Estimating Intertemporal Allocation Parameters using Synthetic Residual Estimation[1]

Sule Alan
Faculty of Economics
University of Cambridge
Sule.Alan@econ.cam.ac.uk

Martin Browning
Department of Economics
University of Oxford
Martin.Browning@economics.ox.ac.uk

March 30, 2009

[1]We thank two referees, the editor, Japp Abbring, Orazio Attanasio, Richard Blundell, Thomas Crossley, Hide Ichimura, Hamish Low, Lonnie Magee, Mike Veall and participants in the 2001 NBER Summer workshop and seminars at UCL and CREST for comments on the earlier work reported in this paper. The first draft of the paper was written during a visit by Sule Alan to the Institute of Economics at the University of Copenhagen. She thanks the Institute and the Centre for Applied Microeconometrics (CAM) and Social Sciences and Humanities Research Council of Canada (SSHRC) for support. We are grateful to the Danish National Research Foundation for support through its grant to CAM.

Abstract

We present a novel structural estimation procedure for models of intertemporal allocation. This is based on modelling expectations errors directly; we refer to it as Synthetic Residual Estimation (SRE). The flexibility of SRE allows us to account for measurement error in consumption and for heterogeneity in intertemporal allocation parameters. An investigation of the small sample properties of the SRE estimator indicates that it dominates GMM estimation of both exact and approximate Euler equations in the case when we have short panels and noisy consumption data. We apply SRE to two panels drawn from the PSID and estimate the joint distribution of the discount factor and the coefficient of relative risk aversion. We reject strongly homogeneity of the discount factor and the coefficient of relative risk aversion. We find that, on average, the more educated are more patient and more risk averse than the less educated. Within education strata, patience and risk aversion are negatively correlated.

1 Introduction

We consider the familiar intertemporal allocation model with iso-elastic preferences. If we have exponential discounting and there are no liquidity constraints, the resulting *exact Euler equation* for consumption growth is:

$$\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}(1+r_{t+1})\,\beta=\varepsilon_{t+1} \tag{1}$$

where C_t is consumption in period t, γ is the coefficient of relative risk aversion[1], β is the discount factor and r_{t+1} is the real rate of interest between periods t and $t+1$. The term ε_{t+1} is a 'surprise' term which satisfies the orthogonality condition:

$$E_t(\varepsilon_{t+1})=1 \tag{2}$$

where $E_t(.)$ denotes the expectation operator conditional on information available at time t.

Over the past quarter century this theoretical framework has been the principal vehicle for estimating preference parameters such as β and γ, and for testing for the validity of the standard orthogonality assumptions in general and for the 'excess sensitivity' of consumption to predictable income growth in particular. GMM estimation is based on the orthogonality condition (2) using instruments dated t or before such as lagged consumption, interest rate and income variables. The attraction of estimation based on equation 1 is that one can estimate the preference parameters without explicitly parameterizing the stochastic environment that agents face.

Browning and Lusardi (1996) discuss the results of 25 studies using Euler equation methods on micro data and conclude that the results are disappointing. A number of subsequent Monte Carlo based papers have investigated why we experience this failure (Carroll 2001, Ludvigson and Paxson, 2001, Attanasio and Low, 2004). The problems identified are manifold but the most

[1] In this framework the elasticity of intertemporal substitution (eis) is the reciprocal of the coefficient of relative risk aversion. Therefore we use these two terms interchangebly throughout the text.

important seems to be the paucity of appropriate data (long panels on demands or consumption) and the substantial measurement error in consumption (see Shapiro (1984), Altonji and Siow (1987) and Runkle (1991)). Regarding the latter, Runkle (1991), for example, estimates that 76% of the variation in the growth rate of food consumption in the PSID is noise.[2] Measurement error of this magnitude means that we cannot use the exact Euler equation for estimation since the equation is nonlinear in parameters (a point first made in the general context of nonlinear GMM by Amemiya (1985)). Generally, the presence of measurement error when estimating non-linear equations raises serious and difficult problems.[3] Alan *et al* (2009) present two estimation strategies that allow for 'classical' measurement error in exact Euler equations, but these require moderate length panels and cannot be extended to allow for heterogeneity in preference parameters.

An early reaction to these problems was to linearize equation 1 by taking logs and approximating $\ln \varepsilon_{t+1}$ in some way. The use of linearized Euler equations (whether first or second order) solves the measurement error problem[4] but the transformation $\ln \varepsilon_{t+1}$ introduces latent variables that lead to violations of the orthogonality conditions exploited by GMM methods. Given these problems, Carroll (2001) concludes that 'empirical estimation of consumption Euler equations should be abandoned'. On the other hand, Attanasio and Low (2004) present results that suggest that Carroll's conclusion is overly pessimistic if we have long panels (40 periods, say) and time series variation in real rates. We do not find this conclusion too comforting for empirical work since we do not have long consumption panels.

Thus the emerging consensus seems to be that we must give up on empirical Euler equations and resort to estimating (or calibrating) structural dynamic programming models based on

[2]The other widely used data resource for consumption studies are quasi-panels. These are constructed from cross-section expenditure survey information by taking within-period means following the same population (for example, means over all the 25-year-olds in one year and all the 26-year-olds in the next year). Although this averaging reduces the effect of measurement error, the construction of quasi-panels from samples which change over time induces sampling error which is very much like measurement error.

[3]In the wider measurement error literature, resolutions of the problem for nonlinear estimators have only been possible in particular circumstances; see Hausman (2001), Schennach (2004), Wansbeek (2001) and Hong and Tamer (2003).

[4]If the measurement is classical in the sense of being multiplicative and independent of everything else.

specifying the environment agents face (see Hubbard, Skinner and Zeldes (1995), Carroll and Samwick (1997), Gourinchas and Parker (2002)). The main problem with this approach is that because of computational complexities it can only accommodate very limited sources of uncertainty and heterogeneity.

In this paper, we propose an alternative approach to estimating the parameters of intertemporal allocation. The key to our approach is that associated with every structural model there is a distribution for the expectations errors (the ε_{t+1}'s in equation 1). If we knew this distribution, then we could identify preference parameters from a path of consumption and interest rates. The problem is that the distribution of expectations errors is not known and may depend on preference parameters in an unknown way. In section 2 we show that for a wide class of models with heterogeneous agents, the distribution of (pooled) expectations errors can be well approximated by a mixture of two lognormals which is independent of preference parameters. Using this result, we show that from equation 1 we can jointly identify the preference parameters and the parameters of the mixture distribution. We term our new procedure *synthetic residual estimation* (SRE) to reflect that it relies on generating synthetic expectations errors, ε_{t+1}. There are a variety of possible estimation methods that could be used; we use Simulated Minimum Distance (SMD)[5] since it allows us to adopt heterogeneity schemes for which it is very difficult to write down the likelihood function. We lay out the details of our simulation based estimation procedure in section 3. In section 4 we compare the small sample properties of SRE with linearized and exact GMM when we do not have any heterogeneity.[6] These Monte Carlo results suggest that even when there is considerable measurement error (for example, half the observed consumption growth variance is due to noise) SRE works well both in absolute terms and relative to GMM, even for moderately short panels.

In the second half of the paper we present an empirical application of SRE. The major inno-

[5] The class of SMD estimators (see Hall and Rust (2002)) includes the Efficient Method of Moments procedure of Gallant and Tauchen (1996) and the Indirect Inference method of Gouriéroux, Monfort and Renault (1993).

[6] Current GMM methods do not provide any way to allow for heterogeneity.

vation in our modelling is that we allow heterogeneity in the discount factor and the coefficient of relative risk aversion. A number of regularities observed in consumption and wealth data can be rationalized by allowing for heterogeneity in the discount factor and/or in risk aversion. The most important of these is the heterogeneity in lifetime wealth accumulation by households with similar earnings profiles. This requires heterogeneity in the discount factor (see Samwick (1998), Krusell and Smith (1998) and Hendricks (2007)). The only estimates of the distribution of discount factors within the context of consumption life cycle models are due to Lawrence (1991), Samwick (1998) and Cagetti (2003). Heterogeneity in risk aversion (or eis) also has great potential for explaining some regularities, particularly for household portfolio allocations. To our knowledge, there are no estimates of the distribution of the eis in the consumption literature. There are, however, several papers in the literature that indicate that the eis is very likely to be heterogenous.[7]

In the empirical application we consider two samples of households drawn from the PSID from 1974 to 1987, based on their broadly defined education group membership. In section 5 we present our sample selection, variable definitions and some of the features of our two samples. In particular, we show that even within education strata there is considerable variation across households in the mean and standard deviation of consumption growth. We use this variation to identify the joint distribution of the discount factor and the coefficient of relative risk aversion. We present our results and their implications in section 6. In line with other studies based on consumption and wealth data, we find that the more educated are more patient than the less educated. The median discount factors are 0.93 and 0.96 for the less educated and the more educated respectively. There is also considerable heterogeneity within education strata with a significant fraction of each stratum having a discount factor below 0.9 and a high proportion of the educated having a value close to unity. We discuss how these estimates should be interpreted

[7]See Attanasio *et al* (2002), Vissing-Jorgensen (2002) and Guvenen (2006) for evidence based on stock market participation and Dohmen *et al* (2005) and Guiso and Piaella (2001) for evidence based on survey responses to risk attitude questions.

in section 5 as our sample selection procedure excludes all liquidity constrained, and potentially high discount rate households.

For the coefficient of relative risk aversion, we find that the less educated households are less risk averse than the more educated households. The medians of the two distributions are 6.2 and 8.4 respectively. These values are higher than those estimated in consumption based studies but closely in line with wealth and portfolio choice based studies. The finding that the less educated have a higher discount rate and a lower coefficient of relative risk aversion than the more educated implies that patience and risk aversion are positively correlated across the two education strata. Within strata, however, we find the opposite result of a negative correlation between patience and risk aversion; this is consistent with experimental evidence, which uses subjects who have the same education level.

2 The Distribution of Expectations Errors

Our estimator is based on sampling from the conditional distribution of the expectations errors (the ε_{t+1}'s in the Euler equation (1)). Our motivation for this is that we found that this distribution displays some strong regularities across many of the simulation models considered in the literature. We illustrate this in this section. The data generating process we use is very standard; details are given in Appendix A. Income and consumption are denoted Y_t and C_t respectively. Assets or debts are carried forward from period $t-1$ to t at a real interest rate of r_t. End of period t assets, A_t, evolve according to:

$$A_t = (1 + r_t)(A_{t-1} + Y_t - C_t) \tag{3}$$

Assuming exponential discounting and an iso-elastic felicity function, this gives the Euler equation (1).

We present simulation results for 17 variants of the standard model. These differ in the curvature of the felicity function (γ); the time discount rate ($\delta = (1-\beta)/\beta$); the income process parameters; whether the interest rate is stochastic; the presence of liquidity constraints and the degree of measurement error.

Our environment has agents with a finite lifetime of 80 periods, with no bequest motive and no initial assets. Agents face two types of income shocks, permanent and transitory. For agent h the assumed income process is:

$$Y_t = P_t u_t \tag{4}$$

where u_t is an iid lognormal shock to transitory income with unit mean and a constant variance $\left(\exp\left(\sigma_u^2\right)-1\right)$. P_t is permanent income which follows a log random walk process:

$$P_t = P_{t-1} z_t \tag{5}$$

where z_t is an iid lognormal shock to permanent income with unit mean and a constant variance $\left(\exp\left(\sigma_z^2\right)-1\right)$. In our simulations we set $\sigma_u = \sigma_z = 0.1$, and also experiment with $\sigma_z = 0.15$. Values such as these are conventional in the consumption and income literature; see Gourinchas and Parker (2002) and Low *et al* (2008). We assume that the innovations to income are independent over time and across individuals so that we assume away aggregate shocks to income. The real interest rate has a mean of 0.03 and is assumed to be the same for everyone between any two periods. For the variants that have stochastic interest rates, the process is an $AR(1)$ with a mean of 0.03, an AR parameter of 0.6 and an error with a standard deviation of 0.025.

For each variant, we first solve the dynamic program and generate a decision rule for each period. Using the decision rules we simulate 80-period consumption paths for each of $10,000$ simulated individuals. We then remove the first 20 and the last 20 periods for each agent to minimize starting and end effects and obtain 40 periods. For each pair of adjacent simulated periods we construct the expectation error (ε_{t+1}) according to equation (1). This gives 39 expectations errors for each of our agents. However, we lose one more period (giving a total of 38 periods) as we want to assess the dependence of the variance of ε_{t+1} on ε_t.

Table 1 presents the features of all 17 variants we consider. The second to fourth columns report the coefficient of relative risk aversion, the discount rate and whether the interest rate is stochastic, respectively. The standard deviation of the logarithm of permanent income shocks (σ_z) is presented in the fifth column. The last column indicates whether we impose a liquidity constraint or not. We take model 1 as our benchmark variant and make changes one at a time. Model 2 lowers the coefficient of relative risk aversion from 4 to 2; model 3 increases the discount rate from 0.05 to 0.15; and model 4 increases the standard deviation of the logarithm of permanent income shocks from 0.1 to 0.15. Models 5 and 6 impose an implicit liquidity constraint; this process is examined under two different impatience levels. For models with an explicit liquidity constraint (models 7, 8 and 11) we have to take account of the fact that the Euler equation does not hold for all periods. To do this, we remove shocks if the end of period assets in the previous period are zero; that is, if the agent does not carry forward assets between t and $t+1$ ($A_t = 0$) then ε_{t+1} for that agent is dropped. We experiment with stochastic interest rates in models 9 to 11. These models are also examined with and without liquidity constraints.

Note that models 6, 8 and 11 simulate 'buffer stock' savers as they generate very little assets due to high impatience and liquidity constraints. To capture the effect of heterogeneity we also experiment with some mixed models. Model 12 is generated by mixing simulated paths of models 1 and 2 with equal probability (heterogeneity in the coefficient of relative risk aversion); model 13 is the mixture of models 1 and 3 (heterogeneity in the discount rate); model 14 is a mixture of models 1 and 4 (heterogeneity in the variance of the income process); and model 15 is a mixture of models $1-4$. Finally, models 16 and 17 add noise to the consumption paths obtained from the baseline model (model 1). In model 16 (respectively, model 17) we introduce moderate (respectively, high) noise so that 30% (respectively, 60%) of the variance of consumption growth is due to measurement error.

The unconditional mean of the expectations errors is unity for all models except for those with measurement error. To test for the functional form of the distribution of errors we first estimate the parameters of a lognormal and then of a mixture of two lognormals for the expectations errors generated by each of the 17 models. We then perform a Kolmogorov-Smirnov goodness of fit test for the error distribution against these estimated parametric distributions.[8] Columns 2 and 3 of Table 2 present the p-values of the test statistics for each model. These indicate that we cannot always fit a lognormal but a mixture of two lognormals always fits well, even for models with heavily skewed distributions and thick tails such as those with implicit liquidity constraints (models 5 and 6) or when we mix homogeneous models (models $12-15$)[9]. It is this regularity that underpins our estimation procedure. The final two columns give the slope parameter and associated t-value from the regression of the square of the current expectations error (minus the mean of unity) on the lagged expectations error:

$$(\varepsilon_{h,t} - 1)^2 = \phi_\varepsilon + \omega_\varepsilon(\varepsilon_{h,t-1} - 1) + \epsilon_{h,t} \tag{6}$$

We run this regression to assess the degree of conditional heteroskedasticity in expectations errors. The t-values in Table 2 indicate strong conditional heteroskedasticity for most models. We have not been able to establish theoretically the sign of the dependence between past shocks and the subsequent variance. It depends on the level of accumulated assets and the marginal propensity to consume out of income. Nevertheless, the simulations suggest that it is important to account for such a dependence. Therefore, in our estimation procedure, we shall allow for this conditional heteroskedasticity and estimate ω_ε.

[8] When performing this test we do not allow the parameters of the lognormal distribution to be estimated. With such a large sample size, this should be largely irrelevant. To make the estimation tractable we only consider 500 households (19,000 observations) for each variant.

[9] We also experimented with Kruskal-Wallis and Wilcoxon signed-rank test. The results are similar; that is, we do not reject the mixture of lognormals for any of the models.

3 Synthetic Residual Estimation (SRE)

3.1 Overview

Our estimation procedure is a variant of Simulated Minimum Distance (SMD) which involves matching statistics from the data with statistics from a simulated model.[10] We define a $J-$vector of statistics (auxiliary parameters) and calculate them from the data, $\boldsymbol{\lambda}^D$. We simulate the model using parameters $\boldsymbol{\theta}$ and calculate the auxiliary parameters for the simulated data, $\boldsymbol{\lambda}^S(\boldsymbol{\theta})$. The final step is to choose parameters that minimize the weighted distance between the sample and simulated auxiliary parameters. To do this we take a $J \times J$ positive definite, data dependent weighting matrix, W, and define the SMD estimator:

$$\boldsymbol{\theta}_{SMD} = \arg\min_{\boldsymbol{\theta}} \left(\boldsymbol{\lambda}^S(\boldsymbol{\theta}) - \boldsymbol{\lambda}^D\right)' W \left(\boldsymbol{\lambda}^S(\boldsymbol{\theta}) - \boldsymbol{\lambda}^D\right) \tag{7}$$

Asymptotic properties of this estimator are given in Gourieroux and Monfort (1993).

The novelty of our approach is that rather than simulating the full model, we simulate the expectations errors and use these to construct consumption paths. For the exposition here we consider a balanced panel with $h = 1, ..., H$ households and $t = 1, ..., T$ periods. In the empirical section we discuss how to deal with the unbalanced panel that we actually use[11]. We allow the discount factor, β, and the coefficient of relative risk aversion, γ, to be heterogenous with some stochastic dependence between the two distributions and the initial values of consumption.

There are four steps for the simulation procedure. In the first step we simulate expectations errors that have the properties identified in the previous section. Thus we simulate mixtures of two unit mean lognormals, allowing for conditional heteroskedasticity. In the second step we simulate values for initial values and preference parameters. In the third step we take the simulated expectations errors, the initial values and the simulated preference parameters and

[10] A detailed description of the general SMD procedure we use is given in Appendix B

[11] We also postpone to the empirical section any discussion of how to allow for time varying observable factors such as household composition.

generate consumption paths using equation 1. Finally we add measurement error.

The simulation procedure takes a set of 15 model parameters. We present a sketch of the parameters here using the notation μ for a location parameter, ϕ for a dispersion parameter and ω for a parameter controlling the dependence between parameters. The parameters for the approximated expectations errors distribution are denoted as $(\phi_{\varepsilon 1}, \phi_{\varepsilon 2}, \omega_{\varepsilon}, \pi)$. The parameters $\phi_{\varepsilon 1}$ and $\phi_{\varepsilon 2}$ are for the dispersions of the two components of the mixture of lognormals (the means are fixed at unity). The parameter ω_{ε} controls the extent of conditional heteroskedasticity and π controls the mixing probabilities. For the distribution of the initial level of consumption we have the location and dispersion parameters (μ_1, ϕ_1). For the discount factor we have three parameters: $(\mu_{\beta}, \phi_{\beta}, \omega_{\beta 1})$. These are, respectively, related to the discount factor location, dispersion and the dependence between the discount factor and initial consumption. The parameters for the coefficient of relative risk aversion are denoted as $(\mu_{\gamma}, \phi_{\gamma}, \omega_{\beta\gamma}, \omega_{\gamma 1})$. These are, respectively, related to the location, dispersion, the dependence between the discount factor and coefficient of relative risk aversion and the dependence between the coefficient of relative risk aversion and initial consumption. The final pair of parameters (μ_m, ϕ_m) are the location and dispersion parameters for the measurement error. We denote the vector of model parameters as:

$$\boldsymbol{\theta} = \left\{\phi_{\varepsilon 1}, \phi_{\varepsilon 2}, \omega_{\varepsilon}, \pi, \mu_1, \phi_1, \mu_{\beta}, \phi_{\beta}, \omega_{\beta 1}, \mu_{\gamma}, \omega_{\beta\gamma}, \phi_{\gamma}, \omega_{\gamma 1}, \mu_m, \phi_m\right\} \tag{8}$$

The simulation procedure takes $\boldsymbol{\theta}$ and returns consumption paths for each household for $t = 1, ..., T$. These parameters are the input for the optimization routine.

3.2 Simulating Expectations Errors

To simulate expectations errors, we draw four sets of mutually independent pseudo-random numbers: $(\nu_{1h,t}, \nu_{2h,t}, \nu_{3h,t}, \nu_{4h,t})$ for all h and $t = 0, 1...T$.[12] The variables $\nu_{1h,t}$, $\nu_{2h,t}$, and $\nu_{4h,t}$

[12] The start at $t = 0$ is to give a first draw $(\nu_{kh,0})$ which is used in generating the conditional heteroskedasticity for the $t = 1$ observation.

are standard normal variables and $\nu_{3h,t}$ is a uniform on $[0,1]$. The expectations errors, $\varepsilon_{h,t}$'s , are simulated recursively. We define two variances by:

$$\sigma_k^2 = \exp(\phi_{\varepsilon k}), \qquad k = 1,2 \tag{9}$$

where the exponential is taken to ensure that the variance is positive. Then we define two initial heterogeneous error terms by:

$$\varepsilon_{kh,0} = \exp\left(-\frac{\ln\left(1+\sigma_k^2\right)}{2} + \sqrt{\ln\left(1+\sigma_k^2\right)}\nu_{kh,0}\right), \qquad k = 1,2 \tag{10}$$

By construction each of these terms has a unit mean. We then mix these distributions with a mixing parameter given by:

$$d_{h,0} = \Phi\left(50 * (\nu_{3h,0} - \pi)\right)$$

where $\Phi\left(.\right)$ is the standard Normal cdf. This is a 'smoothed' indicator function which takes 0 or 1 for values of the uniformly distributed random draws $v_{3h,0}$ that are not very close to π. Such smoothed indicators are routinely used to facilitate derivative based optimization. These values control whether household h draws from the first simulated residual distribution or the second, so that:

$$\varepsilon_{h,0} = d_{h,0}\varepsilon_{1h,0} + (1 - d_{h,0})\varepsilon_{2h,0} \tag{11}$$

These simulated expectations errors for period 0 are used to set up the errors for $t = 1, 2..., T$.

Given expectations errors for period 0, we then continue for the next T periods. For $t > 1$, we define recursively heterogeneous time varying variances by:

$$\sigma_{kh,t}^2 = \exp(\phi_{\varepsilon k} + \omega_\varepsilon\left(\varepsilon_{h,t-1} - 1\right)), \qquad k = 1,2 \tag{12}$$

Thus each variance depends on the lagged realized error term with the same slope coefficient for each variance to capture the conditional heteroskedasticity in expectations errors. Then define two component error terms by:

$$\varepsilon_{kh,t} = \exp\left(-\frac{\ln\left(1+\sigma^2_{kh,t}\right)}{2} + \sqrt{\ln\left(1+\sigma^2_{kh,t}\right)}\nu_{kh,t}\right), \qquad k = 1,2$$

By construction these terms have a unit mean and dispersions governed by the conditionally heteroskedastic terms $\sigma^2_{1h,t}$ and $\sigma^2_{2h,t}$ respectively. We then define a mixing parameter by:

$$d_{h,t} = \Phi\left(50 * \left(\nu_{3h,t} - \pi\right)\right)$$

and define expectations errors by mixing according to:

$$\varepsilon_{h,t} = d_{h,t}\varepsilon_{1h,t} + (1-d_{h,t})\varepsilon_{2h,t} \tag{13}$$

The synthetic expectations errors $\varepsilon_{h,t}$ for $h = 1, ..., H$ and $t = 1, ..., T$ are explicitly designed to capture the features reported in the previous section in that they have a unit unconditional mean and are conditionally heteroskedastic. This step allows us to side-step full simulation of a model with stochastic income, real interest rates and other shocks. We refer to the $\varepsilon_{h,t}$'s as *synthetic residuals* and term our estimation procedure *Synthetic Residual Estimation* (SRE).

3.3 Simulating Preference Parameters

For the simulation of time invariant household specific parameters we draw three H-vectors of standard normal variables with elements a_h, b_h and g_h (for the distributions of initial consumption, β and γ, respectively). We assume that consumption in the first period is lognormally distributed and simulate it by:

$$C^*_{h,1} = \exp\left(\mu_1 + \exp\left(\phi_1\right) a_h\right) \tag{14}$$

When simulating the preference parameters we restrict discount factors and the coefficient of relative risk aversion to be in the intervals $[0.8, 1]$ and $[1, 15]$ respectively.[13] We allow the preference parameters to be correlated with the level of consumption by making them dependent on initial consumption. This method of allowing for correlated latent heterogeneity goes back to Chamberlain (1980), Anderson and Hsiao (1982) and Blundell and Smith (1991); Wooldridge (2005) gives a thorough analysis and eloquent justification for this methodology. We take the following translated logistic model for the discount factor:

$$\beta_h = 0.8 + 0.2 \left(\frac{\exp(\mu_\beta + \exp\left(\phi_\beta\right) b_h + \omega_{\beta 1} \ln\left(C^*_{h,1}\right))}{1 + \exp(\mu_\beta + \exp\left(\phi_\beta\right) b_h + \omega_{\beta 1} \ln\left(C^*_{h,1}\right))} \right) \tag{15}$$

where μ_β and ϕ_β capture the location and dispersion respectively. The parameter $\omega_{\beta 1}$ introduces dependence between the distribution of β and initial consumption. The coefficient of relative risk aversion have the following parameterization:

$$\gamma_h = 1 + 14 \left(\frac{\exp(\mu_\gamma + \omega_{\beta\gamma} b_h + \exp\left(\phi_\gamma\right) g_h + \omega_{\gamma 1} \ln\left(C^*_{h,1}\right))}{1 + \exp(\mu_\gamma + \omega_{\beta\gamma} b_h + \exp\left(\phi_\gamma\right) g_h + \omega_{\gamma 1} \ln\left(C^*_{h,1}\right))} \right) \tag{16}$$

where μ_γ and ϕ_γ capture the location and dispersion respectively. The parameter $\omega_{\beta\gamma}$ captures the dependence between the discount factor and the coefficient of relative risk aversion, and $\omega_{\gamma 1}$ introduces dependence between the coefficient of relative risk aversion and initial consumption.

3.4 Generating Consumption Paths

Given values for $\left(C^*_{h,1}, \beta_h, \gamma_h\right)$ for $h = 1, ..., H$ and synthetic expectations errors, $\varepsilon_{h,t}$, for each household and the real interest rate r_t between periods $t-1$ and t we can construct simulated consumption paths. For $t > 1$ we define consumption values recursively using the inverse of

[13] These intervals are the result of a preliminary search. Other intervals ($[0, 1]$ for the discount factor, for example) give similar results but often resulted in numerical instabilities.

equation 1:

$$C_{h,t}^{*} = C_{h,t-1}^{*} \left\{ \frac{\varepsilon_{h,t}}{\beta_h(1+r_t)} \right\}^{-\frac{1}{\gamma_h}} \tag{17}$$

After generating a consumption path for each household we introduce measurement error.[14] We do this by assuming that the measurement error enters as a multiplicative lognormal variable with idiosyncratic bias and variance. We draw an H-vector of standard normal variables m_h. Then we construct individual standard deviations by:

$$\xi_h = \exp\left(\mu_m + \exp\left(\phi_m\right) m_h\right) \tag{18}$$

We use these to simulate time varying measurement errors using the simulated values $\nu_{4h,t}$ discussed at the beginning of subsection 3.2:

$$\kappa_{h,t} = \exp\left(bias_h + \xi_h \nu_{4h,t}\right) \tag{19}$$

where $bias_h$ is an idiosyncratic bias term (which disappears when we difference consumption). We then define 'observed' simulated consumption as:

$$C_{h,t}^{S} = C_{h,t}^{*} \kappa_{h,t}$$

It is these simulated consumption paths that we use in the SRE optimization step.

[14] Although we call it 'measurement error' throughout the paper this can also be interpreted as an iid 'transitory' taste shocks. The identification of measurement error in the presence of taste shocks is not possible since they both appear in the same way in the auxiliary environment . We acknowledge the fact that, in the empirical work, we recover some sort of noise estimate (combined taste shocks and measurement error) rather than the size of the measurement error in consumption. The real issue is that we control for this noise in order to identify the parameters of interests (the preference parameters).

Taking logs, we have the following expression for simulated observed consumption growth:

$$\Delta \ln C^S_{h,t} = \frac{1}{\gamma_h} \ln(\beta_h) + \frac{1}{\gamma_h} \ln(1 + r_t) + \upsilon_{h,t} \quad (20)$$

$$\upsilon_{h,t} = \left(\xi_h \nu_{4h,t-1} - \frac{1}{\gamma_h} \ln \varepsilon_{h,t}\right) - \xi_h \nu_{4h,t} \quad (21)$$

From this equation we can see the broad outlines of our identification strategy. To identify the distribution of γ_h's we have the heterogeneous responses to real interest rates and the heterogeneous variances of consumption growth (through the factor multiplying $\ln \varepsilon_{h,t}$). The distribution of the discount factors β_h is identified from the distribution of trends (the 'intercept'). Under the model assumptions, measurement error is the only source of auto-correlation in the composite error term $u_{h,t}$. Thus the cross-section variances and auto-covariances of the composite error term determine the variances of the expectations errors $\varepsilon_{h,t}$ and the measurement error, $\kappa_{h,t}$. In the next section we use these considerations to structure our choice of auxiliary parameters.

3.5 Choosing Auxiliary Parameters

We now need to choose statistics of the data - so called, auxiliary parameters (ap's) - that are matched in the SMD step; we denote these $\{\lambda_1, ..\lambda_J\} = \boldsymbol{\lambda}$; we have $J = 24$. As always we have a trade-off between the closeness of the ap's to structural parameters (the 'diagonality' of the binding function, see Gourieroux *et al* (1993)) and the need to be able to calculate the ap's quickly. It should be noted that many of the ap's defined below are closely related; no attempt is made to construct an orthogonal set. None of the ap's are consistent estimators of any parameter of interest; rather, they are chosen to give a good, parsimonious description of the joint distribution of consumption growth and interest rates across the sample.

Our first two ap's relate to the parameters that govern the distribution of initial consumption

(μ_1, ϕ_1). We take the mean and standard deviation of log initial consumption:

$$\lambda_1 = mean\left(\ln\left(C_{h,1}\right)\right),\ \lambda_2 = std\left(\ln\left(C_{h,1}\right)\right) \tag{22}$$

In the empirical section below we discuss how to allow for an unbalanced panel and the fact that the age at which the household is first observed varies across households.

Our next set of ap's ($\lambda_3 - \lambda_{19}$) relate to the preference parameters β and γ. The first of these are the trend and the change in the cross-section dispersion of log consumption. For this, we calculate the cross-section median and interquartile range (iqr) of household log real expenditures in each year[15] and then regress the resulting T values on a constant and a trend. The ap's are the slope coefficients in these regressions (which we scale by 100 for the optimization routine):

$$\begin{aligned} \lambda_3 &= 100 * \text{trend in cross section median ln(consumption)} \\ \lambda_4 &= 100 * \text{trend in cross section iqr ln(consumption)} \end{aligned} \tag{23}$$

The next set of eight ap's are based on regressions of consumption growth on the real interest rate for each household individually[16]:

$$\Delta \log C_{h,t} = \zeta_{0h} + \zeta_{1h} r_t + u_{h,t} \tag{24}$$

Given parameter estimates for each household, define:

$$\hat{u}_{h,t} = \Delta \log C_{h,t} - \hat{\zeta}_{0h} - \hat{\zeta}_{1h} r_t \tag{25}$$

[15] Here and below, we use medians and inter-quartile ranges rather than means and standard deviations to minimise the impact of outliers.

[16] We could take the GMM estimates of these parameters (with the constant and lagged interest rates as instruments) as auxiliary parameters; we prefer the OLS since it is simpler and quicker. In addition, the OLS estimates have a mean which is not the case for just identified GMM estimates.

and the standard deviation of the household specific error and auto-correlation by:

$$\varphi_h = std\left(\hat{u}_{h,t}\right),\ \varsigma_h = corr\left(\hat{u}_{h,t}, \hat{u}_{h,t-1}\right) \tag{26}$$

We then record the following eight cross-section statistics that describe the distribution of the OLS coefficient estimates and the properties of the residuals:

$$\begin{aligned}
\lambda_5 &= median\left(\hat{\zeta}_{0h}\right),\ \lambda_6 = iqr\left(\hat{\zeta}_{0h}\right), \\
\lambda_7 &= median\left(\hat{\zeta}_{1h}\right),\ \lambda_8 = iqr\left(\hat{\zeta}_{1h}\right), \\
\lambda_9 &= mean\left(\varphi_h\right),\ \lambda_{10} = std\left(\varphi_h\right), \\
\lambda_{11} &= mean\left(\varsigma_h\right),\ \lambda_{12} = std\left(\varsigma_h\right)
\end{aligned} \tag{27}$$

The next four ap's are largely complementary to $\lambda_5 - \lambda_8$ but are based on the individual trends and standard deviations (of consumption growth) for individual households. That is, we first calculate the trend τ_h and standard deviation υ_h of $\Delta \log C_{h,t}$ for each household separately. We then record:

$$\begin{aligned}
\lambda_{13} &= mean\left(\tau_h\right),\ \lambda_{14} = std\left(\tau_h\right), \\
\lambda_{15} &= mean\left(\upsilon_h\right),\ \lambda_{16} = std\left(\upsilon_h\right), \\
\lambda_{17} &= corr\left(\tau_h, \upsilon_h\right)
\end{aligned} \tag{28}$$

We then have two ap's that capture the covariances between how wealthy the household is and the trend and standard deviation of log consumption. We denote mean log consumption for household h by ψ_h; this gives a measure of the long run average of the level of consumption and is used to identify the correlation between preference parameters (β, γ) and the initial value.

The ap's are:

$$\lambda_{18} = corr\left(\tau_h, \psi_h\right), \ \lambda_{19} = corr\left(\upsilon_h, \psi_h\right) \tag{29}$$

Between them, ap's $\lambda_5 - \lambda_{19}$ provide a very rich description of the joint distribution of the trends and variability in consumption, reactions to changes in the real interest rate and the persistent variances in observed consumption growth.

Finally we have a series of ap's ($\lambda_{20} - \lambda_{24}$) that are designed to capture the major features of the expectation error distribution which is assumed common to everyone. These are based on the residuals from the following pooled regression:

$$\Delta \log C_{h,t} = \alpha_0 + \alpha_1 r_t + e_{h,t}, \ h = 1, ..., H, \ \ t = 2, ..., T \tag{30}$$

The estimated residuals are:

$$\hat{e}_{h,t} = \Delta \log C_{h,t} - \widehat{\alpha}_0 - \widehat{\alpha}_1 r_t$$

The first three ap's are simply the second to fourth moments of these residuals:

$$\lambda_{20} = std\left(\hat{e}_{h,t}\right), \ \lambda_{21} = skew\left(\hat{e}_{h,t}\right), \ \lambda_{22} = kurt\left(\hat{e}_{h,t}\right) \tag{31}$$

To capture the conditional heteroskedasticity we run an analogue of (6):

$$\left(\hat{e}_{h,t}\right)^2 = \vartheta_0 + \vartheta_1 \hat{e}_{h,t-1} + \text{error}, \ h = 1, ..., H, \ \ t = 3, ..., T \tag{32}$$

This gives the following two ap's:

$$\lambda_{23} = \hat{\vartheta}_0, \ \lambda_{24} = \hat{\vartheta}_1 \tag{33}$$

Thus we have a 24-vector of ap's, $\boldsymbol{\lambda}$, to estimate the 15 model parameters given in (8).

4 Small Sample Properties

In this section we present the small sample performance of SRE in comparison to GMM estimation of exact and approximate Euler equations. We cannot use the same simulation environment as described in the previous section since current GMM Euler equation techniques do not allow for heterogeneity in the parameters. Consequently we take a model with homogeneous preference parameters. We take the same values as the stochastic interest rate model (model 9) in section 2 and add measurement error so that approximately half of the variance in consumption growth is noise.

In our Monte Carlo experiments, we investigate the small sample properties of GMM on the exact Euler equation (EGMM), GMM on the first order approximation (AGMM) and SRE. We perform four sets of experiments. The number of replications in all experiments is 1000. We assume that the econometrician has panel data on consumption and estimates the preference parameters by pooling all households together. The baseline experiment is for 20 ex-ante identical households followed for 40 periods and no measurement error. This is a very favorable environment for GMM. The second experiment takes the baseline case and reduces the number of time periods to 20. This gives some idea of how well the estimators perform in the (fairly realistic) situation in which we have a medium length panel. In the third experiment we add measurement error to the consumption paths in the baseline model so that half of the observed standard deviation of consumption growth becomes noise. In the fourth scenario we consider a case in which households may face binding liquidity constraints in some periods. In this experiment we solve the model with Deaton (1991) type explicit liquidity constraints where households are not allowed to borrow at all. With the parameters used for the baseline case this constraint never binds so we lower the discount factor to 0.87 ($\delta = 0.15$, as in model 8 in section 2). A low discount factor prevents excessive wealth accumulation so that we often observe zero asset levels carried forward from one period to the next. For estimation in this environment we remove periods that correspond to zero asset levels (that is, if the agent does not carry forward assets between t and $t+1$ then consumption growth data observed in periods t *and* $t+1$ are dropped); this selection is standard in the empirical literature.

Our first empirical model is the exact Euler equation (see equation 1). We follow Alan *et al* (2009) and assume a classical multiplicative lognormal measurement error with standard deviation σ_κ. The associated orthogonality conditions are:

$$E_t\left[\left(\frac{C_{h,t+1}}{C_{h,t}}\right)^{-\gamma}(1+r_{t+1})\beta-\exp\{\gamma^2\sigma_\kappa^2\}\right] = 0 \quad (34)$$

$$E_t\left[\left(\frac{C_{h,t+2}}{C_{h,t}}\right)^{-\gamma}(1+r_{t+1})(1+r_{t+2})\beta^2-\exp\{\gamma^2\sigma_\kappa^2\}\right] = 0 \quad (35)$$

The first equation only identifies γ and $\exp\{\sigma_\kappa^2\}/\beta$; the second equation identifies $\exp\{\sigma_\kappa^2\}/\beta^2$ which serves to identify σ_κ and β separately. The instruments taken are the constant and lagged real interest rate for the first equation and the constant for the second, so that we just identify the parameters.[17] Our second empirical model is the first order approximate Euler equation 30. Both exact and approximate equations are estimated as just identified systems using a vector of ones and lagged interest rates as instruments.

For the SRE, we simulate consumption paths using synthetic errors generated by a mixture of two lognormal distributions with the time-varying variance structure described in section 3. We set the mixing probability to 0.5. We thus have six model parameters to estimate: γ, β, σ_κ , $\phi_{\varepsilon 1}$, $\phi_{\varepsilon 2}$ and ω_ε (the last three are the parameters of the two time varying variances of expectations errors). The six auxiliary parameters (ap's) used for estimation are: the constant and slope in the OLS estimation of the approximate Euler equation (equation 30); the standard deviation and auto-regressive coefficient of the OLS residuals and the constant and slope coefficients of

[17] We experimented with different instruments such as lagged consumption growth and lagged income growth and having the same instrument set for both equations. Results with these instruments are worse than the results we present here.

the regression of squared OLS residuals on their lags. The system is just identified (we have six ap's and six parameters), just as in the case of exact GMM and approximate GMM.

Table 3 presents the sampling distributions of the three estimators for our four experiments. Values given are: means, medians (in square brackets) and standard deviations (in brackets). In the absence of measurement error and with a long panel (experiment 1), EGMM and AGMM perform very similarly with both capturing reasonably well the true value of the coefficient of relative risk aversion. EGMM yields a much lower standard deviation than AGMM. SRE performs almost as well in recovering the coefficient of relative risk aversion. The median of the sampling distribution is very close to the true value and the standard deviation is lower than both EGMM and AGMM. Both EGMM and SRE give good estimates for β. Given the SRE parameterization for σ_κ in equation 19 we shall always estimate a positive value but the estimated value is small.

For the second experiment, we see that decreasing the number of time periods from 40 to 20 leads to some substantial changes. First, the standard deviations of all estimators have gone up although not very much for the SRE (from 1.31 to 1.75 for the coefficient of relative risk aversion). Second, both GMM estimators exhibit downward bias in the mean and median estimates of the coefficient of relative risk aversion (the mean of the sampling distribution for the coefficient of relative risk aversion is in fact negative with a very large standard deviation), whereas the SRE yields upward bias. The bias for EGMM and SRE is relatively small. In terms of capturing the true discount factor, EGMM and SRE perform equally well.

In the third experiment, we allow for measurement error in the observation of consumption. The first feature of the estimates given in the Table 3 is that measurement error of this order leads to a downward bias in the AGMM estimator (mean 3.55 and median 3.04). Moreover, the sampling distribution of the estimator is highly dispersed (a standard deviation of 52.3). This result is particularly disappointing for the approximate model since the approximation is chosen to deal with multiplicative measurement error. The SRE is now clearly superior to both AGMM and EGMM; the mean coefficient of relative risk aversion is much closer to the true value (3.97) whilst EGMM exhibits serious downward bias. For the discount factor both EGMM and SRE exhibit downward bias but the EGMM estimates have lower bias. Note also that the SRE estimates the standard deviation of the measurement error more accurately than EGMM.

In the final experiment both SRE and EGMM perform similarly (SRE with some upward bias and EGMM with some downward bias for the coefficient of relative risk aversion). Although the median estimate is very close to the true value of γ, AGMM displays a considerably dispersed sampling distribution. The SRE performs very well in recovering the true discount factor, particularly at the median, but EGMM exhibits a serious upward bias.

The conclusion we draw from these Monte Carlo experiments is that in a very specific context SRE does at least as well as EGMM when there is no measurement error and when long panel data are available. It performs considerably better especially under measurement error. Additionally, SRE almost always dominates AGMM for the estimation of the coefficient of relative risk aversion. This is all in the context of a simple model with homogeneous preference parameters. In the following section, in which we estimate a full-scale structural model of intertemporal consumption choice, it will become obvious to the reader that in addition to doing as well as EGMM in a simple context, SRE has a substantial flexibility for incorporating a wide range of model complexities such as preference heterogeneity and heteroskedastic expectation error distributions.

5 Estimates from the PSID

5.1 Using Food Expenditures

In this section we apply SRE to the Panel Study of Income Dynamics (PSID) to estimate the joint distribution of the discount factor and the coefficient of relative risk aversion. The

PSID contains annual information on food at home and food at restaurants. Despite its shortcomings (no expenditure variable other than food, large measurement error and limited asset information) we chose to work with the PSID because it is the longest available panel survey on consumption and it has been used extensively for Euler equation estimation.

Although the use of food as a 'proxy' for total consumption is common in the empirical literature, it is worth providing a formal justification. This will allow us to relate our estimates based on food expenditures to preferences over all goods ('consumption'). Define two sets of goods: food and other goods. Let c_t^f be the consumption of food in period t and c_t^o be the consumption of other goods.[18] Assume that intertemporal preferences are additive within the period[19] with each sub-utility function taking an iso-elastic form:

$$U_t = E_t \left(\sum_{s=0}^{T-t} \beta^s \left\{ \left(\frac{c_{t+s}^f}{1-\gamma_f} \right)^{1-\gamma_f} + \left(\frac{c_{t+s}^o}{1-\gamma_o} \right)^{1-\gamma_o} \right\} \right) \tag{36}$$

In this formulation we have two coefficients of relative risk aversion parameters, γ_f and γ_o.[20] The within period additivity allows us to break the intertemporal allocation problem into two sub-problems, one for 'other goods' and the other for food:

$$U_t = E_t \left(\sum_{s=0}^{T-t} \beta^s \left(\frac{c_{t+s}^f}{1-\gamma_f} \right)^{1-\gamma_f} \right) + E_t \left(\sum_{s=0}^{T-t} \beta^s \left(\frac{c_{t+s}^o}{1-\gamma_o} \right)^{1-\gamma_o} \right) \tag{37}$$

The Euler equation for food is then given by:

$$\left(\frac{c_{t+1}^f}{c_t^f} \right)^{-\gamma_f} \left(1 + r_{t+1}^f \right) \beta = \varepsilon_{t+1} \text{ with } E_t \left(\varepsilon_{t+1} \right) = 1 \tag{38}$$

[18] To define these two 'consumptions' we require either that preferences within groups are homothetic or that within group relative prices are fixed. Note that this is weaker than the assumptions usually made to justify working with a single commodity, 'consumption', rather than many goods.

[19] Weaker conditions than additivity can be found if we are willing to also use the relative prices of food at home, food outside the home and other goods. This takes us too far from our current focus.

[20] Note we have the same discount factor for both sub-utility functions. If we allowed each good to have its own discount factor then we would no longer have exponential discounting. This is further than we wish to go in this paper.

where r^f_{t+1} is the nominal interest rate between periods t and $t+1$ minus food price inflation (the 'real interest rate for food'). Estimation using the variation in food consumption allows us to recover β and γ_f. The elasticity of intertemporal substitution for food is then given by[21]:

$$\epsilon_f = -\frac{1}{\gamma_f} \tag{39}$$

Although γ_f is of some interest, the primary interest is usually in household's attitudes to intertemporal substitution for *all* goods; Browning and Crossley (2000) term the latter the *total elasticity of intertemporal substitution*, ϵ. They show that if preferences are additive over a good f and other goods then we have:

$$\epsilon_f = \epsilon e_f \tag{40}$$

where e_f is the Marshallian income elasticity for good f. Our food commodity is a composite of food at home (a necessity, with an income elasticity below unity) and food outside the home (a luxury, with an elasticity above unity). We do not know of any estimates of the value of e_f but it is probably a little below unity. If this is the case then our estimates of ϵ_f underestimate the value of the total elasticity, ϵ.

5.2 Sample Selection

Our sample covers the period 1974 to 1987. Although the actual panel length is much longer, some of the food variables are hard to interpret prior to 1974 and food related questions were suspended for two years after 1987. We treat split-ups as separate household units and exclude singles. Our sampling scheme is designed to pick out consecutive periods of five years or more in which: marital status did not change; the age of the head was between 22 and 60 and food expenditures were reported.

We also drop an observation for any period in which the household carries forward assets

[21] Sometimes the negative sign is dropped. Since the elasticity is an own price response, we prefer to retain it.

of less than two months' income from one period to the next. Only household paths of at least five contiguous years of carrying forward assets are used. This restriction is to exclude households that are potentially liquidity constrained. This deselection is conservative in the sense that unconstrained households could carry forward less assets than this (or even, debts). Selecting out households who carry forward low assets will tend to take out those with a low discount factor or a low aversion to risk. All of our results below on the joint distribution of the preference parameters should be viewed in this light. It is an open question as to whether information from periods in which households are constrained can aid point identification of the distribution of preference parameters.

With this sampling scheme we have at least four consecutive years for each sampled household in which we can observe consumption growth and for which the Euler equation should hold. We stratify our sample into two categories: 'less educated' households in which the head has 12 years of education (or less) and 'more educated' for households in which the head has more than 12 years of education. Our final unbalanced panel has a total of 833 households (8116 observations) in the category of 'less educated' and 868 households (9065 observations) in the category of 'more educated'. Table 4 provides the breakdown of the number of consecutive run years for both education strata. We assume that all households face a common interest rate series calculated from the US three-month treasury bill rate and the consumer price index. This amounts to using only the time variation in the construction of intertemporal prices.

In the SRE step we take eight replications of the data, four pseudo-random replications and their antithetic mirror.[22] For example, consider the vector of b_h's in equation (15). With one replication this is an H-vector of pseudo-random draws. In estimation we take a $4H$-vector and then append the negative of these (the antithetic draws) to the vector to give an $8H$-vector. When simulating the consumption growth of households, exact replication of the structure of

[22] The trade-off for the number of replications is between speed and precision. If we have R replications then the covariance matrix for the SMD estimator is $(R+1)/R$ times the covariance for an estimator with analytical auxiliary parameters. A value of 8 gives a factor of 1.125, which is an acceptable loss of precision. The use of antithetic draws makes the factor even closer to unity.

the panel data to hand is crucial. Thus, the lengths of the observed paths are replicated exactly. For example, a household that is observed for 10 consecutive periods (say from 1977 to 1986) has a simulated consumption path for exactly 10 periods corresponding to the interest rates that prevailed between 1977 and 1986. Hence, the auxiliary parameters for the simulated sample are obtained from unbalanced simulated panels of $10 * 833$ 'less educated' and $10 * 868$ 'more educated' households, with $10 * 8116$ and $10 * 9065$ observations respectively.

5.3 The Distribution of Trends and Variances

We estimate the joint distribution of the discount factor and the coefficient of relative risk aversion for each education stratum. Our approach to identifying this distribution begins with the observation that there are marked differences among households in our sample in their consumption growth and their variance of consumption growth. To show this, we take means and standard deviations of consumption growth over time for each household.

In the left panel of Figure 1 we present the distributions of mean consumption growth for our two education strata. Two features of the distribution merit attention. First, the distribution of mean consumption growth for the more educated is to the right of that for the less educated. The mean trends for the less educated and the more educated are -0.8% and 1.3% per year, respectively. This is consistent with the widening gap between education strata in the US that has been observed for earnings and income (see Katz and Autor (1999)). Second, within each education group there is significant heterogeneity. For example, for both educational strata some households have an average consumption growth of more than 10% per year and others have less than -10%. One possible explanation for these within strata differences is that all households have the same preferences but different realizations of the expectations errors, with some having long runs of good or bad shocks. Since these shocks are serially uncorrelated, such runs are improbable and some of the variation can plausibly be attributed to differences in discount factors. In the right hand panel of Figure 1 we present the distribution of the standard deviation of consumption growth for both education groups. Here, we see smaller differences across education groups and substantial variation within each strata.

6 Results

6.1 Choosing a Preferred Model

We first run a pair of first round regressions to take out cohort, family composition and cyclical effects. Details are given in Appendix C. Given these transformations, our model relates to a two person household in which the head of the household is aged 25 in the first year we observe the household. The values of the auxiliary parameters for the two strata are presented in Table 5, in the columns headed 'Data value'. We shall discuss only a subset of the ap values.

1. The more educated have a higher trend than the less educated (ap's λ_3, λ_5 and λ_{13}). The respective values for the λ_{13} (the mean of the trends) are 2.37% and 0.86%. These values are somewhat higher than those displayed in Figure 1; the difference is due to the fact that we account for age, cohort and family size effects. The dispersion of the trends (λ_6 and λ_{14}) is very similar across the two strata.

2. The two education strata have similar distributions for the standard deviation of consumption growth (λ_{15} and λ_{16}).

3. There is an increasing cross-sectional variance (λ_4) for both strata with the more educated having a stronger trend.

4. The coefficients on the real interest rate in the simple regressions (λ_7 and λ_8 are the median and iqr, respectively) are very diverse, with a median close to zero. This is consistent with Euler equation studies on micro data; see Guvenen (2006) for references and discussion.

5. There is a strong negative auto-correlation in the regression residuals (λ_{11}). Although this does not have an immediate structural interpretation, it does lead us to expect to find a

good deal of measurement error. The auto-correlation is not very dispersed (λ_{12}).

6. The correlation between the standard deviation of consumption growth and the trend (λ_{17}) is not significantly different from zero for either strata but there is a positive correlation between the trend and mean log consumption (λ_{18}). There is no significant correlation between the standard deviation of consumption growth and mean log consumption (λ_{19}).

7. The ap λ_{24} indicates negative dependence of the variance of expectations errors on lagged errors for the less educated and positive (albeit, not significantly different from zero) dependence for the more educated.

The most general model we consider has 15 structural parameters; see (8). We also estimate a number of restricted variants of this model. The first row of Table 6 gives the fit for the unrestricted model. The next three rows give versions with the heterogeneity closed down (while still allowing for dependence on initial consumption) for β (row 2); for γ (row3) and for both β and γ (row 4). The fifth row shows the effect of closing down the dependence on initial consumption while still allowing for heterogeneity in β and γ. The final row shows the goodness of fit for the model with homogeneous preference parameters.

For the unrestricted model (row 1), neither strata fits as well as we could hope. An examination of the ap values in Table 5 shows that we fit well all ap's but some have very small standard errors. For example, for the more educated, the worst fit is for λ_{11} (the mean auto-regressive parameter for the residuals) with a t-value of 3.13. However, this high t-value is due more to the precision of the estimated ap rather than the difference; the values for the data and the model are -0.403 and -0.427 respectively, representing an error of about 5%. Given this, we shall take the unrestricted model as acceptable and go on to compare it to more restricted variants.

The most important conclusion that can be drawn from Table 6 is that homogeneity for the preference parameters (row 6 relative to row 1) is decisively rejected; we have $\chi^2(5)$ values of 40.6 and 35.4 for the less educated and more educated respectively. We also reject decisively the model without correlated heterogeneity ($\omega_{\beta 1} = \omega_{\gamma 1} = 0$) - see row 5. Thus it seems that we have to allow that rich and poor have different parameters. Once we allow for correlated heterogeneity, the various models (rows 1 to 4) all have much the same fit. For example, in the case of the less educated, we could close down the heterogeneity in β while still allowing for dependence on initial consumption (row 2). Similarly, for the more educated, both β and γ heterogeneity can be closed down (row 4). Given this mix of results we shall work with the unrestricted model for both strata in all that follows.

We present the parameter estimates for the unrestricted models in Table 7. These estimates are not directly interpretable. The point in presenting them is that interested readers can take these estimates and simulate consumption paths for the two strata. These simulations could be used to investigate issues such as the persistence of low consumption, the change in cross-section dispersion over time (net of the effects we have taken out in the first round regressions) and the evolution of consumption with age (once again, net of life stages effects that were also removed in the first round). We present some of the issues that we consider of interest in the next sub-section.

6.2 The Implications of the Estimates.

In this section we present some implications of our parameter estimates. We consider first the extent of measurement error that we estimate[23]. We compute the extent of the noise by considering the variance of the pooled simulated consumption growth values with and without measurement error, denoted σ^2_{obs} and σ^2_{true}, respectively. The proportion of the observed consumption growth that is due to noise is then given by $\left(\sigma^2_{obs} - \sigma^2_{true}\right)/\sigma^2_{obs}$. The values for the less educated and the more educated are both 0.86. Thus we estimate that 86% of the variance in observed consumption growth is due to measurement error. This is somewhat higher than previous researchers have estimated but it is consistent with the consensus that the PSID food

[23] As already discussed, what we recover here is not the actual measurement error variance but rather some sort of noise estimate (combined taste shocks and measurement error).

expenditure measure is very noisy. We also found considerable cross-section dispersion in the idiosyncratic measurement error variance, with values of 0.012 and 0.073 for the 5th and 95th percentiles for both education strata. There is no indication that the educated do a better or worse job of reporting food expenditures (see the values for μ_m and ϕ_m in Table 7).

As regards the initial distribution of consumption, the less educated have a higher initial value (7% higher) but a lower mean value by 3% per year over the 14 years from age 25 to 38. Thus the more educated overtake the less educated at about age 27.

Turning to the implications for the preference parameters, Figure 2 displays the marginal distributions of β and γ. As can be seen from the left hand panel, the more educated are more patient than the less educated. The median discount factor is 0.93 for the less educated and 0.96 for the more educated (corresponding to discount rates of 7.5% and 4.2% respectively). Some of both strata are very impatient, with first quartile values of 0.88 and 0.91 for the less and more educated respectively. All of this is consistent with the left hand panel in Figure 1. One notable feature of these distributions is the bunching at the top end, particularly for the more educated.

The only estimates of the *distribution* of discount factors in the empirical literature are due to Lawrance (1991), Samwick (1998), Cagetti (2003) and Andersen *et al* (2009). Lawrance (1991) does not allow for heterogeneity within groups. She finds, just as we do, that the discount rate is higher for the less educated. The difference in discount rates between the two education strata is two percentage points whereas we find a difference of 3.3 percentage points. Samwick (1998) backs out the discount factor from simulated wealth at retirement using a standard life-cycle model and the American Survey of Consumer Finances (SCF) 1992. The median discount factor he estimates is quite dependent on model assumptions concerning the elasticity of intertemporal substitution and initial asset holdings. More importantly, he finds wide dispersions. His general finding is that there seem to be three groups: the very impatient with values of the discount rate of more than 20% per year; the moderately impatient with values between zero and 10% per year; and a group (comprising about five to twenty percent of the sample) who have a rate of about −15% per year, so that they discount the present. Our parameterization (which restricts β to be between 0.8 and unity) does not allow for such distributions, but the bunching up of the discount factor at unity for the educated group is consistent with Samwick's findings. Cagetti (2003) estimates the discount factor (homogenous within education strata) to be around 0.98 for those with a college education and about 0.86 for the high school strata. Andersen *et al* (2009) conduct experiments to elicit risk and time discount parameters on a nationally representative sample of Danes. Their mean value for the discount rate, of about 10%, is considerably higher than those we find. They estimate a standard deviation of 2.4% for the heterogeneity distribution which is comparable in magnitude to our findings.

The right hand panel of Figure 2 shows the distributions for the coefficient of relative risk aversion (γ). We find that the less educated households are less risk averse than the more educated households (the median coefficient of relative risk aversion is 6.2 for the less educated and 8.4 for the more educated). One important point to note here is that the iso-elastic form forces a tight link between attitudes to risk and prudence (since there is only one parameter). Thus the finding that the more educated are more risk averse could equally well be interpreted as the more educated being more prudent.

There are several sources of information on risk attitudes in expected utility models[24]: experiments (see Holt and Laury (2005) for a literature review); survey questions (Barsky *et al* (1997), Dohmen *et al* (2005), Guiso and Piaella (2001), Eisenhauer and Ventura (2003)); consumption (expenditure) based empirical studies (Gourinchas and Parker (2002), French and Jones (2004), Blau and Gilleskie (2006) and French (2005)); wealth or portfolio-based empirical studies (Cagetti (2003), Alan (2006) and Kahvecioglu(2005); and empirical studies using other

[24]We do not attempt to relate our results to the vast literature on models that do not assume expected utility maximisation.

contexts such as game show behavior (Beetsma and Schotman (2001)) and auto insurance (Cohen and Einav (2007)). It is important to note that the studies that are directly related to our work (ones that use expenditure and, to some extent, wealth data) assume very little heterogeneity in risk aversion. Nevertheless we can still draw comparisons using our median estimates. In general, estimates based on consumption data generate lower mean coefficients of relative risk aversion as compared to the estimates based on wealth data. Overall, consumption based estimates of the (assumed homogenous) coefficient of relative risk aversion range between unity and 3. Estimates based on wealth and portfolio data are generally higher (between 4 and 18), with the exception of Alan (2006). Our median estimates are much closer to the estimates based on wealth and portfolio allocation data, even though they are consumption based.

As far as we are aware, all studies that allow for heterogeneity in risk tolerance find evidence of substantial differences across people. For example, the widely cited results in Barsky *et al* (1997) indicate considerable risk aversion (the modal group has a value between 4 and 16) but also considerable dispersion (23% have a coefficient of relative risk aversion of less than 2). Similarly, the experiment-based studies such as Andersen *et al* (2009) and Donkers *et al* (2001) all find considerable dispersion in risk tolerance parameters. Using a large representative sample who are asked directly about their attitudes to risk, Dohmen *et al* (2005) find considerable dispersion in responses. Similarly, Guiso and Piaella (2001) find a great deal of heterogeneity in an Italian survey that asks about the willingness to pay for a hypothetical lottery. They estimate a median coefficient of relative risk aversion of 4.8 with 90% of the sample being between 2.2 and 10. Eisenhauer and Ventura (2003) use the same data source and point out an ambiguity in the question, which has a large impact on the estimated values. They find mean values of around 8.6 for the coefficient of relative risk aversion with a strong positive monotonicity in education; the mean values for the two strata that correspond to the less educated and more educated in our study are both above the overall mean: 9.5 and 13.8 respectively.

Non-survey evidence on the extent of heterogeneity in risk preferences is sparse. Cohen and Einav (2007) present a structural model for the choice of the level of deductibility in car insurance. They conclude that 'heterogeneity in risk preferences is rather large'. They also find that more educated people have higher levels of *absolute* risk aversion; this is consistent with our finding if, as seems plausible, the more educated have higher levels of consumption.

As we have seen the more educated have a higher discount factor and a higher coefficient of relative risk aversion than the less educated; that is, there is a positive correlation across the two educational strata. Within strata, however, we find correlations between β and γ of -0.31 and -0.20 for the less and more educated respectively. As far as we are aware there is no non-experimental evidence regarding the joint distribution of the discount factor and risk attitudes. There are, however, a small number of experiments which address this issue (see Anderhub *et al* (2001), Eckel *et al* (2004) and Harrison *et al* (2005)). Anderhub *et al* (2001) find a negative correlation between risk attitudes and the discount factor; Harrison *et al* (2005) find no correlation; Eckel *et al* (2004) conduct experiments with poor people in Montreal and they find negative correlation. Whilst these experimental results are surprisingly consistent with our finding, it is important to emphasize the tentative nature of the conclusions - it is too early to authoritatively state that risk aversion and patience are negatively correlated.

What do our intertemporal allocation parameter estimates imply for household consumption, wealth accumulation and portfolio allocation decisions? More importantly, how can they be used to address the long standing challenges in the literature related to wealth distribution, observed consumption profiles and household portfolio allocations? High risk aversion coupled with high impatience imply a very strong precautionary motive and a weak consumption growth sensitivity to asset returns (given the model specification, in which the elasticity of intertemporal substitution is tied to the coefficient of relative risk aversion). A strong precautionary motive leads to significant buffer stock accumulation especially at young ages. For portfolio holding decisions, these parameters imply lower investment in the stock market. Modest one-time entry costs would lead to most of our sample postponing entry into the stock market and a large proportion never participating. While heterogeneity in risk attitude has potential for explaining the vast heterogeneity observed in household portfolio allocations (even after controlling for demographics, income and wealth), heterogeneity in patience generates wealth heterogeneity and can help us understand the concentration of wealth (see for example Krusell and Smith (1998) and Hendricks (2007)).

7 Conclusions

The current literature, based on empirical, experimental and survey evidence, suggests strongly the need to allow for substantial taste heterogeneity when modelling saving and consumption decisions. Current Euler equation GMM methods are not up to tackling this task. On the other hand, full structural modelling with different simulations for each household is not currently feasible. This paper is motivated by our observation that in simulations of structural consumption models with heterogeneous agents, the distribution of the pooled shocks is well described by a mixture of two lognormals. This is the case even if agents have different variances of income shocks; essentially the differences in variances do not transmit through to the shock distribution when agents are not liquidity constrained. This allows us to devise a novel estimation procedure based on the use of synthetic residuals; we term the method Synthetic Residual Estimation (SRE). SRE has advantages over full structural modelling in that it does not require estimation of income processes and it is does not rule out other sources of shocks. We have shown that in a specific Monte Carlo context SRE performs at least as well as exact GMM in all circumstances and much better when we have short noisy consumption panels.

We do, of course, have some reservations with the current analysis. First, we drop households who do not carry forward assets equal to two months income. This leaves us with a sample who are almost certainly not liquidity constrained in our data period but it excludes many households. Moreover this exclusion is correlated with the parameters of interest so the joint distributions of the preference parameters we derive are no longer representative. Second, we do not make use of observable shocks to income, interest rates, assets or demographics when constructing the synthetic residuals. Resolving shocks into an observable series and an unobservable series would make the analysis more robust (to the distributional assumption for the expectations errors) and give more precise parameter estimates. Finally, we have used a first round regression to remove cyclical, demographic and age effects from the growth of expenditure. Although this is conventional, it would be better to estimate the parameters for macro, cohort, age and family size effects simultaneously with the distributions of preference parameters.

With these caveats, this paper presents the first set of estimates for the joint distribution of heterogeneous discount factors and coefficients of relative risk aversion. We find that both parameters display significant variation across the two educational strata we consider here. More educated households are more patient and more risk averse than the less educated households, but there is considerable overlap between the distributions for the two strata. Within education strata, we find that patience and risk aversion are negatively correlated; that is, the more patient are less risk averse. These patterns are consistent with findings in the recent survey and experimental literature. They can also go some way towards explaining the diverse findings in the literature concerning idiosyncratic consumption growth, portfolio choice and holdings of precautionary savings.

References

[1] Alan, S., (2006), "Entry Costs and Stock Market Participation over the Life Cycle", *Review of Economic Dynamics*, 588-611.

[2] Alan, S., O. Attanasio and M. Browning, (2009), "Estimating Euler equations with Noisy Data: Two Exact GMM Estimators", *Journal of Applied Econometrics*, 24: 309-324.

[3] Altonji, J. G. and A. Siow, (1987), "Testing the Response of Consumption to Income Changes with Noisy Panel Data", *Quarterly Journal of Economics*, 102, 293-328.

[4] Amemiya, Y., (1985), "Instrumental Variable Estimator for the Nonlinear Errors-in-variables model", *Journal of Econometrics*, 28(3), 273-289

[5] Anderhub, V., W. Guth, U. Gneezy and D. Sonsino, (2001), "On the Interaction of Risk and Time Preferences: An Experimental Study" *German Economic Review*, 2(3).

[6] Andersen, S., W. Harrison, M.I. Lau, and E.E.Rutstrom (2009), "Preference Heterogeneity in Experiments: Comparing the Field and Lab," *Journal of Economic Behavior & Organization*, Forthcoming.

[7] Anderson, T. W. and Hsiao, C., (1982), "Formulation and Estimation of Dynamic-Models Using Panel Data." *Journal of Econometrics*, 18(1), pp. 47-82.

[8] Attanasio, O. P. and H. Low, (2004), "Estimating Euler equations", *Review of Economic Dynamics*, 406-435.

[9] Attanasio, O. P., J. Banks and S. Tanner, (2002) "Asset Holding and Consumption Volatility", *Journal of Political Economy*, 110(4), 771-792.

[10] Barsky, R. B., F. T. Juster, M. S. Kimball and M. D. Shapiro, (1997), "Preference parameters and behavioral heterogeneity: An experimental approach in the health and retirement study", *Quarterly Journal of Economics* 112(2): 537-579.

[11] Beetsma, R.M.W.J. and P.C. Schotman (2001), "Measuring risk attitudes in a natural experiment: Data from the television game show Lingo", *Economic Journal*, 111(474), 821-848.

[12] Blau, D., and D. Gilleskie, (2006), "Health Insurance and Retirement of Married Couples", *Journal of Applied Econometrics,* 21(7), 935-953.

[13] Blundell, R. and R. Smith (1991), "Initial conditions and efficient estimation in dynamic panel data models", *Annales d'Economie et de Statistique*, 20/21, 109-123.

[14] Browning, M. and T. Crossley, (2000), "Luxuries are Easier to Postpone: A Proof", *Journal of Political Economy*, 108(5):1064-1068.

[15] Browning, M., M. Ejrnæs and J. Alvarez, (2006), "Modelling income processes with lots of heterogeneity", University of Oxford Discussion Paper.

[16] Browning, M. and A. Lusardi, (1996), "Household Saving: Micro Theories and Micro Facts", *Journal of Economic Literature*, December, 1797-1855.

[17] Cagetti, M. (2003), "Wealth Accumulation Over the Life Cycle and Precautionary Savings", *Journal of Business and Economic Statistics*, 21(3), 339-353.

[18] Carroll, C., (2001), "Death to the Log-Linearized Consumption Euler equation!", *Advances in Macroeconomics,* 1(1), Article 6.

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

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

[21] Cohen, A. and L. Einav (2007), "Estimating risk preferences from deductible choice", *American Economic Review* 97(3): 745-788.

[22] Deaton, A., (1991), "Saving and Liquidity Constraints", *Econometrica*, 59(5), 1221-1248.

[23] Dohmen, T., Falk, A., Huffman, D. Sunde, U. Schupp, J. and G. G. Wagner, (2005), "Individual Risk Attitudes: New Evidence Form a Large, Representative, Experimentally-Validated Survey", IZA Discussion Paper 1730.

[24] Donkers, B., B. Melenberg, and A. van Soest (2001), "Estimating Risk Attitudes using Lotteries; a large Sample Approach", *Journal of Risk and Uncertainty*, 22(2), 165-195.

[25] Eckel, Catherine; Johnson, Cathleen and Montmarquette, Claude, (2004) "Saving Decisions of the Working Poor: Short-and Long-Term Horizons," CIRANO Working Papers 2004s-45. 2004.

[26] Eisenhauer, J. G. and L. Ventura, (2003), "Survey measures of risk aversion and prudence", *Applied Economics* 35(13): 1477-1484.

[27] French, E. (2005), "The Effects of Health, Wealth, and Wages on Labor Supply and Retirement Behavior", *Review of Economic Studies*, 72, 395-427.

[28] French, E. and J. B. Jones, (2004), "The Effects of Health Insurance and Self Insurance on Retirement Behavior", Federal Reserve Bank of Chicago Working Paper.

[29] Gallant, A. R. and G. Tauchen, (1996), "Which Moments to Match?", *Econometric Theory*, 12:657-681.

[30] Gourieroux, C., A. Monfort and E. Renault, (1993), "Indirect Inference", *Journal of Applied Econometrics*, 1993, 8:S85-S118.

[31] Gourinchas, P. O. and J. A. Parker, (2002), "Consumption over the life cycle", *Econometrica* 70(1): 47-89.

[32] Guiso, L. and M. Paiella, (2001), "Risk aversion, wealth and background risk", CEPR Working Paper 2728.

[33] Guvenen, F., (2006), "Reconciling conflicting evidence on the elasticity of intertemporal substitution: A macroeconomic perspective" *Journal of Monetary Economics* 53(7): 1451-1472.

[34] Hall, G. and J. Rust, (2002), "Econometric Methods for Endogenously Sampled Time Series: The Case of Commodity Price Speculation in the Steel Market", Cowles Discussion Paper 1376, University of Yale.

[35] Harrison, G. W., M. I. Lau, E. Rustrom and M. Sullivan, (2005), "Eliciting Risk and Time Preferences Using a Field Experiment: Some Methodological Issues" J. Carpenter, G. Harrison and J. List, *Field Experiments in Economics.* Greenwich, CT: JAI Press.

[36] Hausman, J., (2001), "Mismeasured Variables in Econometric Analysis: Problems from the Right and Problems from the left ", *Journal of Economic Perspectives,* 15(4), 57-67.

[37] Hendricks, L. (2007), "How Important is Discount Rate heterogeneity for Wealth Inequality", *Journal of Economic Dynamics and Control*, 31(9), 3042-3068.

[38] Holt, C. A. and S. K. Laury, (2005), "Risk aversion and incentive effects: New data without order effects", *American Economic Review* 95(3): 902-904.

[39] Hong, H. and E. Tamer, (2003), "A Simple Estimator for Nonlinear Error in Variable Models", *Journal of Econometrics*, 117, 1-19.

[40] Hubbard, R. G., J. Skinner and S. P. Zeldes, (1995), "Precautionary Saving and Social Insurance", *Journal of Political Economy,* vol. 103(2), 360-399.

[41] Kahvecioglu, D. C., (2005), "Asset Allocation, Bequests and Wealth Dynamics of the Elderly", unpublished manuscript.

[42] Katz, L. F. and Autor, D., (1999), "Changes in the Wage Structure and Earnings Inequality", O. Ashenfelter and D. Card, *Handbook of Labor Economics.* New York: North Holland.

[43] Krusell, P. and A. Smith, (1998), "Income and Wealth Heterogeneity in the Macroeconomy", *Journal of Political Economy,* 2:1, 245-272.

[44] Lawrence E. C., (1991), "Poverty and the Rate of Time Preference: Evidence from Panel Data", *Journal of Political Economy,* 99(1), 54-77.

[45] Low, H., C. Meghir, L. Pistaferri (2008), "Wage risk and employment risk over the life cycle", Institute for Fiscal Studies Working Papers, W06/27.

[46] Ludvigson, S. and C. Paxson, (2001), "Approximation Bias in Linearized Euler equations", *Journal of Economics and Statistics,* vol. 83(2), 242-256.

[47] Runkle, D. E., (1991), "Liquidity Constraints and The Permanent Income Hypothesis", *Journal of Monetary Economics*, 27, 73-98.

[48] Samwick, A. A., (1998), "Discount Rate heterogeneity and Social Security Reform", *Journal of Development Economics*, 57, 117-146.

[49] Schennach, S. M., (2004), "Estimation of Nonlinear Models with Measurement Error", *Econometrica,* 72, 33-75.

[50] Shapiro, M. D., (1984), "The Permanent Income Hypothesis and the Real Interest Rate", *Economics Letters,* 14, 93-100.

[51] Vissing-Jorgensen, A., (2002), "Limited Asset Market Participation and the Elasticity of Intertemporal Substitution", *Journal of Political Economy*, 110(4), 825-853.

[52] Wansbeek, T., (2001), "GMM Estimation in Panel Data Models with Measurement Error." *Journal of Econometrics*, 104, 259-268.

[53] Wooldridge, J. M., (2005), "Simple Solutions to the Initial Conditions Problem in Dynamic, Nonlinear Panel Data Models with Unobserved Heterogeneity", *Journal of Applied Econometrics*, 20(1), pp. 39-54.

Appendix

A The Consumption Function

We assume that the utility function is intertemporally additive and the sub-utilities are iso-elastic. The problem of the generic consumer h at time t is:

$$
\begin{aligned}
& \max E_t \left[\sum_{j=0}^{T-t} \frac{(C_{h,t+j})^{1-\gamma}}{1-\gamma} \frac{1}{(1+\delta)^j} \right] \\
s.t.\ A_{h,t+j+1} &= (1+r_{h,t+j})A_{h,t+j} + Y_{h,t+j} - C_{h,t+j}
\end{aligned}
$$

where C is non-durable consumption (separable from durables) , A is assets, Y is stochastic labor income and r is the stochastic real interest rate. We assume a finite life and that end of life, T, is certain. The discount rate δ and the coefficient of risk aversion γ are positive. Our generic consumer has no bequest motive so that $A_{T+1} = 0$. The stochastic process driving labor income is taken to be that described in section 3. We assume that the innovations to income are independent overtime and across individuals; that is, we assume away aggregate shocks to income. individuals use only one asset to smooth their consumption against these idiosyncratic income shocks. The return on this asset (interest rate) is generated by a stationary AR(1) process:

$$r_{h,t+1} = (1-\rho)\mu + \rho r_{h,t} + \epsilon_{h,t+1} \tag{41}$$

where μ is the unconditional mean, ρ is the AR(1) coefficient with $0 < \rho < 1$, and ϵ_{t+1} is assumed to be *iid* Normal with mean zero and standard deviation σ_ϵ.

Following Deaton(1991), the budget constraint is re-defined as

$$X_{h,t+j+1} = (1 + r_{h,t+j+1})(X_{h,t+j} - C_{h,t+j}) + Y_{h,t+j+1} \tag{42}$$

where $X_{h,t+j} = A_{h,t+j} + Y_{h,t+j}$ (cash on hand). The income process is nonstationary which makes the problem harder to solve since the range of possible income values is large. Instead, we redefine all the relevant variables in terms of their ratios to permanent income and solve for the consumption to income ratio. By doing this we reduce the number of state variables to two, namely the cash on hand to income ratio and the interest rate. Moreover, we obtain an iid income process which can be approximated by standard Quadrature methods. Given this redefinition of the relevant variables, the Euler equation can be written as

$$\theta_t(w_t, r_t)^{-\gamma} - \frac{1}{(1+\delta)} E_t \left[(1 + r_{t+1}) \theta_{t+1}(w_{t+1}, r_{t+1})^{-\gamma} z_{t+1}^{-\gamma} \right] = 0 \tag{43}$$

where $\theta_t = \frac{C_t}{P_t}$, $w_t = \frac{X_t}{P_t}$. The dynamic program is solved via policy function iteration using the terminal value condition. At the terminal date T, consumption is a function of only cash on hand and since the bequest motive is assumed away we have $\theta_T = w_T$.

For the income process, we use a 10 point Gaussian Quadrature and we approximate the interest rate process by forming a 10 point first order discrete Markov process. We use a cubic spline to approximate the consumption function at each iteration. Since we solve a finite life problem, we obtain T consumption-to-income ratio functions $\{\theta_1(w_1, r_1), ..., \theta_T(w_T)\}$.

We initialize the algorithm with the consumption rule at the end of life $c_T(x_T) = x_T$. The constraint on borrowing is that at the end of the life, the agent person has to pay pay back all their outstanding debt. In practice this constraint will never bind since the utility function satisfies the Inada conditions which implies that zero consumption is never chosen. Instead we will observe very impatient individuals getting very close to the borrowing limit, whereas it will be irrelevant for the patient ones. Since we do not assume an explicit borrowing limit as in Deaton (1991), the consumption functions are continuously differentiable. In fact, in our case where agents have iso-elastic preferences and income uncertainty, consumption functions are strictly concave. In order to solve the problem, we define an exogenous grid for the cash on hand to income ratio: $\{x_j\}_{j=1}^{J}$. It is important to adjust the grid as the solution goes back in time. The algorithm finds the consumption that makes the standard Euler equation hold for each value of x and r. In practice, we took 500 points for x and 10 points for r. After obtaining c_{T-1}, we use a cubic spline to approximate $c_{T-1}(x_{T-1})$ for each r. After obtaining the consumption functions for each age, we simulate life time consumption paths using the intertemporal budget constraint and generating random draws for income and interest rate. Generated paths differ due to different realizations of income and interest rates for each individual.

For our Monte Carlo experiments we generate 80 period consumption paths for *ex ante* identical consumers. Individuals are assumed to face the same interest rate series. Therefore individuals' consumption paths differ due only to different income realizations.

B Simulated Minimum Distance

Our estimation procedure is simulation based. Following Hall and Rust (2002) we refer to the general technique as Simulated Minimum Distance (SMD) since it is based on matching (minimizing the distance between) statistics from the data and from a simulated model. The class of SMD estimators includes the EMM procedure of Gallant and Tauchen (1996) and the Indirect Inference methods of Gourieroux, Monfort and Renault (1993). Here we present a short account of the method as applied generally to panel data; see Hall and Rust (2002) and Browning, Ejrnæs and Alvarez (2006) for details.

Suppose that we observe $h = 1, 2..H$ units over $t = 1, 2...T$ periods recording the values on a set of Y variables that we wish to model and a set of X variables that are to be taken as conditioning variables. Thus we record $\{(Y_1, X_1), ...(Y_H, X_H)\}$ where Y_h is a $T \times l$ matrix and X_h is a $T \times k$ matrix. For modelling we assume that Y given X is identically and independently distributed over units with the parametric conditional distribution $F(Y_h|X_h; \theta)$, where θ is an m-vector of parameters.[25] If this distribution is tractable enough we could derive a likelihood

[25] This could be generalised to allow for correlated heterogeneity by allowing for dependence on the initial

function and use either maximum likelihood estimation or simulated maximum likelihood estimation. Alternatively, we might derive some moment implications of this distribution for observables and use GMM to recover estimates of a subset of the parameter vector. Sometimes, however, deriving the likelihood function is extremely onerous; in that case, we can use SMD if we can simulate Y_h given the observed X_h and parameters for the model. To do this, we first choose an integer S for the number of replications and then generate $S * H$ simulated outcomes $\left\{(Y_1^1, X_1), ...(Y_H^1, X_H), (Y_1^2, X_1), ...(Y_H^S, X_H)\right\}$; these outcomes, of course, depend on the model chosen $(F(.))$ and the value of θ taken in the model.

Thus we have some data on H units and some simulated data on $S * H$ units that have the same form. The obvious procedure is to choose a value for the parameters which minimizes the distance between some features of the real data and the same features of the simulated data. To do this, define a set of *auxiliary parameters* that are used for matching. Gallant and Tauchen (1996) suggest first finding a 'score generator' (flexible quasi-likelihood function) which nests the true model, and then using the score vector from this as auxiliary parameters. In the Gourieroux *et al.* (1993) Indirect Inference procedure, the auxiliary parameters are maximizers of a given data dependent criterion which constitutes an approximation to the true DGP. Both of these approaches are motivated by attempts to derive estimators that have efficiency properties that are close to MLE. In Hall and Rust (2002), the auxiliary parameters are simply statistics that describe important aspects of the data; this is very close to calibration. We follow this approach. Thus we first define a set of J auxiliary parameters:

$$\gamma_j^D = \frac{1}{H} \sum_{h=1}^{H} g^j \left(Y_h, X_h\right), \; j = 1, 2...J \tag{44}$$

where $J \geq m$ so that we have at least as many auxiliary parameters as model parameters. Denote the J-vector of auxiliary parameters derived from the data by $\boldsymbol{\gamma}^D$. Using the same

values, as in Chamberlain (1980), Blundell and Smith (1991) and Wooldridge (2000).

functions $g^j(.)$ we can also calculate the corresponding values for the simulated data:

$$\gamma_j^S = \frac{1}{S*H}\sum_{s=1}^{S}\sum_{h=1}^{H} g^j\left(Y_h^s, X_h\right), \ j = 1, 2...J \tag{45}$$

and denote the corresponding vector by $\boldsymbol{\gamma}^S(\theta)$. Identification follows if the Jacobian of the mapping from model parameters to auxiliary parameters has full rank:

$$rank\left(\nabla_\theta \boldsymbol{\gamma}^S(\theta)\right) = m \text{ with probability } 1 \tag{46}$$

This effectively requires that the model parameters be 'relevant' for the auxiliary parameters.

Given sample and simulated auxiliary parameters we take a $J \times J$ positive definite matrix W and define the SMD estimator:

$$\hat{\theta}_{SMD} = \arg\min_\theta \left(\boldsymbol{\gamma}^S(\theta) - \boldsymbol{\gamma}^D\right)' W \left(\boldsymbol{\gamma}^S(\theta) - \boldsymbol{\gamma}^D\right) \tag{47}$$

The choice we adopt is the (bootstrapped) covariance matrix of $\boldsymbol{\gamma}^D$. Typically we have $J > m$; in this case the choice of weighting matrix gives a criterion value that is distributed as a $\chi^2(J-m)$ under the null that we have the correct model.

C First Round Regressions

The first round regression is designed to take out cyclical, family size and age effects from the *growth* of expenditure. We take as raw data the level of expenditure on food by household h at time t, X_{ht}, the household size, Z_{it} and the age of the head of household (minus 25) when first observed in the data, A_h ($= 0, ..32$ in our data). We construct the first differences of log expenditure and household size, Δx_{ht} and Δz_{ht} respectively. We also construct year dummies d_{st} for the years 1975 to 1986 (the first and last year dummies are not included) which equal unity if $s = t$ and zero otherwise. We then run the regression:

$$\Delta x_{ht} = \alpha + \beta \Delta z_{ht} + \gamma A_h + \sum_{s=1975}^{1986} \delta_s d_{ts} + u_{ht} \tag{48}$$

Note that the inclusion of the age level in a first differenced equation allows that the growth of consumption changes over the life-cycle (strictly that the levels age effect is quadratic). We denote the mean of the $T - s + 1$ estimated time dummy coefficients, $\hat{\delta}_s$, by $\bar{\delta}$. We then predict consumption growth as:

$$\hat{p}_{ht} = \Delta x_{ht} - \hat{\beta} \Delta z_{ht} - \hat{\gamma} A_h - \sum_{s=1975}^{1986} \hat{\delta}_s d_{ts} + \bar{\delta} \tag{49}$$

The inclusion of the mean time effect is to replace the common trend that the time dummies take out.

We now convert the differences to levels which requires an initial observation. Since we do not observe all households from the 'initial age', here taken as 25, we have to make an adjustment to the initial observed value in the data to give paths that notionally start at age 25. To do this we regress the log of the first observation (which may not be at $t = 1$) for each household on age (A_h), age squared, first year observed and the latter variable crossed with age. This allows for flexible cohort effects on the level of expenditure at age 25 ($A_h = 0$). To predict first period log consumption we take the residuals plus the estimated intercept, denoted $\hat{x}_{h1}$. Finally we construct paths of adjusted log expenditures recursively by:

$$\hat{x}_{ht} = \hat{x}_{h,t-1} + \hat{p}_{ht} \tag{50}$$

This is done for t running from the second year the household is observed to the final year it is observed. These adjusted paths are the paths used in the analysis.

	Coeff. RRA	Discount rate	real interest rate	Income	Liquidity
Model	γ	δ	**stochastic**	**process,** σ_z	**constraint**
1	4	0.05	No	0.1	No
2	2	0.05	No	0.1	No
3	4	0.15	No	0.1	No
4	4	0.05	No	0.15	No
5	4	0.05	No	Carroll	Implicit
6	4	0.15	No	Carroll	Implicit
7	4	0.05	No	0.1	Yes
8	4	0.15	No	0.1	Yes
9	4	0.05	Yes	0.1	No
10	4	0.15	Yes	0.1	No
11	4	0.15	Yes	0.1	Yes
12 (1&2)	4/2	0.05	No	0.1	No
13 (1&3)	4	0.05/0.15	No	0.1	No
14 (1&4)	4	0.05	No	0.1/0.15	No
15 (1&2&3&4)	4/2	0.05/0.15	No	0.1/0.15	No
16	Model 1 with moderate measurement error (30% noise)				
17	Model 1 with high measurement error (60% noise)				

Table 1: Simulated models

The interest rate is 0.03 for constant interest rate models. In models with stochastic interest rates (models $9-11$) the interest rate is assumed to have a mean of 0.03, a standard deviation of 0.025 and an AR(1) coefficient of 0.6. The standard deviation of (the logarithm of) shocks to transitory income, σ_u, is set to 0.1 for all models and the standard deviation of (the logarithm of) shocks to permanent income is denoted as σ_z. 'Carroll' income process refers to the assumption that transitory shocks take the value zero with a 1% probability. All models are solved for $T = 80$ periods and simulated for $N = 10,000$ agents.

	Test for Equality-of-Distributions		Heteroskedasticity	
Model	**L-test**	**M-test**	**Coefficient (ω_ε)**	**t-ratio**
1	17.4	53.8	−.010	−2.7
2	45.5	20.6	−.011	−0.3
3	0.01	78.6	.079	5.4
4	0.00	13.1	−.195	−13
5	15.1	60.6	−.011	−3.0
6	0.00	28.8	3.05	39.7
7	18.9	52.2	−.010	−2.6
8	41.6	42.0	.003	0.7
9	14.1	54.7	−.013	−3.1
10	0.01	48.9	.053	2.5
11	68.2	48.9	−.004	−.74
12	0.00	65.8	−.005	−1.5
13	0.00	39.1	.049	4.4
14	0.00	15.8	−.086	−6.7
15	0.00	79.4	−.015	−2.2
16	51.9	50.5	−.536	−122
17	14.8	45.0	−1.11	−95.1

Table 2: Tests for expectations errors

L-test refers to p-values obtained from Kolmogorov-Smirnov equality-of-distributions test under the lognormality assumption. M-test refers to p-values obtained from the same test under mixture of two lognormals. The last two columns give the slope parameter (ω_ε) and associated t-value from the following regression: $(\varepsilon_{h,t} - 1)^2 = \phi_\varepsilon + \omega_\varepsilon(\varepsilon_{h,t-1} - 1) + \epsilon_{h,t}$ to assess the conditional heteroskedasticity in expectations errors.

	True values				EGMM			AGMM	SRE		
Experiment	T	γ	β	σ_κ	$\hat{\gamma}$	$\hat{\beta}$	$\hat{\sigma}_\kappa$	$\hat{\gamma}$	$\hat{\gamma}$	$\hat{\beta}$	$\hat{\sigma}_\kappa$
1	40	4	0.952	0	4.17	.952	.011	4.08	4.21	.945	.017
					[3.51]	[.960]	[.011]	[3.75]	[4.09]	[.948]	[.017]
					(2.04)	(.031)	(.009)	(19.3)	(1.31)	(.021)	(.009)
2	20	4	0.952	0	3.76	.958	.020	−1.27	4.48	.951	.015
					[3.13]	[.963]	[.021]	[2.75]	[4.59]	[.957]	[.015]
					(3.30)	(.029)	(.011)	(102.7)	(1.75)	(.025)	(.009)
3	40	4	0.952	0.15	2.97	.936	.210	3.55	3.97	.927	.145
					[2.99]	[.961]	[.230]	[3.04]	[3.45]	[.946]	[.145]
					(2.67)	(.024)	(.012)	(52.3)	(2.12)	(.056)	(.011)
4	40	4	0.87	0	3.70	.920	.055	27.2	4.43	.857	.016
					[3.68]	[.931]	[.026]	[3.92]	[4.37]	[.869]	[.015]
					(3.71)	(.055)	(.027)	(610)	(1.62)	(.067)	(.010)

Table 3: Small sample distributions of EGMM, AGMM and SRE

Values are means, [medians] and (standard deviations) of sampling distributions. The number of Monte Carlo replications is 1000. For experiment 3, the standard deviation of the measurement error ($\hat{\sigma}_\kappa$) is 0.15, which amounts to 50 percent noise in the consumption growth variance. For all experiments, the true value of the coefficient of relative risk aversion is 4. For experiments 1, 2 and 3, the true value of the discount factor is 0.952 and for experiment 4 it is 0.87.

Number of consecutive run years	Number of Households	
	Less educated	More educated
5	124	93
6	81	84
7	7	57
8	72	58
9	63	62
10	46	52
11	61	50
12	43	57
13	29	49
14	235	306
Total Households	833	868

Table 4: Distribution of unbalanced PSID samples

	Less Educated					**More Educated**				
Ap	Data Value	Confidence interval 2.5%	Confidence interval 97.5%	Model Value	$\|t\|$	Data Value	Confidence interval 2.5%	Confidence interval 97.5%	Model Value	$\|t\|$
λ_1	1.335	1.312	1.358	1.335	0.44	1.273	1.249	1.298	1.260	0.81
λ_2	0.364	0.347	0.382	0.367	0.24	0.368	0.349	0.387	0.373	0.48
λ_3	0.783	0.526	1.042	0.892	0.67	1.981	1.663	2.166	2.252	1.71
λ_4	0.212	−0.115	0.525	0.152	0.06	0.393	0.107	0.630	0.402	0.04
λ_5	0.700	0.313	1.267	0.369	1.17	2.117	1.741	2.489	1.819	1.45
λ_6	0.101	0.095	0.108	0.104	0.74	0.099	0.093	0.107	0.103	0.71
λ_7	0.020	−0.127	0.124	0.098	0.87	−0.023	0.167	0.133	0.125	1.72
λ_8	3.931	3.688	4.363	4.165	1.11	3.562	3.351	3.775	3.593	0.23
λ_9	0.258	0.252	0.264	0.255	0.64	0.255	0.250	0.260	0.254	0.36
λ_{10}	0.118	0.113	0.122	0.113	1.56	0.104	0.100	0.108	0.105	0.71
λ_{11}	−0.419	−0.431	−0.407	−0.436	2.32	−0.403	−0.415	−0.391	−0.427	3.13
λ_{12}	0.254	0.243	0.266	0.252	0.34	0.257	0.247	0.269	0.247	1.48
λ_{13}	0.863	0.608	1.115	0.828	0.21	2.367	2.120	2.610	2.235	0.84
λ_{14}	0.055	0.052	0.058	0.054	0.30	0.053	0.051	0.056	0.052	0.77
λ_{15}	0.278	0.272	0.284	0.274	0.83	0.272	0.267	0.278	0.270	0.69
λ_{16}	0.123	0.118	0.127	0.116	2.39	0.109	0.104	0.113	0.107	0.55
λ_{17}	−0.033	−0.086	0.018	0.029	1.94	0.018	−0.051	0.082	0.011	0.20
λ_{18}	0.099	0.053	0.151	0.129	0.99	0.073	0.028	0.117	0.117	1.63
λ_{19}	−0.285	−0.083	0.019	−0.025	0.10	−0.032	−0.078	0.016	−0.439	0.41
λ_{20}	0.282	0.276	0.288	0.283	0.48	0.275	0.270	0.280	0.276	0.26
λ_{21}	0.027	−0.012	0.067	0.031	0.19	0.028	−0.071	0.015	0.014	1.58
λ_{22}	3.873	3.757	3.990	3.862	0.17	3.798	3.685	3.902	3.747	0.76
λ_{23}	0.079	0.076	0.083	0.081	0.65	0.076	0.073	0.079	0.076	0.01
λ_{24}	−1.238	−2.443	−0.102	−1.353	0.16	0.138	−0.861	1.142	−0.297	0.70

Table 5: Auxiliary parameters for the two samples.
See section 3.5 for the definitions of ap's

Parameter restrictions	Degrees of freedom	Less educated χ^2	More educated χ^2
$-$	9	33.76	36.84
$\phi_\beta = \omega_{\beta\gamma} = 0$	11	36.33	37.69
$\phi_\gamma = \omega_{\beta\gamma} = 0$	11	45.50	37.93
$\phi_\beta = \phi_\gamma = \omega_{\beta\gamma} = 0$	12	44.67	38.83
$\omega_{\beta 1} = \omega_{\gamma 1} = 0$	11	60.37	68.14
$\phi_\beta = \phi_\gamma = \omega_{\beta\gamma} = \omega_{\beta 1} = \omega_{\gamma 1} = 0$	14	74.40	72.26

Table 6: Goodness of fit

The number of ap's equals 24. The first row gives the fit for the unrestricted model. The next three rows give versions with the heterogeneity closed down for β (row 2); for γ (row3); and for both β and γ (row 4) while still allowing for dependence on initial consumption. The fifth row shows the effect of closing down the dependence on initial consumption while still allowing for heterogeneity in β and γ. The final row shows the goodness of fit for the model with homogeneous preference parameters.

Parameter group	Parameter name	Education Low	High
Distribution of expectations errors	$\phi_{\varepsilon 1}$	-1.465	-0.621
	$\phi_{\varepsilon 2}$	-1.593	-0.818
	ω_{ε}	1.688	-0.160
	π	0.827	0.186
Distribution of initial consumption	μ_1	1.335	1.259
	ϕ_1	-1.172	-1.095
Distribution of β	μ_β	7.039	7.389
	ϕ_β	-2.340	-0.232
	$\omega_{\beta 1}$	-4.824	-4.725
Distribution of γ	μ_γ	-2.100	-1.803
	$\omega_{\beta\gamma}$	0.139	0.556
	ϕ_γ	0.037	-0.479
	ω_{γ_1}	1.183	-1.528
Distribution of measurement error	μ_m	-1.764	-1.768
	ϕ_m	-1.228	-1.366

Table 7: Estimates of model parameters

The parameters for the expectations errors distribution are denoted as $(\phi_{\varepsilon 1}, \phi_{\varepsilon 2}, \omega_\varepsilon, \pi)$. The parameters $\phi_{\varepsilon 1}$ and $\phi_{\varepsilon 2}$ are for the dispersions of the two components of the mixture of log-normals (the means are fixed at unity). The parameter ω_ε controls the extent of conditional heteroskedasticity and π controls the mixing probabilities. For the distribution of the initial level of consumption, the location and dispersion parameters are (μ_1, ϕ_1). The discount factor parameters are denoted as $(\mu_\beta, \phi_\beta, \omega_{\beta 1})$. These are, respectively, related to the discount factor location, dispersion and the dependence between the discount factor and initial consumption. The parameters for the coefficient of relative risk aversion are denoted as $(\mu_\gamma, \phi_\gamma, \omega_{\beta\gamma}, \omega_{\gamma 1})$. These are, respectively, related to the location, dispersion, the dependence between the discount factor and coefficient of relative risk aversion and the dependence between the coefficient of relative risk aversion and initial consumption. The final pair of parameters (μ_m, ϕ_m) are the location and dispersion parameters for the measurement error.

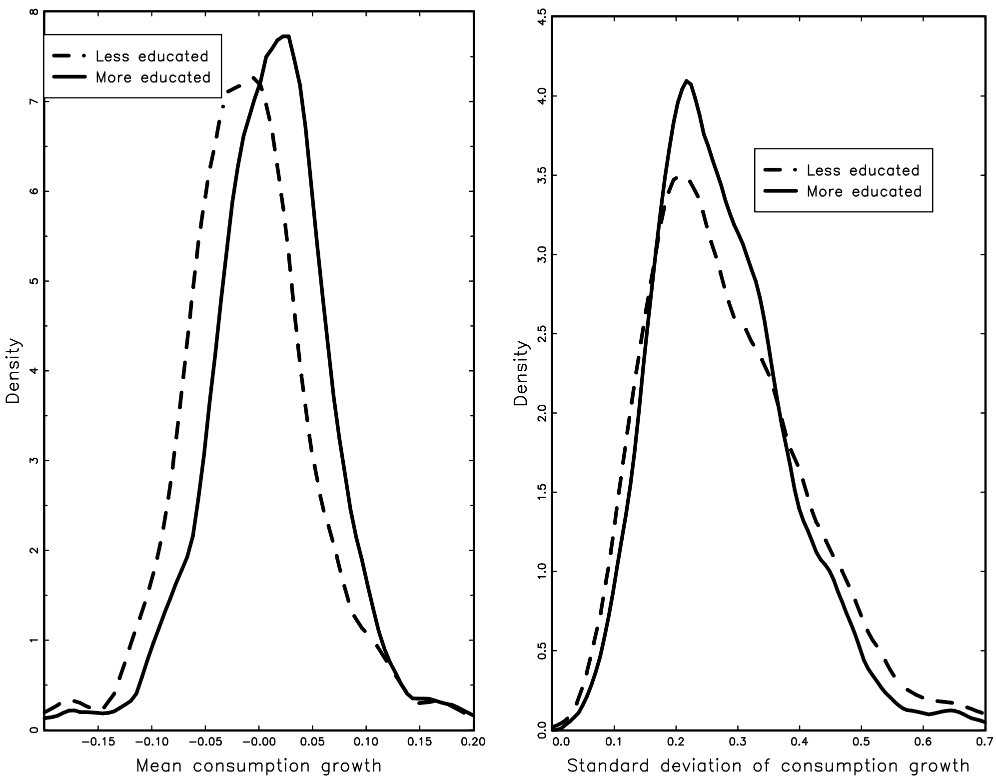

Figure 1: Distribution of means and standard deviations of consumption growth

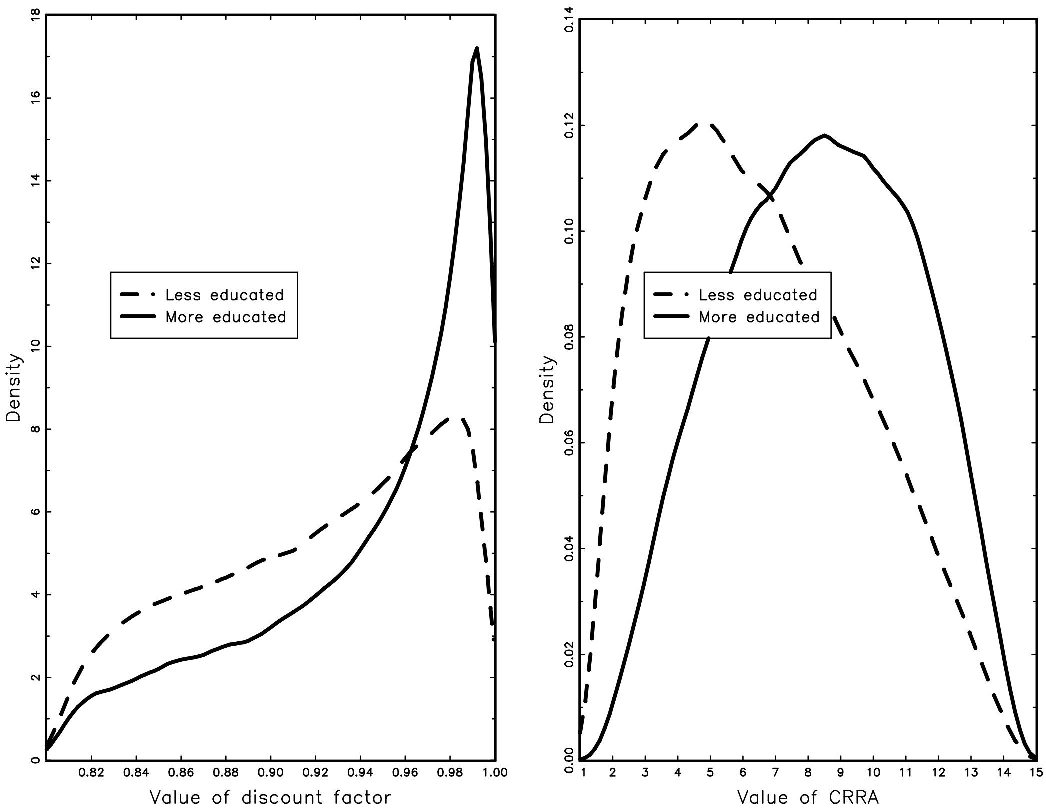

Figure 2: Marginal distributions of β and γ