The implications of richer earnings dynamics for consumption and wealth

Mariacristina De Nardi, Giulio Fella, and Gonzalo Paz-Pardo*

June 16, 2016

Abstract

Earnings dynamics are much richer than those typically used in macro models with heterogenous agents. This paper provides multiple contributions. First, it proposes a simple non-parametric method to model rich earnings dynamics that is easy to estimate and introduce in structural models. Second, it estimates our non-parametric earnings process using two data sets: the Panel Study of Income Dynamics and a large, synthetic, data set that matches the dynamics of the U.S. tax earnings. Third, it uses a life cycle model of consumption to compare the consumption and saving implications of the two estimated processes to those of a standard AR(1). We find that, unlike the standard AR(1) process, our estimated, richer earnings process generates an increase of consumption inequality over the life cycle that is consistent with the data and better fits the savings of the households at the bottom 60% of the wealth distribution.

Keywords: Earnings risk, savings, consumption, inequality, life cycle.
JEL Classification: D14, D31, E21, J31.

*Mariacristina De Nardi: UCL, Federal Reserve Bank of Chicago, IFS, and NBER, denardim@nber.org. Giulio Fella: Queen Mary University of London, CFM, and IFS, g.fella@qmul.ac.uk. Gonzalo Paz Pardo: UCL, gonzalo.pardo.13@ucl.ac.uk. De Nardi gratefully acknowledges support from the ERC, grant 614328 "Savings and Risks" and from the ESRC through the Centre for Macroeconomics. Fella is grateful to UCL for their generous hospitality while he was working on this paper. We thank Serdar Ozkan for assistance in generating the synthetic W2 data set and Marco Bassetto, Richard Blundell, Tony Braun, Jeremy Lise, Fabrizio Perri, Fabien Postel-Vinay, Ananh Sehsadri and Gustavo Ventura for helpful comments and suggestions. We are grateful to Moritz Kuhn for providing us with additional statistics from the Survey on Consumer Finances data set. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research, any agency of the Federal Government, or the Federal Reserve Bank of Chicago.

1 Introduction

Macroeconomic models with heterogeneous agents are ideal laboratory economies to quantitatively study a large set of issues that include household behavior under uncertainty, inequality, and the effects of taxes, transfers, and social insurance reforms. For instance, Scholz, Seshadri and Khitatrakun (2006) study the adequacy of savings at retirement, Storesletten, Telmer and Yaron (2004) study the evolution of consumption inequality over the life cycle, and Conesa, Kitao and Krueger (2009) study the optimal taxation of capital. Many of these quantitative models adopt earnings processes that imply that the persistence of earnings shocks is independent of age and earnings histories, and that positive and negative income changes are equally likely (for instance, a common assumption is that earnings follow a linear process with normal innovations).

Earnings risk, consumption, and wealth accumulation are tightly linked. The magnitude and persistence of earnings shocks determine how saving and consumption adjust to buffer their impact on current and future consumption and the extent to which people can self-insure by using savings. Appropriately capturing earnings risk is therefore important to understand consumption and wealth accumulation decisions as well as the welfare implications of income fluctuations and the potential role for social insurance.

A growing body of empirical work provides evidence that households' earnings dynamics feature substantial asymmetries and non-linearities and devises flexible statistical models that can allow for these features. For instance, recent work takes advantage of new methodologies (Arellano, Blundell and Bonhomme, 2015) or large data sets (Guvenen, Karahan, Ozkan and Song, 2015) to show that earnings changes display substantial skewness and kurtosis and that the persistence of shocks depends both on age and current earnings.[1]

Unfortunately, the complexity of these rich earnings processes makes it computationally

[1]Guvenen et al. (2015) document these features using Social Security Administration tax earnings (W2) data for a large panel of individuals, while Arellano et al. (2015) show that similar features hold in the PSID too.

very costly to incorporate them in the rich structural models of household's decision making that are needed to study consumption and saving decisions and their implications for consumption, wealth, and welfare inequality, both in the cross-section and over the life cycle. For this reason, rich earnings processes matching a large number of observed conditional and unconditional earnings moments have not been introduced in structural models of household's decision making over the life cycle so far.

This paper provides multiple contributions. First, it proposes a new, non-parametric, way to model life-cycle earnings dynamics that is both consistent with the new empirical findings and easy to estimate and introduce in structural quantitative models of optimal households' decision making. The method is sufficiently flexible to accommodate non-linearities, heteroskedasticity and deviations from normality. While the usual approach involves first estimating a **parametric, linear Markov process** for (the stochastic component of) earnings and then discretizing it, our proposed method estimates non-parametrically an **age-specific Markov chain** directly from the data. More specifically, in our main case—that assumes that earnings follow a Markov chain of order one—our method starts by computing the age-specific transition matrix from the percentile rank of the earnings distribution at age t to that at age $t+1$. Then, it discretizes the marginal distribution of earnings at each age by replacing the (heterogeneous) values of earnings in each rank percentile with their average. The result is a non-parametric representation of the earnings process that follows a Markov chain with an age-dependent transition matrix and a fixed number of age-dependent earnings states. Our method can be generalized to allow for Markov chains of order higher than one.[2]

Second, this paper applies our method both to the PSID and to a large synthetic panel generated simulating the parametric earnings process estimated by Guvenen et al. (2015) on W2 data.[3] The advantage of the PSID is that it is a well known data set, that includes more information about earners than the W2 data. Conversely, unlike the PSID, the W2 data

[2]In Section 8 we show that our findings are robust to allowing for a Markov chain of order two.

[3]Our synthetic data set matches, by construction, all of the moments that these authors use in their estimation procedure and allows us to sidestep the problem that accessing the W2 tax data is very involved.

set is extremely large and does not feature top-coding or under-representation of very high earners. For this reason we apply our non-parametric estimation method to both data sets. We find that our approximation replicates very well the first four unconditional moments of log earnings and the conditional moments of log earnings growth in both datasets. In both cases the estimated earnings process features important non-linearities, heteroskedasticity, negative skewness and kurtosis. In particular, we find that that our non-linear earnings process implies much lower persistence at low and high earnings levels and for the young than implied by the (near-) unit root processes commonly used in the quantitative macro literature with heterogeneous agents (e.g. Huggett, 1996).

Third, and importantly, this paper studies the implications of these features of earnings dynamics for consumption and wealth in the context of a standard life cycle model of savings and consumption with incomplete markets. More specifically, we compare the implications of our flexible estimated process, featuring an age-specific Markov chain, to those of a discretized AR(1) process calibrated to match the US earnings Gini and to approximate the life-cycle profile of earnings inequality.

Our main findings are the following. Unlike the AR(1), our non-linear earnings process generates an increase in consumption inequality over the life cycle that is in line with the observed data. We show that more than half of the improvement in fitting the growth of the consumption variance between ages 25 and 55 is due to the combination of age-dependent persistence and innovation variances, as well as skewness and kurtosis. The remaining fraction is accounted for by the dependence of conditional on current earnings.

In addition, our rich earnings process generates a better fit of the wealth holdings of the bottom 60% of individuals. Even our process, however, fails to improve on the standard AR(1) process in terms of fitting the right tail of the wealth distribution. This is a frequent drawback of life-cycle models with no entrepreneurs or transmission of bequests (De Nardi, 2004), (Cagetti and De Nardi, 2006), and (Cagetti and De Nardi, 2009).[4] Our findings are

[4]In computations available upon request, we verify that this result is not sensitive to increasing the value

robust to allowing earnings to follow a Markov chain of order two.

Finally, we compare the forecasting performance of our flexible Markov chain to that of standard AR(1) processes including the random walk process commonly used in the literature. The first-order Markov chain estimated on the W2 data has a lower forecast error up 3 to 4 years ahead. Even 5 to 6 years out its forecast error never exceeds that of the best AR(1) process by more than 10%. We also look at the implications of estimating a Markov Chain of order 2 and we find that outperforms the forecasting ability of all first-order processes at all horizons. Thus, we conclude that our flexible earnings representation of the synthetic W2 data implies a good representation of the data even at long horizons. Furthermore, our main results for consumption and wealth inequality are very similar when we use a Markov chain of order 1 or 2.

The rest of the paper is organized as follows. Section 2 discusses the related literature. Section 3 describes the main features of the data on earnings dynamics and inequality in consumption and wealth that our model seeks to match. Section 4 explains how we estimate our earnings process. Section 5 highlights the implications of our estimated earnings process over the life cycle and in the cross-section, both for the W2 tax data and the PSID. Section 6 specifies our standard quantitative model of savings and consumption over the life cycle and its calibration. Section 7 discusses the main implications of various earnings processes in terms of consumption and savings in this model and decomposes its determinants. Section 8 shows that allowing earnings to follow a second-order Markov chain does not affect our results. Section 9 concludes. Appendix A discusses key features of both the PSID and our synthetic W2 data set. Appendix B shows how well our non-parametric earnings processes match the PSID and W2 data.

of the CRRA coefficient up to a rather high value of 4.

2 Related literature

Our goal is to take a standard life cycle model with exogenously incomplete markets (as in Huggett (1996)) and study how the various earnings processes that we consider affect consumption, wealth, and their inequality, both over the life cycle and in the cross-section. Thus, our work connects with two branches of the literature. First, it relates to the literature on quantitative models with heterogeneous agents, which have been used, for instance, to study inequality and the effects of government policy reforms. Second, it relates to the literature on earnings dynamics and its effects on consumption choices. Both branches of the literature are vast.

The literature on quantitative models with heterogeneous agents typically assumes a very parsimonious specification of the earnings process; namely that the logarithm of earnings follow a linear process with Gaussian innovations. This process is then typically discretized using some variant of the methods described in Tauchen (1986) or Tauchen and Hussey (1991). An important set of applications of the literature on quantitative models with heterogeneous agents studies consumption (e.g., Storesletten et al., 2004; Krueger and Perri, 2006; Jonathan Heathcote, 2010) and wealth (e.g., Castañeda, Díaz-Giménez and Ríos-Rull, 2003; De Nardi, 2004; Cagetti and De Nardi, 2009) inequality and their evolution over the life cycle.[5] It has been argued, however, that incorporating non-Gaussian shocks is important to properly capture earnings dynamics. Geweke and Keane (2000) show that allowing for non-Gaussian innovations to log-earnings results in better accounting of transitions between low and high earnings states relative to the AR(1) model with Gaussian innovations in Lillard and Willis (1978).[6] Bonhomme and Robin (2009) also adopt a specification of the (marginal) earnings distribution that allows for non-Gaussian innovations to the first-

[5]See Quadrini and Ríos-Rull (2014), Cagetti and De Nardi (2008) and De Nardi (2015) for a discussion of what features of the wealth data various versions of these models are able to match.

[6]They find that the distribution of log earnings shocks is leptokurtic, which means that it has fat tails: there is a small but relevant number of individuals who suffer large short-run earnings shocks. A Gaussian AR(1) fails to capture this feature of the data, thus overestimating persistence of log earnings.

order Markov stochastic component of earnings, with transition probabilities derived from a one-parameter Plackett copula. They too find evidence against normality in the marginal distribution of earnings shocks, which is leptokurtic. Our empirical method is closely related to theirs in that it does not impose Gaussianity and features first-order Markov dynamics. Castañeda et al. (2003) were the first to study the implications for wealth inequality of an earnings process that, like ours, does not impose normality and state-independent persistence. Differently from ours, though, their earnings process is calibrated to match certain moments of the wealth distribution, rather than only estimated from earnings data, and is age-invariant while ours is allowed to be age-dependent.[7]

Meghir and Pistaferri (2004) relax the assumption of i.i.d. log earnings innovations and model the conditional variance of log earnings shocks by allowing for autoregressive conditional heteroskedasticity (ARCH) in the variance of log earnings. In this context, the current realizations of the variance are thus informative about future earnings.[8] They find evidence of ARCH-type variances for both the permanent and transitory shocks and for individual-specific variances. Blundell, Graber and Mogstad (2015) use Norwegian panel data and find that a better description of earnings dynamics requires allowing for heterogeneity by education levels and accounting for non-stationarity. Their preferred specification for log earnings includes an idiosyncratic constant term, an idiosyncratic experience profile, an AR(1) persistent shock component and an MA(1) transitory shock component. All components considered are allowed to have a skill-dependent distribution and the variances of shocks are allowed to be age-dependent and hence non-stationary. Our process also allows for heteroskedasticity, though not of the ARCH type, and non-stationarity.

Arellano et al. (2015) model a log earnings process composed by the sum of a transitory stochastic component and a first-order Markov process which, like ours, allows for significant non-linearities in age and previous earnings levels. Guvenen et al. (2015) study the evolution

[7]Age dependence is irrelevant anyway in Castañeda et al. (2003) as their model is infinite horizon.

[8]They also allow the variance of log earnings innovations to depend on year effects, the education of the individual, and unobservable individual factors.

of male earnings over the life cycle using a large Social Security Administration panel data set. Using non-parametric methods they find that labor earnings do not conform to the standard assumptions in most of the empirical literature surveyed in Blundell (2014); namely, that the stochastic component of earnings is the sum of linear processes with i.i.d, normal innovations. In particular, they find that earnings shocks display strong negative skewness and very high kurtosis. They also show that cross-sectional moments change with age and previous earnings levels. Finally, they estimate a complex set of parametric processes that match a large number of moments in the data, including higher order moments.

Finally, Browning, Ejrnaes and Álvarez (2010) and Altonji, Smith and Vidangos (2013) estimate very flexible earnings models. Both models are not tractable for our purpose. Browning et al. (2010) show that it is necessary to include individual-level heterogeneity in earnings processes to account for many important features of the data. They compare a standard unit-root model for log earnings with a more complex ARMA(1,1) process with large heterogeneity, where individual-specific parameters are derived from three stochastic latent factors. This richer process significantly improves the fit of the earnings process to the data. Altonji et al. (2013) consider a multivariate model of earnings and jointly model log wages, job changes, unemployment transitions and hours worked by explicitly allowing these processes to interact (for instance, wages depend on job tenure, while job changes depend on the expected wage should the individual remain in the same job). Their model provides important insights on the evolution of earnings and job tenure. Furthermore, it further stresses the importance of allowing for heterogeneity and state dependence when approximating earnings dynamics.

Recent developments in this literature are discussed in Meghir and Pistaferri (2011). The consequences of these richer earnings processes on consumption, savings and welfare remain, however, an important open question. Unfortunately, the number of state variables needed to include these kind of processes in a structural model is very large and generates both

computational and modeling issues because it is not obvious how the discretization of each of these variables should be performed to preserve their relationships and dynamics.

Finally, a recent paper by Civale, Díez-Catalán and Fazilet (2016) adapts Tauchen's (1986) method to discretize stationary AR(1) processes with non-Gaussian innovations and explores the implications for the equilibrium capital stock in an economy à la Aiyagari (1994*a*).

3 Facts about earnings, consumption and wealth

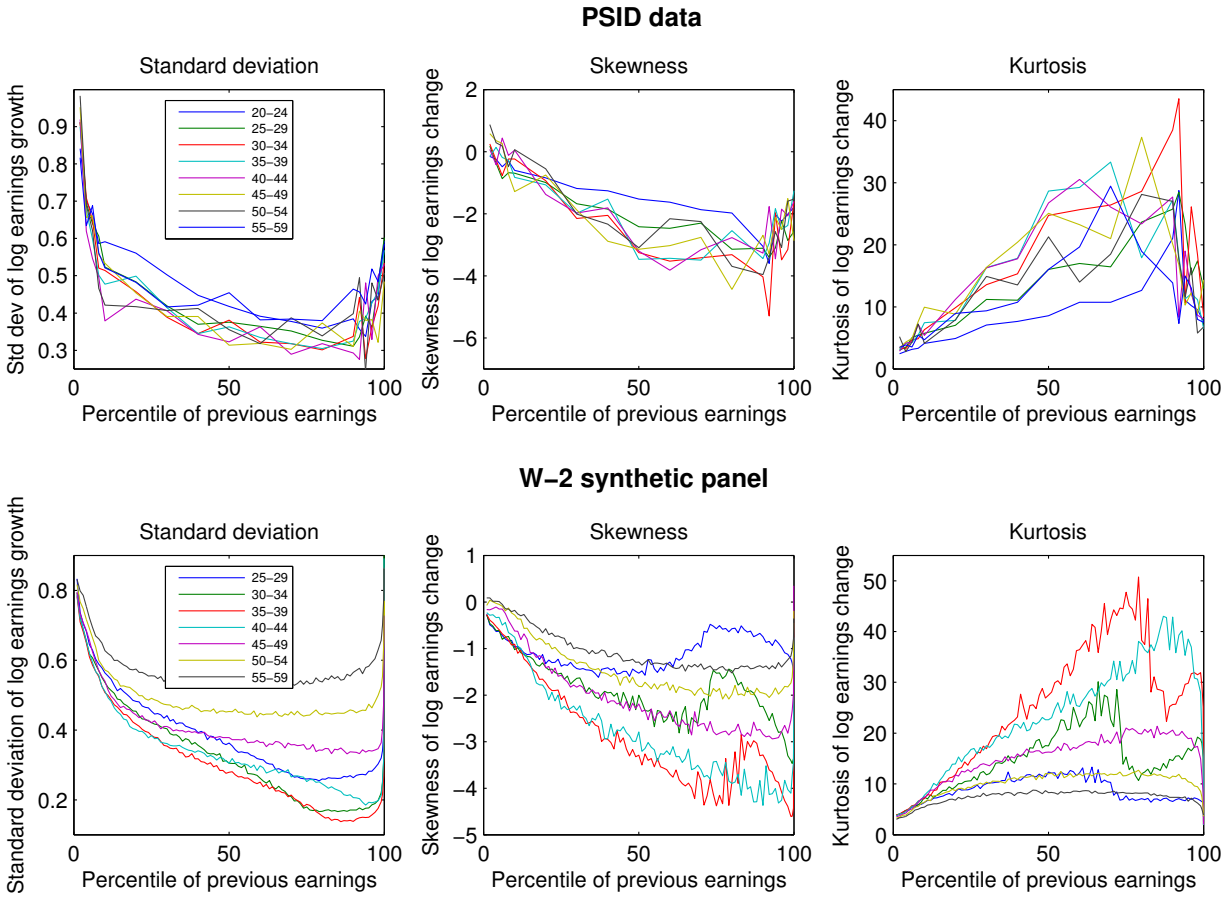
The earnings shocks experienced by US workers display important deviations from the assumptions of log-normality and independence from age and earnings realizations as documented by Arellano et al. (2015) for the PSID and Guvenen et al. (2015) for W2 tax data. The top panel of Figure 1 reports our computations for the conditional moments of log earnings growth from the PSID data. The bottom panel reports the same moments estimated on the synthetic panel dataset that we generate by simulating the parametric process estimated by Guvenen et al. (2015) on W2 data. To capture the non-linearity and asymmetry in the data, Guvenen et al. (2015) fit a flexible process consisting of a mixture of AR(1) plus a heterogeneous income profile and estimate it by simulated method of moments.[9]

The figure shows that the conditional variance of log earnings growth is U-shaped across all age groups: individuals with the largest and smallest earnings are the ones that suffer from higher earnings risk. Furthermore, the variance declines until age 35 and starts increasing after age 45.

The figure also shows that log earnings growth has strong negative skewness and very high kurtosis, and that these moments depend both on age and previous earnings. A negative skewness means that individuals face a higher risk of negative relative to positive changes in

[9]The parameterization is the one denoted "benchmark" or "AR(1)" and reported in column (2) in Table III of their paper. Further details about this process, as well as the PSID data, are provided in Appendix A.

Figure 1: Conditional moments of log earnings changes by age cohort (PSID and Synthetic W2 data)



earnings. The skewness is more negative for individuals in higher earnings percentiles and for individuals between 35 and 45 years of age. This indicates that individuals face a larger downward risk as they approach middle age.[10]

The kurtosis is a measure of the peakedness of the distribution of log earnings changes. A high kurtosis means that most of the people experience relatively small earnings shocks in a given year but that at the same time a small proportion of individuals face very large earnings shocks. The kurtosis of earnings growth gets as large as 30 (compared to 3 for a standard normal distribution). Kurtosis is hump-shaped by earnings percentile and increases until age 35-45 to then decrease thereafter. Our graphs thus indicates large deviations between the moments observed in the data and those implied by the linear earnings processes.

Overall, the moments in the PSID data are both qualitatively and quantitatively very close to those computed from our synthetic W2 panel. However, it should be noted that the W2 synthetic panel is much larger and is not affected by top-coding or differential survey responses, and thus provides better information on the earnings-rich. For instance, the increase in the variance of earnings beyond the 95th percentile is more pronounced in the W2 data set. Similarly, the negative skewness and the kurtosis are much more pronounced for the highest percentiles in the W2 data set.

The decrease in the conditional variance as individuals get older in the W2 data is also documented in Sabelhaus and Song (2010), who in addition find a substantial decline over time, particularly after 1990.

Turning to wealth and consumption inequality, the top line of Table 1 summarizes the main statistics for the US.[11] Overall, wealth is very unequally distributed, with a Gini coefficient of 0.72 (the corresponding value for the earnings distribution is 0.51 (Quadrini and Ríos-Rull, 1997)). Using different time periods yields a slightly higher concentration of wealth

[10] Graber and Lise (2015) account for this kind of earnings behavior in the context of a search and matching model with a job ladder.

[11] The data on wealth are from Wolff (1987) and come from the 1983 Survey of Consumer Finances. We use the consumption data from the CEX sample derived by Heathcote, Perri and Violante (2010), and we use their sample A and variable definitions.

in the hands of the richest few.

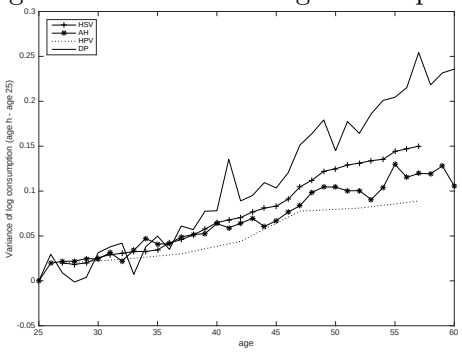
As discussed by Quadrini and Ríos-Rull (2014), Huggett (1996), Cagetti and De Nardi (2008), and De Nardi (2015), the standard life cycle model with incomplete markets and discretized AR(1) earnings shock cannot match the large concentration of wealth in the hands of the richest few and generates too many people at zero (or negative) wealth. An important question is whether a better representation of earnings risk can help match wealth inequality and along which dimensions.

Table 1: Wealth and consumption distribution statistics, U.S. data

		Percentage held by the top						At negative or
	Gini	1%	5%	20%	40%	60%	80%	zero
Wealth	.72	28	49	75	89	96	99	5.8-15.0%
Non-durable consumption	.31	5	15	40	62	79	92	—

The second line in Table 1 displays the distribution of equivalized consumption of non durables and services for the entire US adult population in 1989. The consumption Gini that we compute is in line with the estimates in the literature (around 0.29 for 1989 in Fisher, Johnson and Smeeding (2013)) and so is the shape of the distribution (our implied 90/10 and 50/10 ratios are 4.23 and 2.12 respectively, which are in line with those in Meyer and Sullivan (2010)). Consumption inequality is thus significant, but much lower in magnitude than both earnings and wealth inequality. Relatedly, Figure 2 reports the estimated age profile of the cross-sectional variance of log, adult-equivalent consumption from a number of studies. As is well known from the seminal work of Deaton and Paxson (1994), it is increasing though flatter than the profile of the variance of log earnings. There are multiple estimates in the literature of the age profile for the variance of consumption. The steeper profile in Figure 2 (DP) corresponds to the estimate from CEX data over the the period 1980-90 computed by Deaton and Paxson (1994). As first pointed out by Slesnick and Ulker (2004) and Heathcote, Storesletten and Violante (2005) (HSV) the rise in inequality by age is substantially smaller when estimated from CEX data but over a longer time period. Heathcote et al. (2010) (HPV)

Figure 2: Variances of log consumption



and Aguiar and Hurst (2013) (AH) confirm these findings.[12]

Storesletten et al. (2004) showed that the canonical heterogeneous agent model with a (near-) unit root in earnings and pay-as-you-go social security generates a rise in consumption dispersion consistent with the estimates from Deaton and Paxson (1994). Yet, such a model cannot match the much lower increase documented by later studies unless the process for earnings has an idiosyncratic deterministic time trend, or Heterogeneous Income Profile (Guvenen, 2007; Primiceri and van Rens, 2009). Huggett, Ventura and Yaron (2011) show that heterogeneity in earnings growth rates can be explained by the endogenous response of life-cycle, human-capital investment to heterogeneity in initial human capital levels. Intuitively, heterogeneity in individual, life-cycle trend growth generates a substantially smaller rise in consumption dispersion as the individual-specific trend growth is known to consumers, though not to the econometrician.

4 The discretized earnings processes

The most common earnings process used in quantitative models of consumption and saving is a discrete approximation (based on Tauchen, 1986, and its variants) to an AR(1) process

[12] All the estimates reported in Figure 2 control for cohort effects, with the exception of Aguiar and Hurst (2013) that control for cohort and normalized time effects. Heathcote et al. (2005) and Heathcote et al. (2010) find a marginally smaller increase when controlling for time effects.

for the logarithm of earnings (e.g. Huggett, 1996). We are going to compare the implications of such a process with those of two alternative earnings processes estimated by applying our non-parametric methodology on the PSID and the synthetic W2 data. In a nutshell, our methodology estimates a discrete Markov chain directly from the data and is much more flexible; i.e., it does not impose symmetry or linearity.

4.1 Benchmark earnings process

The benchmark earnings process is based on Huggett (1996)'s calibration, where the labor endowment process is an AR(1) with persistence parameter $\gamma = 0.96$ and Gaussian shocks with variance $\sigma_\epsilon^2 = 0.045$. The initial endowment of the first cohort of agents is also normally distributed, with variance $\sigma_{y_0}^2 = 0.38$. Huggett's choice of value for the variance of shocks σ_ϵ^2 is based on similar estimates in previous literature (e.g. Lillard and Willis (1978) or Carroll, Hall and Zeldes (1992)). The variance of the initial condition $\sigma_{y_1}^2$ is chosen to match the earnings Gini for young agents (Lillard (1977), Shorrocks (1980)). Given both variances, γ is calibrated to match an overall earnings Gini for US males of 0.42.

We then follow the discretization strategy applied by Huggett (1996), which is based on Tauchen (1986): The state space for log earnings is divided in 18 equidistant points, that range between $-4\sigma_{y_1}$ and $4\sigma_{y_1}$. To better approximate the upper tail of the earnings distribution, a further point is included ad hoc, situated at $6\sigma_{y_1}$. The small number of individuals that are situated in this upper point earn about 40 times median earnings. The rest of the grid ranges between 8% of median earnings and 11 times median earnings. The computation of the transition matrix relies on the fact that, conditional on earnings at age $h-1$, earnings at age h are drawn from a normal distribution with mean γy_{h-1} and standard deviation σ_ϵ.

There are two main differences between our benchmark earnings process and the original one in Huggett (1996). First, Huggett studies agents from age 20 to 65, while we define

working the life as spanning from age 25 to 60 for comparability with our W2 synthetic data. Second, we borrow the life-earnings profile from Hansen (1993), since the exact values used by Huggett into his model are not readily available.

4.1.1 Non-parametrically estimated processes for the PSID and W2 earnings data

We estimate two non-parametric earnings processes, one from the PSID data and one from the Social Security W2 synthetic panel that we generate from the empirical processes estimated by Guvenen et al. (2015). Appendix A discusses the PSID data and our synthetic W2 data. The main difference with respect to the discretization method in the previous section is that the alternative discretization we propose is very flexible and therefore capable of matching the asymmetries and non-linearities in the PSID and W2 data.

We first purge the original earnings data from time and age effects and then discretize the residual stochastic component of earnings. Let y_{ht}^i denote the logarithm of labor earnings for an individual i at time h and age t. We assume the process for y_{ht}^i takes the following form

$$y_{ht}^i = d_h + f(\theta, t) + \eta_{ht}^i, \tag{1}$$

where d_h is a yearly dummy and $f(\theta, t)$ is a quartic function of age. The term η_{ht}^i captures the stochastic component of earnings.[13]

We assume that the distribution of the stochastic component of earnings η_{ht}^i is i.i.d. across individuals but do not impose any additional restriction other than assuming that conditionally on age t, η_{ht}^i follows a Markov chain of order one, with age-dependent state space $Z^t = \{\bar{z}_1, \ldots, \bar{z}_N\},\ t = 1, \ldots, T$ and an age dependent transition matrix Π^t, which has size $(N \times N)$. That is, we assume that the dimension N of the state space is constant across

[13]We do not need to include yearly dummies when using W2 synthetic data because they are already extracted in the original estimation procedure.

ages but we allow its possible realizations and its transition matrix to be age-dependent.

We determine the points of the state-space and the transition matrices at each age in the following way.

1. We recover the stochastic component of earnings as the residual of running the regression associated with equation (1).

2. Fix the number of bins, N, at each age. At each age, we order the realized log earnings residuals by their size and we group them into bins, each of which contains $1/N$ of the number of observations at that age.

   Because the PSID and the our synthetic W2 data greatly differ in their sample size, we choose N in the two data sets as follows.

   - Due to the limited sample size of the PSID data, we have to strike a balance between a rich approximation of the actual earnings dynamics by earnings level (that is, a large number of bins) and keeping the sample size in each bin sufficiently large. We have thus evaluated many possibilities. In our main specification we report the results for bins representing deciles, with the exception of the top and bottom deciles, that we split in 5. Therefore, bins 1 to 5 and 14 to 18 include 2% of the agents at any given age, while bins $n = 6, \ldots, 13$ include 10 % of the agents at any given age. This implies a total of 18 bins.
   - Our synthetic W2 dataset is simulated to contain 18 million observations (500,000 individuals over 36 years), which implies that we are not constrained by issues of insufficient sample size. For this data set, we thus report results for a discretization with 103 gridpoints, which aims at accurately capturing the earnings dynamics of the earnings-rich. The bottom 99 gridpoints correspond to the bottom 99 percentiles of the earnings distribution, while the top 1% is divided into 4 bins. More specifically, we separate in a special bin the top 0.01%, we create another

bin for the rest of the top 0.05% and we divide the remaining people of the top percentile in two bins.

3. The points of the state space at each age t are chosen so that point z_t^n is the mean in bin n at age t. We have considered specifications in which the summary statistic of each bin is the median instead of the mean, with no impact on the results.

4. The initial distribution at model age 0, is the empirical distribution at the first age we consider.

5. The elements π_{mn}^t of the transition matrix Π^t between age t and $t+1$ are the proportion of individuals in bin m at age t that are in bin n at age $t+1$. The use of transition matrices is well established in the study of income mobility (e.g. Jäntti and Jenkins (2015)). The main difference is that while studies of income mobility are usually concerned about intra- or inter-generational mobility across relative rankings in the earnings distribution, we are interested in capturing mobility across earnings levels.

For this procedure to provide consistent estimates of the population earnings distribution over the life cycle, we need a large enough number of individuals in the sample for every age group. This is not a problem for our W2 synthetic data, that we simulate, but is an issue for PSID data. To solve this problem, we assume that age t actually includes people aged $t-1$, t and $t+1$. Specifically, we create, for every age t in the sample, a fictitious t-year-old cohort which is formed of all individuals in the sample who are $t-1$, t and $t+1$ years old. We then apply the method described above to this fictitious cohort to derive the state space for agents of age t. Since we repeat this for every age, most observations in the original data base are used three times.

To keep comparability between the AR(1) earnings process and our estimated processes, we use the same age-efficiency profile in all cases. Hence, we discard the age-efficiency profile that we extract from the PSID and W2 data and we use those data sets to estimate earnings

mobility. Furthermore, average earnings in each of the three economies are normalized to 1 so that the total amount of resources that are exogenously entering the economies are the same.

Appendix B extensively discusses how well our non-parametric estimation method matches the observed moments both in the PSID and the synthetic W2 data, including for the same number of earnings bin. The main conclusion from that comparison is that our method does a very good job of matching the vast majority of moments in the data, and especially so for the earnings process with 103 bins.

5 What are the implications of our earnings processes?

5.1 Moments across the life-cycle

Figure 3 shows the average earnings profile that we assume for all three processes, calibrated using the age-efficiency profile in Hansen (1993). The earnings process is calibrated so that average earnings across all agents (including retired individuals) are normalized to 1.

Figure 3: Average earnings by age

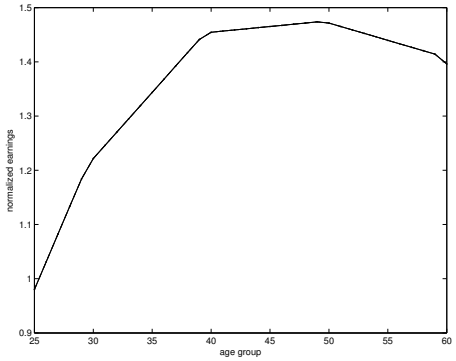


Table 2 reports the cross-sectional Gini coefficient of earnings for the overall working population, and the Gini coefficients for the first and last cohort of the simulated processes (25 and 60 years old).

Table 2: Earnings process statistics

Process	Overall Gini	Gini at 25	Gini at 60
Benchmark AR(1)	0.4160	0.3371	0.4398
NL PSID Process	0.3307	0.2637	0.3870
NL Synthetic W2 Process	0.4120	0.3456	0.4980

The table shows that, compared to the benchmark AR(1) economy, whose earnings process is chosen to match the level and the rise of the Gini coefficient by age, the NL earnings process estimated on the PSID generates a significantly lower Gini coefficient both in the whole cross-section of working ages and at younger and older ages. This is likely due to the features of the PSID data we discussed earlier (lack of over-sampling of high earners, lower response rate of high earners, etc.). The NL synthetic W2 process, instead, is much more successful in terms of replicating the actual level of overall inequality and generates earnings inequality around the age of retirement that is higher than that generated by the other two processes.

Figure 4 reports the earnings Gini by age computed in our synthetic W2 and the 1989 Survey of Consumer Finances (SCF). The SCF oversamples and re-weights the rich and thus provides a more accurate representation of inequality (Bricker, Krimmel, Henriques and Sabelhaus (2016)) than many survey data sets. The graph shows that the earnings Gini by age generated by our W2 data set is consistent with the one in the SCF data set.

Figure 4: Earnings Gini by age, W2 data and SCF data

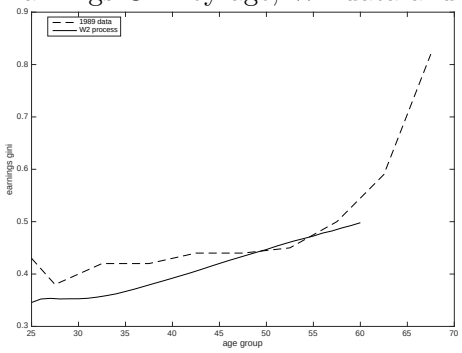


Figure 5: Variance of the earnings processes by age

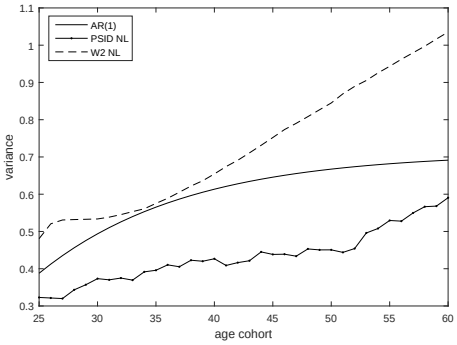


Figure 5 reports the variances of log earnings by age cohort for the 25-60 years old workers. This figure confirms the findings implied by the Gini coefficients. The NL PSID process generates less inequality than the benchmark, but displays, consistently with the data, a significant variance increase across the life-cycle. It should be noted, in fact, that in the benchmark AR(1), the variance of shocks during the lifetime is constant and that the realized increase in the variance of the earnings process over the life cycle is driven by the fact that the variance of initial conditions is assumed to be larger than the variance of the process itself. The NL synthetic W2 process generates a higher amount of inequality at every age and a very significant variance growth towards later ages (which is consistent with the one in our synthetic W2 data, see Appendix B). This explains the large Gini coefficient of earnings at age 60 that we observed earlier.

5.2 Earnings persistence and mobility

Despite showing a level of overall earnings inequality similar to the benchmark AR(1), the NL W2 earnings process displays lower persistence and higher levels of mobility, particularly at earlier ages. The NL PSID, which is less unequal and known to be affected by measurement error (Meghir and Pistaferri, 2004), generates even less persistent earnings.

The left panel in Figure 6 plots the autocorrelation coefficient of earnings by age. In the AR(1) this coefficient is constant across the life-cycle by construction, but our flexible

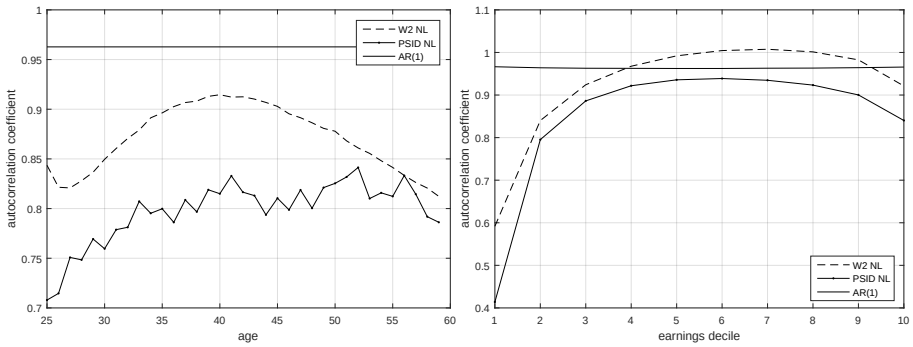


Figure 6: Autocorrelation coefficient by age (left panel) and previous earnings decile (right panel)

earnings processes capture significant changes in persistence as individuals get older. Namely, both PSID and W2 point to a lower level of persistence when individuals are young. In addition, earnings become less persistent at later ages in the W2 process.

Our processes also allow persistence to depend on the level of previous earnings. In both the PSID and W2 NL processes, persistence is significantly lower for earnings in the lowest decile (right panel in Figure 6) and largest for earners between the median and the 90th percentile. The interaction between age and previous earnings decile is reported in Figure 7.[14]

The lower persistence displayed by the NL PSID is consistent with previous studies that point to the existence of both a transitory component and noticeable measurement error in PSID data (for example, Bound, Brown, Duncan and Rodgers (1994) find that measurement error explains 22 percent of the variance of the rate of growth of earnings). Figure 8 shows that the implied persistence by age of the NL PSID process can be reconciled with a NL W2 process with a transitory shock/measurement error i.i.d. component of standard deviation equal to 0.25 (8 times larger than in our standard synthetic data set), which is the estimate

[14]Given that in our process there is no within-bin variation of earnings, persistence parameters conditional on earnings are very imprecisely estimated and parameters conditional on age and earnings cannot be computed.

We overcome this by fitting a polynomial $y_t \simeq f(y_{t-1}, t)$ and approximating the persistence parameter $\rho = \frac{\partial y_t}{\partial y_{t-1}}$ with its derivative $f'(y_{t-1}, t)$ evaluated at average earnings for each decile. We use a 6-degree Hermitian polynomial when conditioning only on y_{t-1}, and tensor products of a (3,2)-degree Hermitian polynomial on earnings and age when conditioning on both.

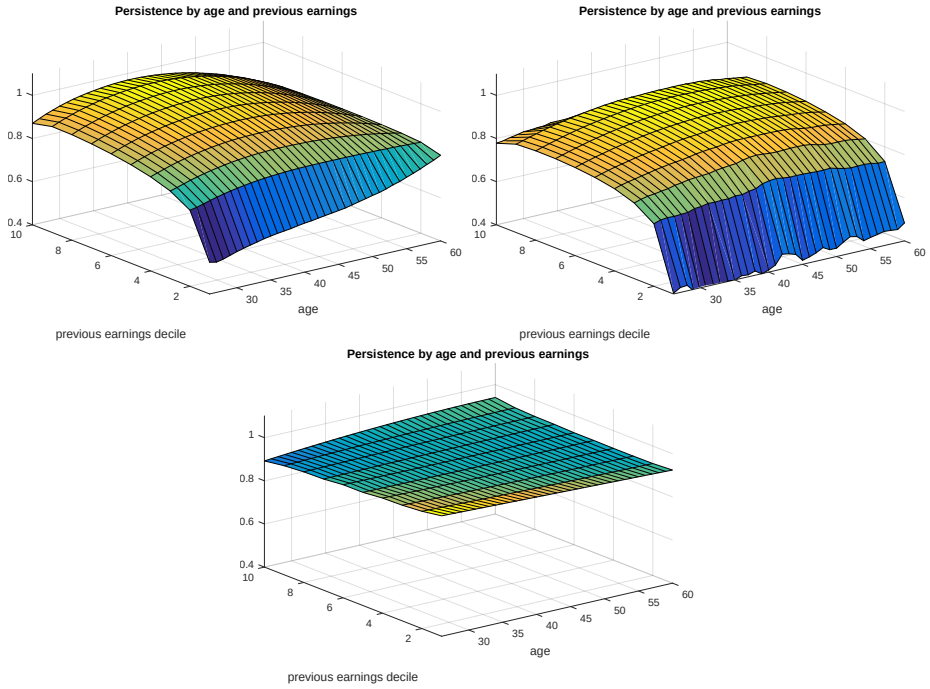


Figure 7: Autoregressive coefficient by previous earnings decile and age (top left: W2; top right: PSID; bottom: AR(1))

reported by Storesletten et al. (2004). Appendix B.3 provides further details about the performance of our processes in the presence of measurement error while Section 8 relaxes the assumption that earnings follow a first-order Markov process and shows that our allowing for second-order Markov structure does not significantly alter our findings.

5.3 Skewness

An important feature of earnings risk in the data is that it is not symmetrically distributed. As we have seen earlier, earnings shocks display substantial negative skewness, implying that large earnings drops are more likely than large earnings increases. Instead, a standard AR(1) with Gaussian innovations has constant, zero, skewness over the life cycle. This is reflected in Figure 9, which plots the conditional distributions of earnings faced by a person with average earnings at age 25 as this person ages. For instance, the top left panel shows the

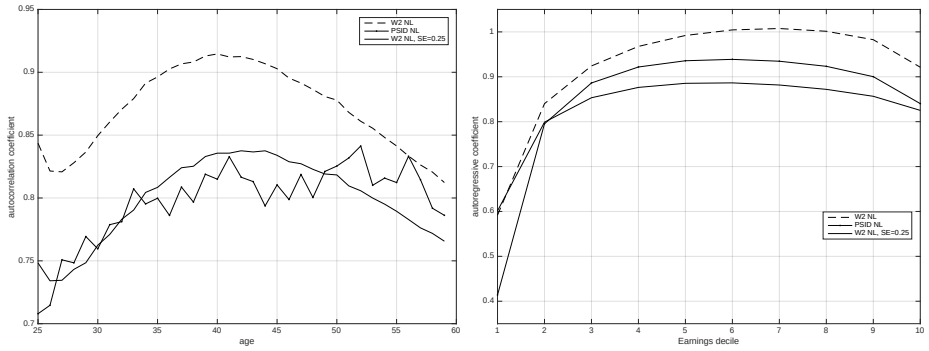


Figure 8: Autoregressive coefficient by age (left panel) and by previous earnings decile (right panel), including large ME process

future distributions of earnings that a 25-year-old individual will face throughout his lifetime, conditional on being born with average earnings. At any given age, by construction, there is a 90% probability that this individual will have earnings between both prediction bands, and therefore also a 5% probability of earning more than the upper bound and a 5% probability of earning less than the lower bound. As the figure illustrates, the probability of getting a very low earnings realization is larger in the W2 economy, while the probability of drawing a very high earnings realization is similar in both economies.

Skewness is significantly higher, in absolute value, before age 50.

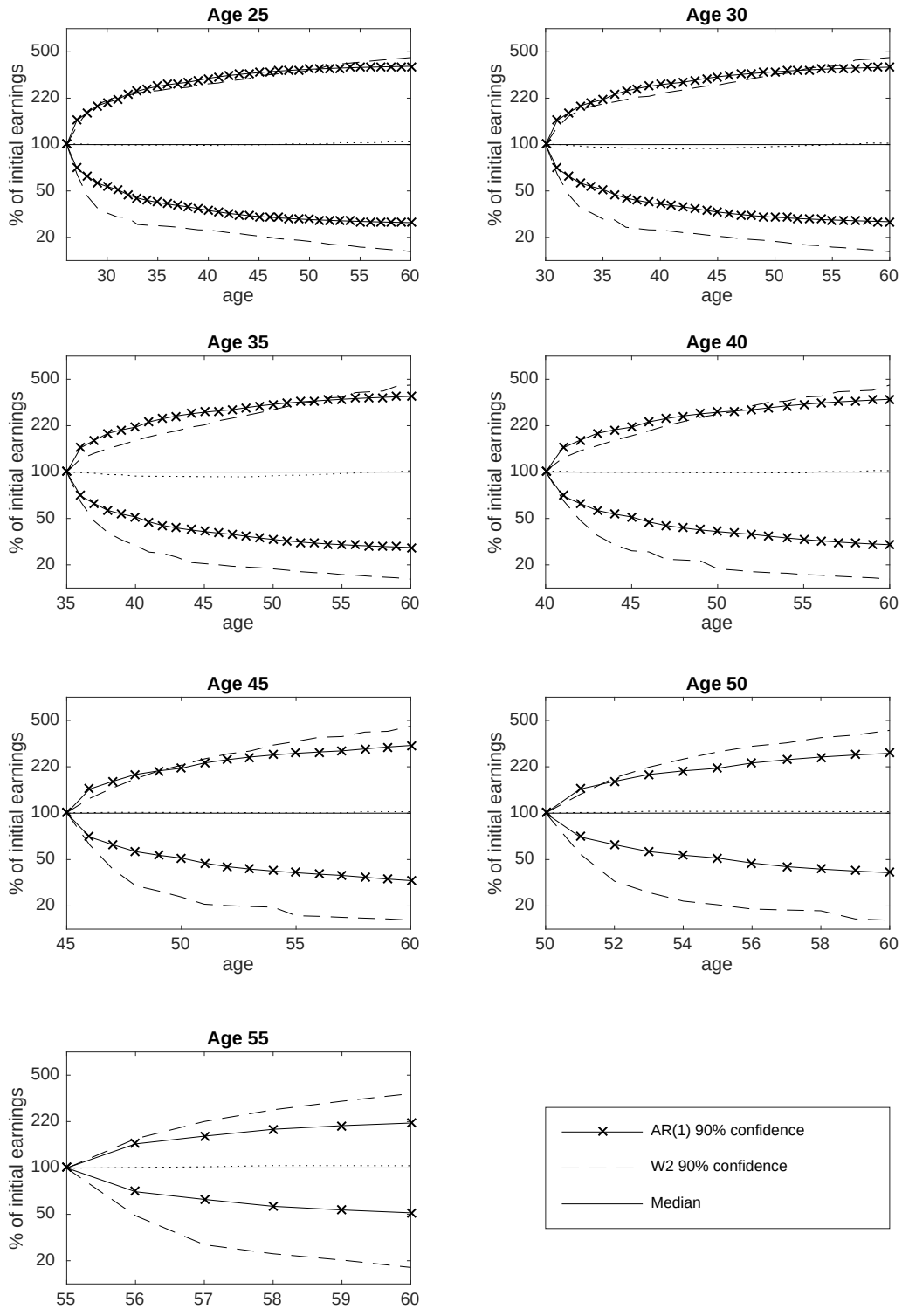


Figure 9: Distributions of future earnings, conditional on average earnings at a certain age (log scale for vertical axis)

6 The model

The model is based on Huggett (1996)'s paper. There is an infinitely lived government and overlapping generations of individuals who are equal at birth but receive idiosyncratic shocks to labor income throughout their working lives. The model period is one year.

6.1 Demographics

Each year, a positive measure of agents is born. Agents start working life at age 25 with no assets and a random productivity draw. At 30, each agent has $(1+n)$ children. Working life ends at 60 when agents retire.

Agents face a positive probability of dying throughout their lifetimes. This probability grows with age and is 1 at age 85; i.e., agents die for sure before turning 86 years old.

We restrict attention to stationary equilibria, hence variables are only indexed by age t.

6.2 Preferences and technology

Preferences are time separable, with a constant discount factor. The intra-period utility is CRRA: $u(c_t) = c_t^{1-\sigma}/(1-\sigma)$.

Agents are endowed with one indivisible unit of labor which they supply inelastically at zero disutility. The efficiency of their labor supply is subject to random shocks and follows a Markov chain of order 1 with (possibly) age-dependent state spaces and transition matrices.

We consider an open economy where prices (the risk free rate of return r and the wage per efficiency unit of labor w) are fixed.[15]

[15]The general equilibrium effects when prices are endogenized as in Aiyagari (1994*b*) are minor. For this reason, we report only partial equilibrium results.

6.3 Markets and the government

Asset markets are incomplete. Agents can only invest in the risk-free asset and cannot borrow. Since, there are no annuity markets to insure against the risk of premature death, there is a positive flow of accidental bequests in each period. These are distributed in equal amount b among all agents alive in the economy.

The government taxes labor earnings and capital income to finance an exogenous stream of public expenditure and a pension system. Income from capital and labor income are taxed at flat rates τ_k and τ_l respectively. Retired agents receive a lump-sum pension p from the government until they die.

6.4 The household's problem

For any given period, a t-year old agent chooses consumption c and risk-free asset holdings for the next period a', as a function of the state variables $x = (t, a, y)$. The term t indicates the agent's age, a indicates current asset holdings of the agent, and y stands for the earnings process realization. For given prices, the optimal decision rules are functions ($c(x)$ and $a'(x)$) that solve the dynamic programming problem described below.

(i) From age $t = 1$ to age $t = 36$ (from 25 to 60 years of age), the agent is working and has a probability of dying before the next period $1 - s_t$. The problem he solves is:

$$V(t,a,y) = \max_{c,a'} \Big\{ u(c) + \beta s_t E_t V(t+1, a', y') \Big\} \tag{2}$$
$$\text{s.t. } a' = \Big[1 + r\,(1-\tau_k)\Big]\, a - c + (1-\tau_l) w y + b, \quad a' \geq 0.$$

The evolution of y' follows the stochastic process described above. At every age t, y can lie in an age-specific grid y_t and its evolution towards y_{t+1} is determined by an age-specific transition matrix Q_{y_t}.

(ii) From t_r to T (from 61 to 86) agents no longer work and live off pensions and interest.

Their value function satisfies:

$$W(t,a) = \max_{c,a'}\Big\{u(c) + s_t\beta W(t+1,a')\Big\} \tag{3}$$
$$\text{s.t. } a' = \Big[1 + r\,(1-\tau_a)\Big]a - c + p + b, \quad a' \geq 0.$$

The terminal period value function $W(T+1,a)$ is set to equal 0 (agents do not derive utility from bequeathing).

The definition of equilibrium is in Appendix C.

6.5 The model calibration

A model period is one year. The coefficient of relative risk aversion is set to 1.5. The discount factor β is calibrated as the average of the discount factors that match a capital to income ratio of 3 in each of the three economies[16].

The fixed interest rate is 6% and the wage per efficiency unit of labor is normalized to 1.

The population growth rate n is set to 1.2% per year. The survival probabilities s_t are from Bell, Wade and Goss (1992). Government spending is 19% of GDP (g) (as in the the Council of Economic Advisors (1998) data), while capital tax rate τ_a is taken from Kotlikoff, Smetters and Walliser (1999), (see Table 3). The labor tax rate adjusts to balance the government budget constraint.

Table 3: Calibration parameters

σ	n	g	τ_k
1.5	0.012	0.19	0.2

[16]We also computed the results for a closed economy where the discount rate is calibrated independently in each economy to match a K/Y ratio of 3 with a standard Cobb-Douglas production function. Naturally, in the riskier W2 economy the calibrated discount rate, or equivalently the equilibrium interest rate, is slightly lower to discourage saving and keep the capital/output ratio constant. However, changes are small and do not significantly alter any of our conclusions.

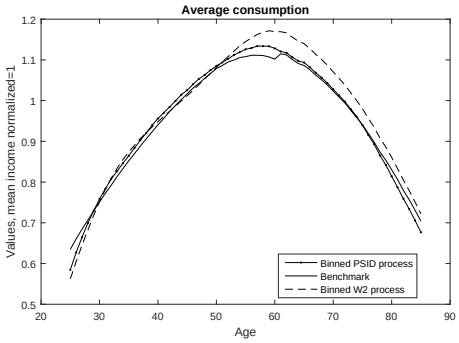
The social security pension benefit p equals 40% of the average earnings of a person in the economy (as in De Nardi (2004)).

7 The results from the model

7.1 Consumption

7.1.1 The consumption implications

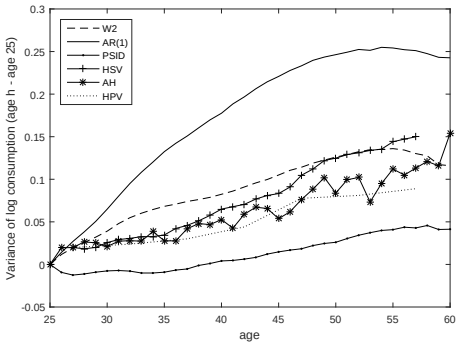
Figure 10: Average consumption profiles



The model generates a hump shaped pattern of **average consumption over the lifetime** (Figure 10), as in the data for the US (Carroll and Summers, 1991) and the UK (Attanasio and Weber, 2010). Given that we have imposed a common average income profile for the three earnings processes, any differences in the average consumption profiles stem from differences in precautionary saving due to differences in the riskiness of the respective processes. The average consumption profile has a higher initial level, grows at a slower rate and peaks at a lower level for the benchmark AR(1) process, reflecting the lower riskiness of the process and therefore lower precautionary saving. Precautionary saving in middle life (between age 45 and 60) seems also more important for the W2 process, likely reflecting the much higher level and rate of growth of earnings uncertainty it implies over those ages (see Figure 5).

Figure 11 plots the evolution the **variance of log consumption over the working life** implied by our three earnings processes and structural model, compared with the more recent estimates from the literature reported in Figure 2. The rate of increase of consumption inequality over the working life is informative about agents' ability to insure against earnings risk. For this reason, it provides a useful benchmark against which to assess the ability of the model to capture the degree of insurability of earnings shocks in the data.

Figure 11: Variances of log consumption



The benchmark AR(1) earnings process fails to generate a realistic age profile of consumption dispersion. Its growth rate is roughly double that implied by most of the estimates in the literature. At the other extreme, the NL PSID process fails to generate a sufficient increase in dispersion over the life-cycle. The NL synthetic W2 process, on the other hand, accurately replicates the overall increase in consumption dispersion, though it overestimates its rate of increase between age 30 and 35. In particular, it matches very closely the overall increase in the data reported in Heathcote, Storesletten and Violante (2014) (HSV) and is very similar to that from the other sources.

The finding that estimated richer earnings processes imply a profile of consumption dispersion in line with the data is quite remarkable. As we have discussed in Section 3, standard models with linear earnings processes generate a profile similar to the one for the benchmark AR(1) in Figure 11 in the absence of heterogeneity in life cycle earnings profiles. Our findings reveal that the age profile of the variance of consumption can alternatively be accounted for

by the response of saving to a richer earnings dynamics.

Table 4: Consumption distribution, data vs model

	Percentage held by the top					
Gini	1%	5%	20%	40%	60%	80%
U.S. Data (Non-durable consumption)						
.31	5	15	40	62	79	92
Benchmark AR(1)						
.35	5	17	43	66	81	92
NL Synthetic W2 Process						
.33	4	15	41	64	80	92
NL PSID Process						
.25	4	13	35	58	76	90

Finally, Table 4 reports the cross-sectional consumption distribution both in the data and generated by the model for different earnings processes. The distribution generated by the NL W2 process is remarkably close to that in the data. On the other hand, the distributions generated by the benchmark AR(1) and the NL PSID process respectively significantly overstate and understate the extent of overall inequality in the data.

7.1.2 Accounting for consumption inequality

Our rich earnings process deviates from the standard linear process for earnings, such as the benchmark AR(1), commonly used in the heterogeneous-agents literature in a number of ways. In particular, it has:

1. Different parameter values, in particular a lower persistence—$\rho = 0.87$ versus 0.96 for the benchmark AR(1).

2. Age-dependent persistence;

3. Age-dependent variance of innovations;

4. Non-normal innovations (negative skewness and high kurtosis);

5. State-dependent second and higher moments.

In order to understand the contribution of each of these factors to the substantially lower growth in consumption variance over the life cycle we conduct a series of counterfactual experiments, simulating the model under progressively richer stochastic processes for earnings. In particular, we start from the benchmark AR(1) process with $\rho = 0.96$ and then add sequentially each of the above extra properties. Note, that the first four properties can be introduced while maintaining the AR(1) structure as they do not imply any non-linearity in earnings (properties 2 and 3 introduce non-linearities only in age/time). Conversely, property 5 cannot be accounted for within an AR(1) framework. So, the contribution of property 5 obtains as the residual of our findings for our NL W2 process and a generalized AR(1) process encompassing properties 1 to 4.

We implement our series of experiments in the following way. First, we estimate progressively more flexible parametric AR(1) processes for earnings on our synthetic W2 data, introducing incrementally properties 1 to 4 above. Second, we use each of these estimated AR(1) processes to simulate a large panel of individual earnings histories. This panel constitutes our counterfactual dataset for the experiment in question. Third, we apply our method and estimate a first-order, age-dependent Markov chain on the panel thus generated to obtain a discretized process that matches the features of the data. Finally, we feed the estimated Markov chain to the model. In all cases, we recalibrate the discount factor to maintain the wealth-income ratio at its target value of 3. Figure 12 plots the consumption variance profile for the various experiments.

For our first experiment we estimate a standard AR(1) process with normal initial condition and innovations on our synthetic dataset. The only difference compared to our benchmark AR(1) is due to different estimated parameter values; namely, the parameter estimates are $\sigma^2_{y_0} = 0.48,\ \gamma = 0.87$ and $\sigma^2_\epsilon = 0.19$ against $\sigma^2_{y_0} = 0.38,\ \gamma = 0.96$ and $\sigma^2_\epsilon = 0.045$ for the benchmark AR(1).

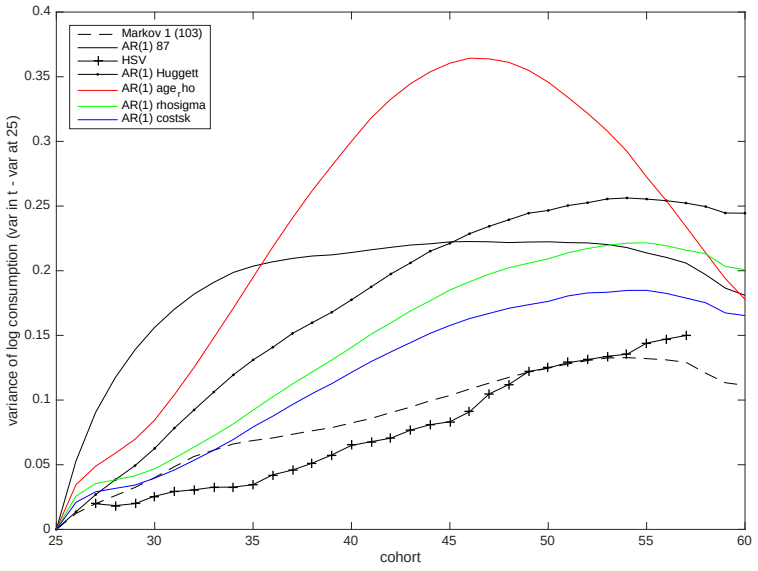


Figure 12: Age profile of consumption inequality for different specifications of the earnings process.

The black solid line in Figure 12 plots the consumption variance profile for this experiment. Comparing it to the profile for the benchmark AR(1) (black solid line with circles) and for our flexible case (dashed line) makes clear that the different evolution of the consumption variance cannot be accounted for by differences in parameter values for the same parametric functional form. If anything, the parameter values estimated on our synthetic dataset imply an even higher growth in the consumption variance until age 45 relative to the benchmark AR(1).

In our second experiment, we estimate an AR(1) process with the same variances as in the previous case but with the age-dependent autoregressive coefficients plotted in Figure 6. The red line in Figure 12 the consumption variance profile for this case. It is clear, that while the below-average persistence early in life flattens the profile until about age 30, the above-average persistence between age 30 and 50 implies a dramatically higher rate of growth in that age bracket.

The third experiment relaxes the assumption of age-independent variances for the earnings innovation while maintaining normality. The green line in Figure 12 plots the consumption variance profile for this case. Relative to the previous experiment, allowing for

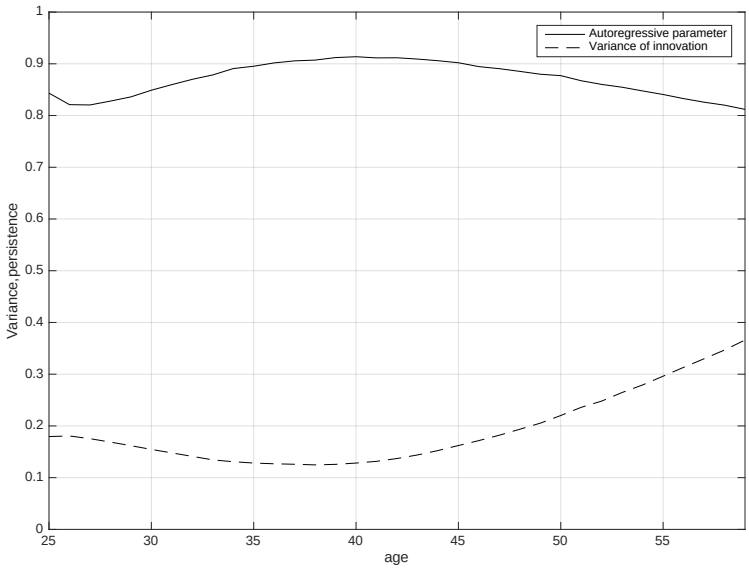


Figure 13: Age-dependent values of ρ (solid line) and σ_ε^2 (dashed line).

heteroskedasticity dramatically reduces the rate of growth of the consumption variance. Intuitively, the standard permanent income mechanism implies that the variance of consumption is an increasing function of earnings persistence—which affects the response of consumption to an earnings innovation—and the variance of earnings innovations itself. In the data, earnings persistence and the variance of earnings innovations are negatively correlated over the life cycle, as illustrated by Figure 13 that plots the estimated autoregressive coefficient and the innovation variances at the different ages. Comparing the red and green lines in Figure 12 reveals that, until age 57, the effect of the U-shaped pattern of the innovation variance over the life cycle more than offsets the effect of the hump-shaped pattern of earnings persistence. As a result, the rate of growth of consumption dispersion between age 25 and its peak at age 55 falls by about 3 log points relative to the benchmark AR(1) cased. Finally, our fourth experiment introduces skewness and kurtosis of the innovation. Their estimates are respectively -1.13 and 11.31. The blue line in Figure 12 plots the associated consumption variance profile. The fact that shocks have negative skewness and high kurtosis induces individuals to increase their precautionary saving and further reduces the rate of growth of the consumption variance. Relative to the benchmark AR(1) the overall rate of

growth between age 25 and 55 falls by about 8 percentage points. This accounts for more than half of the difference between our flexible process and the benchmark AR(1). Allowing for age-varying skewness and kurtosis does not change this pattern in a meaningful way. For this reason we do not report results for that case.[17]

The remaining 6 percentage point difference in the rate of growth between age 25 and 55 is due to the fact that our flexible process allows conditional moments to vary with the previous earnings realization, as documented in Figure 1.

7.2 Wealth

7.2.1 The wealth implications

We now compare the key statistics of the wealth distributions generated by our benchmark AR(1) earnings process with those of our non-parametrically estimated earnings processes[18].

Table 5: Wealth distribution statistics

Capital-output ratio	Bequests-output ratio	Wealth Gini	Percentage wealth in the top 1%	5%	20%	40%	60%	80%	Percentage with negative or zero wealth
U.S. data									
3.0	2.6%	.72	28	49	75	89	96	99	5.8–15.0
Benchmark AR(1)									
3.0	2.8%	.72	11	35	74	93	99	100	19
NL Synthetic W2 Process									
3.0	2.7%	.67	10	31	68	89	97	99.75	8
NL PSID Process									
2.9	2.7%	.63	9	28	64	86	96	99.54	6

The benchmark AR(1) calibration suffers from the limitations that have already been analyzed in the literature. Although it matches the wealth Gini coefficient we observe in the data, it does so by generating too many people with zero assets (the left tail is too fat) and

[17]The results are available upon request.

[18]Source for US data: De Nardi (2004), except for bequest-output ratio, which is taken from Villanueva (2005)

not enough concentration at the top (the right tail is too thin). For instance, the number of people with zero assets[19] is (19%), which is even higher than the highest estimate in the data, while the richest 1% only hold 12% of total wealth, compared with 28% in the data.

Both the NL PSID process and the NL synthetic W2 better match the left tail of the distribution. They generate a substantially smaller amount of agents with zero assets (6% and 8% respectively)) and better fit the holdings of the poorest 60% of people: in the distribution generated by the NL W2 process, the poorest 60% hold 11% of total net worth, which matches the actual data, and in the distribution derived from the NL PSID process they hold 14%, which is still much closer to the 11% observed in the data, compared to only 7% implied by the benchmark AR(1) earnings process.

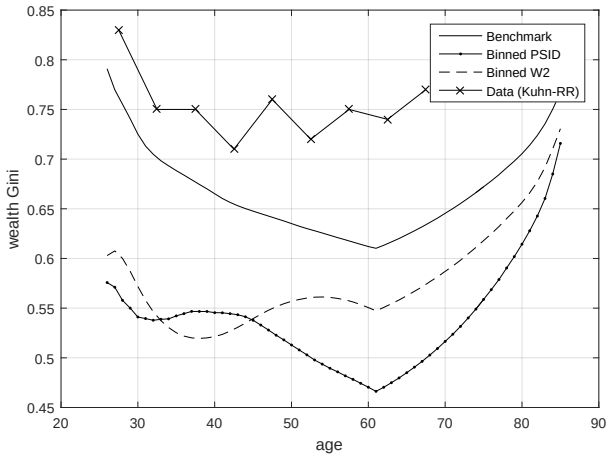
However, our computations show that even our most more flexible earnings processes do not help better match the right tail of the distribution. Although Castañeda et al. (2003) have shown that a highly skewed earnings process can generate a long right tail for the wealth distribution, it turns out that even the NL synthetic W2 process, which as shown earlier has a higher degree of earnings inequality and negative skewness than our NL PSID process, generates too little wealth concentration among top earners.

Summarizing, our empirical method, applied to synthetic W2 data, performs approximately as well as the benchmark AR(1) in the right tail, outperforming it in the left tail. This is remarkable given that neither of our empirical processes have been calibrated to match any particular moments of the wealth distribution. The inclusion of inter-generational links and bequest transmission, as in De Nardi (2004), is likely to take these results closer to the actual data, particularly by improving the fit of the right tail of the distribution. It should also be considered that we are not taking inter-vivos transfers into account, which may represent an additional source of wealth, particularly at earlier ages.

The absence of these important mechanisms in our model can also explain why the W2

[19]Given that all newborns in the model start life with zero wealth by assumption, we do not include them in any of the model computations of the fraction of agents with zero wealth.

Figure 14: Wealth Gini by cohort



Source for the data (1989): Kuhn and Ríos-Rull (2015)

process, although it generates a qualitatively reasonable wealth Gini age profile, underestimates the actual level of inequality at each age (Figure 14). In this case, it is the AR(1) benchmark which delivers the largest wealth Gini at every age, but, as argued earlier, this is generated by pushing too many agents towards zero. Besides, it implies a very large drop in wealth Gini from ages 20 to 60.

7.2.2 Accounting for the wealth distribution

As for consumption, one wants to understand the contribution of the various novel features of the earnings process to the differences in the wealth distributions relative to the benchmark AR(1) process. For this reason, we conduct the same set of experiments as in Section 7.1.2. Table 6 reports wealth distribution statistics for each of these counterfactual simulations. For ease of comparison, the first three rows in the table are the same as in Table 5.

We have already seen that both the benchmark AR(1) and our NL W2 process, even though the latter includes realistic earnings risk for top earners, underestimate the wealth holdings of the richest 5%. This naturally extends to all the decomposition experiments. The better fit of the the bottom 60% of the wealth distribution by the NL W2 is due to the

Table 6: Wealth distribution statistics

Capital-output ratio	Bequests-output ratio	Wealth Gini	Percentage wealth in the top 1%	5%	20%	40%	60%	80%	Percentage with negative or zero wealth
U.S. data									
3.0	2.6%	.72	28	49	75	89	96	99	5.8–15.0
Benchmark AR(1), $\rho = 0.96$									
3.0	2.8%	.72	11	35	74	93	99	100	19
NL Synthetic W2 Process									
3.0	2.7%	.67	10	31	68	89	97	99.75	8
AR(1) with ρ=0.87									
3.0	2.4%	.70	10	33	71	91	98	99.98	16
AR(1) with age-varying ρ and σ									
3.0	2.4%	.69	10	32	70	91	98	99.97	16
AR(1) with age-varying ρ, σ, skewness and kurtosis									
3.0	2.5%	.68	11	33	70	90	98	99.93	14

following factors. Two percentage points of the fall in wealth holdings by the top 40% (from 93 to 91 percent) are due to the lower average persistence; i.e., $\rho = 0.87$ instead of $\rho = 0.96$ for the benchmark AR(1). However, this earnings process still generates a counter-factually large amount of individuals with zero assets. Introducing both age-dependent persistence and heteroskedasticity of the innovations does not change these findings. Introducing also skewness and kurtosis increases precautionary saving by the poorest individuals and reduces the share of wealth held by the top 40% by 1 percentage points and the share of individuals with zero wealth by 2 percentage points. The remaining difference is accounted for by the fact that our flexible process allows moments to vary with the previous earnings realization. While this feature accounts for only 1 percentage point in the share of wealth held by the top 40%, it generates a large reduction—from 14 to 8 per cent—in the share of individuals with zero wealth.

7.2.3 Saving functions

Given the differences in earnings risk faced by households, saving functions differ across economies. Figure 15 shows the saving rates of a 30-year-old individual with zero wealth conditional on his last earnings realization. Under both earnings processes, agents with

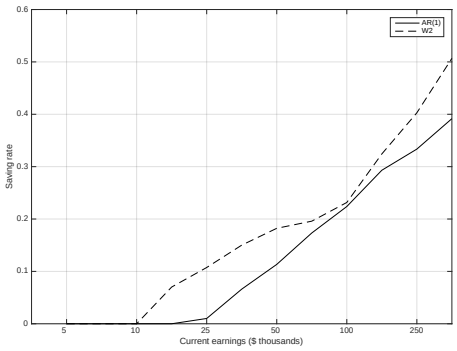


Figure 15: Saving rate by previous earnings for a 30-year-old with no wealth.

zero wealth that receive a very low earnings realization do not save as they expect their earnings will be higher next period. In contrast, agents with larger earnings realizations save significantly more in the W2 than in the AR(1) economy (an individual with average earnings, for example, saves 16% of his earnings in the former and 7% in the latter). This differential behaviour is due to the earnings implications that the W2 and AR(1) individuals face when they receive a positive earnings shock and, in particular, to the large negative skewness and high kurtosis of the W2 process (see Figure 1 and the discussion in Section 5.3).

Exogenous differences in the earnings processes account for a very small part of the variations observed. In order to disentangle them from the endogenous effect of agents' decicion, we counter-factually introduce the W2 savings function into the model with the AR(1) earnings process, and viceversa, still respecting both the no-borrowing and budget constraints. Table 7 shows that the number of individuals at the borrowing constraint varies very little as we change the earnings process but not the policy function. On the other hand,

the share of borrowing constrained individuals does change significantly when changing the savings function the earnings process. As a result of the earnings risk and especially the earnings skewness, individuals engage in more precautionary saving under the W2 earnings process, which implies that fewer households are close to the borrowing constraint in the W2 economy.

	AR(1) earnings process	W2 earnings process
AR(1) savings function	17%	14%
W2 savings function	5%	3%

Table 7: Proportion of working-age individuals with zero wealth

8 Second-order Markov process for earnings

8.1 Relative persistence

The main limitation of our approach is that it does not allow for a purely transitory shock, while MaCurdy (1982) and Abowd and Card (1989) document, using survey data sets, that the process for earnings is well approximated as a linear combination of a unit-root and a transitory process. Following this result, several papers in the heterogeneous-agent literature on consumption inequality (starting with Storesletten et al. (2004)) model earnings as the sum of a permanent and a transitory component, or use a (near unit-root) AR(1) process whose persistence comes from an estimated process that allows for a purely transitory shock (while the transitory component is ignored in the model).

By construction the forecasting performance of our non-linear, first-order Markov process is better than that of a linear AR(1) at one-period horizon. Yet, by not allowing for a purely transitory component, our process may underestimate earnings persistence, raising the question of how well it forecasts at longer horizons. In this section, we examine this question and compare the forecasting ability of the unit-root AR(1) process that is commonly

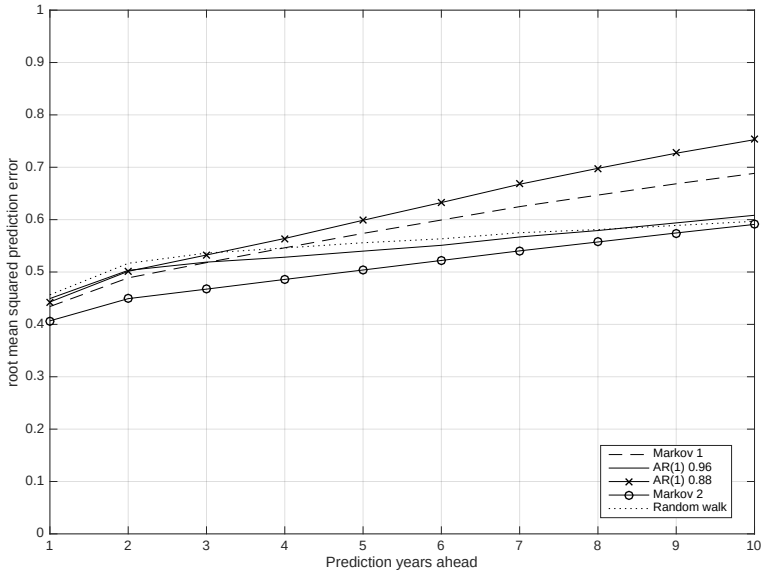


Figure 16: Root Mean Squared Forecast Error by process and horizon

used in the literature with those of our Markov chain of order one and with a Markov chain of order 2.

Apart from the longer memory, our method is unchanged when we allow for a Markov chain of order 2. Namely, denoting by N the number of quantiles partitioning the marginal earnings distribution at each age, the age-dependent transition probability matrices have dimension $(N^2 \times N)$. Their elements are constructed as the proportion of individuals moving from any possible two-period history sequence of quantiles into a quantile next period. For tractability, we partition the marginal earnings distribution into N=18 quantiles, namely the top and bottom five 2% percentiles and the 9 intermediate deciles (this is the same partition we adopted for the PSID). As before, we attribute to each individual in a particular bin the average earnings level in that bin.

Figure 16 shows the root mean forecast errors for log earnings at different horizons for the various processes we consider; namely, (1) a random walk; (2) an AR(1) estimated on our synthetic dataset with $\rho = 0.88$; (3) our discretized first-order process with 103 grid points (Markov 2); (4) a discretized second-order Markov process with 18 grid points (Markov 2).

To interpret the values reported, remember that a mean absolute forecast error of x for

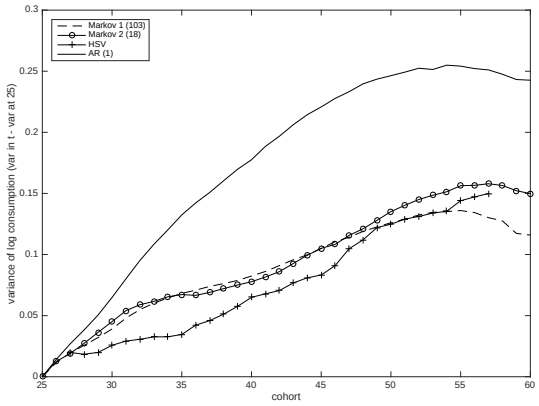


Figure 17: Life-cycle growth of the variance of log consumption

log earnings has the same order of magnitude as the percentage forecast error for the level of earnings. The AR(1) with $\rho = 0.88$ has by far the worse performance. The discretized first-order process perform better than the AR(1) with $\rho = 0.96$ and the random walk up to respectively 3 and 4 periods ahead. Even 5 to 6 years out its performance is never worse that 10% compared to the best of the other two processess.

Finally, the Markov 2 process outperforms all others processes at all horizons.

8.2 Model simulation

This section reports the implications of the second-order Markov process for consumption and wealth and compares them to their counterpart for the first-order process.

8.2.1 The consumption and implications

Figure 17 plots the age profile of log consumption inequality for the M2(18) process against the profile for the M1(103) and the empirical estimates in Figure 2. It is apparent that the growth rate of consumption dispersion over the life-cycle is similar for both processes and in line with the literature.

Finally, Table 8 compares the implications of the two alternative discretizations for the wealth distributions. Again, the differences between the first and second-order Markov pro-

cess are negligible.

Due to these findings on consumption and wealth and to the one that our proposed discretization method perform well also at longer horizons (on the W2 synthetic data set), we conclude that the differences in the consumption and wealth distributions implied by our estimated non-linear earnings process relative to the benchmark AR(1) are driven mainly by its non-linearity and asymmetry and not by a misspecified autocorrelation structure.

Table 8: Wealth distribution statistics

Capital-output ratio	Bequests-output ratio	Wealth Gini	Percentage wealth in the top 1%	5%	20%	40%	60%	80%	Percentage with negative or zero wealth
U.S. data									
3.0	2.6%	.72	28	49	75	89	96	99	5.8–15.0
NL Synthetic W2 Process (Markov 1, 103)									
3.0	2.7%	.67	10	31	68	89	97	99.75	8
Markov 2 with 18 gridpoints									
3.0	2.8%	.66	10	31	68	88	97	99.69	8

9 Conclusions

Our main findings are the following. First, allowing for non-linearity and asymmetry in earnings reveals that earnings are significantly less persistent at low and high earnings levels relative to an AR(1) process, like the one used by Huggett (1996), which is rather typical of the quantitative macro literature with heterogeneous agents. Second, our flexible earnings process, when introduced in a standard quantitative model of consumption and savings over the life cycle, generates a much better fit of the growth in consumption dispersion as a function of age and of the wealth holdings of the bottom 60% of people. On the other hand, the model still generates too little wealth concentration at the top of the wealth distribution, just as the the model with AR(1) earnings process. This is a frequent drawback of life-cycle

models with no entrepreneurship or transmission of bequests (De Nardi, 2004), (Cagetti and De Nardi, 2006), and (Cagetti and De Nardi, 2009).

Finally, and importantly, our proposed discretization method also perform well at longer horizons. In addition, our results are robust to allowing for a second order Markov chain that outperforms, both in the short and the long run, the earnings predictive ability of all other earnings processes typically used in quantitative macro model.

References

Abowd, John and Card, David (1989), 'On the covariance structure of earnings and hours changes', *Econometrica* **57**(2), 411–45.

Aguiar, Mark and Hurst, Erik (2013), 'Deconstructing life cycle expenditure', *Journal of Political Economy* **121**(3), 437–492.

Aiyagari, S. R. (1994*a*), 'Uninsured idiosyncratic risk, and aggregate saving', *Quarterly Journal of Economics* **109**(3), 659–684.

Aiyagari, S. Rao (1994*b*), 'Uninsured idiosyncratic risk and aggregate saving', *Quarterly Journal of Economics* **109**(3), 659–684.

Altonji, Joseph G., Smith, Anthony A. and Vidangos, Ivan (2013), 'Modeling earnings dynamics', *Econometrica* **81**(4), 1395–1454.

Arellano, Manuel, Blundell, Richard and Bonhomme, Stéphane (2015), Earnings and consumption dynamics: A non-linear panel data framework. unpublished manuscript.

Attanasio, Orazio P. and Weber, Guglielmo (2010), 'Consumption and saving: Models of intertemporal allocation and their implications for public policy', *Journal of Economic Literature* **48**, 693–751.

Bell, Felicitie C., Wade, Alice H. and Goss, Stephen C. (1992), 'Life tables for the United States Social Security Area: 1900-2080', (Social Security Administration, Office of the Actuary).

Blundell, Richard (2014), 'Income Dynamics and Life-cycle Inequality: Mechanisms and Controversies', *The Economic Journal* **124**(576), 289–318.

Blundell, Richard, Graber, Michael and Mogstad, Magne (2015), 'Labor income dynamics and the insurance from taxes, transfers and the family', *Journal of Public Economics* **127**, 58–73.

Bonhomme, Stéphane and Robin, Jean-Marc (2009), 'Assessing the equalizing force of mobility using short panels: France, 1990–2000', *The Review of Economic Studies* **76**(1), 63–92.

Bound, John, Brown, Charles, Duncan, Greg J and Rodgers, Willard L (1994), 'Evidence on the validity of cross-sectional and longitudinal labor market data', *Journal of Labor Economics* pp. 345–368.

Bricker, Jesse, Krimmel, Jacob, Henriques, Alice and Sabelhaus, John (2016), 'Measuring income and wealth at the top using administrative and survey data', Mimeo.

Browning, Martin, Ejrnaes, Mette and Álvarez, Javier (2010), 'Modelling income processes with lots of heterogeneity', *Review of Economic Studies* **77**, 1353–1381.

Budria, Santiago, Diaz-Gimenez, Javier, Quadrini, Vincenzo and Rios-Rull, Jose-Victor (2002), 'New facts on the distribution of earnings, income and wealth in the U.S.', *Federal Reserve Bank of Minneapolis Quarterly Review* **26**, 2–35.

Cagetti, Marco and De Nardi, Mariacristina (2006), 'Entrepreneurship, frictions and wealth', *Journal of Political Economy* **114**(5), 835–870.

Cagetti, Marco and De Nardi, Mariacristina (2008), 'Wealth inequality: Data and models', *Macroeconomic Dynamics* **12**(S2), 285–313.

Cagetti, Marco and De Nardi, Mariacristina (2009), 'Estate taxation, entrepreneurship, and wealth', *American Economic Review* **99**(1), 85–111.

Carroll, Christopher D. and Summers, Lawrence H. (1991), Consumption growth parallels

income growth: Some new evidence, *in* B. DouglasBernheim and John B.Shoven, eds, 'National Saving and Economic Performance', University of Chicago Press, pp. 305–348.

Carroll, Christopher D, Hall, Robert E and Zeldes, Stephen P (1992), 'The buffer-stock theory of saving: Some macroeconomic evidence', *Brookings papers on economic activity* pp. 61–156.

Castañeda, Ana, Díaz-Giménez, Javier and Ríos-Rull, José-Víctor (2003), 'Accounting for the U.S. Earnings and Wealth Inequality', *Journal of Political Economy* **111**(4), pp. 818–857.

Civale, Simone, Díez-Catalán, Luis and Fazilet, Fatih (2016), Discretizing a process with non-zero skewness and high kurtosis. Mimeo, University of Minnesota.

Conesa, Juan Carlos, Kitao, Sagiri and Krueger, Dirk (2009), 'Taxing capital? not a bad idea after all!', *American Economic Review* **99**, 25–48.

Council of Economic Advisors (1998), 'Economic Report of the President', (Washington: United States Government Printing Office).

De Nardi, Mariacristina (2004), 'Wealth inequality and intergenerational links', *Review of Economic Studies* **71**, 743–768.

De Nardi, Mariacristina (2015), 'Quantitative models of wealth inequality: A survey', National Bureau of Economic Research working paper no. 21106.

Deaton, Angus and Paxson, Christina (1994), 'Intertemporal choice and inequality', *Journal of Political Economy* **102**(3), 437–67.

Fisher, Jonathan D, Johnson, David S and Smeeding, Timothy M (2013), 'Measuring the trends in inequality of individuals and families: Income and consumption', *The American Economic Review* **103**(3), 184–188.

Geweke, John and Keane, Michael (2000), 'An empirical analysis of earnings dynamics among men in the psid: 1968-1989', *Journal of Econometrics* **96**, 293–356.

Graber, Michael and Lise, Jeremy (2015), 'Labor market frictions, human capital accumulation and consumption inequality', mimeo, University College London.

Guvenen, Fatih (2007), 'Learning your earning: Are labor income shocks really very persistent?', *American Economic Review* **97**(3), 687–712.

Guvenen, Fatih, Karahan, Fatih, Ozkan, Serdar and Song, Jae (2015), What do data on millions of U.S. workers reveal about life-cycle earnings risk? Working paper, University of Minnesota.

Haider, S. J. (2001), 'Earnings instability and earnings inequality of males in the united states: 1967-1991', *Journal of Labor Economics* **19**(4), 799–836.

Hansen, Gary D. (1993), 'The cyclical and secular behaviour of labour input: Comparing efficiency units and hours worked', *Journal of Applied Econometrics* **8**(1), 71–80.

Heathcote, Jonathan, Perri, Fabrizio and Violante, Giovanni L (2010), 'Unequal we stand: An empirical analysis of economic inequality in the united states, 1967-2006', *Review of Economic dynamics* **13**(1), 15–51.

Heathcote, Jonathan, Storesletten, Kjetil and Violante, Giovanni L (2005), 'Two views of inequality over the life cycle', *Journal of the European Economic Association* **3**(2-3), 765–775.

Heathcote, Jonathan, Storesletten, Kjetil and Violante, Giovanni L (2014), 'Consumption and labor supply with partial insurance: An analytical framework', *American Economic Review* **104**(7), 2075–2126.

Huggett, Mark (1996), 'Wealth distribution in life-cycle economies', *Journal of Monetary Economics* **38**(3), 469–494.

Huggett, Mark, Ventura, Gustavo and Yaron, Amir (2011), 'Sources of lifetime inequality', *American Economic Review* **101**(7), 2923–54.

Jäntti, Markus and Jenkins, Stephen P. (2015), Income mobility, *in* Anthony B.Atkinson and FrançoisBourguignon, eds, 'Handbook of Income Distribution', Vol. 2 of *Handbook of Income Distribution*, Elsevier, pp. 807–935.

Jonathan Heathcote, Kjetil Storesletten, Giovanni L. Violante (2010), 'The macroeconomic implications of rising wage inequality in the united states', *Journal of Political Economy* **118**(4), 681–722.

Kotlikoff, Laurence J., Smetters, Kent and Walliser, Jan (1999), 'Privatizing social security in the united states - comparing the options', *Review of Economic Dynamics* **2**(3), 532–574.

Krueger, Dirk and Perri, Fabrizio (2006), 'Does income inequality lead to consumption inequality? evidence and theory', *The Review of Economic Studies* **73**(1), 163–193.

Kuhn, Moritz and Ríos-Rull, José-Víctor (2015), '2013 update on the us earnings, income and wealth distributional facts: a view from macroeconomics', Working paper.

Lillard, Lee A. (1977), 'Inequality: earnings vs. human wealth', *American Economic Review* **67**, 42–53.

Lillard, Lee A. and Willis, Robert J. (1978), 'Dynamic aspects of earning mobility', *Econometrica* **46**(5), 985–1012.

MaCurdy, Thomas E. (1982), 'The use of time series processes to model the error structure of earnings in a longitudinal data analysis', *Journal of Econometrics* **18**, 83 – 114.

Meghir, Costas and Pistaferri, Luigi (2004), 'Income variance dynamics and heterogeneity', *Econometrica* **72**(1), 1–32.

Meghir, Costas and Pistaferri, Luigi (2011), 'Earnings, consumption and life cycle choices', *Handbook of Labor Economics* **4**, 773–854.

Meyer, Bruce D and Sullivan, James X (2010), 'Consumption and income inequality in the us since the 1960s', University of Chicago manuscript.

Primiceri, Giorgio E. and van Rens, Thijs (2009), 'Heterogeneous life-cycle profiles, income risk and consumption inequality', *Journal of Monetary Economics* **56**(1), 20 – 39.

Quadrini, Vincenzo and Ríos-Rull, José-Víctor (1997), 'Understanding the US distribution of wealth', *Federal Reserve Bank of Minneapolis Quarterly Review* **21**(2), 22–36.

Quadrini, Vincenzo and Ríos-Rull, José-Víctor (2014), Inequality in macroeconomics, *in* A.B.Atkinson and F.Bourguignon, eds, 'Handbook of Income Distribution, vol. 2B', Elsevier, chapter 15.

Sabelhaus, John and Song, Jae (2010), 'The great moderation in micro labor earnings', *Journal of Monetary Economics* **57**(4), 391–403.

Scholz, John Karl, Seshadri, Ananth and Khitatrakun, Surachai (2006), 'Are americans saving optimally for retirement?', *Journal of Political Economy* **114**, 607–643.

Shorrocks, Antony F. (1980), Income stability in the United States, *in* N.Klevmarken and J.Lybeck, eds, 'The statics and dynamics of income', Tieto Ltd.

Slesnick, Daniel T. and Ulker, Aydogan (2004), Inequality and the life cycle: Age, cohort effects and consumption. Mimeo.

Storesletten, Kjetil, Telmer, Christopher I. and Yaron, Amir (2004), 'Consumption and risk sharing over the life cycle', *Journal of Monetary Economics* **51**(3), 609–633.

Tauchen, George (1986), 'Finite State Markov-Chain Approximations to Univariate and Vector Autoregressions', *Economics Letters* **20**(2), 177–181.

Tauchen, George and Hussey, Robert (1991), 'Quadrature-Based Methods for Obtaining Approximate Solutions to Nonlinear Asset Pricing Models', *Econometrica* **59**(2), 371–96.

Villanueva, Ernesto (2005), 'Inter vivos transfers and bequests in three OECD countries.', *Economic Policy* **20**(43), 506–565.

Wolff, Edward N. (1987), 'Estimates of household wealth inequality in the u.s., 1962-1983', *Review of Income and Wealth* **33**(3), 251–257.

A Appendix for online publication only: The PSID and synthetic W2 data

A.0.2 PSID data

The Panel Study of Income Dynamics (PSID) follows a large number of US households over time and reports information about their demographic characteristics and sources of income. The PSID was initially composed of two major subsamples. The first of them, the SRC (Survey Research Center) or core subsample, was designed to be representative of the US population and is a random sample itself. The second, the SEO (Survey of Economic Opportunity) subsample, was created to study the characteristics of the most deprived households. Later, Immigrant and Latino subsamples were also added to the PSID.

From 1968 to 1997, the survey was yearly. After 1997, it started having a biennial structure. Since the model period is one year, we restrict ourselves to the yearly part of the survey (1968-1997). We only consider the SRC or core subsample because the SEO oversamples the poor. After dropping the SEO and Latino samples we are left with a random sample, which makes computations simpler since weights are not needed Haider (2001) [20].

Because our binning strategy requires a large amount of data per age group, we drop individuals still working after age 60 and we assume that people retire at 60. For comparability with the W2 synthetic dataset that we derive from the econometric processes in Guvenen et al. (2015), that start age age 25, we also drop individuals below age 25.

In line with previous literature, we restrict our study to male heads of household. A valid observation is formed by two values for individual labor income: this year's and next year's. Neither of the two income values can be missing. We deflate labor income by CPI-U (taken from FRED economic data, where index 100=1982-1984). We do not consider

[20] In our case, we are creating a dataset of pairs of income observations from the PSID panel. It must be taken into account that the weighting of our final dataset can be affected by attrition and by the fact that we are neglecting observations of yearly income under 900 $ (expressed in 1982 dollars)

earnings observations with values under 900$. This avoids the problem of taking logs of very small numbers and is in line with usual practice in previous literature (De Nardi (2004) or Guvenen et al. (2015), for instance).

Labor income does not include unemployment or workers' compensation, nor earnings of the spouse, because adding them would reduce the sample period to 1977-1992, due to data availability issues.

The major shortcomings the PSID data are that the sample size is not very large (and the earnings-rich are not oversampled), that they are less likely to respond to surveys, and that earnings are top-coded. As a result, our right tail of the income distribution from this data set is not accurately estimated (Budria, Diaz-Gimenez, Quadrini and Rios-Rull (2002)). This is an important limitation because the earnings risk faced by the high earners is potentially an important determinant of their saving behavior and wealth concentration in the upper tail (Castañeda et al. (2003)).

A.0.3 W2 synthetic data

The original data used by Guvenen et al. (2015) come from the Master Earnings File (MEF) of the U.S. Social Security Administration (SSA) records and contain information for every individual in the US who has ever been issued a Social Security number. They also include some demographic information (sex, race, date of birth...) and earnings from the employee's W2 forms, that U.S. employers are legally required to report to the SSA since 1978. Labor earnings are measured annually, and they include all wages, salaries, bonuses and exercised stock options. The data are uncapped. Therefore, due to their very nature, the W2 tax data, in contrast to the PSID data, does not suffer from small sample size, under-sampling of the rich, differential survey response, or top-coding. It does not, however, contain good information on the earnings of the self-employed, due to both the nature of the W2 data (that caps the recorded amount of wages that the self-employed pays to himself) and the fact

that Guvenen et al. (2015) drop all observations that derive more than 10% of their earnings from self-employment.

Unfortunately, the original W2 data are not publicly available, but Guvenen et al. (2015) have estimated a flexible, and rather complex, parametric process using the simulated method of moments and shown that it fits the data well. We use their estimates to simulate a synthetic W2 data set. More specifically, the parameterization that we simulate is their "benchmark model" reported in column (2) of Table III in their paper. The model assumes that the stochastic component of earnings is the sum of: (i) an individual-specific linear time trend (heterogeneous income profile); (ii) a mixture of AR(1) processes where each component receives an innovation in a given year with age-dependent probability; and (iii) an i.i.d. transitory shock.

Let $\tilde{y}_t^i$ denote the stochastic component of earnings for individual i at age $t = 1, 2, \ldots$. The specification of the parametric model is:

$$
\begin{aligned}
\tilde{y}_t^i &= (\alpha^i + \beta^i t) + y_t^i \\
y_t^i &= z_t^i + x_t^i + \nu_t^i + \varepsilon_t^i \\
z_t^i &= .259 z_{t-1}^i + \eta_{zt}^i, \quad z_0^i \sim N(0, .133) \\
x_t^i &= .425 x_{t-1}^i + \eta_{xt}^i, \quad x_0^i \sim N(0, .089) \\
\nu_t^i &= \nu_{t-1}^i + \eta_{\nu t}^i, \qquad \nu_0^i = 0 \\
\varepsilon_t^i &\sim N(0, .029).
\end{aligned}
$$

The vector (α, β) has a bivariate normal distribution with zero mean and the covariance matrix implied by $\sigma_\alpha = .552$, $\sigma_\beta \times 10 = .13$ and $\text{corr}_{\alpha\beta} \times 100 = -.49$. The innovations $\eta_{jt}^i, j = z, x, v$ to the three AR(1) components are mutually exclusive with respective probabilities p_j. More formally, $\eta_{jt}^i = \eta_{jt}^{*i} \times \mathbb{I}\{s_{it} \in I_{p_j}\}$ with s_{it} a standard uniform random variable. The indicator function $\mathbb{I}\{s_{it} \in I_{p_j}\}$ takes value one if s_{it} falls in the interval I_{p_j}, with $I_{p_z} = [0, p_z]$, $I_{p_x} = (p_z, p_z + p_x], I_{p_\nu} = (p_z + p_x, 1]$, where $p_z + p_x \leq 1$ and $p_\nu = 1 - p_z - p_x$.

The (mixing) probabilities depend on age and on the previous realization of the stochastic component of earnings according to

$$p_j(y^i_{t-1}, t) = .067 + .038 \times \frac{t}{10} - .203 \times y^i_{t-1} + .071 \times \frac{t}{10} \times y^i_{t-1}$$

and are truncated when they fall outside the $[0, 1]$ interval.

Finally, the innovations to the AR(1) components have distribution $\eta^{*i}_{jt} \sim N(\mu_j, \sigma^i_j)$, with individual-specific standard deviation distributed according to $\log \sigma^i_j \sim N(\bar{\sigma}_j - \frac{\sigma^2_{jj}}{2}, \sigma_{jj}), j = z, x$ and $\sigma^i_\nu = \sigma_\nu$. The parameter values are $\mu_z = -.426$, $\mu_x = .021$, $\mu_\nu = .060$, $\bar{\sigma}_z = .847$, $\bar{\sigma}_x = .361$, $\sigma_{zz} = .107$, $\sigma_{xx} = 1.94$ and $\sigma_\nu = .087$.

Our synthetic W2 panel is obtained by simulating the above process to generate a panel of 500,000 individuals histories between ages 25 and 60.

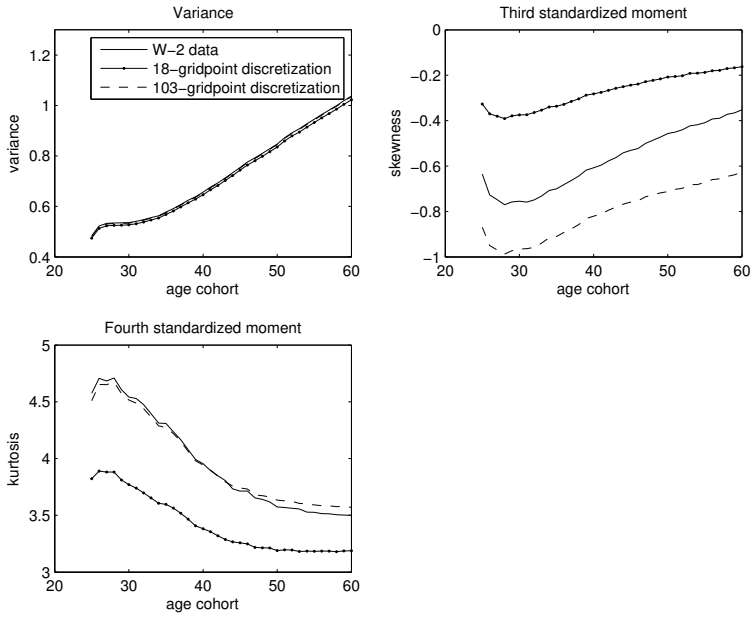
B Appendix for online publication only: How well does our non-parametric estimation of earnings processes fit the PSID and W2 data?

In this section we analyze the fit of our process of several features of the data. We start by analyzing how well we approximate one-year empirical moments and then discuss the consequences of our binning procedure and our one-step Markov chain assumption in presence of measurement error and when considering longer time horizons.

B.1 Fit of empirical moments

Because we allowed for a different number of earnings bins for the two data sets (18 for the PSID and 103 for the synthetic W2 data set), here we also report results for the case in which we allow for a smaller bin size in the synthetic W2 data set (18 points) so that

Figure 18: W2: Unconditional moments of log earnings



the results across the two data sets can be compared keeping the coarseness of the earnings binning constant.

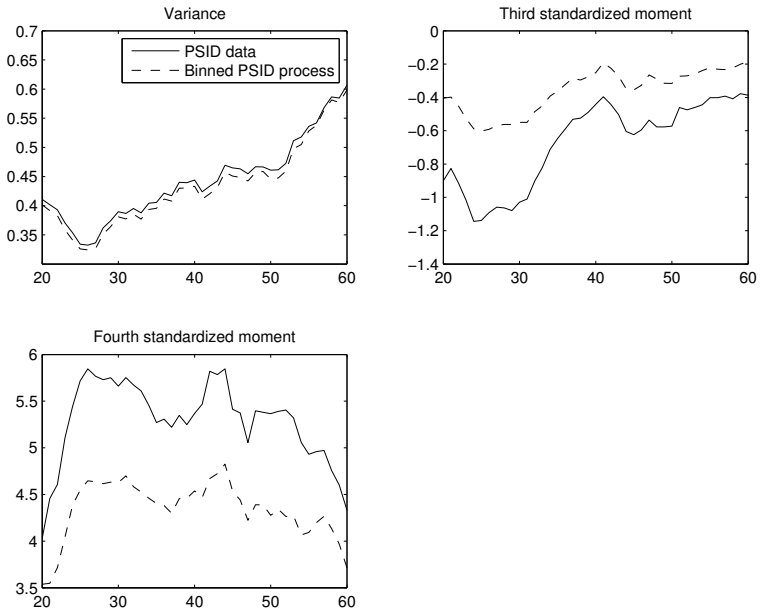
We first show results for the unconditional moments of log earnings by age and then for conditional moments of earnings growth, also by age.

We do not report results for the unconditional first moment because, since the data is detrended, it is an average of OLS residuals and therefore mean zero by definition.

Both the 18-gridpoint and 103-gridpoint discretizations successfully replicate unconditional variances over the life cycle and generate a negative skew and high kurtosis for log earnings (Figure 18). The 18-gridpoint discretization understimates both the negative skew and the kurtosis of the log earnings distribution, while the 103-gridpoint discretization matches very closely the kurtosis profile, but overestimates the negative skew. The comparison of our NL PSID process with PSID data yields very similar results (see Figure 19).

In order to analyze the conditional moments of log earnings and for comparability with Guvenen et al. (2015), we group individuals in 5-year cohorts. Still, our graphs are not fully

Figure 19: PSID: unconditional moments of log earnings



comparable to Section 3 of their paper, since we condition on the earnings of the previous year instead of the earnings of the previous 5 years. As in their paper, log earnings growth is defined as the difference between detrended earnings in two consecutive years.

Now the division in quantiles now serves two purposes. As earlier, it defines the size of the bins, but now it also determines the conditioning variable for the computation of the moments. In the "103-gridpoint discretization", we compute all moments conditional on each of the 103 bins in which we divide the overall population, while in the "18-gridpoint discretization" they are conditional on the 18 bins.

We first consider the conditional mean (Figures 20, 21 and 22). Both discretizations replicate the fact that individuals who are situated in a low earnings quintile can expect larger earnings increases.

With respect to the conditional variance (Figures 23, 24 and 26), the empirical process is also successful. For clearer illustration, we have included how our processes replicate the standard deviation we observe for individuals at ages 35-40 (Figure 25).

With respect to the third standardized moment, all discretizations closely match the

Figure 20: 18-gridpoint discretization: conditional mean of earnings growth

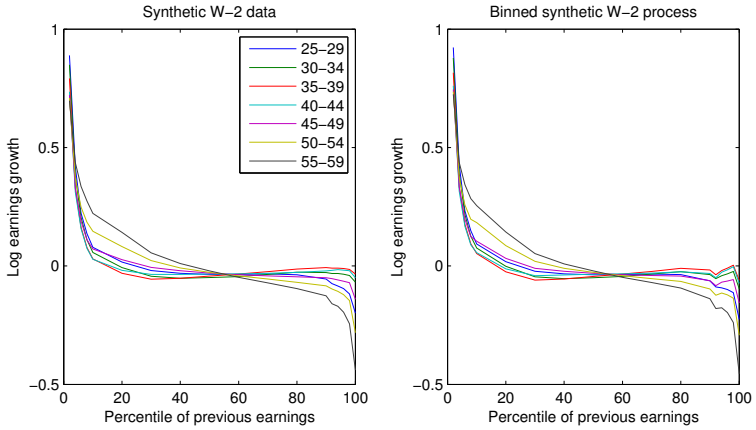


Figure 21: 103-gridpoint discretization: conditional mean of earnings growth

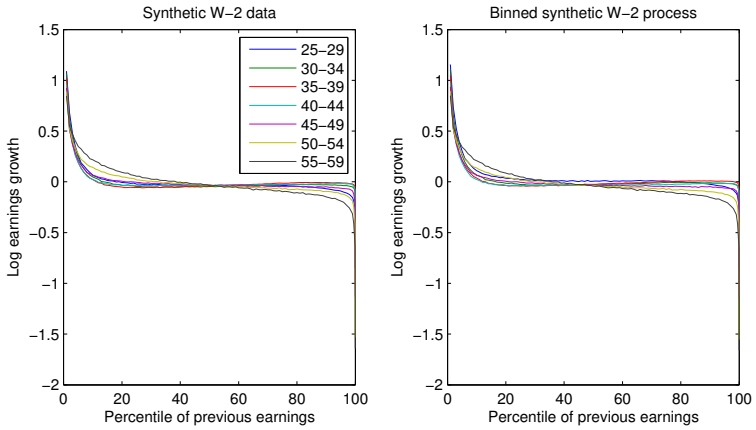


Figure 22: PSID: conditional mean of earnings growth

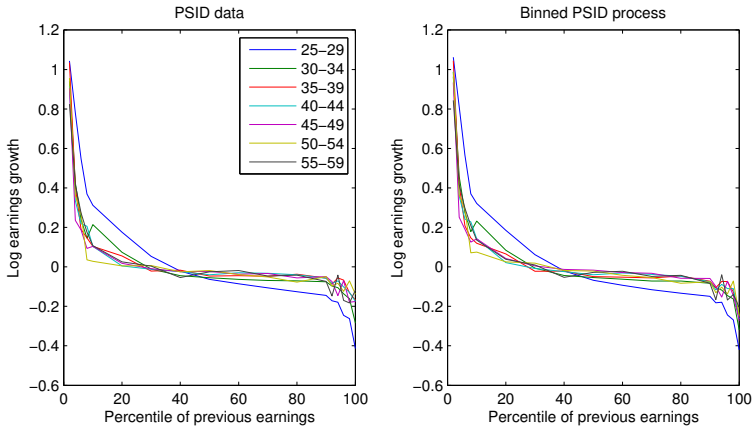


Figure 23: 18-gridpoint discretization: conditional standard deviation of earnings growth

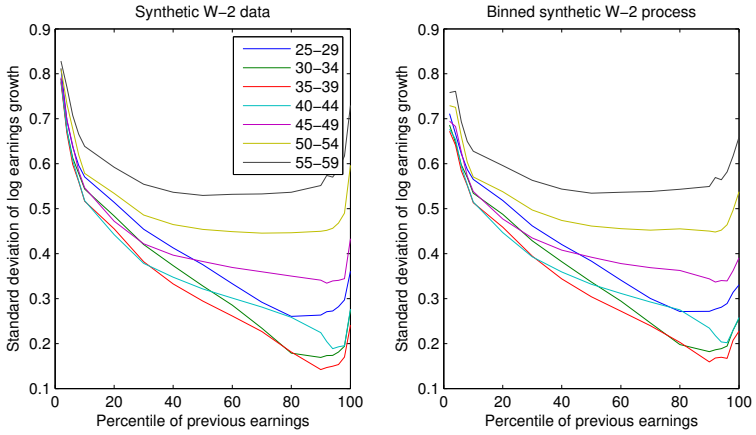


Figure 24: 103-gridpoint discretization: conditional standard deviation of earnings growth

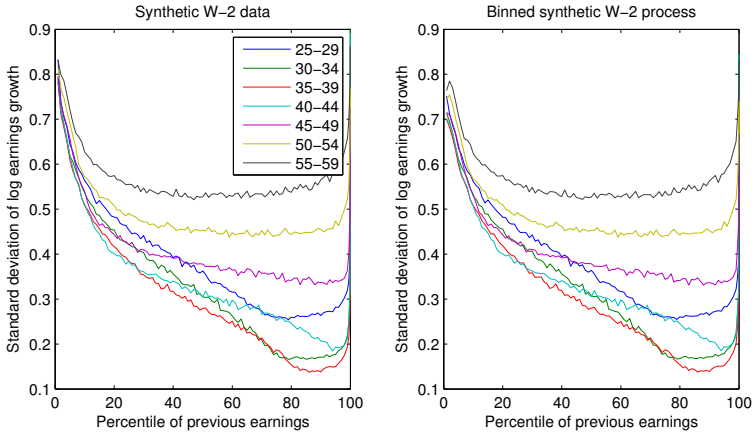


Figure 25: Conditional standard deviation for 35-40 year-old cohort

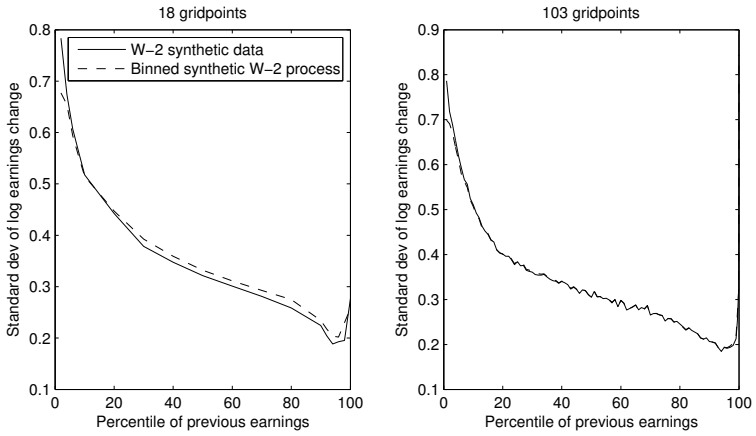


Figure 26: PSID: conditional standard deviation of earnings growth

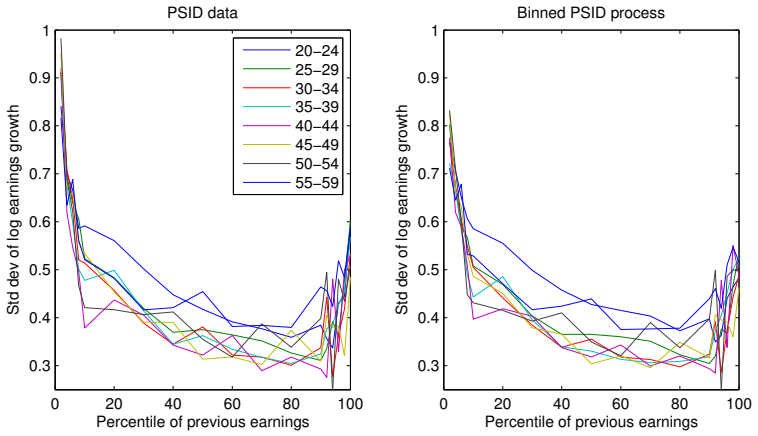
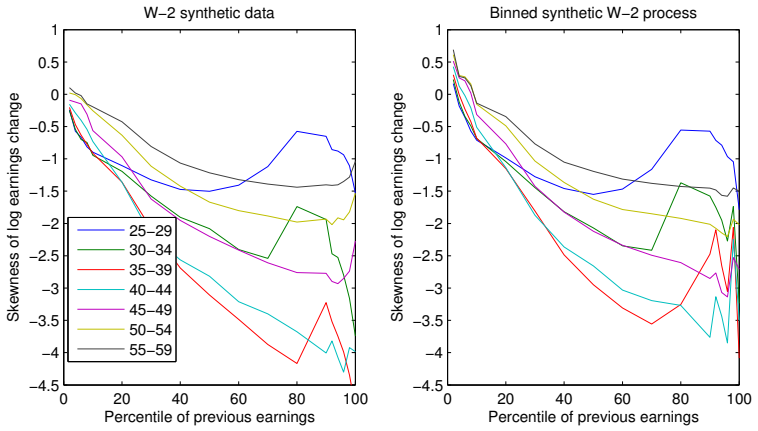


Figure 27: 18-gridpoint discretization: conditional third moment of earnings growth



negative skew that we observe in the data and the fact that it is larger for individual with higher previous earnings (Figures 27 28, 30). This can also be seen in the example in Figure 29.

Finally, both synthetic processes (see Figures 31, 32, 33, 34) match the high kurtosis in the data, with just a slight underestimation in the 18-gridpoint case.

Therefore, we conclude that our procedure generates a reasonably good approximation up to the fourth moment of the actual earnings process, particularly for discretizations with a large number of gridpoints, when applied to a sufficiently large dataset. This represents a significant improvement with respect to the use of a Gaussian AR(1) process, and is one of

Figure 28: 103-gridpoint discretization: conditional third moment of earnings growth

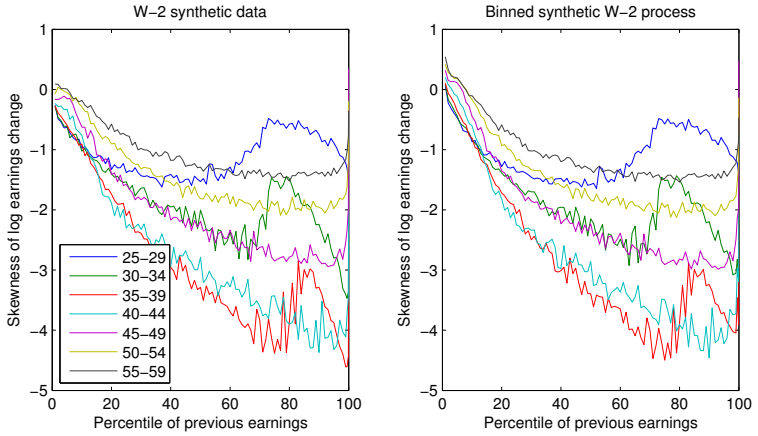


Figure 29: Conditional third moment for 35-40 year-old cohort

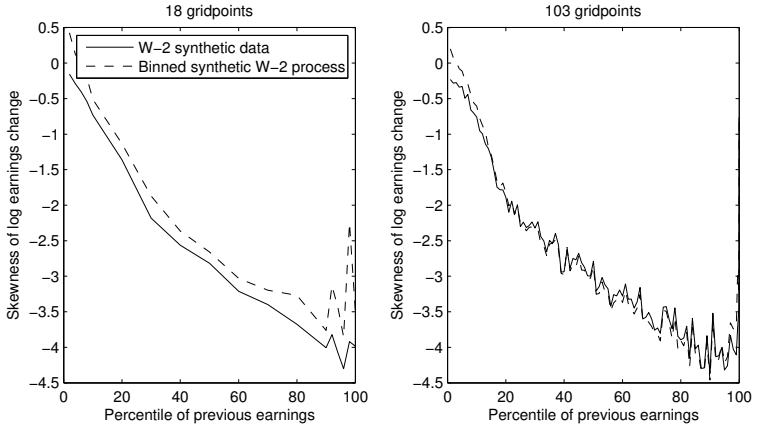


Figure 30: PSID: conditional third moment of earnings growth

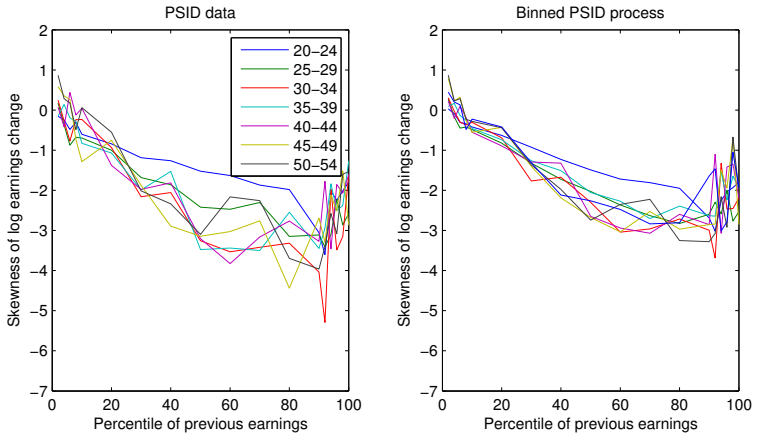


Figure 31: 18-gridpoint discretization: conditional fourth moment of earnings growth

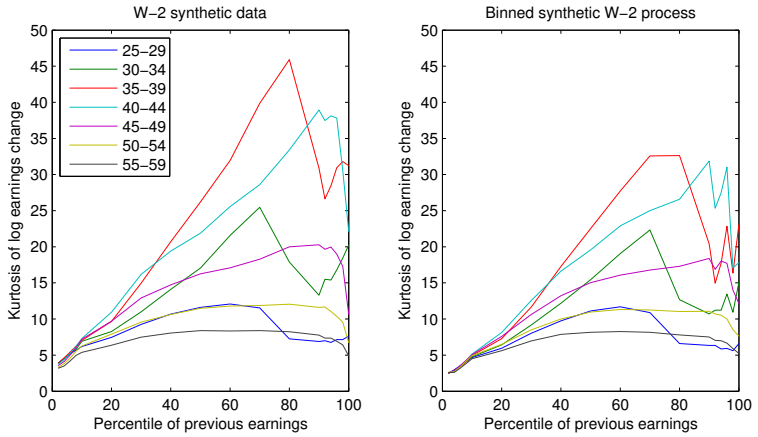


Figure 32: 103-gridpoint discretization: conditional fourth moment of earnings growth

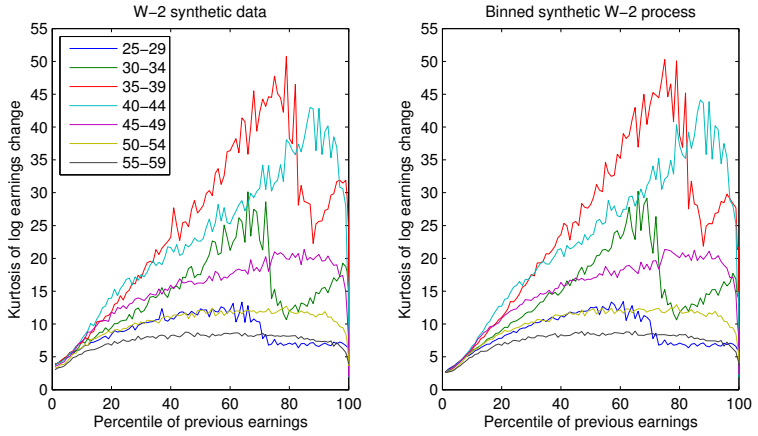


Figure 33: Conditional fourth moment for 35-40 year-old cohort

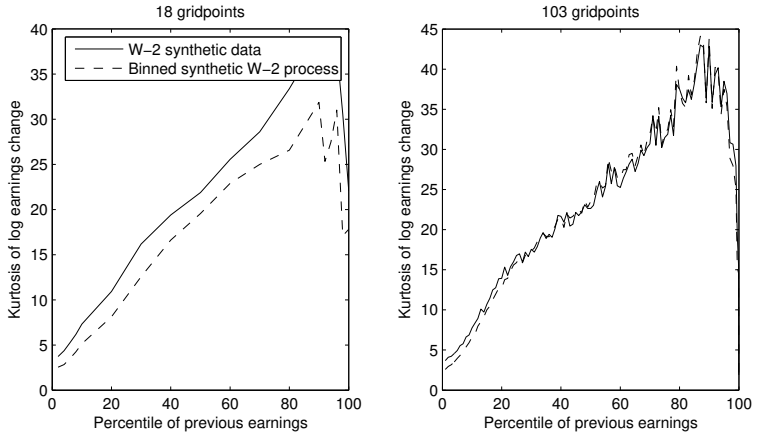
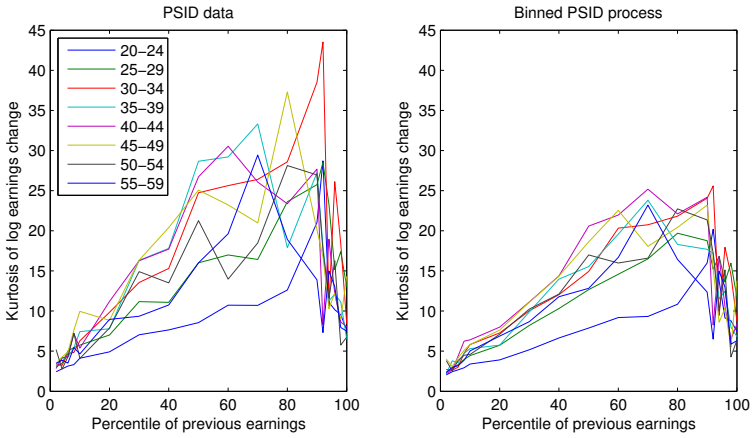


Figure 34: PSID: conditional fourth moment of earnings growth



the key contributions of this empirical method.

When we refer to "NL synthetic W2 process", unless specifically mentioned, we imply the 103-gridpoint discretization. Using the 18-gridpoint discretization instead has no significant impact on the results (not reported).

B.2 Fanning out of gridpoints

Given that the variance of earnings increases by age, it is natural that gridpoints fan out in our NL W2 process. Figures 35 and 36 depict the fanning out of gridpoints. Each point represents a percentile of earnings (we have averaged the top 4 gridpoints, that represent the top percentile).

The graph in logs shows directly the binned residuals (Figure 35), while the graph in levels (Figure 36) shows the same values after exponentiating and controlling to keep the average at each age at 1, but before introducing the age component.

This fanning out generates an increase in within-bin variance, particularly for very high earnings realizations. However, our close fit of the empirical moments as described in Appendix B.1 suggests that this does not fundamentally affect any of our results.

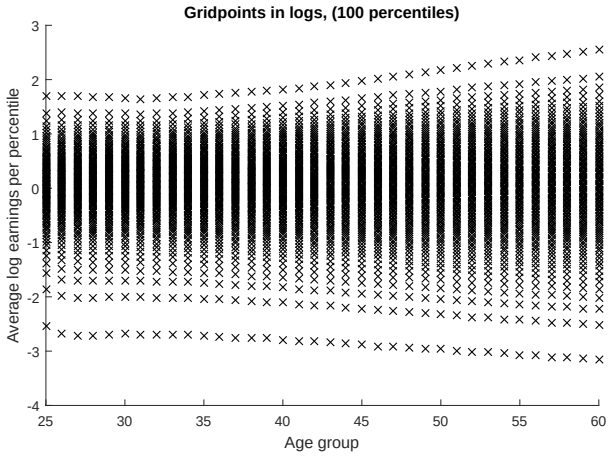


Figure 35: Gridpoints, logs

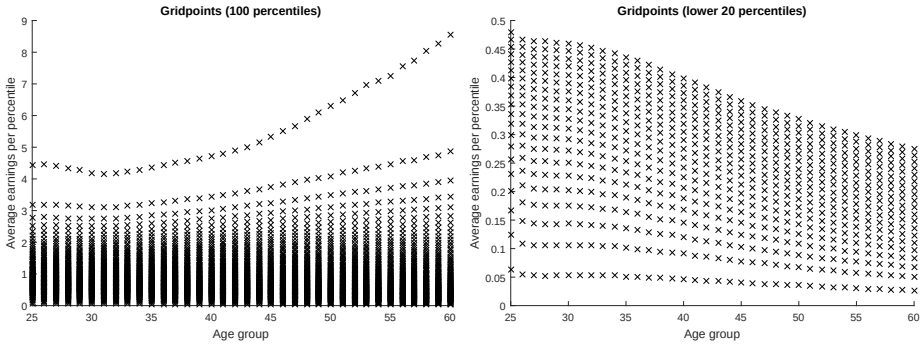


Figure 36: Gridpoints, values (left: all of them; right: just the bottom 20)

B.3 Transitory shocks and measurement error

This section further analyzes the effects that measurement error, which is known to be an issue in survey data like the PSID, can have in our processes. Naturally, without any other information, a pure transitory shock and classical measurement error are not separately identifiable, so we consider both together here.

The original econometric process that generates our W2 synthetic dataset (from Guvenen et al. (2015)) includes an i.i.d. transitory shock with standard deviation 0.029. In order to analyze its effects, we have re-estimated the process without a transitory shock and with a larger transitory shock (SE=0.25).

As we can see in the graphs of implied persistences by age (Figure 37) or by earnings decile (Figure 38), the existence of a SE=0.029 shock causes virtually no difference with respect to considering permanent income alone. However, the presence of measurement error with SE=0.25 generates a decrease in the autocorrelation coefficient of around 0.1 at each age. Besides, the inclusion of a larger transitory shock implies larger variances of earnings at every age (Figure 39).

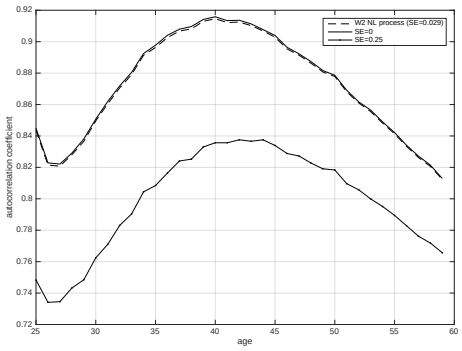


Figure 37: Persistence by age (permanent and total income)

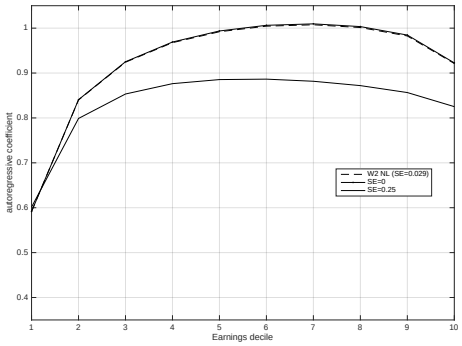


Figure 38: Persistence by decile (permanent and total income)

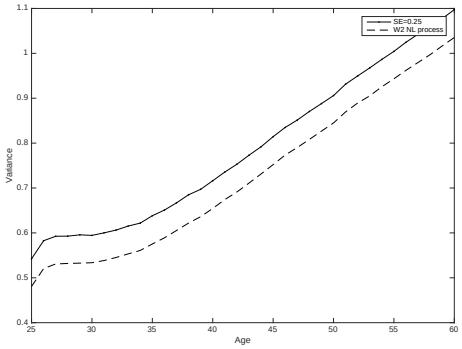


Figure 39: Variance of log earnings by age

B.4 Long-term persistence of earnings

Our NL earnings processes rely on a one-period memory Markov assumption. Earnings are a one-dimensional state: once an individual falls into an earnings bin, all of his past history is erased. This implies individuals of very different past histories and characteristics, which may condition their future ability to generate earnings, are pooled together in the same bin. Since it is expectable that an individual with a good history who suffered a bad shock has a better chance of jumping back up that an individual with a poor earnings history, it is possible that we overestimate the idiosyncratic probability of earnings changes and subsequently underestimate the long-term persistence of earnings realizations.

In order to measure how relevant this effect is in our data, we have computed the implied 2, 5 and 10-year autocorrelation parameters of the NL W2 process and compared it with the actual persistences directly estimated from our synthetic panel. [21] Results are represented in Figure 40. For comparison, we have included the implied persistence of our benchmark AR(1) process (for which $\rho = 0.96$) and the implied persistence of an AR(1) estimated in the synthetic W2 dataset (for which $\rho = 0.88$).

As expected, we do underestimate persistence compared to the actual data, but that is a feature of the assumption that the process is Markov of order 1, not of our methorodlogy.

[21] For example, the 2-year parameter is the result of estimating ρ in $y_t = \rho y_{t-2} + \epsilon_t$, where y_t is log detrended earnings.

In fact, our process dominates the performance of the standards AR(1)s at both assumed persistence levels in that we do match the life-cycle profile of earnings persistence and do no worse on the average level of persistence.

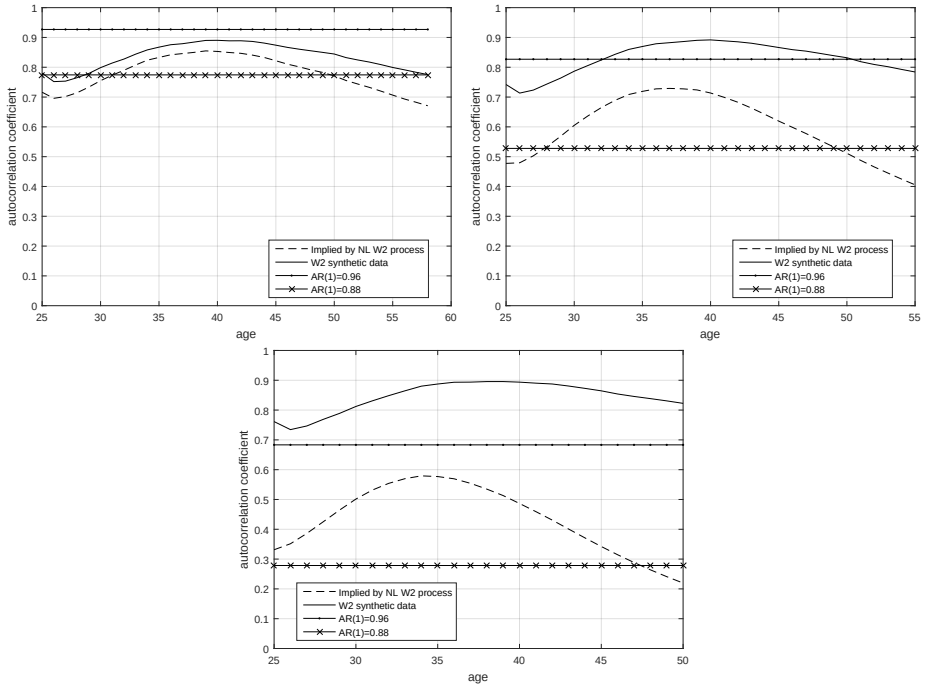


Figure 40: Persistence over a 2- (top left panel), 5- (top right panel) and 10- (bottom panel) year period.

C Appendix for online publication only: Definition of Stationary Equilibrium

A stationary equilibrium is composed of

$$\begin{cases} \text{allocations } c(x), a'(x), \\ \text{government tax rates and transfers } (\tau_a, \tau_l, p), \\ \text{and a distribution } \psi_t(x) \text{ over state variables } x \text{ for every age } t \end{cases}$$

such that, given an interest rate r and a wage rate w, the following hold (let μ_t be the proportion of agents of age t):

(i) Given prices and government tax rates and transfers, the functions $c(x)$ and $a'(x)$ solve the described maximization problem for a household with state variables x.

(ii) The tax rate τ_l is chosen so that the government budget constraint balances at every period:

$$g = \sum_t \mu_t \int_X \Big[\tau_a \, r \, a + \tau_l y I_{t<t_r} - p \, I_{t\geq t_r}\Big] d\psi_t(x).$$

(iii) Distributions are consistent with individual behavior and the exogenous process for labor productivity (the transition function P is described in detail below):

$$\psi_t(B) = \int_X P(x, t-1, B) d\psi_{t-1}(x) \qquad \forall B \in B(X). \tag{4}$$

The probability measure ψ_t, together with $(X, B(X))$, forms a probability space. $X = [0, \text{inf}) \times Y$ is the state space (of assets and exogenous labor productivity) and $B(X)$ is the Borel σ-algebra on X. [22]

The function $P(x, t, B)$ is a transition function which gives the probability that an age t agent transits to B next period given that his current state is x. This transition function is determined by the decision rule on asset holding and the exogenous transition probabilities on the labor productivity shock. Survival probabilities are exogenous and orthogonal with respect to decision rules and the exogenous labor productivity process, so they are not included in the transition function as it is described here.

The distribution over states of age 1 agents is determined by the exogenous initial distribution of labor productivity, since agents start life with no assets.

[22] Therefore, for a given set $B(X)$, $\psi_t(B)$ is the proportion of individuals of age t that lie in B. $\mu_t \psi_t(B)$ indicates the fraction of total population that agents of age t in B represent.