



Centre for Financial Analysis & Policy

Working Paper no. 32:

Estimating Intertemporal Allocation Parameters using Synthetic Residual Estimation

Sule ALAN & Martin BROWNING

January 2008

Centre for
**Financial
Analysis &
Policy**



CAMBRIDGE

The Working Paper is intended as a mean whereby researchers' thoughts and findings may be communicated to interested readers for their comments. The paper should be considered preliminary in nature and may require substantial revision. Accordingly, a Working Paper should not be quoted nor the data referred to without the written consent of the author(s). All rights reserved.

© 2008 Sule Alan & Martin Browning

Comments and suggestions would be welcomed by the authors.

Estimating Intertemporal Allocation Parameters using Synthetic Residual Estimation¹

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

January 2008

¹We 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 expectation 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 discount factors and coefficients of relative risk aversion. 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 with 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 factors 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

Over the past quarter century many attempts have been made to estimate the parameters governing intertemporal allocation using Euler equation techniques applied to micro data; Browning and Lusardi (1996) discuss the results of 25 studies using 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 important seem to be the paucity of appropriate data (long panels on consumption) and the substantial measurement error in consumption (see Shapiro (1984), Altonji and Siow (1987) and Runkle (1991)). The latter means that we cannot use the exact Euler equation for estimation if the equation is non-linear in parameters (a point first made in the general context of nonlinear GMM by Amemiya (1985)). The use of ‘approximate’ Euler equations (whether first order or second order) ‘solve’ the measurement error problem but bring with them new problems in that they introduce latent variables that lead to violations of the orthogonality conditions exploited by GMM methods. Thus 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 the Carroll 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 return to estimating consumption functions (‘structural models’) based on specifying the environment agents face (see Carroll and Samwick (1997), Gourinchas and Parker (2002) and Attanasio, Banks, Meghir and Weber (1999)). In practice these methods are very similar to calibration (as used in, for example, Hubbard, Skinner and Zeldes (1995)). The problems with this approach are that it is very cumbersome and can only accommodate very limited sources of uncertainty and heterogeneity. Moreover, results may not be robust to small changes in the specification of the structural model. For example, Browning and Ejrnaes (2006) show that the Gourinchas and Parker (2002) and Attanasio *et al.* (1999) results are sensitive to how we account for family composition.

In this paper we focus on estimating the parameters of intertemporal allocation using an alternative approach to GMM estimation of Euler equations. This estimation method is based on modelling the distribution of expectations errors. We term our new procedure ‘synthetic residual estimation’ (SRE). The key to our approach is that associated with every structural model there is a expectations error distribution. We show that if we know the form of this distribution, up to a finite set of parameters, and observe consumption paths and interest rates, then we can identify utility parameters (the discount factor and the elasticity of intertemporal elasticity (EIS)) without having to specify the underlying stochastic environment. Without extra information the underlying models of income, health or other driving variables is not identified, but this is a strength rather than a weakness if we are only interested in preference parameters, since it gives the method robustness as compared with full-fledged structural estimation. Moreover, SRE is extremely flexible so that preference heterogeneity can easily be incorporated.

After summarizing the problems associated with the conventional Euler equation methods in the next section, we present an analysis of the distributions of expectations errors associated with models that are widely used in the literature in section 3. For example, we examine nearly patient agents with unit root income processes, Deaton’s (1991) buffer stock model with explicit liquidity constraints and models with impatient agents with self-imposed liquidity constraints. This serves to develop intuition and to illustrate many of the points we wish to make. The main conclusion from our investigations is that all models that have been suggested in the literature give an expectations error distribution that can be adequately modelled as a mixture of two lognormal distributions. This finding provides the basis for our estimation technique.

In section 4 we present our estimator. It is a simulation based method that is in the class of Simulated Minimum Distance (SMD) estimators.¹ The estimation involves the specification of

¹The class of SMD estimators (see (Hall and Rust (2002)) includes the EMM procedure of Gallant and Tauchen (1996) and the Indirect Inference method of Gouriéroux, Monfort and Renault (1993).

‘auxiliary parameters’ (ap’s) which are then matched to their theoretical predictions to estimate the parameters of interest. The conventional linearized Euler equation provides a very simple and convenient vehicle to do this. The method suggested is many orders of magnitude faster than full structural estimation. Above we stated that we can recover the utility parameters if we know the expectations error distribution. Since we never do know the distribution, we address the problem of testing whether the distribution chosen for the estimation procedure is a good approximation using goodness-of-fit tests applied to the predicted distribution.

Since we develop a fully parametric model, the obvious alternative to SMD is to use maximum likelihood estimation (MLE). For simpler variants of our model (for example, without allowing for measurement error, conditional heteroskedasticity in expectation errors or heterogeneity in preference parameters) this would be feasible but for the more complex model we adopt here this is very difficult. There are also substantive reasons for using SMD rather than MLE. Good MLE practice consists in specifying a fully parametric model, estimating by MLE and then applying of goodness of fit tests to a wide range of statistics to ensure that the model really does capture the data generating process. If the goodness of fit (GF) tests reject, then the researcher goes back to the parametric model and relaxes it in the direction suggested by the failures. This continues until the model fits well all of the statistics used in the GF tests.² In practice, this is close to estimation by SMD which effectively requires maximisation of the GF statistics. In SMD the GF statistics are known as auxiliary parameters.

In section 5 we present Monte Carlo evidence on our estimator and Euler equation based GMM estimators. We take as designs for these simulations the designs used in the recent papers alluded to above. We find that if consumption is measured with even moderate error, exact Euler equation estimation performs poorly. We also replicate the previous finding that approximate methods do poorly if we have short panels. By contrast, our SRE estimator works well when other estimators do not. In particular, when there is considerable measurement error (for example, half the observed consumption growth variance is due to noise) our estimator works well even for moderate sample sizes.

In section 6, we present an empirical application of SRE to data drawn from the PSID. This requires the modelling of ‘food’ rather than total consumption. We provide an explicit justification for using food alone and provide conditions under which using food gives estimates of the objects of interest. In particular, we show that we can identify the joint distribution of discount factors and the coefficients of relative risk aversion for food. The latter is not a primary object of interest but we can use the results of Browning and Crossley (2001) to show the coefficient of relative risk for food gives a bound for the elasticity of intertemporal substitution for all goods.

A number of regularities observed in consumption and wealth data can be rationalized by allowing for heterogeneity in the discount factor and/or the elasticity of intertemporal substitution (EIS) (the inverse of the coefficient risk aversion in the case of iso-elastic utility). The most important of these are the heterogeneity in lifetime wealth accumulation by households with similar earnings profiles. This seems to require 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 life cycle models are due to Lawrence (1991), Samwick (1998) and Cagetti (2003). Even though heterogeneity in risk aversion also has a great potential explaining such regularities (in particular, household portfolio allocations), to our knowledge, estimates of the distribution of the coefficient of relative risk aversion do not exist in the literature. The literature, however, produced several important papers that show that this parameter is very likely to be heterogenous. 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. Furthermore, to our knowledge, estimates of joint distributions of the discount factor and the coefficient of relative risk aversion do not exist in the literature within the context of a life cycle model. Due to the flexibility of the SRE we are able to estimate such distributions by also allowing heterogeneity to depend on initial consumption levels without solving

²Of course, this procedure raises issues of pre-test bias. This does not vitiate the need for goodness of fit tests, it simply makes their application more difficult.

the underlying structural model.

In the empirical application we consider two samples of households in the PSID based on their broadly defined education group membership. Using the SRE we estimate the joint distribution of the coefficient of relative risk aversion and the discount factor for each group separately. In line with studies based on consumption and wealth data, we find that the more educated are more patient than the less educated. The median discount rates are 7.5% and 4.2% for the less educated and the more educated respectively. We also find that especially for the more educated group, the estimated marginal discount factor distribution bunches up at the top end, which suggests that some households have discount rates less than zero (or discount factors above unity). For the coefficient of relative risk aversion, we find that the lower educated households are less risk averse than more educated households. The medians of the two distributions are 6.2 and 8.4 respectively. These values are higher than those estimated by consumption based studies but perfectly 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 educational strata. Within strata, however, we find the opposite result of a negative correlation between patience and risk averse; this is consistent with experimental evidence, which uses subjects who have the same education level.

2 Euler Equation Empirical Approaches

The basic model we use is the familiar intertemporal allocation model with iso-elastic preferences for each household represented by:

$$u_t = \frac{c_t^{1-\gamma}}{1-\gamma} \quad (1)$$

where γ is the (positive) coefficient of relative risk aversion. This particular utility function has the property that the single parameter γ governs both the attitude toward risk and intertemporal allocation; the elasticity of intertemporal substitution (EIS) is the reciprocal of the coefficient of relative risk aversion ($\frac{1}{\gamma}$). Due to this property, we will use the terms coefficient of relative risk aversion and EIS interchangeably throughout the text. If there are no liquidity constraints, this function yields the *exact Euler equation* for consumption growth given by:

$$\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma} (1 + r_{t+1})\beta = \varepsilon_{t+1} \text{ with } E_t(\varepsilon_{t+1}) = 1 \quad (2)$$

where β is the discount factor and r_{t+1} is the real rate between periods t and $t + 1$ and ε_{t+1} is a ‘surprise’ term. This relationship has been the basis of very many estimates of the preference parameters (β, γ) and tests 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 assumed orthogonality of the error term ε_{t+1} to all variables dated t or before, such as lagged consumption, interest rate and income variables. The attraction of estimation based on 2 is that one can estimate the preference parameters without explicitly parameterizing the stochastic environment that agents face.

Problems for GMM estimation on micro data arise if the consumption data are measured with error. It is now widely accepted that household level consumption data information is likely to be very noisy. For example, Runkle (1991) estimates that 76% of the variation in the growth rate of food consumption in the PSID is noise. Dynan (1993) reports that the standard deviation of changes in log consumption in the CEX (American Consumer Expenditure Survey) is 0.2, which seems too large for ‘true’ variations. The other widely used data resource are quasi-panels, constructed from cross-section expenditure survey information by taking within-period means following the same population (e.g. 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. The presence of measurement error when estimating non-linear equations raises serious and

difficult problems³. In our context, GMM applied to the nonlinear consumption Euler equation would yield inconsistent estimates of the discount factor even though the estimates of the coefficient of relative risk aversion is not asymptotically biased. However, as demonstrated in section 5, in the finite sample, measurement error can cause significant bias in this parameter.

One way to address the measurement error problem is to assume that it is multiplicative and to work with a log-linearisation of equation 2. Carroll (2001), Ludvigson and Paxson (2001) and Attanasio and Low (2004) discuss the problems with such linearisation schemes. Our reading of these papers (and the results we present in section 5) is that the econometric methods we currently have in hand are not up to estimating the EIS (or the discount rate) on short and noisy panels. This is the case even if when we are willing to assume that everyone has the same parameters; as we shall demonstrate below this is not a plausible assumption.

What alternatives remain? One is to revert to old style consumption studies that are only loosely linked to conventional life-cycle theory. This is not a very attractive option given the modern emphasis on empirical modelling that stays close to the theory. A second alternative is to move to estimation based on structural models. This involves the numerical solution of the dynamic programming problem for every parameter value that the estimation procedure considers. An obvious problem regarding this approach is the fact that one needs to specify the underlying stochastic process (usually the income process) which is not necessary for Euler equation estimation (whether exact or approximate). It is not clear whether a slight misspecification of the income process will not completely change the results. To examine this would require that the estimation procedure be analyzed under misspecification, which would be extremely time consuming. Moreover, this type of estimation is currently not feasible if one wants to incorporate a plausible preference heterogeneity into the model. Although full structural modelling is potentially promising, an alternative is needed that reduces substantially the computational burden without sacrificing the close link to the theory. We present here an alternative that relies on simulating the distribution of expectations errors directly.

3 The Distribution of Expectations Errors

Our estimator is based on sampling from the conditional distribution of the expectations errors. 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. We have considered a great number of standard models and we present the results for a total of 17 scenarios. Our basic environment has agents with a finite lifetime of 80 periods with no bequest motive. We take the iso-elastic utility function with exponential discounting. We assume that agents face two types of income shocks, permanent and transitory. For agent h the assumed income process is:

$$Y_{h,t} = P_{h,t}u_{h,t} \quad (3)$$

where $u_{h,t}$ is an iid lognormal shock to transitory income with unit mean and a constant variance ($\exp(\sigma_u^2) - 1$) and $P_{h,t}$ is permanent income which follows a log random walk process:

$$P_{h,t} = P_{h,t-1}z_{h,t} \quad (4)$$

where $z_{h,t}$ is an iid lognormal shock to permanent income with unit mean and a constant variance ($\exp(\sigma_z^2) - 1$). In our simulations we set $\sigma_u = \sigma_z = 0.1$; these values are in line with those used in the literature and experiments with other values give qualitatively similar results (see Alan and Browning (2003)). We assume that the innovations to income are independent over time and across individuals so that we assume away aggregate shocks to income⁴. The real rate has a mean of 0.03

³In 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), Hong and Tamer (2003), Alan, Attanasio and Browning (2007))

⁴We allow for macro shocks in our empirical work below.

and is assumed to be the same for everyone between any two periods. For the models that have a 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; this is similar to empirical estimates for the US.

For each model 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 21 and the last 20 periods for each agent to minimise starting and end effects and obtain 40 periods. Table 1 presents the features of all 17 models we consider; models differ in the degree of curvature of the felicity function (γ), degree of impatience ($\delta - r$), income process parameters and the presence of liquidity constraints. The second-fourth columns report the coefficient of relative risk aversion, the discount rate ($= \frac{1-\beta}{\beta}$) and whether the interest rate is stochastic, respectively. The permanent income variance (σ_z^2) and transitory income variance (σ_u^2) are presented in column 5. The last column indicates whether we impose a liquidity constraint or not. The range of models in the Table covers all models considered in the literature under the constant relative risk aversion utility assumption.

For each pair of adjacent simulated periods we construct the expectations error (ε_{t+1}) according to equation (2). 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 ε_{t+1} on ε_t . For the models with a liquidity constraint (models 7, 8 and 11), after simulating consumption paths 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) as the Euler equation does not hold for these periods. To capture the effect of heterogeneity we also experiment with some mixed models. Model 12 is generated mixing simulated paths of models 1 and 2 (coefficient of relative risk aversion heterogeneity) with equal probability, model 13 is the mixture of models 1 and 3 (discount rate heterogeneity), model 14 is a mixture of models 1 and 4 (income process heterogeneity) and model 15 is a mixture of models 1, 2, 3 and 4 (full preference heterogeneity). Finally, models 16 and 17 add noise to the consumption paths from the baseline model, 1. In model 16 (respectively, model 17) we introduce moderate (respectively, high) noise with 30% (respectively, 60%) of the variance of consumption growth being due to measurement error.

The unconditional mean of the expectations is unity for all models except for those with measurement error. To test for the functional form for the distribution of errors we first estimate the parameters for a lognormal model and then for a mixture of two lognormals for each model. We then perform a Kolmogorov-Smirnov goodness of fit test for the error distribution against these estimated parametric distributions.⁵ Columns 2 and 3 of Table 2 present the p-values for the test statistics for each model. These indicate that we cannot always fit with a lognormal but a mixture of two lognormals always fits well, even for models with heavily skewed tails such as those with a Carroll income process or when we mix homogeneous models. It is this regularity (which also extends to other environments with different income processes) that underpins our estimation procedure. The final two columns give the slope parameter and t-value from the regression of the square of the current expectations error on the deviation from the mean of the lagged expectations error:

$$\varepsilon_{h,t}^2 = \pi_0 + \pi_1(\varepsilon_{h,t-1} - 1) + u_{t+1} \quad (5)$$

The t-values in Table 2 indicate that for some models there is strong conditional heteroskedasticity. We shall allow for this in our estimation procedure.

⁵When performing this test we do not allow that the parameters of the lognormal distribution are 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 model.

	CRRA	Discount rate	Real 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)				
Notes: Mean interest rate is 0.03 for all models. Standard deviation of 0.025 for stochastic rate models (9, 10 and 11).					
Income process: σ_z : std of permanent income shocks; $\sigma_u = 0.1$.					
* Model 1 with a 1% probability of zero income in any period.					

Table 1: Simulated Model Environments

Model	Distribution test		heteroskedasticity	
	L-test	M-test	Coeff	t-ratio
1	17.4	70.2	-.010	-2.7
2	45.5	32.1	-.011	-0.3
3	0.01	31.5	.079	5.4
4	0.00	11.1	-.195	-13
5	15.1	86.2	-.011	-3.0
6	0.00	24.8	3.05	39.7
7	18.9	68.4	-.010	-2.6
8	41.6	23.2	.003	0.7
9	14.1	35.6	-.013	-3.1
10	0.01	45.1	.053	4.8
11	68.2	84.8	-.004	-.74
12	0.00	49.9	-.005	-1.5
13	0.00	26.6	.049	4.4
14	0.00	55.8	-.086	-6.7
15	0.00	62.8	-.015	-2.2
16	51.9	73.9	-.536	-122
17	14.8	24.6	-1.11	-95.1
L-test is p-value (in %) for lognormality.				
M-test is p-value for mixture of two lognormals.				

Table 2: Distribution Tests for Expectational Errors

4 Synthetic Residual Estimation (SRE)

4.1 Overview

Our estimation procedure is a variant of Simulated Minimum Distance (SMD) which involves matching statistics from the data and from a simulated model.⁶ 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 $h = 1, \dots, H$ households and $t = 1, \dots, T$ periods. In the empirical section we discuss how to deal with the unbalanced panel that we actually use⁷. 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. *A priori*, it is very likely that patience and risk aversion will be heterogeneous; in the empirical section we give our specific motivation for allowing for heterogeneity in the current context.

The simulation procedure takes a set of 15 parameters:

$$\begin{aligned}
 &(\phi_{10}, \phi_{20}, \phi_1, \phi_2) \text{ for the expectation errors} \\
 &(\theta_0, \theta_1) \text{ for the initial values} \\
 &(\pi_{\beta_0}, \pi_{\beta_1}, \pi_{\beta_2}) \text{ for the discount factor} \\
 &(\pi_{\gamma_0}, \pi_{\gamma_1}, \pi_{\gamma_2}, \pi_{\gamma_3}) \text{ for the coefficient of relative risk aversion} \\
 &(\eta_1, \eta_2) \text{ for measurement error}
 \end{aligned} \tag{6}$$

and returns consumption paths for each household for $t = 1, \dots, T$. The parameters in (6) are the input for the optimization routine.

There are three steps for the simulation procedure. In the first step we simulate expectation 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 observations, preference parameters and measurement error realizations. In the third step we take the simulated expectations errors, the initial values and the simulated preference parameters and generate consumption paths using equation 2. Finally we add measurement error.

4.2 Simulating Expectation Errors

To simulate expectation errors, we draw four sets of mutually independent pseudo-random numbers: $(\nu_{1h,t}, \nu_{2h,t}, \nu_{3h,t}, \mu_{h,t})$ for all h and t . The variables $\nu_{1h,t}$, $\nu_{2h,t}$, and $\mu_{h,t}$ are standard normal variables and $\nu_{3h,t}$ is a uniform on $[0, 1]$. We then take four parameters $(\phi_{10}, \phi_{20}, \phi_1, \phi_2)$ from (6).

The expectations error terms, $\varepsilon_{h,t}$, are simulated recursively. For $t = 0$, we define two variances, $\sigma_{1h,t}^2$ and $\sigma_{2h,t}^2$, by:

$$\sigma_k^2 = \exp(\phi_{k0}), \quad k = 1, 2 \tag{7}$$

Then define two heterogeneous component error terms by:

$$\varepsilon_{k,h,0} = \exp\left(-\frac{\ln(1 + \sigma_k^2)}{2} + \sqrt{\ln(1 + \sigma_k^2)}\nu_{k,h,0}\right), \quad k = 1, 2 \tag{8}$$

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

$$d_{h,0} = \Phi(50 * (\nu_{3h,0} - \phi_2))$$

where $\Phi(\cdot)$ is the standard Normal cdf. This is a ‘smoothed’ indicator function which takes 0 or 1 for values of the uniformly distributed random draws $\nu_{3h,0}$ that are not very close to ϕ_2 . Such

⁶A description of the general SMD procedure we use is given in Appendix B.

⁷We also postpone to the empirical section any discussion of how to allow for time varying observable factors such as household composition.

smoothed indicators are routinely used to facilitate derivative based optimization. These values control whether household h in period t draws from the first component distribution or the second, so that:

$$\varepsilon_{h,0} = d_{h,0}\varepsilon_{1,h,0} + (1 - d_{h,0})\varepsilon_{2,h,0} \quad (9)$$

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

Given expectations errors for period 0, we then continue for the next T periods. For $t > 1$, define recursively heterogeneous time varying variances, $\sigma_{1,h,t}^2$ and $\sigma_{2,h,t}^2$, by:

$$\sigma_{k,h,t}^2 = \exp(\phi_{k0} + \phi_1 (\varepsilon_{h,t-1} - 1)), \quad k = 1, 2 \quad (10)$$

Thus each variance depends on the lagged realized error term with the same slope coefficient for each variance. This is designed to capture the dependence in equation 5. Then define two component error terms by:

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

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

$$d_{h,t} = \Phi(50 * (\nu_{3h,t} - \phi_2))$$

and define expectations errors by:

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

The synthetic expectations errors $\varepsilon_{h,t}$ for $h = 1, \dots, H$ and $t = 1, \dots, T$ 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 and real rates and other shocks. We refer to the $\varepsilon_{h,t}$'s as *synthetic residuals* and term our estimation procedure *Synthetic Residual Estimation* (SRE).

4.3 Simulating Preference Parameters

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

$$C_{h,1}^* = \exp(\theta_0 + \exp(\theta_1) c_h) \quad (12)$$

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.⁸ We also allow that these preference parameters may be correlated with the level of consumption by allowing that they are correlated with the initial value of 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 logistic model for the discount factor:

$$\beta_h = 0.8 + 0.2 \left(\frac{\exp(\pi_{\beta 0} + \exp(\pi_{\beta 1}) b_h + \pi_{\beta 2} \ln(C_{h,1}^*))}{1 + \exp(\pi_{\beta 0} + \exp(\pi_{\beta 1}) b_h + \pi_{\beta 2} \ln(C_{h,1}^*))} \right) \quad (13)$$

⁸These 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.

where π_{β_0} and π_{β_1} capture the location and dispersion respectively ($\exp(\pi_{\beta_1})$ is to ensure that the dispersion parameter is positive). The parameter π_{β_2} introduces dependence between the distribution of β and the initial consumption values. The coefficients of relative risk aversion have the following parameterization:

$$\gamma_h = 1 + 14 \left(\frac{\exp(\pi_{\gamma_0} + \pi_{\gamma_1} b_h + \exp(\pi_{\gamma_2}) g_h + \pi_{\gamma_3} \ln(C_{h,1}^*))}{1 + \exp(\pi_{\gamma_0} + \pi_{\gamma_1} b_h + \exp(\pi_{\gamma_2}) g_h + \pi_{\gamma_3} \ln(C_{h,1}^*))} \right) \quad (14)$$

where π_{γ_0} and π_{γ_2} capture the location and dispersion respectively. The parameter π_{γ_1} captures dependence between the distributions of the two parameters and π_{γ_3} introduces dependence between the distribution of γ and the initial values.

4.4 Generating Consumption Paths

Given values for $(C_{h,1}^*, \beta_h, \gamma_h)$ for $h = 1, \dots, H$ and synthetic expectation 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 equation 2:

$$C_{h,t}^* = C_{h,t-1}^* \left\{ \frac{\varepsilon_{h,t}}{\beta_h(1+r_t)} \right\}^{-\frac{1}{\gamma_h}} \quad (15)$$

After generating a full path of consumption for each household we introduce measurement error.⁹ We do this 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:

$$\xi_h = \exp(\eta_1 + \exp(\eta_2) m_h) \quad (16)$$

We use these to simulate time varying measurement errors according to:

$$\kappa_{h,t} = \exp(\omega_h + \xi_h \mu_{h,t}) \quad (17)$$

where ω_h is an idiosyncratic bias term (which disappears when we difference; see below). We then define ‘observed’ simulated consumption as:

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

It is these simulated consumption paths that are used in the SRE optimization step¹⁰.

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

$$\Delta \ln C_{h,t}^S = \frac{1}{\gamma_h} \ln(\beta_h) + \frac{1}{\gamma_h} \ln(1+r_t) + \left(\xi_h \mu_{h,t} - \frac{1}{\gamma_h} \ln \varepsilon_{h,t} \right) - \xi_h \mu_{h,t-1} \quad (18)$$

From this equation we can see the broad outlines of our identification strategy. The responses to real interest rates identify the distribution of the γ_h 's. The distribution of the discount factors β_h is

⁹ Although we call it ‘measurement error’ throughout the paper this can also be interpreted as 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 (preference parameters).

¹⁰ One can also incorporate a multi-asset framework into the SRE. For example, in a standard two-asset case (a risk-free bond and a risky stock) we would have two exact Euler equations. Expectations errors using these equations could be approximated parametrically (by taking into account the dependence between them). For this though one needs to take careful account of stock market non-participants. We leave this extension to future research.

then identified from this and the distribution of trends. Under the model assumptions, measurement error is the only source of auto-correlation in the composite error term:

$$u_{h,t} = \left(\xi_h \mu_{h,t} - \frac{1}{\gamma_h} \ln \varepsilon_{h,t} \right) - \xi_h \mu_{h,t-1} \quad (19)$$

Thus the cross-section variances and auto-covariances of the composite error term determines the variances of the expectation errors $\varepsilon_{h,t}$ and the measurement error, $\kappa_{h,t}$. In the next subsection we use this to structure our choice of auxiliary parameters that we match between the data and the simulated sample.

4.5 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 $\lambda_1, \dots, \lambda_K$. As always we have a trade-off between the closeness of the ap's to structural parameters (the 'diagonality' of the binding function, see Gouriéroux *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 and interest rates across the sample.

Our first two ap's relate to the initial distribution parameters (θ_0, θ_1) in (6). We take the mean and standard deviation of log initial consumption:

$$\begin{aligned} \lambda_1 &= \text{mean}(\ln(C_{h,1})) \\ \lambda_2 &= \text{std}(\ln(C_{h,2})) \end{aligned} \quad (20)$$

In the empirical section below we discuss how to allow for an unbalanced panel and the fact that the initial observation period varies across households.

Our next six ap's are closely related to the preference parameters $(\pi_{\beta_0}, \pi_{\beta_1}, \pi_{\gamma_0}, \pi_{\gamma_1}, \pi_{\gamma_2})$ in (6). The first ap is to match the trend and the change in the cross-section dispersion of log consumption. To do this, we calculate the cross-section median and interquartile range (iqr) of household log real expenditures in each year¹¹ 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):

$$\begin{aligned} \lambda_3 &= 100 * \text{trend in cross section median } \ln(\text{consumption}) \\ \lambda_4 &= 100 * \text{trend in cross section iqr } \ln(\text{consumption}) \end{aligned} \quad (21)$$

The next four ap's are based on regressions of consumption growth on the real rate for each household individually:

$$\Delta \log C_{h,t} = \zeta_{0h} + \zeta_{1h} r_t + \epsilon_{h,t} \quad (22)$$

Note that we could equally well 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. Given parameter estimates for each household, define:

$$\hat{\epsilon}_{h,t} = \Delta \log C_{h,t} - \hat{\zeta}_{0h} - \hat{\zeta}_{1h} r_t \quad (23)$$

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

$$\begin{aligned} \varphi_h &= \text{std}(\hat{\epsilon}_{h,t}) \\ \varsigma_h &= \text{corr}(\hat{\epsilon}_{h,t}, \hat{\epsilon}_{h,t-1}) \end{aligned} \quad (24)$$

¹¹Here and below, we use medians and inter-quartile ranges rather than means and standard deviations to minimise the impact of outliers.

We then record the following eight cross-section statistics:

$$\begin{aligned}
\lambda_5 &= \text{median}(\hat{\zeta}_{0h}) \\
\lambda_6 &= \text{iqr}(\hat{\zeta}_{0h}) \\
\lambda_7 &= \text{median}(\hat{\zeta}_{1h}) \\
\lambda_8 &= \text{iqr}(\hat{\zeta}_{1h}) \\
\lambda_9 &= \text{mean}(\varphi_h) \\
\lambda_{10} &= \text{std}(\varphi_h) \\
\lambda_{11} &= \text{mean}(\varsigma_h) \\
\lambda_{12} &= \text{std}(\varsigma_h)
\end{aligned} \tag{25}$$

The next four ap's are based on the individual trends and standard deviations (of log consumption) for individual households. That is, we first calculate the trend τ_h and standard deviation v_h for each household. We then record:

$$\begin{aligned}
\lambda_{13} &= \text{mean}(\tau_h) \\
\lambda_{14} &= \text{std}(\tau_h) \\
\lambda_{15} &= \text{mean}(v_h) \\
\lambda_{16} &= \text{std}(v_h) \\
\lambda_{17} &= \text{corr}(\tau_h, v_h)
\end{aligned} \tag{26}$$

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:

$$\begin{aligned}
\lambda_{18} &= \text{corr}(\tau_h, \psi_h) \\
\lambda_{19} &= \text{corr}(v_h, \psi_h)
\end{aligned} \tag{27}$$

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 rate and persistent variances in observed consumption growth.

Finally we have a series of ap's that are based on pooled regressions. These are designed to capture the major features of the expectation error distribution which is assumed common to everyone. The ap's are based on the residuals from the following pooled regression:

$$\Delta \log C_{h,t} = \alpha_0 + \alpha_1 r_t + e_{h,t}, \quad h = 1, \dots, H, \quad t = 2, \dots, T \tag{28}$$

The estimated residuals are:

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

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

$$\begin{aligned}
\lambda_{20} &= \text{std}(\hat{e}_{h,t}) \\
\lambda_{21} &= \text{skew}(\hat{e}_{h,t}) \\
\lambda_{22} &= \text{kurt}(\hat{e}_{h,t})
\end{aligned} \tag{29}$$

To capture the conditional heteroskedasticity we run an analogue of (5):

$$(\hat{e}_{h,t})^2 = \vartheta_0 + \vartheta_1 \hat{e}_{h,t-1} + \text{error}, \quad h = 1, \dots, H, \quad t = 3, \dots, T \tag{30}$$

This gives the following two ap's:

$$\begin{aligned}\lambda_{23} &= \hat{\vartheta}_0 \\ \lambda_{24} &= \hat{\vartheta}_1\end{aligned}\tag{31}$$

Thus we have 24 ap's to match to estimate 15 parameters (see.(6)).

5 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 real rate model 9 in section 3 and add measurement error that implies 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 the case where households may face binding liquidity constraints in some periods. In this experiment we solve the model with Deaton (1991) type explicit liquidity constraint 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 3). 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 2). We follow Alan, Attanasio and Browning (2007) and assume a classical multiplicative lognormal measurement error with standard deviation σ_κ . The associated orthogonality conditions are:

$$E_t \left[\left(\frac{C_{h,t+1}}{C_{h,t}} \right)^{-\gamma} (1 + r_{t+1})\beta - \exp\{\gamma^2 \sigma_\kappa^2\} \right] = 0\tag{32}$$

$$E_t \left[\left(\frac{C_{h,t+2}}{C_{h,t}} \right)^{-\gamma} (1 + r_{t+1})(1 + r_{t+2})\beta^2 - \exp\{\gamma^2 \sigma_\kappa^2\} \right] = 0\tag{33}$$

The first equation (the 'conventional' exact Euler 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 rate for the first equation and the constant for the second, so that we just identify the parameters.¹² Our second empirical model is the approximate Euler

¹²We 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.

Experiment	T	γ	β	σ_κ	EGMM			AGMM	SRE		
					$\hat{\gamma}$	$\hat{\beta}$	$\hat{\sigma}_\kappa$	$\hat{\gamma}$	$\hat{\gamma}$	$\hat{\beta}$	$\hat{\sigma}_\kappa$
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)

Values are means, [medians] and (standard deviations) of sampling distributions.
Number of Monte Carlo replications = 1000. $\sigma_\kappa = 0.15$ amounts to 50% noise.

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

equation (equation 28 without the family size variable). 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 4. We set the mixing probability to 0.5. We thus have six structural parameters to estimate: γ , β , σ_η , ϕ_{10} , ϕ_{20} and ϕ_1 (the parameters of two time varying variances of expectation errors). The six auxiliary parameters used for estimation are: the constant and slope in the OLS estimation of the approximate Euler equation (equation 28); the standard deviation and auto-regressive coefficient of the OLS residuals and the constant and slope coefficients of the regression of squared OLS residuals on their lags (not squared). The system is just identified (we have six ap's and six structural 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 17 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 does perform as well as the SRE.

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 very similarly (SRE with some upward 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 considerable sampling distribution. The SRE performs very well at 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 and using the same model for the SRE as in actually generating the data, 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.

6 Estimates from the PSID

6.1 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, representativeness issues due to attrition) 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 previously.

Although the use of food as ‘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 consumption of food in period t and c_t^o be consumption of other goods.¹³ Assume that intertemporal preferences are additive within the period with each sub-utility function taking an iso-elastic form:

$$U_t = E_t \left(\sum_{s=0}^{T-t} \beta^s \left\{ \left(\frac{c_{t+s}^f}{1-\gamma_f} \right)^{1-\gamma_f} + \left(\frac{c_{t+s}^o}{1-\gamma_o} \right)^{1-\gamma_o} \right\} \right) \quad (34)$$

In this formulation we have two coefficients of relative risk aversion parameters, γ_f and γ_o .¹⁴ The additivity allows us to break the intertemporal allocation problem into two sub-problems, one for food and the other for ‘other goods’.

$$U_t = E_t \left(\sum_{s=0}^{T-t} \beta^s \left(\frac{c_{t+s}^f}{1-\gamma_f} \right)^{1-\gamma_f} \right) + E_t \left(\sum_{s=0}^{T-t} \beta^s \left(\frac{c_{t+s}^o}{1-\gamma_o} \right)^{1-\gamma_o} \right) \quad (35)$$

¹³To 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.

¹⁴Note 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.

The Euler equation for food is given by:

$$\left(\frac{c_{t+1}^f}{c_t^f}\right)^{-\gamma_f} (1 + r_{t+1}^f) \beta = \varepsilon_{t+1} \text{ with } E_t(\varepsilon_{t+1}) = 1 \quad (36)$$

where r_{t+1}^f is the nominal rate between periods t and $t + 1$, discounted by food inflation (the ‘real rate for food’). Estimating using the variation in food consumption allows us to recover β and γ_f . The elasticity of intertemporal substitution for food is then given by¹⁵:

$$\phi_f = -\frac{1}{\gamma_f} \quad (37)$$

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

$$\phi_f = \phi \cdot e_f \quad (38)$$

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 ϕ_f under-estimate the value of the total elasticity, ϕ .

6.2 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 several years after 1987. We treat split-ups as separate household units and exclude single headed households. 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; food expenditures were reported and assets, at the level of at least two months’ income, were carried forward from one period to the next. The latter restriction is to exclude households that are potentially liquidity constrained¹⁶. Thus we have at least four consecutive years for each sampled households in which we can observe consumption growth and for which the Euler equation should hold. We stratify into two categories: ‘less educated’ households in which the head received less than 12 years (inclusive) and ‘more educated’ are the households whose head received more than 12 years. 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 ‘more educated’. Table 6 in Appendix C shows the exact distribution for years of observation. 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; details are given below).¹⁷ Exact replication of the structure of the panel data to hand when simulating consumption growth of households is crucial in the estimation procedure. In the simulations, the consumption path of each household is imitated by setting the initial consumption to the one observed in the data for the corresponding household. The lengths of the observed

¹⁵Sometimes the negative sign is dropped. Since the elasticity is an own price response, we prefer to retain it.

¹⁶Selecting 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.

¹⁷The 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 the *MLE* estimator. 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.

paths are also replicated exactly. For example, a household that is observed for 8 consecutive periods (say from 1979 to 1986) has a simulated consumption path for exactly 8 periods corresponding to the interest rates that prevailed between 1979 and 1986. Hence, the auxiliary parameters for the simulated sample are obtained from unbalanced simulated panels of 8×833 ‘less educated’ and 8×868 ‘more educated’ households (with 8×8116 and 8×9065 observations respectively).

6.3 The Distribution of Trends and Variances

Of all the features that empirical analysis using micro data has to address, heterogeneity is the most important. In this section we present a method to identify the heterogeneity in the discount factor and coefficient of relative risk aversion within each education group. Our approach to identifying the distribution of discount factors begins with the observation that there are marked differences among households in our sample in their consumption growth and their variance of consumption growth. These may indicate heterogeneity in discount factors and the EIS across households.¹⁸ To show this, we take means over time of consumption growth and consumption growth variance for each household.

In the left panel of Figure 1 we present the distributions of mean consumption growth for our two education groups. 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 reason for these within strata differences is that different households have different realizations of the expectations errors and some have 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 strata.

7 Results

7.1 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 D. 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 7 in Appendix E. 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 and λ_{13}). The respective values for the latter are 2.37% and 0.86% . These values are somewhat higher than those displayed in figure 1; the difference is due to the accounting for cohort and family size effects. The dispersion of the trends (λ_{14}) is very similar across the two strata.
2. There is an increasing cross-section variance (λ_4) for both strata with the more educated having a stronger trend.
3. The coefficients on the real interest rate changes 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.

¹⁸The ‘may’ is there since these differences could, of course, be generated in a model with cross-section differences in income realisations, even if discount factors and attitudes to risk are homogeneous.

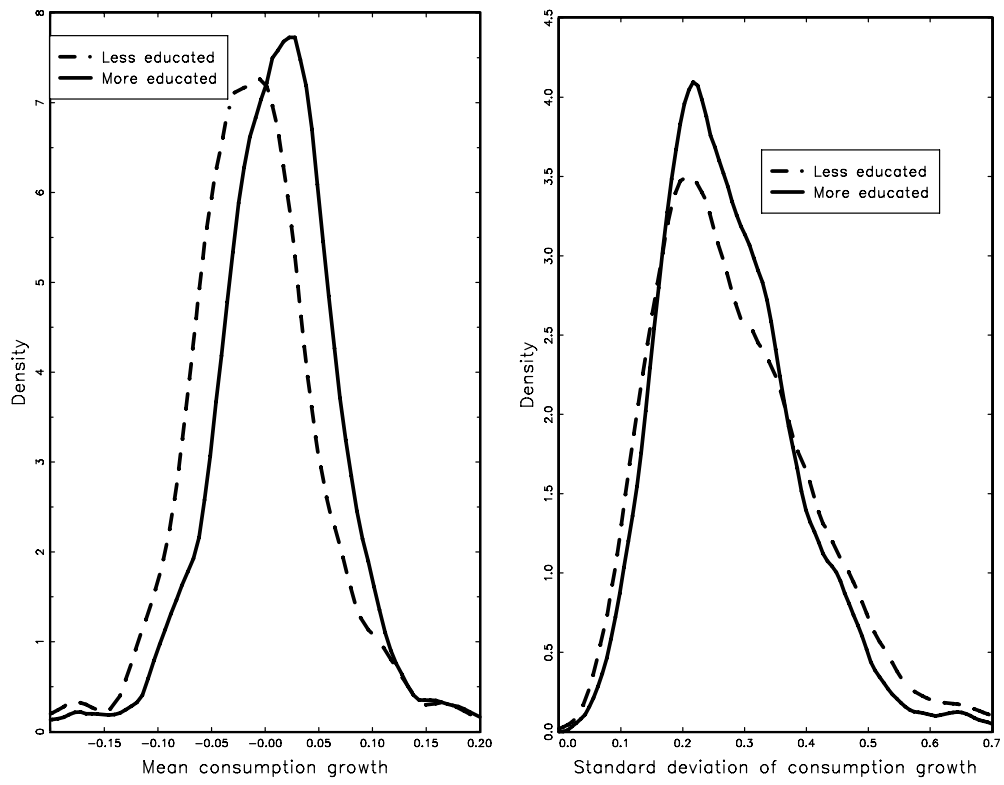


Figure 1: Distribution of means and standard deviations of consumption growth

4. 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. On the other hand, the auto-correlation is not very dispersed (λ_{12}).
5. The correlation between the starting value 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 growth variance and mean log consumption (λ_{19}).
6. The ARCH ap (λ_{24}) indicates negative dependence 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 (6). We also estimated a number of restricted variants of this model, see Table 4. The first row gives the fit for the unrestricted model. The next three rows give versions with the heterogeneity closed down for β , γ and both β and γ (except for the dependence on the initial value); see equations 13 and 14. The fifth row shows the effect of closing down the dependence on the initial value. The final row shows the goodness of fit for the model with homogeneous preference parameters. For the unrestricted models (row 1), neither strata fits as well as we could hope. An examination of the ap values 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 go on and use the unrestricted model and compare it to more restricted variants.

The most important conclusions that can be drawn from Table 4 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 ($\pi_{\beta 2} = \pi_{\gamma 3} = 0$) - 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. In the case of the less educated we could close down the uncorrelated heterogeneity in β (see the second row) whereas the model with no uncorrelated heterogeneity does well for the more educated (row 4). Given this mix of results we shall work with the unrestricted model for both strata in all that follows.

Parameter restrictions	Degrees of freedom	Less educated	More educated
		χ^2	χ^2
—	9	33.76	36.84
$\pi_{\beta 1} = \pi_{\gamma 1} = 0$	11	36.33	37.69
$\pi_{\gamma 1} = \pi_{\gamma 2} = 0$	11	45.50	37.93
$\pi_{\beta 1} = \pi_{\gamma 1} = \pi_{\gamma 2} = 0$	12	44.67	38.83
$\pi_{\beta 2} = \pi_{\gamma 3} = 0$	11	60.37	68.14
$\pi_{\beta 1} = \pi_{\gamma 1} = \pi_{\gamma 2} = \pi_{\beta 2} = \pi_{\gamma 3} = 0$	14	74.40	72.26
Notes. Number of ap's equals 24			

Table 4: Goodness of fit

We present the parameter estimates for the unrestricted models in Table 5; see (6) for details of the parameters. These estimates are of little intrinsic interest since they 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 family effects that were also removed in the first round). We present some of the objects that we consider of interest in the next sub-section.

Parameter group	Parameter name	Education	
		Low	High
Distribution of shocks	ϕ_{10}	-1.465	-0.621
	ϕ_{20}	-1.593	-0.818
	ϕ_1	0.827	0.186
	ϕ_2	1.688	-0.160
Initial value	θ_0	1.335	1.259
	θ_1	-1.172	-1.095
Distribution of β	π_{β_0}	7.039	7.389
	π_{β_1}	-2.340	-0.232
	π_{β_2}	-4.824	-4.725
Distribution of γ	π_{γ_0}	-2.100	-1.803
	π_{γ_1}	0.139	0.556
	π_{γ_2}	0.037	-0.479
	π_{γ_3}	1.183	-1.528
Measurement error	η_1	-1.764	-1.768
	η_2	-1.228	-1.366

Table 5: Structural Estimates

7.2 The Implications of the Estimates.

In this sub-section we present some implications of our parameter estimates. We consider first the extent of measurement error that we estimate. We compute the extent of the noise by considering the variance of the pooled simulated consumption growth values without and with measurement error, denoted σ_{true}^2 and σ_{obs}^2 , respectively. The proportion of the observed consumption growth that is due to noise is then given by $(\sigma_{obs}^2 - \sigma_{true}^2) / \sigma_{obs}^2$. 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 (Runkle (1991) estimates about 76% noise in the PSID food data), but it is consistent with the consensus that the PSID food measure is very noisy. We also found considerable dispersion in the idiosyncratic measurement error variance, with values for the standard deviations of 0.11 to 0.27 for the 5th and 95th deciles 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 η_1 and η_2 in Table 5).

As regards the starting distribution of consumption, the less educated have a higher starting 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 for β and γ . As can be seen from the left hand figure, the more educated are more patient than the less educated. The medians for β are 0.93 and 0.96 for the less educated and the more educated respectively (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 figure 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 literature are due to Lawrance (1991), Cagetti (2003) and Samwick (1998). Lawrance (1991) uses an Euler equation approach on the PSID to examine the dependence of the discount rate on permanent factors such as cohort, education, race and income averaged over several years. She 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 those with a college education and those without such an education is two percentage points whereas we find a difference of 3.3 percentage points. She also finds that the discount rate is lower for higher income households but the differences between the top and

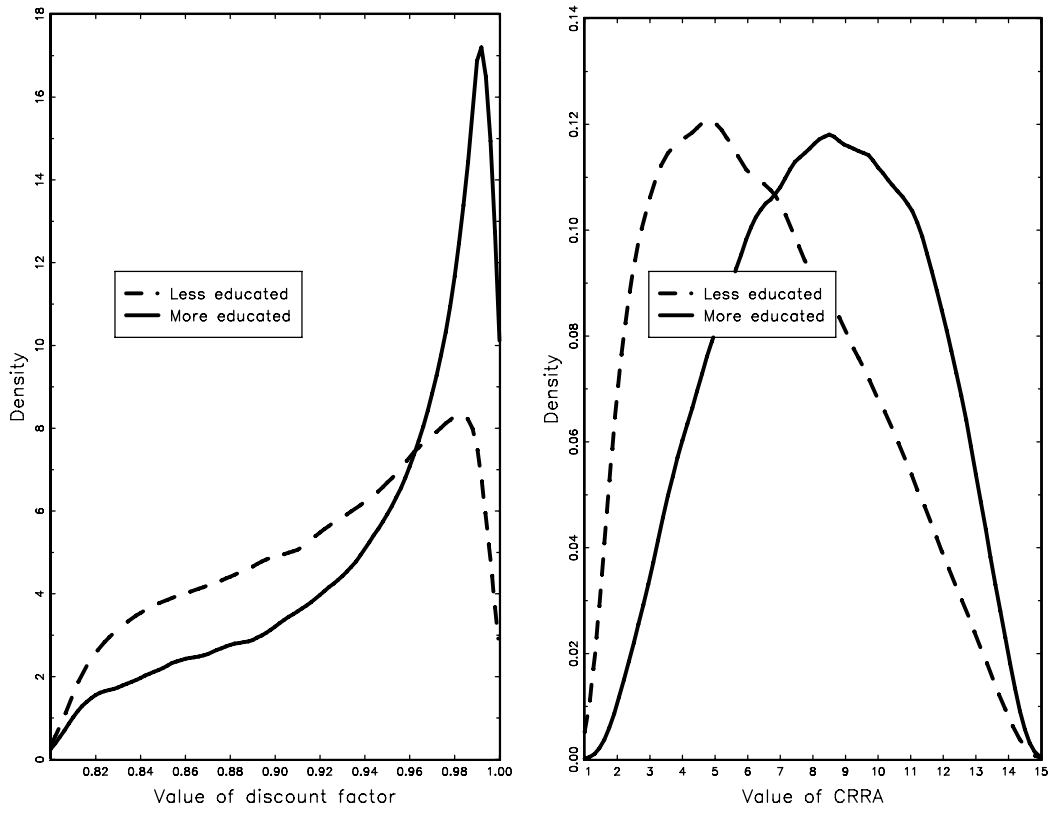


Figure 2: Marginal distributions of β and γ .

bottom decile income households is modest: about one to two percentage points. Using a standard life-cycle model, Samwick (1998) backs out the discount factor from simulated wealth at retirement (without using any formal estimation method) using 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, with median values from 0.93. More importantly, he finds quite wide dispersions. The general pattern seems to be that there are three groups: the very impatient (values of the discount rate of more than 20% per year), the moderately impatient (values between zero and 10%) and a group (comprising about five to twenty percent of the sample) who have a rate of about -15% (so that they discount the present). It will be clear that 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) matches the median net worth in the SCF and the PSID, and estimates that the (homogenous within education strata) discount factor is around 0.98 for those with a college education and about 0.86 for the high school strata.

The right hand panel of figure 2 shows the distributions for the coefficient of relative risk aversion (γ). These high values are being driven by the fact that the estimated responses to real rate changes when we generate the ap's are generally small (see the discussion of ap's λ_7 and λ_8 in subsection 7). We find that the lower educated households are less risk averse than more educated households. The medians for the two distributions are 6.2 and 8.4 respectively. 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 models¹⁹: experiments (see Holt and Laury (2005)); survey questions (Barsky et al (1997), Dohmen *et al* (2005), Guiso and Piaella (2001), Eisenhauer and Ventura (2003)); consumption based empirical studies (for example, Gourinchas and Parker (2002), French and Jones (2004), Blau and Gilleskie (2005) and French (2005)); wealth or portfolio based empirical studies and empirical studies using other contexts such as game show behavior (see Cohen and Einav (2007) for a striking example and references). In general, estimates based on consumption data generate lower 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. Cagetti (2003) estimates the coefficient of relative risk aversion to be around 4 (for all 3 education categories he considers). When he matches the mean net worth instead of median the estimates of the coefficient of relative risk aversion goes up to 8. Estimates based on asset allocation data are generally even higher. Thus Cocco, Gomes and Maenhout (2005) show, via simulations of a standard portfolio allocation model, that the share of wealth invested in the stock market generated by the model can be reconciled with data if one assumes a risk aversion coefficient of around 10 (with a discount factor such as 0.96). Kahvecioglu (2005) goes beyond simulations and estimates the parameters by matching the allocations and the stock market participation (by adding a one-time fixed entry cost) and finds a very high coefficient of relative risk aversion of around 18. The exception in this branch of the literature is Alan (2006) who matches only the stock market participation profile in the PSID, allowing for costs to stock market participation. She estimates the coefficient of relative risk aversion to be 1.6 and concludes that such a small value has no hope of matching the observed portfolio shares. It will be clear that the means of our (consumption based) estimates are much closer to the portfolio conclusions than to the consumption based estimates, even though our estimates are themselves, consumption based.

Turning to the estimates of heterogeneity in risk attitudes, for comparison, we have to rely mainly on survey data (with direct risk attitude questions) and analysis of other contexts as, to our knowledge, estimates of such does not exist in the context of consumption and portfolio choice. Even though survey questions typically involve wealth or income rather than consumption, they have some relevance for preferences over risky consumption. In particular, all surveys show strong

¹⁹We do not attempt to relate our results to the vast literature on models that do not assume expected utility maximisation.

evidence of heterogeneity. Dohmen *et al* (2005) use a large German representative sample who are asked directly about their attitudes to risk. They find that even when we allow for differences in gender, age and parental education background there is considerable dispersion in responses; they report an R^2 of about 0.05 for a regression of their ‘financial matters’ risk measure on observable characteristics. Although some of the unexplained variance is probably due to measurement error, this provides strong evidence for unobserved heterogeneity. 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. These individual responses are weakly correlated with observables (not including own education); Guiso and Piaella (2001) conclude that risk attitudes ‘are characterized by massive unexplained heterogeneity’. Eisenhauer and Ventura (2003) use the same data source as Guiso and Piaella (2001) 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 higher educated people have higher levels of *absolute* risk aversion; this is consistent with our finding if, as seems plausible, the more educated having 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, 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; Harrison *et al* (2005) provide an insightful discussion of the relevant measurement issues. The results to date are rather mixed. Anderhub *et al* (2001) (using a sample of Israeli students) find a negative correlation between risk attitudes and the discount factor. Harrison *et al* (2005) present results for a representative sample drawn from the Danish population and find no correlation. Eckel *et al* (2004) conduct experiments with poor people in Montreal. They find that ‘risk averse individuals are also more present-oriented’; that is, risk aversion and the discount factor are negatively correlated. 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 (the case for most of the less educated and a majority of the highly educated) 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 a great potential 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 understanding the concentration of wealth (see for example Krusell and Smith (1998) and Hendricks (2007)).

8 Conclusions

The current literature, based on empirical, experimental and survey evidence, suggests strongly the need to allow for substantial heterogeneity when modelling saving and consumption. Current Euler equation GMM methods are not able to allow for heterogeneity if we have noisy consumption data. On the other hand, full structural modelling with different simulations for each household are 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 log Normals. This is the case even if agents have different income processes; essentially the differences in variances of the latter 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 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 those 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 who are liquidity constrained. For example, we could use those with low assets or debts and allow that they face a higher interest rate. 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 expectation 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 age, cohort 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. We also find that the more educated are more patient and more risk averse than the less educated, but there is considerable overlap between the distributions for the two strata. Within 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 to 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, (2007), “Estimating Euler equations with Noisy Data: Two Exact GMM Estimators”, unpublished manuscript.
- [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] Anderson, T. W. and Hsiao, C. "Formulation and Estimation of Dynamic-Models Using Panel Data." *Journal of Econometrics*, 1982, 18(1), pp. 47-82.
- [7] Attanasio, O. P., J. Banks, C. Meghir, G. Weber, (1999), "Humps and Bumps in Lifetime Consumption", *Journal of Business and Economic Statistics*, 17(1), 22-35.
- [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] Blau, D., and D. Gilleskie, (2006), "Health Insurance and Retirement of Married Couples", *Journal Of Applied Econometrics*, 21(7), 935-953.
- [12] 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.
- [13] Browning, M. and T. Crossley, (2000), "Luxuries are Easier to Postpone: A Proof", *Journal of Political Economy*, 108(5):1064-1068.
- [14] Browning, M. and T. Crossley, (2001), "The Life-Cycle Model of Consumption and Savings," *Journal of Economic Perspectives*, 15(3):3-22.
- [15] Browning, M., M. Ejrnæs, (2007), "Consumption and Children", *Review of Economics and Statistics*, forthcoming.
- [16] Browning, M., M. Ejrnæs and J. Alvarez, (2006), "Modelling income processes with lots of heterogeneity", University of Oxford Discussion Paper.
- [17] Browning, M. and A. Lusardi, (1996), "Household Saving: Micro Theories and Micro Facts", *Journal of Economic Literature*, December, 1797-1855.
- [18] Cagetti, M. (2003), "Wealth Accumulation Over the Life Cycle and Precautionary Savings", *Journal of Business and Economic Statistics*, 21(3), 339-353.
- [19] Carroll, C., (2001), "Death to the Log-Linearized Consumption Euler equation!", *Advances in Macroeconomics*, 1(1), Article 6.
- [20] Carroll, C. and A. Samwick, (1997) "The Nature of Precautionary Wealth", *Journal of Monetary Economics*, 40, 41-71.
- [21] Chamberlain, G. (1980), "Analysis of Covariance with Qualitative Data", *Review of Economic Studies*, 47, 225-238.
- [22] Cocco, J.F., F.J. Gomes, and P.J. Maenhout (2005), "Consumption and Portfolio Choice over the Life Cycle", *The Review of Financial Studies*, 18(2).
- [23] Cohen, A. and L. Einav (2007), "Estimating risk preferences from deductible choice", *American Economic Review* 97(3): 745-788.
- [24] Deaton, A., (1991), "Saving and Liquidity Constraints", *Econometrica*, 59(5), 1221-1248.
- [25] Dohmen, T., Falk, A., Huffman, D. Sunde, U. Schupp, J. and G. G. Wagner, (2005), "Individual Risk Attitudes: New Evidence From a Large, Representative, Experimentally-Validated Survey", IZA Discussion Paper 1730.

- [26] Dynan K. E., (1993), "How Prudent Are Consumers?", *Journal of Political Economy*, vol. 101, no:6, 1104-1113.
- [27] 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.
- [28] Eisenhauer, J. G. and L. Ventura, (2003), "Survey measures of risk aversion and prudence", *Applied Economics* 35(13): 1477-1484.
- [29] French, E. (2005), "The Effects of Health, Wealth, and Wages on Labor Supply and Retirement Behavior", *Review of Economic Studies*, 72, 395-427.
- [30] 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.
- [31] Gallant, A. R. and G. Tauchen, (1996), "Which Moments to Match?", *Econometric Theory*, 12:657-681.
- [32] Gourieroux, C., A. Monfort and E. Renault, (1993), "Indirect Inference", *Journal of Applied Econometrics*, 1993, 8:S85-S118.
- [33] Gourinchas, P. O. and J. A. Parker, (2002), "Consumption over the life cycle", *Econometrica* 70(1): 47-89.
- [34] Guiso, L. and M. Paiella, (2001), "Risk aversion, wealth and background risk", CEPR Working Paper 2728.
- [35] Guvenen, F., (2006), "Reconciling conflicting evidence on the elasticity of intertemporal substitution: A macroeconomic perspective" *Journal of Monetary Economics* 53(7): 1451-1472.
- [36] 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.
- [37] Harrison, G. W., M. I. Lau, E. Ruström 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.
- [38] 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.
- [39] Hendricks, L. (2007), "How Important is Discount Rate heterogeneity for Wealth Inequality", *Journal of Economic Dynamics and Control*, 31(9), 3042-3068.
- [40] 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.
- [41] Hong, H. and E. Tamer, (2003), "A Simple Estimator for Nonlinear Error in Variable Models", *Journal of Econometrics*, 117, 1-19.
- [42] Hubbard, R. G., J. Skinner and S. P. Zeldes, (1995), "Precautionary Saving and Social Insurance", *Journal of Political Economy*, vol. 103(2), 360-399.
- [43] Kahvecioglu, D. C., (2005), "Asset Allocation, Bequests and Wealth Dynamics of the Elderly", unpublished manuscript.
- [44] 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.

- [45] Krusell, P. and A. Smith, (1998), "Income and Wealth Heterogeneity in the Macroeconomy", *Journal of Political Economy*, 2:1, 245-272.
- [46] Lawrence E. C., (1991), "Poverty and the Rate of Time Preference: Evidence from Panel Data", *Journal of Political Economy*, 99(1), 54-77.
- [47] Ludvigson, S. and C. Paxson, (2001), "Approximation Bias in Linearized Euler equations", *Journal of Economics and Statistics*, vol. 83(2), 242-256.
- [48] Runkle, D. E., (1991), "Liquidity Constraints and The Permanent Income Hypothesis", *Journal of Monetary Economics*, 27, 73-98.
- [49] Samwick, A. A., (1998), "Discount Rate heterogeneity and Social Security Reform", *Journal of Development Economics*, 57, 117-146.
- [50] Schennach, S. M., (2004), "Estimation of Nonlinear Models with Measurement Error", *Econometrica*, 72, 33-75.
- [51] Shapiro, M. D., (1984), "The Permanent Income Hypothesis and the Real Interest Rate", *Economics Letters*, 14, 93-100.
- [52] Vissing-Jorgensen, A., (2002), "Limited Asset Market Participation and the Elasticity of Intertemporal Substitution", *Journal of Political Economy*, 110(4), 825-853.
- [53] Wansbeek, T., (2001), "GMM Estimation in Panel Data Models with Measurement Error." *Journal of Econometrics*, 104, 259-268.
- [54] 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:

$$\max E_t \left[\sum_{j=0}^{T-t} \frac{(C_{h,t+j})^{1-\gamma}}{1-\gamma} \frac{1}{(1+\delta)^j} \right]$$

$$s.t. A_{h,t+j+1} = (1+r_{h,t+j})A_{h,t+j} + Y_{h,t+j} - C_{h,t+j}$$

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:

$$r_{h,t+1} = (1-\rho)\mu + \rho r_{h,t} + \epsilon_{h,t+1} \quad (39)$$

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

$$X_{h,t+j+1} = (1+r_{h,t+j+1})(X_{h,t+j} - C_{h,t+j}) + Y_{h,t+j+1} \quad (40)$$

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

$$\theta_t(w_t, r_t)^{-\gamma} - \frac{1}{(1+\delta)} E_t [(1+r_{t+1})\theta_{t+1}(w_{t+1}, r_{t+1})^{-\gamma} z_{t+1}^{-\gamma}] = 0 \quad (41)$$

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 $\{\theta_1(w_1, r_1), \dots, \theta_T(w_T)\}$.

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 Gouriéroux, 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, Ejrnaes and Alvarez (2006) for details.

Suppose that we observe $h = 1, 2, \dots, H$ units over $t = 1, 2, \dots, T$ 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 $\{(Y_1, X_1), \dots, (Y_H, X_H)\}$ 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.²⁰ If this distribution is tractable enough we could derive a likelihood 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 $\{(Y_1^1, X_1), \dots, (Y_H^1, X_H), (Y_1^2, X_1), \dots, (Y_H^S, X_H)\}$; these outcomes, of course, depend on the model chosen ($F(\cdot)$) 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 Gouriéroux *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:

$$\gamma_j^D = \frac{1}{H} \sum_{h=1}^H g^j(Y_h, X_h), \quad j = 1, 2, \dots, J \quad (42)$$

²⁰This could be generalised to allow for correlated heterogeneity by allowing for dependence on the initial values, as in Chamberlain (1980), Blundell and Smith (1991) and Wooldridge (2000).

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 γ^D . Using the same functions $g^j(\cdot)$ we can also calculate the corresponding values for the simulated data:

$$\gamma_j^S = \frac{1}{S * H} \sum_{s=1}^S \sum_{h=1}^H g^j(Y_h^s, X_h), \quad j = 1, 2, \dots, J \quad (43)$$

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

$$\text{rank}(\nabla_{\theta} \gamma^S(\theta)) = m \text{ with probability 1} \quad (44)$$

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:

$$\hat{\theta}_{SMD} = \arg \min_{\theta} (\gamma^S(\theta) - \gamma^D)' W (\gamma^S(\theta) - \gamma^D) \quad (45)$$

The choice we adopt is the (bootstrapped) covariance matrix of γ^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 Panel Distribution

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 6: Distribution of Unbalanced PSID Samples

D 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, $A_h (= 0, \dots, 32$ 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:

$$\Delta x_{ht} = \alpha + \beta \Delta z_{ht} + \gamma A_h + \sum_{s=1975}^{1986} \delta_s d_{ts} + u_{ht} \quad (46)$$

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:

$$\hat{p}_{ht} = \Delta x_{ht} - \hat{\beta} \Delta z_{ht} - \hat{\gamma} A_h - \sum_{s=1975}^{1986} \hat{\delta}_s d_{ts} + \bar{\delta} \quad (47)$$

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:

$$\hat{x}_{ht} = \hat{x}_{h,t-1} + \hat{p}_{ht} \quad (48)$$

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.

E Auxiliary Parameter Values

Ap	Less Educated					More Educated				
	Data	Confidence interval		Model	t	Data	Confidence interval		Model	t
		Value	2.5%				97.5%	Value		
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 7: Auxiliary Parameters for the two samples.



The Centre for Financial Analysis and Policy
Cambridge Judge Business School
University of Cambridge
Trumpington Street
Cambridge, CB2 1AG

www.cfap.jbs.cam.ac.uk
cfap-admin@cerf.cam.ac.uk

Centre for
**Financial
Analysis &
Policy**



CAMBRIDGE