y1)	0.195	0.194	0.202	0.187	0.205	0.166
	(0.004)	[0.155]	[-1.103]	[1.234]	[-1.586]	[4.568]
S(IN)	0.171	0.171	0.174	0.180	0.183	0.175
	(0.003)	[0.001]	[-0.605]	[-1.844]	[-2.348]	[-0.796]
S(AR)	0.239	0.240	0.232	0.250	0.249	0.247
	(0.004)	[-0.180]	[1.294]	[-1.954]	[-1.772]	[-1.436]
S(TR)	0.311	0.301	0.284	0.299	0.273	0.323
	(0.006)	[1.107]	[2.986]	[1.359]	[4.180]	[-1.274]
S(QU)	0.088	0.085	0.078	0.090	0.079	0.089
	(0.002)	[1.177]	[3.599]	[-0.488]	[3.462]	[-0.255]
S(LV)	0.043	0.043	0.043	0.042	0.038	0.043
	(0.001)	[-0.025]	[0.256]	[1.095]	[5.015]	[-0.069]
S(AU1)	0.138	0.130	0.127	0.121	0.120	0.112
	(0.003)	[1.595]	[2.289]	[3.446]	[3.593]	[5.282]
S(AU2)	0.193	0.195	0.184	0.199	0.186	0.200
	(0.003)	[-0.266]	[1.941]	[-1.142]	[1.528]	[-1.374]
S(SKW)	0.837	0.843	0.857	0.849	0.827	0.831
	(0.016)	[-0.260]	[-0.838]	[-0.512]	[0.409]	[0.243]
S(KUR)	1.936	2.002	2.050	2.039	2.003	1.991
	(0.072)	[-0.617]	[-1.061]	[-0.953]	[-0.621]	[-0.508]

Values in (.) are standard deviations, Values in [.] are t-values

Table C.3: Auxiliary parameters for estimation

Table C.3 *Continued*

ap name	Data	Preferred	Hom. AR	Hom. trend	No co-dep.	Unit root
C(Y1IN)	7.530	7.806	8.390	7.828	8.355	8.319
	(0.122)	[-1.506]	[-4.694]	[-1.628]	[-4.504]	[-4.305]
C(Y1AR)	0.246	0.175	-0.499	-0.792	0.033	-0.724
	(0.206)	[0.229]	[2.410]	[3.360]	[0.689]	[3.141]
C(Y1TR)	-3.347	-3.540	-4.122	-4.307	-5.047	-3.901
	(0.211)	[0.608]	[2.446]	[3.031]	[5.366]	[1.747]
C(Y1QU)	2.567	2.690	2.966	3.287	4.101	2.867
	(0.227)	[-0.359]	[-1.170]	[-2.111]	[-4.500]	[-0.880]
C(Y1LV)	0.780	0.842	1.031	0.702	0.063	0.712
	(0.205)	[-0.201]	[-0.813]	[0.256]	[2.331]	[0.222]
C(Y1AU1)	-1.418	-1.186	-0.297	-0.698	0.252	-0.212
	(0.258)	[-0.600]	[-2.902]	[-1.864]	[-4.323]	[-3.122]
C(Y1AU2)	-0.653	-0.567	-0.544	-0.496	-0.254	-0.405
	(0.217)	[-0.266]	[-0.338]	[-0.483]	[-1.230]	[-0.764]
C(INAR)	-0.161	-0.228	-0.121	-0.738	-0.179	-0.771
	(0.227)	[0.198]	[-0.119]	[1.698]	[0.053]	[1.794]
C(INTR)	-5.373	-5.498	-5.463	-5.210	-6.478	-5.252
	(0.176)	[0.472]	[0.340]	[-0.617]	[4.194]	[-0.461]
C(INQU)	4.122	4.232	4.105	3.720	5.247	4.031
	(0.205)	[-0.358]	[0.057]	[1.309]	[-3.657]	[0.296]
C(INLV)	0.900	0.766	1.138	1.639	0.190	0.588
	(0.219)	[0.410]	[-0.723]	[-2.249]	[2.164]	[0.951]
C(INAU1)	-0.358	-0.389	0.294	-0.078	0.400	-0.040
	(0.239)	[0.088]	[-1.821]	[-0.781]	[-2.117]	[-0.888]
C(INAU2)	-0.072	-0.405	-0.582	-0.148	-0.155	-0.312
	(0.214)	[1.037]	[1.590]	[0.238]	[0.259]	[0.749]
C(ARTR)	1.286	1.096	0.626	0.399	0.072	1.019
	(0.218)	[0.579]	[2.013]	[2.706]	[3.705]	[0.814]
C(ARQU)	-1.201	-0.941	-0.573	-0.244	-0.055	-0.952
	(0.222)	[-0.778]	[-1.882]	[-2.868]	[-3.433]	[-0.746]
C(ARLV)	0.557	0.625	0.192	0.387	0.794	-0.462
	(0.215)	[-0.212]	[1.132]	[0.527]	[-0.738]	[3.164]
C(ARAU1)	2.321	3.031	4.337	3.545	3.923	3.434
	(0.232)	[-2.043]	[-5.800]	[-3.522]	[-4.609]	[-3.202]
C(ARAU2)	1.048	1.148	0.716	1.211	1.400	1.507
	(0.215)	[-0.308]	[1.031]	[-0.505]	[-1.092]	[-1.422]
C(TRQU)	-9.153	-9.026	-8.919	-9.225	-9.157	-9.110
	(0.047)	[-1.801]	[-3.333]	[1.037]	[0.061]	[-0.613]
C(TRLV)	3.169	3.147	2.826	1.600	0.128	2.734
	(0.212)	[0.068]	[1.079]	[4.943]	[9.578]	[1.372]
C(TRAU1)	0.176	0.031	0.182	0.361	-0.305	0.780
	(0.219)	[0.440]	[-0.017]	[-0.562]	[1.462]	[-1.834]
C(TRAU2)	0.165	0.395	0.229	-0.036	0.175	-0.463
	(0.222)	[-0.691]	[-0.193]	[0.604]	[-0.028]	[1.888]

Values in (.) are standard deviations, Values in [.] are t-values

Table C.3 *Continued*						
ap name	Data	Preferred	Hom. AR	Hom. trend	No co-dep.	Unit root
C(LVAU1)	-0.781	-1.022	-0.424	0.003	-0.055	-0.158
	(0.213)	[0.755]	[-1.121]	[-2.458]	[-2.277]	[-1.954]
C(LVAU2)	0.090	0.169	0.179	-0.112	0.505	-1.264
	(0.233)	[-0.226]	[-0.255]	[0.580]	[-1.188]	[3.875]
C(AU1AU2	-4.679	-5.116	-5.265	-4.915	-4.858	-4.796
	(0.196)	[1.489]	[1.994]	[0.804]	[0.608]	[0.397]
SKW(Y1)	-0.017	-0.005	0.169	0.396	0.582	0.308
	(0.126)	[-0.066]	[-0.983]	[-2.181]	[-3.161]	[-1.717]
KUR(Y1)	1.859	1.666	2.305	1.546	4.156	1.080
	(0.348)	[0.370]	[-0.856]	[0.600]	[-4.403]	[1.494]
M(Y2)	12.487	12.489	12.488	12.496	12.490	12.491
	(0.004)	[-0.361]	[-0.222]	[-1.530]	[-0.626]	[-0.678]
S(Y2)	0.170	0.167	0.176	0.170	0.176	0.165
	(0.005)	[0.408]	[-0.806]	[-0.084]	[-0.890]	[0.618]
C(Y1Y2)	0.605	0.631	0.718	0.711	0.723	0.792
	(0.021)	[-0.827]	[-3.566]	[-3.334]	[-3.710]	[-5.902]
C(Y1Y3)	0.475	0.507	0.569	0.600	0.609	0.708
	(0.020)	[-1.059]	[-3.145]	[-4.187]	[-4.480]	[-7.795]
C(Y2Y3)	0.743	0.771	0.769	0.831	0.803	0.818
	(0.029)	[-0.642]	[-0.594]	[-2.056]	[-1.385]	[-1.752]

Values in (.) are standard deviations, Values in [.] are t-values

ap name	Data	Preferred	Hom. AR	Hom. trend	No co-dep.	Unit root
$\hat{\kappa}_0$ (CS variance)	-0.147	-0.389	-0.478	-0.069	-0.457	-0.104
	(0.082)	[1.962]	[2.684]	[-0.637]	[2.520]	[-0.353]
$\hat{\kappa}_1$(CS variance)	0.389	0.799	0.841	0.381	0.645	0.557
	(0.119)	[-2.295]	[-2.530]	[0.048]	[-1.435]	[-0.938]
$\hat{\kappa}_2$(CS variance)	-0.087	-0.323	-0.343	-0.177	-0.246	-0.216
	(0.064)	[2.453]	[2.653]	[0.941]	[1.647]	[1.335]
$\hat{\kappa}_3$(CS variance)	-0.001	0.044	0.049	0.026	0.032	0.025
	(0.011)	[-2.618]	[-2.916]	[-1.559]	[-1.934]	[-1.540]
$Corr(\tilde{y}_{it}, \tilde{y}_{it-1}) \cdot 10$	8.984	8.893	8.889	8.897	8.784	9.002
	(0.026)	[2.341]	[2.434]	[2.235]	[5.168]	[-0.484]
$Corr(\tilde{y}_{it}, \tilde{y}_{it-2}) \cdot 10$	8.473	8.341	8.317	8.367	8.164	8.543
	(0.037)	[2.373]	[2.820]	[1.915]	[5.576]	[-1.262]
$Corr(\tilde{y}_{it}, \tilde{y}_{it-3}) \cdot 10$	8.026	7.843	7.784	7.931	7.570	8.232
	(0.045)	[2.680]	[3.549]	[1.390]	[6.692]	[-3.028]
$V(\Delta\tilde{y}_{it}) \cdot 10$	0.108	0.113	0.113	0.111	0.103	0.113
	(0.002)	[-1.464]	[-1.476]	[-0.941]	[1.250]	[-1.333]
$Corr(\Delta\tilde{y}_{it}, \Delta\tilde{y}_{it-1}) \cdot 10$	-2.324	-2.340	-2.394	-2.553	-2.446	-2.597
	(0.061)	[0.175]	[0.770]	[2.504]	[1.336]	[2.996]
$Corr(\Delta\tilde{y}_{it}, \Delta\tilde{y}_{it-1}) \cdot 10$	-0.177	-0.065	0.006	-0.301	-0.022	-0.648
	(0.062)	[-1.206]	[-1.976]	[1.346]	[-1.671]	[5.108]
$C(\tilde{y}_{it+2} - \tilde{y}_{it-3}, \tilde{y}_{it} - \tilde{y}_{it-1})$	0.499	0.564	0.569	0.453	0.506	0.393
	(0.016)	[-2.752]	[-2.978]	[1.913]	[-0.313]	[4.457]
$\hat{\tau}_0(autoc.1$,level $\tilde{y}_{it})$	0.969	0.960	0.948	0.937	0.924	0.956
	(0.003)	[1.673]	[3.999]	[6.232]	[8.921]	[2.508]
$\hat{\tau}_1(autoc.1$,level $\tilde{y}_{it}$)	-0.382	-0.370	-0.298	-0.241	-0.249	-0.227
	(0.018)	[-0.456]	[-3.127]	[-5.232]	[-4.936]	[-5.744]
$\hat{\tau}_0(autoc.2$,level $\tilde{y}_{it})$	0.959	0.948	0.937	0.915	0.894	0.949
	(0.005)	[1.448]	[2.943]	[5.768]	[8.539]	[1.335]
$\hat{\tau}_0(autoc.2$,level $\tilde{y}_{it})$	-0.519	-0.506	-0.462	-0.343	-0.370	-0.330
	(0.023)	[-0.383]	[-1.641]	[-5.104]	[-4.313]	[-5.467]
$\hat{\tau}_0(autoc.1$,growth $\Delta\tilde{y}_{it})$	-0.258	-0.250	-0.262	-0.281	-0.264	-0.264
	(0.008)	[-0.725]	[0.317]	[1.976]	[0.473]	[0.497]
$\hat{\tau}_1(autoc.1$,growth $\Delta\tilde{y}_{it})$	0.183	0.120	0.162	0.188	0.139	0.031
	(0.029)	[1.444]	[0.476]	[-0.116]	[1.012]	[3.494]
$\hat{\tau}_0(autoc.2$, growth $\Delta\tilde{y}_{it})$	-0.020	-0.006	-0.001	-0.039	-0.010	-0.068
	(0.008)	[-1.203]	[-1.620]	[1.652]	[-0.863]	[4.115]
$\hat{\tau}_1(autoc.2$, growth $\Delta\tilde{y}_{it})$	0.011	-0.005	0.009	0.064	0.050	0.025
	(0.021)	[0.499]	[0.057]	[-1.643]	[-1.209]	[-0.445]

Values in (.) are standard deviations, Values in [.] are t-values

Table C.4: Auxiliary parameters for Goodness of fit

Table C.4 *Continued*

ap name	Data	Preferred	Hom. AR	Hom. trend	No co-depend	Unit root
P. A. lag 1	-0.289	-0.264	-0.280	-0.304	-0.279	-0.314
	(0.010)	[-1.700]	[-0.629]	[1.016]	[-0.648]	[1.682]
P. A. lag 2	-0.114	-0.094	-0.097	-0.134	-0.096	-0.174
	(0.010)	[-1.368]	[-1.102]	[1.374]	[-1.224]	[4.097]
P. A. lag 3	-0.066	-0.104	-0.117	-0.076	-0.084	-0.083
	(0.010)	[2.584]	[3.460]	[0.694]	[1.256]	[1.175]
P. A. lag 4	-0.038	-0.051	-0.074	-0.028	-0.050	-0.022
	(0.007)	[1.322]	[3.569]	[-0.900]	[1.240]	[-1.555]
P. A. lag 7	-0.005	-0.006	-0.007	-0.008	-0.017	0.006
	(0.006)	[0.072]	[0.176]	[0.231]	[1.185]	[-1.179]
P.A. lag 10	0.007	0.004	0.012	-0.007	-0.008	0.006
	(0.007)	[0.288]	[-0.494]	[1.417]	[1.554]	[0.144]
P.A. lag 12	0.013	0.016	0.023	0.004	0.006	0.017
	(0.006)	[-0.244]	[-1.035]	[0.983]	[0.782]	[-0.383]
$C(\Delta\tilde{y}_{it+4}, \tilde{y}_{it})$,$t \in [1;9]$	-0.345	-0.327	-0.379	-0.385	-0.569	-0.226
	(0.038)	[-0.311]	[0.600]	[0.708]	[3.895]	[-2.060]
$C(\Delta\tilde{y}_{it+4}, \tilde{y}_{it})$,$t \in [10;18]$	-0.007	0.006	0.000	-0.181	-0.218	0.051
	(0.029)	[-0.311]	[-0.171]	[4.034]	[4.886]	[-1.362]
$C(\Delta\tilde{y}_{it+4}, \tilde{y}_{it})$,$t \in [19;26]$	-0.197	-0.097	-0.044	-0.204	-0.181	-0.071
	(0.029)	[-2.300]	[-3.530]	[0.171]	[-0.359]	[-2.902]
P(t,t+1)	0.746	0.738	0.788	0.771	0.741	0.773
	(0.005)	[0.998]	[-5.139]	[-2.990]	[0.678]	[-3.251]
P(t,t+10)	0.485	0.442	0.499	0.437	0.418	0.427
	(0.011)	[2.550]	[-0.803]	[2.830]	[3.976]	[3.455]

Values in (.) are standard deviations, Values in [.] are t-values

P.A: Partial autocorrelation based on $\Delta\tilde{y}_{it}$

Name	Estimate
c (skewness)	0.1268
d (kurtosis)	0.8125
ϕ_1	12.4543
ϕ_2	-2.6884
ϕ_3	0.0316
ϕ_4	0.7573
ϕ_5	0.5383
ϕ_6	0.0429
ϕ_7	0.2020
ψ_{11}	-2.8759
ψ_{21}	0.0655
ψ_{22}	-1.0307
ψ_{31}	-0.0152
ψ_{32}	0.0139
ψ_{33}	-5.2245
ψ_{41}	0.1650
ψ_{42}	-0.5372
ψ_{43}	-0.3669
ψ_{44}	-18.9666
ψ_{51}	0.3825
ψ_{52}	0.5080
ψ_{53}	1.5193
ψ_{54}	-0.0140
ψ_{55}	0.1816
ψ_{61}	-0.1034
ψ_{62}	-0.1571
ψ_{63}	-0.0748
ψ_{64}	0.1189
ψ_{65}	-0.4770
ψ_{71}	-0.0735
ψ_{72}	-0.0478
ψ_{73}	-0.0800
ψ_{74}	0.0791
ψ_{75}	-0.0189
m_1	-0.0137
$\tilde{c}_0$	-0.1044
$\tilde{c}_1$	-2.0452
$\tilde{c}_2$	-0.055
ψ_{81}	0.0242
ψ_{83}	0.0479
ψ_{84}	-0.1315

Table C.5: Estimates of Distribution parameters