Stanford Econ 237 & MgtEcon 617: Lecture 2[1]

Profs. Martin Schneider and Chris Tonetti

1 Overview

Why Labor Income? Incomplete Markets and Non-tradable Assets.

Why spend so much time on labor income risk? Utility is over consumption (and maybe leisure). Labor supplied is the overwhelming majority of income used to finance consumption for the overwhelming majority of people (and leisure is obviously affected by labor supply). There are two somewhat separate ways of looking at this fact. The first is to go into detail trying to understand and model what determines labor income and income risk (much of this approach uses search and matching technology and might involve things like heterogeneity in firms and/or workers, assortative matching, production externalities, management hierarchies, human capital accumulation (with or without externalities), etc.). The second approach is to take as given a good statistical representation of stochastic labor income, and see how labor income and associated risks affects all other economic decision making. Essential to both studies is the realization that the world does not seem to be a complete markets economy in which there is perfect insurance over labor income risk. Rather, various forms of evidence supports that the world appears to be an incomplete market world, in which people can only partially insure against their labor income risk. Much of this partial insurance takes the form of self-insurance through wealth (mostly accumulated labor income with some bequests and financial income for some).

We will spend much of this course thinking about the second approach. In particular we will often work with Bewley models (Bewley, Aiyagari, Hugget, Imrohoroglu, etc.), which tend to take income risk as given. We will study the implications of stochastic income on saving and consumption behavior, as well as implications for asset demand. Ultimately, in equilibrium models, we can also talk about asset supply to be able to study asset prices (supply + demand).

Keep in mind during this lecture that ultimately, we are working towards solving and understanding the income fluctuation problem of a household:

$$
\begin{aligned}
V(a,y) = &\max_{a'} \; u(c) + \beta\, \mathbb{E}[V(a',y')] \\
\text{s.t.} & \\
& c + a' = (1+r)a + y \\
& a' \geq a_{min}
\end{aligned}
$$

with r exogenous and a given stochastic process for y. Will solve numerically for optimal policies on a grid over the state space. First we proceed to study statistical error component models of stochastic labor income, then we will study finite-state Markov model approximations to these ECMs, and finally will solve the IFP.

A key interpretation of labor income from a finance perspective is that labor income is like a non-tradable asset. That is, there is no market in which you can trade your future earnings. Also note that a similar problem can be interpreted in the context of a firm instead of a consumer, with appropriate modifications. For example, flow utility from consumption becomes the value of issuing dividends, continuation utility is interpreted as capital gains/losses, savings is a firm capital, and interpreting labor income risk as idiosyncratic earnings risk (perhaps due to productivity/cost shocks or changes in demand for the firms product).

[1]This lecture note builds on course material from Chris Tonetti's Ph.D. education. In particular, this lecture draws on class notes by Gianluca Violante, sometimes verbatim.

2 Labor Income Processes

2.1 Earnings Process Estimation Strategies

The first major choice in estimating a statistical process for labor earnings is to decide whether to use data on consumption or restrict the estimation to using only earnings data. Models of consumption, whether statistical or structural, have strong predictions on how consumption should respond to earnings shocks. It is therefore feasible to use data on changes in consumption to gain inference on the earnings process. The main advantage of this strategy is to introduce more data, but the main drawback is that the estimation relies heavily on the proposed model of consumption.

An alternative is to only use earnings data. Free from any structural modeling assumptions, specify and estimate a statistical process for earnings. This is the approach covered in these notes.

2.2 Data

For the estimation we require a panel of earnings data. A repeated cross section is not enough. In the U.S., this leads many people to use the Panel Study of Income Dynamics (PSID). Most studies apply sample selection criteria to the data to remove outliers and achieve a more homogeneous population. Following common practice, we will think of an individual in the sample as a male head of household between the ages of 25 and 60 with non-zero annual labor earnings. Data are annual. Typically, we work with short panels, e.g., the PSID has a significantly larger cross sectional dimension than time dimension.

2.3 Obtaining Residual Earnings

One approach to estimating an earnings process is to model residual earnings (conditioning on observable variables). Then, one can represent the earnings process using the residual earnings stochastic process as fluctuations around a deterministic function of observables. This is the approach covered in these notes. One could, however, estimate a stochastic earnings process as a function of these observables.

Assume a competitive model in the labor market, yielding a wage per efficiency unit of labor, w_t. Let i index individual, j age, and t time.

$$Y_{i,j,t} = w_t \exp(f(X_{i,j,t}) + y_{i,j,t})\bar{h} \tag{1}$$

- $Y_{i,j,t}$ - measured annual disposable labor income
- $\bar{h}$ - exogenous number of hours worked[2]
- $\exp(f(X_{i,j,t}))$ - predictable individual labor efficiency
- $X_{i,j,t}$ - demographic observables and predictable variables [age, gender, edu, race, etc.]
- $y_{i,j,t}$ - idiosyncratic stochastic component of earnings
- f - time invariant function of observables $X_{i,j,t}$

Note:

$$\ln Y_{i,j,t} = \beta_t + f(X_{i,j,t}) + y_{i,j,t} \tag{2}$$

where β is the price of labor.

To complete the first step, run a regression on Equation 2 to obtain residuals $y_{i,j,t}$.

[2]With an elastic labor supply, estimate a wage process by using earnings divided by hours.

2.4 Model Specification

Let $j \in \{0, 1, 2, \dots, J\}$ be an individual's effective age, defined as $age - 21$.

$$
\begin{aligned}
y_{i,j} &= \alpha_i + \epsilon_{i,j} + \nu_{i,j} && (3)\\
\epsilon_{i,j} &= \rho\epsilon_{i,j-1} + \eta_{i,j} && (4)
\end{aligned}
$$

where

$$
\begin{aligned}
&\alpha_i, \nu_{i,j}, \eta_{i,j}, \epsilon_{i,0} \sim \, ?\\
&E[y_{i,j}] = E[\alpha_i] = E[\nu_{i,j}] = E[\eta_{i,j}] = E[\epsilon_{i,0}] = 0\\
&Var(\alpha_i) = \sigma^2_\alpha, \quad Var(\nu_{i,j}) = \sigma^2_\nu, \quad Var(\eta_{i,j}) = \sigma^2_\eta, \quad Var(\epsilon_{i,0}) = \sigma^2_{\epsilon_0}\\
&\epsilon_{i,0} \quad \perp \quad \nu_{i,j} \quad \perp \quad \eta_{i,j} \quad \perp \quad \alpha_i \quad \forall\, j
\end{aligned}
$$

The goal is to estimate $\theta = [\rho, \sigma^2_\nu, \sigma^2_\eta, \sigma^2_\alpha, \sigma^2_{\epsilon_0}]$.

2.5 Level Moments

There are $\frac{J(J+1)}{2}$ covariance moments.

$$
M(\theta) = vec\begin{pmatrix}
E[y_{i,0}y_{i,0}] & ... & ... & ... & ... & ...\\
E[y_{i,1}y_{i,0}] & E[y_{i,1}y_{i,1}] & ... & ... & ... & ...\\
E[y_{i,2}y_{i,0}] & E[y_{i,2}y_{i,1}] & E[y_{i,2}y_{i,2}] & ... & ... & ...\\
E[y_{i,j}y_{i,0}] & ... & ... & E[y_{i,j}y_{i,j}] & ... & ...\\
... & ... & ... & ... & ... & ...\\
E[y_{i,J}y_{i,0}] & ... & ... & ... & ... & E[y_{i,J}y_{i,J}]
\end{pmatrix}
$$

Note we have already imposed that the expectation of each innovation is zero.

$$
\begin{aligned}
E[y_{i,j}y_{i,j}] &= E[(\alpha_i + \epsilon_{i,j} + \nu_{i,j})(\alpha_i + \epsilon_{i,j} + \nu_{i,j})]\\
&= E[\alpha_i^2 + 2\alpha_i\epsilon_{i,j} + 2\alpha_i\nu_{i,j} + \epsilon^2_{i,j} + 2\epsilon_{i,j}\nu_{i,j} + \nu^2_{i,j}]\\
&= E[\alpha_i^2] + E[\epsilon^2_{i,j}] + 2E[\epsilon_{i,j}]E[\nu_{i,j}] + E[\nu^2_{i,j}]\\
&= \sigma^2_\alpha + E[\epsilon^2_{i,j}] + \sigma^2_\nu
\end{aligned}
$$

Note that $E[\epsilon_{i,j}] = \rho^j E[\epsilon_{i,0}] = 0$.

It remains to derive $E[\epsilon^2_{i,j}]$.

$$
\begin{aligned}
E[\epsilon^2_{i,j}] &= E[(\rho\epsilon_{i,j-1} + \eta_{i,j})^2]\\
&= \rho^2 E[\epsilon^2_{i,j-1}] + \sigma^2_\eta\\
&= \rho^2(\rho^2 E[\epsilon^2_{i,j-2}] + \sigma^2_\eta) + \sigma^2_\eta\\
&= \rho^{2t}\sigma^2_{\epsilon_0} + \sum_{k=1}^{j} \rho^{2(j-k)}\sigma^2_\eta && (5)
\end{aligned}
$$

Now to derive the autocovariance across time periods:

$$
\begin{aligned}
E[y_{i,j}y_{i,j+h}] &= E[(\alpha_i + \epsilon_{i,j} + \nu_{i,j})(\alpha_i + \epsilon_{i,j+h} + \nu_{i,j+h})] \\
&= E[\alpha_i^2 + \alpha_i\epsilon_{i,j+h} + \alpha_i\nu_{i,j+h} + \alpha_i\epsilon_{i,j} + \epsilon_{i,j}\epsilon_{i,j+h} + \epsilon_{i,j}\nu_{i,j+h} + \epsilon_{i,j+h}\nu_{i,j} + \nu_{i,j}\nu_{i,j+h}] \\
&= E[\alpha_i^2] + E[\epsilon_{i,j}\epsilon_{i,j+h}] \\
&= \sigma_\alpha^2 + \rho^h E[\epsilon_{i,j}^2] \qquad (6)
\end{aligned}
$$

Note we have imposed that the ν are independently distributed across time. Finally, using 5,

$$
\boxed{E[y_{i,j}y_{i,j+h}] =} \begin{cases} \sigma_\alpha^2 + E[\epsilon_{i,j}^2] + \sigma_\nu^2 & \text{if } h = 0 \\ \sigma_\alpha^2 + \rho^h E[\epsilon_{i,j}^2] & \text{if } h > 0 \end{cases} \qquad (7)
$$

where

$$
E[\epsilon_{i,j}^2] = \rho^{2t}\sigma_{\epsilon_0}^2 + \sum_{k=1}^{j} \rho^{2(j-k)}\sigma_\eta^2.
$$

2.6 Exact Identification

Recall,

$$
\begin{aligned}
Var(y_0) &= \sigma_\alpha^2 + \sigma_{\epsilon_0}^2 + \sigma_\nu^2 \qquad (8) \\
Var(y_1) &= \sigma_\alpha^2 + \rho^2\sigma_{\epsilon_0}^2 + \sigma_\eta^2 + \sigma_\nu^2 \qquad (9) \\
Cov(y_0, y_h) &= \sigma_\alpha^2 + \rho^h\sigma_{\epsilon_0}^2 \qquad (10) \\
Cov(y_1, y_h) &= \sigma_\alpha^2 + \rho^h(\rho^2\sigma_{\epsilon_0}^2 + \sigma_\eta^2) \qquad (11)
\end{aligned}
$$

Then,

$$
\rho = \frac{Cov(y_0, y_3) - Cov(y_0, y_2)}{Cov(y_0, y_2) - Cov(y_0, y_1)} \qquad (12)
$$

$$
\sigma_{\epsilon_0}^2 = \frac{Cov(y_0, y_2) - Cov(y_0, y_1)}{\rho(\rho - 1)} \qquad (13)
$$

$$
\sigma_\nu^2 = Var(y_0) + \frac{(Cov(y_0, y_2) - (1+\rho)Cov(y_0, y_1))}{\rho} \qquad (14)
$$

$$
\sigma_\alpha^2 = Cov(y_0, y_1) - \frac{Cov(y_0, y_2) - Cov(y_0, y_1)}{\rho - 1} \qquad (15)
$$

$$
\sigma_\eta^2 = Var(y_1) - Var(y_0) - \left(\frac{\rho + 1}{\rho}\right)(Cov(y_0, y_2) - Cov(y_0, y_1)) \qquad (16)
$$

We have full identification from the autocovariance function. Obviously, the model is overidentified.

2.7 Estimation

The standard estimation strategy in the literature is to use a Minimum Distance Estimator. The goal is to choose the parameters that minimize the distance between empirical and theoretical moments.

Recall the definition of $M(\theta)$, which stacked all of the non-redundant autocovariance moments into a vector and let $\hat{M}$ be the empirical counterpart.

The estimated parameters, $\widehat{\theta}$, are the solution to

$$\min_{\theta} \left[M(\theta) - \hat{M}\right]' W \left[M(\theta) - \hat{M}\right] \tag{17}$$

where W is a weighting matrix.

2.8 Weighting Matrix

To implement the estimator, we need to choose W. Altonji and Segal (1996) show that the Optimal Minimum Distance (OMD) estimator, where W is set to the optimal weighting matrix, introduces significant small sample bias. Many papers in the literature use the Equally Weighted Minimum Distance (EWMD) estimator, where W is the identity matrix, as a result. An alternative, employed by Blundell, Pistaferri, and Preston (2008) is to use Diagonally Weighted Minimum Distance (DWMD), where W is set to the diagonal elements of the optimal weighting matrix with off-diagonal elements set to zero.

2.9 Standard Errors

Chamberlain (1984) shows standard errors can be obtained as

$$\widehat{var(\widehat{\theta})} = (G'WG)^{-1}G'WVWG(G'WG)^{-1} \tag{18}$$

where V^{-1} is the optimal weighting matrix and G is the Jacobian matrix evaluated at the estimated parameters, $\frac{\partial \bar{M}(\theta)}{\partial \theta}|_{\theta=\widehat{\theta}}$. Recall the data were originally obtained as residuals from a first stage regression. See Murphy and Topel (1985) for adjusting second stage standard errors.

Alternatively, standard errors can be computed by bootstrap. Bootstrap samples are drawn (with replacement) at the household level with each sample containing the same number of households as the original sample. Then apply the first stage regression on each sample, estimate the parameters of interest on the residual for each sample, and compute statistics using cross-sample variations. The resulting confidence intervals thus account for arbitrary serial dependence, heteroskedasticity, and additional estimation error induced by the use of residuals from the first stage regressions. Bootstraping is a computationally intensive technique. Run as many samples as is computationally feasible, with a rule of thumb being 500.

2.10 Transitory Effects

2.10.1 Measurement Error

Micro data, especially those based on surveys, have measurement error, $\tau_{i,t}$. The typical assumption is that it is i.i.d across agents and over time. With this assumption, measurement error is indistinguishable from ν in our specification. Econometricians should thus be aware when interpreting parameter estimates of the transitory component. The transitory component of earnings could be modeled as an MA(q), with q >0, in which case the variance of the transitory component can be estimated separately from classical measurement error.

The PSID ran validation studies in 1982 and 1986 where they confirm the earnings and hours data from employer records. They found a small error in earnings (10-20 percent) but larger error in hours worked (20-40 percent). Depending on the question, measurement error may not be that important because so much action comes from the fixed and persistent effects, α_i and η_i.

2.11 Extensions

2.11.1 Transitory Shocks

Just because earnings dynamics are largely driven by fixed and persistent effects, that does not mean we can omit the transitory shock from our specification.

Assume the true specification is that of Equation (3) but was modeled as

$$y_{i,j} = \alpha_i + \epsilon_{i,j} \tag{19}$$
$$\epsilon_{i,j} = \rho\epsilon_{i,j-1} + \eta_{i,j} \tag{20}$$

then

$$\boxed{E[y_{i,j}y_{i,j+h}] =} \begin{cases} \sigma_\alpha^2 + E[\epsilon_{i,j}^2] & \text{if } h = 0 \\ \sigma_\alpha^2 + \rho^h E[\epsilon_{i,j}^2] & \text{if } h > 0 \end{cases} \tag{21}$$

where

$$E[\epsilon_{i,j}^2] = \rho^{2t}\sigma_{\epsilon_0}^2 + \sum_{k=1}^{j} \rho^{2(j-k)}\sigma_\eta^2.$$

To understand the effect on ρ of omitting ϵ let's analyze $\frac{m_{0,2}-m_{0,1}}{m_{0,1}-m_{0,0}}$. Under the misspecified model:

$$\frac{Cov(y_0, y_2) - Cov(y_0, y_1)}{Cov(y_0, y_1) - Var(y_0)} = \frac{\rho^2\sigma_{\epsilon_0}^2 - \rho\sigma_{\epsilon_0}^2}{\rho\sigma_{\epsilon_0}^2 - \sigma_{\epsilon_0}^2} = \rho$$

however, in the true model:

$$\frac{Cov(y_0, y_2) - Cov(y_0, y_1)}{Cov(y_0, y_1) - Var(y_0)} = \frac{\rho^2\sigma_{\epsilon_0}^2 - \rho\sigma_{\epsilon_0}^2}{\rho\sigma_{\epsilon_0}^2 - \sigma_{\epsilon_0}^2 - \sigma_\nu^2} = \rho\left[\frac{1}{1 + \frac{\sigma_\nu^2}{(1-\rho)\sigma_\eta^2}}\right] < \rho$$

So, under the misspecified model the estimate would be to set ρ equal to the empirical counterpart to $\frac{m_{0,2}-m_{0,1}}{m_{0,1}-m_{0,0}}$. However, we can see that this empirical counterpart is less than ρ in the true model. Thus, omitting the transitory component introduces a downward bias in the estimate of the persistence of earnings.

The intuition for the downward bias is that the transitory variance introduces a big drop in the autocovariance function between lag zero and one, which is misinterpreted as a low autocorellation in the persistent shock. This explains many of the low estimates in the literature, such as Heaton and Lucas (1996) who estimate $\rho = 0.6$ when they specify a process with only an AR(1) component.

2.11.2 Time Variation in Parameters

There is extensive evidence that there exists time variation in the variance of persistent and transitory shocks. Authors have estimated both how risk evolves over the business cycle, as well as the long term trends in idiosyncratic earnings risk over the past few decades.

Cyclicality in Risk

Storesletten, Telmer, and Yaron (2001) allow for the conditional variance of the shocks to be different in times of expansions (σ_E^2) versus contractions (σ_C^2). They find $(\sigma_C^2) > (\sigma_E^2)$, which has asset pricing implications, as well as, implications for the welfare cost of business cycles. See Constantinides and Duffie (1996) for a classic description of how the conditional variance of earnings can affect asset prices. See Heathcote, Storesletten, and Violante (2014) for an extension of the framework suitable for quantitative analysis.

There has been a recent revival of interest in this topic largely driven by availability of large-scale administrative data. See Guvenen, Ozkan, and Song (2014) for evidence that labor income shocks have very large kurtosis, variance is acyclical, but left-skewness is strongly countercyclical.

Long-run Trends in Risk

There have been multiple papers that have analyzed the evolution of the conditional variance of persistent and transitory shocks since the formation of the PSID: 1968-2007. To model time variation, we would need to make some changes. We could either replace age with time in the model (an different assumption on time vs. cohort vs. age effects) or we could make sure to align the variances faced by people of different ages at the same calendar time.

If we wanted to allow for age varying variances, we could maintain the same basic specification, but allowing for heteroskedasticity:

$$y_{i,j} = \alpha_i + \epsilon_{i,j} + \nu_{i,j} \tag{22}$$

$$\epsilon_{i,j} = \rho\epsilon_{i,j-1} + \eta_{i,j} \tag{23}$$

where

$$\alpha \sim (0, \sigma_\alpha^2), \quad \eta_{i,j} \sim (0, \sigma_{\epsilon,j}^2), \quad \nu_{i,j} \sim (0, \sigma_{\nu,j}^2)$$

Identification proceeds in a similar, but more complicated, manner to the homoskedastic case and the same moments and estimator can be used to estimate the variance for each shock at each point in time. Often, ρ is assumed to be unity, but that is not necessary.

Alternatively, the variances can be modeled as evolving according to a process. See Meghir and Pistaferri (2004) for evidence of a GARCH component in variance terms.

2.12 Alternative Specifications: HIP vs. RIP

Lillard and Weiss (1979) present a model in which the life cycle earnings process is no longer stochastic, but rather deterministic and individual specific. Although the deterministic model was largely abandoned in favor of the stochastic models discussed above, recently the idea of heterogeneous income profiles has been revived.[3] In particular, Guvenen (2009) develops a hybrid model, where there is an individual specific age profile **and** stochastic shocks to income.

Let log labor income deviations from a common age profile be

$$y_{i,j} = \alpha_i + \beta_i j + \epsilon_{i,j} + \nu_{i,j} \tag{24}$$

$$\epsilon_{i,j} = \rho\epsilon_{i,j-1} + \eta_{i,j} \tag{25}$$

Note this model provides variation around a common age profile for 3 reasons. The α_i and β_i are a deterministic individual specific intercept and slope. $\epsilon_{i,j}$ is a stochastic persistent component and

[3]To contrast with the models with ex-ante heterogeneous income profiles (HIP models) sometimes the stochastic process with common expected income profiles are called "restricted income profiles" (RIP) models.

$\nu_{i,j}$ is a transitory component.

Instead of estimating a separate α_i and β_i for each individual, estimate $(\sigma^2_\alpha, \sigma^2_\beta, \sigma^2_{\alpha,\beta})$, where

$$(\alpha_i, \beta_i) \sim \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_\alpha & \sigma^2_{\alpha,\beta} \\ \sigma^2_{\alpha,\beta} & \sigma^2_\beta \end{bmatrix} \right).$$

Often $\sigma^2_{\alpha,\beta}$ is set to 0.

The HIP and RIP models have very different implications for the labor income risk individuals face over the life cycle. The idiosyncratic age profile generates persistent deviations from the common trend, without introducing any risk from the perspective of the agent. From the perspective of the econometrician, ignoring this variable would lead to an upward bias in the estimation of the autocovariance parameter in the persistent shock. Guvenen (2009) estimates $\rho = 0.99$ when σ^2_β is restricted to 0, while $\rho = 0.82$ when σ^2_β is unrestricted. Guvenen (2007) explores the case where agents have to learn their profile over time. Learning occurs slowly, as the agent, just like the econometrician, has difficulty distinguishing between the income profile slope and persistent shocks. This allows the HIP model to produce a rise in consumption inequality over the lifecycle.

MaCurdy (1982) proposed a test (and Abowd and Card (1989) performed a variant of this test), for HIP models based on the sign of the implied autocovariance of income growth. Both are often cited as supporting RIP models to the exclusion of HIP. However, Guvenen (2009) shows this test has low power, especially because it relies on many period (∼10) lags of covariances which are noisy in the data.

To test the HIP vs. RIP model it is imperative to compare the model implications to the data. HIP and RIP have very different predictions for earnings variance and covariances as a function of age. See Guvenen (2009) for details.

2.13 Likelihood Based Methods

Although it has been traditional to estimate statistical earnings processes with minimum distance estimation, there have been some examples of using likelihood based techniques. One major advantage to certain likelihood based estimators is the ability to estimate more complex models, while a downside is the reliance on distributional assumptions on error terms. To my knowledge, Geweke and Keane (2000) was the first attempt, focusing on jointly estimating earnings process parameters and marital status to analyze the transition probabilities between income quartiles over the life cycle. They used the Gibbs sampler and Bayesian techniques, allowing the error terms to be distributed according to a mixture of Normal distributions for better model fit. Nakata and Tonetti (2015) explores the small sample properties of many different estimators of the standard earnings processes. In addition to minimum distance estimation, they examine the Maximum Likelihood estimator built with the Kalman filter, an estimator that uses Metropolis-Hastings, and a Bayesian routine using Gibbs sampling. They test the performance of these estimators on different specifications of the income process, including time variation and shocks from mixtures of Normals. See Nakata and Tonetti (2015) for more information on the construction and performance of likelihood based estimators of income processes.

3 Finite-State Markov Approximations to Persistent Labor Earnings Processes

Ultimately, we are building towards numerically solving for optimal policies in the income fluctuation problem, in which income risk is an input estimated outside the model. Depending on the numerical methods used to solve the IFP, one may want to represent the income process differently. In these notes, we will solve the recursive income fluctuation problem on a grid, i.e., we will solve for the optimal saving policy at a finite number of grid points. Thus, to evaluate the optimal saving policy off the grid requires interpolation. Alternatively, we could assume that the saving policy belongs to a certain class of functions (e.g. polynomials of order X) and solve for the parameters of the function that best approximate the optimal saving policy.

Because we want to solve the recursive problem on a grid, we want to have a recursive finite-state representation of the stochastic income process that is Markov. One option is to just estimate non-parametrically a Markov process directly using micro data (see De Nardi, Fella, and Pardo (2016)). Another option is to estimate a finite-state Markov process that approximates a continuous-state income process. For example, there is a large literature in labor economics that estimates ARMA income processes, and one can approximate this ARMA process with a discrete-state Markov process without ever working with the micro data.

Here we explore the classic Tauchen (1986) method, as well as the Rouwenhurst method, which, as discussed in Kopecky and Suen (2010), has much better properties when approximating highly persistent processes.[4] The Tauchen (1986) method is parametric, in that it assumes a functional form for the distribution of the innovation. Instead, the Rouwenhurst method iteratively constructs a Markov chain that is binomially distributed, such that parameters of the binomial distribution can be set so that the Markov chain as the same unconditional mean, unconditional variance, conditional mean, conditional variance, and autoregressive parameter as the target AR(1) exactly.

For exposition, let's approximate $\tilde{y}$ that follows an AR(1) with a finite state Markov process. Let $y_j \in \{y_1, y_2, \ldots, y_N\}$.

$$\tilde{y}_j = \rho \, \tilde{y}_{j-1} + \tilde{\epsilon}_j \tag{26}$$

$$\tilde{\epsilon} \sim G \text{ i.i.d. s.t. } \mathbb{E}[\tilde{\epsilon}] = 0 \text{ and } Var(\tilde{\epsilon}) = \sigma_{\epsilon}^2 \tag{27}$$

3.1 Tauchen Method

The main idea is to choose the elements of the state space and the transition probabilities.

1. Using conditional distribution to calculate transition probabilities instead of choosing transition probabilities to target moments of the conditional distribution.

To choose the state space, choose max value y_N to be a multiple m of the unconditional standard deviation:

$$y_N = m \left(\frac{\sigma_{\epsilon}^2}{1 - \rho^2} \right)^{0.5} \tag{28}$$

let minimum value $y_1 = -y_N$, and let $\{y_2, y_3, \ldots, y_{N-1}\}$ be equally spaced on interior interval with distance $d := y_n - y_{n-1}$. A typical recommendation for m is something like 3 or 4.

To calculate the transition probabilities, calculate the probability the innovation will send the

[4] See also Tauchen and Hussey (1991) for another classic quadrature based technique.

random variable from its starting point h closest to a particular point k. Let

$$\pi_{hk} := p(y_j = y_k | y_{j-1} = y_h) = p(y_k - d/2 < \rho y_h + \epsilon_j \leq y_k + d/2) \quad (29)$$

$$= p(y_k - d/2 - \rho y_h < \epsilon_j \leq y_k + d/2 - \rho y_h) \quad (30)$$

Define the standardized distribution of the innovation with unit variance F:

$$p(\tilde{\epsilon} < \bar{\epsilon}) = G(\bar{\epsilon}) = F\left(\frac{\bar{\epsilon}}{\sigma_{\tilde{\epsilon}}}\right) \quad (31)$$

In practice, F assumed to be standard normal.

Then choose the transition probabilities such that for interior nodes $1 < k < N-1$ for each h

$$\pi_{hk} = F\left(\frac{y_k + d/2 - \rho y_h}{\sigma_\epsilon}\right) - F\left(\frac{y_k - d/2 - \rho y_h}{\sigma_\epsilon}\right) \quad (32)$$

while for the boundaries $k = 1$ and $k = N$

$$\pi_{h1} = F\left(\frac{y_1 + d/2 - \rho y_h}{\sigma_\epsilon}\right) \quad (33)$$

$$\pi_{hN} = F\left(\frac{y_N - d/2 - \rho y_h}{\sigma_\epsilon}\right). \quad (34)$$

Note as $N \to \infty$, approximation improves.

Other rules for choosing the state space, such as quadrature, are likely to provide better approximations through more efficient placement of points in the interior interval. For example, the state space can be determined by $y_n = \sqrt{2}\sigma x_i$, where x_i are the Gauss–Hermite nodes defined on $[-\infty, \infty]$. Then the elements of the transition matrix are appropriately weighted ratios of normal densities. See Tauchen and Hussey (1991) for details.

Accuracy of the approximation

$$y_j = \rho \ y_{j-1} + \epsilon_j \quad (35)$$

and

$$\rho = Cov(y_j, y_{j-1})/Var(y_j) \quad (36)$$

$$\sigma_\epsilon^2 = (1 - \rho^2)Var(y_j) \quad (37)$$

We can compute the unconditional second moments of the y_j distribution directly. Let P be the transition matrix with elements π_{jk}. Then, the stationary distribution π^* solves

$$\pi^* = P\pi^* \quad (38)$$

$$(I - P)\pi^* = 0. \quad (39)$$

This is an $N \times N$ system of linear equations. One can now directly compare ρ to $\tilde{\rho}$ and σ_ϵ^2 to $\sigma_{\tilde{\epsilon}}^2$. Alternatively, a long sample can be simulated for the Markov process and key statistics can be compared to those of the original AR(1) process.

Why does the Tauchen method fail?

1. As $\rho \to 1$ the transition matrix goes to the identity matrix too fast.

2. Too fast means exponentially faster relative to Rowenhourst

3. Thus, Tauchen's method overshoots persistence and has bad properties matching the underlying conditional distribution

3.2 Rouwenhorst Method

How does the Rowenhourst Method work? The big picture is to choose transition probabilities to target moments of the conditional distribution.

1. Use a symmetric and evenly spaced state space with N grid points and $-\Psi, \Psi$ endpoints.

2. Construct the transition matrix via recursion on the number of grid points. It is a function of 2 parameters, p and q.

3. Start with Θ_2, a 2×2 transition matrix with $(1,1)$ element p and $(2,2)$ element q

$$\Theta_2 = \begin{bmatrix} p & 1-p \\ 1-q & q \end{bmatrix} \tag{40}$$

4. For $N \geq 3$, construct the $N \times N$ matrix as the sum of 4 components

   - p times Θ_{N-1} stacked in the upper left corner of an $N \times N$ matrix of zeros plus
   - 1-p times Θ_{N-1} stacked in the upper right corner of an $N \times N$ matrix of zeros plus
   - 1-q times Θ_{N-1} stacked in the bottom left corner of an $N \times N$ matrix of zeros plus
   - q times Θ_{N-1} stacked in the bottom right corner of an $N \times N$ matrix of zeros

   Let $\mathbf{0}$ be an $N-1 \times 1$ column vector of zeros.

$$\Theta_N = p \begin{bmatrix} \Theta_{N-1} & \mathbf{0} \\ \mathbf{0}' & 0 \end{bmatrix} + (1-p) \begin{bmatrix} \mathbf{0} & \Theta_{N-1} \\ 0 & \mathbf{0}' \end{bmatrix} + (1-q) \begin{bmatrix} \mathbf{0}' & 0 \\ \Theta_{N-1} & \mathbf{0} \end{bmatrix} + q \begin{bmatrix} 0 & \mathbf{0}' \\ \mathbf{0} & \Theta_{N-1} \end{bmatrix} \tag{41}$$

5. The invariant distribution of the Markov chain defined by this method is a binomial distribution, with binomial parameters related to N, p, q, Ψ. It is straightforward to calculate conditional and unconditional moments of the Markov chain as functions of N, p, q, Ψ.

6. Conditional on a realization of the underlying process $\tilde{y}_{j-1}$, the mean of $\tilde{y} = \rho\tilde{y}_{j-1}$, and the conditional variance is σ^2_ϵ. Note the unconditional variance is $\sigma^2_{\tilde{y}} = \sigma^2_\epsilon/(1-\rho^2)$.

7. Then, with a particular choice of $p = q = \frac{1+\rho}{2}$ and $\Psi = \sqrt{N-1}\sigma_{\tilde{y}}$ we can match the unconditional mean, unconditional variance, and auto-regressive parameters exactly

8. It can be shown that the conditional mean and conditional variance also match exactly

The Rowenhourst method does not create a transition matrix that is a discretized version of the conditional distribution. It uses information only on the persistence and the variance of the innovation to construct the transition matrix, not the conditional distribution of the transitions of the continuous process. This is fundamentally different to the Tauchen and Tauchen and Hussey methods. It should be clear that the Rowenhourst method can always match the unconditional variance and the persistence of the underlying process, regardless of the underlying persistence.

See Galindev and Lkhagvasuren (2010) for extensions to VAR processes, which uses a transformation of the VAR into a system of cross-correlated AR(1) processes with the useful structure that they are either independent of each other or perfectly correlated with one of the independent univariate processes. Then univariate methods can be applied.

4 The Income Fluctuation Problem

$$V(a,y) = \max_{a'} \; u(c) + \beta \; \sum_{y'} \pi_{y'y} V(a',y')]$$

s.t.

$$c + a' = (1+r)a + y$$
$$a' \geq a_{min}$$

with r exogenous and a given Markov process for y.

4.1 Theory

Four important saving motives are

1. Intertemporal saving motive: $\beta(1+r) > 1$
2. Smoothing: Permanent income hypothesis and concave utility
3. Precautionary saving motive due to uncertainty: Prudence in preferences ($u''' > 0$; DARA is sufficient) or borrowing constraints
4. Non-stationary environment: Life-cycle considerations, e.g., saving for retirement.

The Permanent Income Hypothesis Much can be learned about the income fluctuation problem from studying the strict version of the permanent income setting with quadratic utility and $\beta R = 1$.

We will first go over some simple settings to get intuition on the saving/consumption problems in endowment economies. Index different agents by i.

Autarky (no financial markets): $c_t^i = y_t^i$.

Complete markets.
Arrow-Debreu Budget:

$$\sum_{t=1}^{\infty} \sum_{s^t \in S^t} p_t(s^t)[c_t^i(s^t) - y_t^i(s^t)] = 0.$$

Using the First Welfare theorem, characterize complete markets competitive equilibrium allocations as solution to Pareto problem

$$\max_{c_t^i(s^t)} \sum_{t=1}^{\infty} \sum_{s^t \in S^t} \beta^t \pi(s^t) \sum_i \alpha^i u(c_t^i(s^t))$$

s.t.

$$\sum_i c_t^i(s^t) = Y_t(s^t) \; \forall \; t \;\; s^t$$

which has solution implying

$$\frac{u'(c_t^i(s^t))}{u'(c_t^j(s^t))} = \frac{\alpha^j}{\alpha^i} \; \forall \; s^t, (i,j).$$

This is exactly the definition of full insurance, since the ratio of marginal utilities across all agents is constant. Letting utility be CRRA and summing over individuals shows

$$c_t^j(s^t) = \left[\frac{(\alpha^j)^{\frac{1}{\gamma}}}{\sum_i (\alpha^i)^{\frac{1}{\gamma}}}\right] C_t(s^t)$$

Thus individual consumption tracks aggregate consumption for all households and if there is no aggregate income fluctuation, consumption is constant for all individuals.

Let us now consider an intermediate case, the Permanent Income Hypothesis economy. First consider the sequential form of the complete markets economy. The budget constraint becomes

$$c_t^i(s^t) + \sum_{s_{t+1}} q_t(s_{t+1}, s^t) a_{t+1}^i(s_{t+1}, s^t) = y_t^i(s^t) + a_t^i(s^t) \quad \forall\ t,\ s^t \tag{42}$$

where $q_t(s_{t+1}, s^t)$ is the price at date t and history s^t of an Arrow security, i.e., a security that pays out one unit of consumption if state s_{t+1} realizes in period $t+1$. Now, in the PIH economy, the set of securities available for trade is restricted.

$$c_t^i(s^t) + q_t(s^t) a_{t+1}^i(s^t) = y_t^i(s^t) + a_t^i(s^t) \quad \forall\ t,\ s^t \tag{43}$$

where $q_t(s^t)$ is the price at date t of an asset that pays one unit of consumption at date $t+1$ independent of the realization of the state s_{t+1}—a one period risk free bond.

As we will see, the absence of insurance induces the consumer to hold some of the bond in order to smooth consumption. Lets assume there are no aggregate fluctuations for now, so

$$q_t(s^t) = q = 1/(1+r) \tag{44}$$

$$a_{t+1} = (1+r)(y_t + a_t - c_t) \tag{45}$$

Finally, impose a No-Ponzi condition

$$\mathbb{E}\left[\lim_{\tau\to\infty}\left(\frac{1}{1+r}\right)^{\tau} a_{t+\tau}\right] \geq 0 \tag{46}$$

and note

$$a_t = \sum_{j=0}^{\infty}\left(\frac{1}{1+r}\right)^{t} \mathbb{E}_t[c_{t+j} - y_{t+j}]. \tag{47}$$

Strict PIH: Quadratic Utility and $\beta R = 1$ Let

$$u(c) = b_1 c_T - 0.5 b_2 c_t^2 \tag{48}$$

Then, from the consumption Euler equation

$$b_1 - b_2 c_t = \beta R \mathbb{E}_t[b_1 - b_2 c_{t+1}] \tag{49}$$

$$\mathbb{E}_t[c_{t+1}] = c_t \tag{50}$$

i.e., consumption is a martingale. Furthermore, using the law of iterated expectations and the martingale property

$$\mathbb{E}_t[c_{t+j}] = c_t \quad \forall\ j \geq 0 \tag{51}$$

Iterating forward on the budget constraint (i.e., a_{t+1} as a function of a_{t+2}, etc.) using conditional expectations and the martingale property of consumption yields

$$\sum_{j=0}^{\infty}\left(\frac{1}{1+r}\right)^{j}\mathbb{E}_t c_{t+j} = a_t + \sum_{j=0}^{\infty}\left(\frac{1}{1+r}\right)^{j}\mathbb{E}_t y_{t+j} \tag{52}$$

$$c_t = \frac{r}{1+r}\left[a_t + \sum_{j=0}^{\infty}\left(\frac{1}{1+r}\right)^{j}\mathbb{E}_t y_{t+j}\right] \tag{53}$$

$$c_t = \frac{r}{1+r}\left[a_t + H_t\right] \tag{54}$$

where we have defined human wealth H_t to be the expected discounted value of future earnings. Finally, we can define permanent income as the annuity value $(r/(1+r))$ of total wealth including human and financial $W_t := a_t + H_t$.

Thus, we have proved that in the strict PIH environment, consumption is a martingale and equals permanent income, the annuity value of human and financial wealth.

Also note that in the deterministic case, from the Euler equation, $c_t = c_{t+1}$ and iterating on the budget constraint yields

$$c_t = \frac{r}{1+r}\left[a_t + \sum_{j=0}^{\infty}\left(\frac{1}{1+r}\right)^{j} y_{t+j}\right] \tag{55}$$

This highlights a certainty equivalence property, in that c_t in the stochastic environment is just c_t derived in the deterministic case in which conditional expectations of income are used instead of income realizations. That is, the optimal consumption policy is independent of the variance and higher moments of the income process.

So, what do consumption dynamics look like, given realizations of the income shocks? Using that consumption is a martingale,

$$\Delta c_t = c_t - c_{t-1} = c_t - \mathbb{E}_{t-1}[c_t] = \frac{r}{1+r}(W_t - \mathbb{E}_{t-1}W_t) \tag{56}$$

The change in consumption is the unexpected change in total wealth. I.e.,

$$W_t - \mathbb{E}_{t-1}W_t = \sum_{j=0}^{\infty}\left(\frac{1}{1+r}\right)^{j}(\mathbb{E}_t - \mathbb{E}_{t-1})y_{t+j} \tag{57}$$

thus, the change in consumption between period t and $t-1$ is proportional to the change in expected earnings due to new information that arrives in period t.

$$\Delta c_t = \frac{r}{1+r}\sum_{j=0}^{\infty}\left(\frac{1}{1+r}\right)^{j}(\mathbb{E}_t - \mathbb{E}_{t-1})y_{t+j} \tag{58}$$

Strict PIH with Permanent and Transitory Income Components Let income follow

$$y_{i,t} = \epsilon_{i,t} + \nu_{i,t} \tag{59}$$

$$\epsilon_{i,t} = \epsilon_{i,t-1} + \eta_{i,t} \tag{60}$$

It is easy to see that

$$(\mathbb{E}_t - \mathbb{E}_{t-1})y_t = \eta_t + \nu_t \tag{61}$$

That is, given innovations are distributed with mean zero, the difference between expected income at time $t-1$ and realized income at t is the value of the innovations. Furthermore, since the transitory innovation is i.i.d. over time, it is easy to see

$$(\mathbb{E}_t - \mathbb{E}_{t-1})y_{t+j} = \eta_t \quad \forall\ j > 1 \tag{62}$$

Thus, going back to the expression for consumption dynamics

$$\Delta c_t = \frac{r}{1+r}\left[\nu_t + \eta_t + \frac{1}{1+r}\eta_t + \left(\frac{1}{1+r}\right)^2 \eta_t + \dots\right] \tag{63}$$

$$\Delta c_t = \frac{r}{1+r}\left[\nu_t + \eta_t \sum_{j=0}^{\infty}\left(\frac{1}{1+r}\right)^j\right] \tag{64}$$

$$\Delta c_t = \frac{r}{1+r}\nu_t + \eta_t \tag{65}$$

Thus, households adjust their consumption weakly to transitory shocks and completely for permanent shocks.

Binding Borrowing Constraints When do borrowing constraints bind and what impact does the existence of a potentially binding borrowing constraint have on consumption and wealth dynamics?

Note from the budget constraint

$$a_{t+1} = (1+r)(y_t + a_t - c_t) \tag{66}$$

$$\Delta a_{t+1} = (1+r)y_t + ra_t - (1+r)c_t \tag{67}$$

Using the optimal consumption choice derived above it can be shown after some algebra that

$$\Delta a_{t+1} = -\sum_{j=0}^{\infty}\left(\frac{1}{1+r}\right)^{j-1} \mathbb{E}_t \Delta y_{t+j} \tag{68}$$

If income follows a random walk with innovation η, then $\Delta y_{t+j} = \eta_{t+j}$, $\mathbb{E}_t \Delta y_{t+j} = 0$, and $\Delta a_{t+1} = 0$. Wealth only changes when some of income is consumed and some is saved, and in the PIH world with RW shocks, all labor income is consumed each period (all shocks in income are met one for one into shocks in consumption). Thus, if the individual starts with wealth at time zero $a_0 > a_{min}$, then the borrowing constraint will never bind.

If income is i.i.d., with innovation ν_t, then $\Delta y_{t+j} = \Delta \nu_{t+j}$, and $\Delta a_{t+1} = \nu_t$. That is, financial wealth follows a random walk. Thus, wealth will hit the borrowing constraint almost surely.

Suppose there is an exogenously specified borrowing constraint such that $a_{min} = 0$. Then the Euler equation needs to be modified to account for the Lagrange multiplier on the borrowing constraint, and thus instead of $\mathbb{E}_t(c_{t+1}) = c_t$, we have

$$c_t = \begin{cases} \mathbb{E}_t(c_{t+1}) & \text{if } a_{t+1} > 0 \\ y_t + a_t & \text{if } a_{t+1} = 0 \end{cases} \tag{69}$$

That is, when the household wants to consume more than its wealth allows and it cannot borrow, then it consumes all of its income and wealth, i.e., it goes as far as it can (recall this is in the quadratic utility case, so there is no prudence inducing precautionary savings).

Note, this can be rewritten as

$$c_t = \min\{y_t + a_t, \mathbb{E}_t[c_{t+1}]\} \tag{70}$$
$$= \min\{y_t + a_t, \mathbb{E}_t[\min\{y_{t+1} + a_{t+1}, \mathbb{E}_{t+1}[c_{t+2}]\}] \tag{71}$$

Imagine a mean preserving spread in income. Very low realizations of y_{t+1} become more likely, which makes the borrowing constraint in the future more likely to bind, which reduces the value of $\mathbb{E}_t[\min\{y_{t+1} + a_{t+1}, \mathbb{E}_{t+1}[c_{t+2}]\}]$, which reduces the value of $\mathbb{E}_t[c_{t+1}]$. So, even if the borrowing constraint is not binding at time t the fact that it might be binding in the future makes agents consume less today. This is how a borrowing constraint can both directly decrease consumption, but also indirectly decrease consumption through precautionary saving motives driven by forward looking considerations and risk. Thus, prudence is not necessary for there to be a precautionary saving motive, which can be induced with a borrowing constraint even with quadratic utility. Furthermore, this force of a borrowing constraint inducing precautionary saving is present even when prudence is present, e.g., CRRA utility.

Compact State Space Also, in a variety of settings, we can show conditions on βR such that the optimal consumption sequence is bounded above, which implies a compact asset space. This is very important, since this is a necessary condition for existence of stationary equilibria.

Let λ be the Lagrange multiplier on the borrowing constraint. FOC:

$$u'(c_t) = \beta \ R \ \mathbb{E}_t[u'(c_{t+1})] + \lambda_t \tag{72}$$
$$u'(c_t) \geq \beta \ R \ \mathbb{E}_t[u'(c_{t+1})] \tag{73}$$

For example, consider the deterministic case if $\beta R > 1$. Note that $u'(c) \geq \beta R u'(c_{t+1}) > u'(c_{t+1}) \rightarrow c_{t+1} > c_t$ by concavity. Thus, c_t is unbounded and no stationary equilibrium with finite consumption and assets exists.

In the stochastic income case, it can be shown a necessary condition for existence of equilibrium is $\beta R < 1$. See the Ljungqvist and Sargent chapter on self-insurance for details.

4.2 Numerical Methods

Recall,

$$V(a, y) = \max_{a'} \ u(c) + \beta \sum_{y'} \pi_{yy'} V(a', y')$$

s.t.

$$c + a' = (1 + r)a + y$$
$$a' \geq a_{min}$$

This problem yields the Euler equation

$$u'(c(a, y)) \geq \beta \ (1 + r) \sum_{y'} \pi_{yy'}[u'(c(a', y'))]$$

Note, equality of the Euler equation holds if $a' > a_{min}$. We are trying to solve for the optimal consumption policy $c(a, y)$, which is an invariant function of the states, satisfies the Euler equation, and does not violate the borrowing constraint.

The Endogenous Grid Method. Start from a guess $c_0(a, y)$ and iterate on the Euler equation until consecutive iterations are close enough.

1. Construct a grid on (a', y) (not (a, y)) with $a' \in G_A = \{a_1, a_2, \ldots, a_{N_a}\}$ s.t. $a_1 = a_{min}$ and $y \in G_y = \{y_1, y_2, \ldots, y_N\}$. Choose $\bar{a}$ so that it is large enough that in practice $a' < \bar{a}$ always. Given the non-linearity of the optimal policy driven by the borrowing constraint, place more grid points closer to a_{min}. In practice, N_a is substantially larger than N.

2. Guess $c_0(a, y)$. A decent guess is $c_0(a_i, y_j) = ra_i + y_j$.

3. Label the RHS of the Euler equation $B(a'_i, y_j)$. For all pairs (a'_i, y_j) on the mesh $G_a \times G_y$ calculate and store $B(a'_i, y_j)$.

$$B(a'_i, y_j) = \beta\ (1 + r)\ \sum_{y'} \pi_{yy'}[u'(c_0(a'_i, y'))] \tag{74}$$

4. Solve for the value $\hat{c}(a'_i, y_j)$ that satisfies the Euler equation given by

$$u'(\hat{c}(a'_i, y_j)) = B(a'_i, y_j) \tag{75}$$

   Note that this can be done analytically. For example, if $u'(c) = c^{-\gamma}$ then $\hat{c}(a'_i, y_j) = [B(a'_i, y_j)]^{-\frac{1}{\gamma}}$. This is the essential element of the endogenous grid method that provides such increased computational efficiency. If the grid was over a instead of a', then a non-linear equation solver must be used to solve the Euler equation. Also, the expectation in the Euler equation is computed once and stored, whereas it would be computed each time inside the nonlinear equation solver in the traditional approach, which would also require interpolation.

5. Use the budget constraint to solve for $a^*(a'_i, y_j)$:

$$\hat{c}(a'_i, y_j) + a'_i = (1 + r)a^*(a'_i, y_j) + y_j \tag{76}$$

   This implicitly defines the function $c(a^*_i, y_j) = \hat{c}(a'_i, y_j)$. $a^*(a'_i, y_j)$ is the value of assets today that would optimally result in a'_i assets tomorrow given an income shock of y_j today. These a^* are the elements of the endogenous grid, which change on each iteration and are not generally elements of G_A. Let a^*_1 be the value for current assets that just induces the borrowing constraint to bind next period. That is, let a^*_1 be the a^* that solves the budget constraint equation for a'_1.

6. The next step is to update the guess to $c_1(a_i, y_j)$.

   (a) For grid points $a'_i > a^*_1$, use interpolation (e.g., linear) using values $c(a^*_i, y_j)$ on the two adjacent values (a^*_n, a^*_{n+1}) that straddle a given grid point a_i.

   (b) For grid points $a'_i < a^*_1$ the borrowing constraint is binding, so instead of using the Euler equation, use the budget constraint:

$$c(a_i, y_j) = (1 + r)a_i + y_j - a'_1 \tag{77}$$

7. Continuing iterating until convergence. Check convergence by comparing $c_{n+1}(a'_i, y_j)$ to $c_n(a'_i, y_j)$. For example, stop at iteration $n + 1$ when

$$\max_{i,j} |c_{n+1}(a'_i, y_j) - c_n(a'_i, y_j)| < \epsilon \tag{78}$$

   for some small ϵ.

4.2.1 Accuracy of Numerical Solution

One method to check the accuracy of the numerical solution is to analyze the Euler equation error.

$$u'(c(a,y)) - \beta\ (1+r)\ \sum_{y'} \pi_{yy'}[u'(c(a',y')] \approx 0 \tag{79}$$

When evaluated on the grid points, this approximation should be quite precise, but the error could be large when evaluated off the grid.

Let ϵ be the approximation error defined such that the equation holds exactly at (a,y)

$$u'(c(a,y)(1-\epsilon(a,y))) - \beta\ (1+r)\ \sum_{y'} \pi_{yy'}[u'(c(a',y')]. \tag{80}$$

Then

$$\epsilon(a,y) = 1 - \frac{u'^{-1}\left(\beta\ (1+r)\ \sum_{y'} \pi_{yy'}[u'(c(a',y'))]\right)}{c(a,y)}. \tag{81}$$

$\epsilon(a,y) = 0.01$ means that the consumer is making a mistake equal to \$1 for each \$100 consumed when choosing c in the state (a,y). Santos (2000) shows the cost of a mistake in consumption in terms of welfare is of the order ϵ^2. See Aruoba, Fernandez-Villaverde, and Rubio-Ramirez (2006) for an analysis of various numerical solution methods and a comparison of accuracy.

References

Abowd, J. M., and D. Card (1989): "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, 57(2), 411–445.

Altonji, J. G., and L. Segal (1996): "Small-Sample Bias in GMM Estimation of Covariance Structures," *Journal of Business and Economic Statistics*, 14(3), 353–366.

Aruoba, S. B., J. Fernandez-Villaverde, and J. F. Rubio-Ramirez (2006): "Comparing Solution Methods for Dynamic Equilibrium Economies," *Journal of Economic Dynamics and Control*, 30, 2477–2508.

Blundell, R., L. Pistaferri, and I. Preston (2008): "Consumption Inequality and Partial Insurance," *American Economic Review*, 98:5, 1887–1921.

Chamberlain, G. (1984): "Panel Data," in *Handbook of Econometrics*, ed. by Z. Griliches, and M. D. Intriligator, vol. 2, pp. 1247–1318. North-Holland.

Constantinides, G. M., and D. Duffie (1996): "Asset Pricing with Heterogeneous Consumers," *Journal of Political Economy*, 104(2), 219–240.

De Nardi, M., G. Fella, and G. P. Pardo (2016): "The Implications of Richer Earnings Dynamics for Consumption and Wealth," *NBER WP 21917*.

Galindev, R., and D. Lkhagvasuren (2010): "Discretization of Highly Persistent Correlated AR(1) Shocks," *Journal of Economic Dynamics and Control*, 34(7), 262–280.

Geweke, J., and M. Keane (2000): "An Empirical Analysis of Earnings Dynamics Among Men in the PSID: 1968-1989," *Journal of Econometrics*, 96, 293–356.

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

——— (2009): "An empirical investigation of labor income processes," *Review of Economic Dynamics*, 12(1), 58–79.

Guvenen, F., S. Ozkan, and J. Song (2014): "The Nature of Countercyclical Income Risk," *Journal of Political Economy*.

Heathcote, J., K. Storesletten, and G. L. Violante (2014): "Consumption and Labor Supply with Partial Insurance: An Analytical Framework," *American Economic Review*, 104(7), 2075–2126.

Heaton, J., and D. J. Lucas (1996): "Evaluating the Effects of Incomplete Markets on Risk Sharing and Asset Pricing," *The Journal of Political Economy*, 104(3), 443–487.

Kopecky, K. A., and R. M. H. Suen (2010): "Finite State Markov Chain Approximations to Highly Persistent Processes," *Review of Economic Dynamics*, 13(3), 701–714.

Lillard, L. A., and Y. Weiss (1979): "Components of Variation in Panel Earnings Data: American Scientists 1960-1970," *Econometrica*, 47(2), 437–454.

MaCurdy, T. E. (1982): "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis," *Journal of Econometrics*, 18(1), 83–114.

Meghir, C., and L. Pistaferri (2004): "Income Variance Dynamics and Heterogeneity," *Econometrica*, 72(1), 1–32.

Murphy, K. M., and R. Topel (1985): "Estimation and Inference in Two-Step Econometric Models," *Journal of Business and Economic Statistics*, 3(4), 88–97.

Nakata, T., and C. Tonetti (2015): "Small Sample Properties of Bayesian Estimators of Labor Income Processes," *Journal of Applied Economics*, 18(1), 122–148.

Santos, M. S. (2000): "Accuracy of Numerical Solutions Using the Euler Equation Residuals," *Econometrica*, 68(6), 1377–1402.

Storesletten, K., C. I. Telmer, and A. Yaron (2001): "The welfare cost of business cycles revisited: Finite lives and cyclical variation in idiosyncratic risk," *European Economic Review*, 45(7), 1311–1339.

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

Tauchen, G., and R. Hussey (1991): "Quadrature-Based Methods for Obtaining Approximate Solutions to Linear Asset Pricing Models," *Econometrica*, 59(2), 371–396.