Matching CEX Consumption Moments With Bayesian Learning of Heterogeneous Income Profiles

EC475 - Quantitative Economics

MSc in Econometrics and Mathematical Economics

April 28, 2008

Word count: 5686. (Please note that predominantly figures, not equations, cause the large page count.)

Matching CEX Consumption Moments With Bayesian Learning of Heterogeneous Income Profiles

Ryan Giordano

April 30, 2008

Abstract

In this paper, we use the method of simulated moments (MSM) to investigate the ability of a lifecycle model with heterogeneous income processes and Bayesian learning to reproduce the first two moments of log consumption data from the CEX. We focus on dovetailing two papers by Gourinchas and Parker [?] and Guvenen [?] that propose two different models of consumption over the lifecycle.

We perform five numerical experiments, treating various model parameters estimated or assumed in Gourinchas and Parker [?] or Guvenen [?] as fixed and optimizing with respect to the remaining parameters. Though the model was able to provide a reasonable fit to the target moments, in every case it either estimated counterintuitive parameter values or produced unrealistic predictions about savings behavior. This could be the result of two factors. First, due to limited computing power, we were unable to solve the model on a very fine grid or make a strong case that we found global not local minima. Second, we find that the data derived from the CEX after accounting for family size gives a much smaller increase in consumption variance over time than measures that do not control for family size, such as those used in Guvenen [?], and that this may have caused the heterogeneous income profile parameters estimated in Guvenen [?] to provide a poor fit for consumption data.

1 Introduction

In this paper, we use the method of simulated moments (MSM), as described in [?] to investigate the ability of a lifecycle model with heterogeneous income processes and Bayesian learning to reproduce the first two moments of log consumption data from the CEX. We focus on dovetailing two papers by Gourinchas and Parker [?] (henceforth referred to as G&P) and Guvenen [?] that propose two different models of consumption over the lifecycle. The models of consumer decisions in the two papers are nearly identical, but they differ in the way they treat the income process. G&P models the log income process with a traditional combination of a fully observed persistent AR(1) shock (in G&P , it is a unit root, that is, permanent) and a transitory shock. In their model, every individual has the same income process, and heterogeneity is generated by different income draws, particularly of the unit root process. Guvenen analyzes a model in which each individual has a different income process, indexed by a single parameter that controls income growth, so that within-cohort heterogeneity is created not only by different draws of the error terms, but also by differences between individuals. In order to make this realistic, the agents in Guvenen do not observe their parameter perfectly, but learn about it in a Bayesian way throughout their life. The model is motivated by a need to account for the way in-cohort consumption and income variance changes over the lifecycle. In particular, in American data income variance is dramatically increasing and has a non-concave shape [**?**]. Though the magnitude of the increase can be accounted for with extremely persistent AR shocks , the shape of the profile cannot be reproduced with the type of income process in G&P [**?**]. The Guvenen model does not precisely nest the G&P model, since in Guvenen the exact magnitude of the permanent shock, like the growth parameter, is not observed. However, it can be thought of as having more descriptive power than the G&P model.

G&P estimate the parameters in their model – in particular, the future discounting rate and risk aversion – by matching the evolution of within-cohort average consumption over the lifecycle. Guvenen, in contrast, estimates the parameters of his income process by using lagged autocovariances in panel income data to match the moments of his theoretical income process [**?**]. In , another paper, he also uses CEX consumption data to estimate the amount of learning that must take place to account for the variance in the data[**?**]. However, he does not try to match the precise shape of the mean and variance of consumption as is done in G&P. This paper will attempt to do that.

In particular, we will find, under various configurations, the parameters of the model that minimizes the quadratic distance between the first and second (non central) moments of the model's and the CEX's consumption over a working lifecycle. [1] Since we have taken a number of shortcuts (in particular, we have used ad-hoc rather than data-driven approaches to modeling the retirement function), and had a number of computational difficulties, including estimated parameters on the boundary of the domain, possible local minima, and a relatively coarse grid, this estimation should be considered more of an exploration of the descriptive powers of the model rather than estimates of any "true" underlying parameters. That is, this should be considered a first step towards a more robust estimation.

[1] Note that we will model the second non central moment, not variance, in order to avoid the theoretical difficulty of having a sample mean inside the simulated moment. However, since the mean and second non-central moment imply a particular value for the variance, we are in fact matching the variance indirectly.

2 Theoretical Framework

2.1 Income Process

Following [?], our income process is defined as

$$
\begin{aligned}
&y_{it} = ln(Y_{it}) = g_t + z_{it} + u_{it} + \beta_i t + \omega_i \\
&z_{it} = \alpha z_{i(t-1)} + v_{it} \\
&\quad g_t \text{ is a common deterministic time-dependent expected income growth} \\
&\quad u_{it} \text{ and } v_{it} \text{ are classical Gaussian errors with variances } \sigma_u^2 \text{ and } \sigma_v^2 \text{ respectively} \\
&\quad \beta_i \text{ is a quantity fixed throughout the life for each individual} \\
&\quad \omega_i \text{ is a (possibly) idiosyncratic income scale factor}
\end{aligned}
\tag{1}
$$

Throughout the paper we will suppress the individual indices i when the meaning is clear. We will also denote logs of behavior variables (consumption, income, and cash-on-hand variables) with lower-case letters and levels with upper-case. Within this equation, the individual observes only y_t, ω_i, and g_t. From these observations, the individual makes inference about the values of z_t and β_i in a Bayesian fashion, using this inference to calculate expectations of functions of his future income. Note that Guvenen [?] treats ω_i as unknown, though individuals learn to a high degree of accuracy within a few periods. Since including it adds another state variable without adding much to the descriptive power of the model, I will treat ω_i as known. However, I retain it to allow myself to match the starting levels of consumption as described in section ??.

As in [?], we will treat this as a Kalman filtering problem, with

$$
\begin{aligned}
&\text{State vector } S_t = \begin{pmatrix} \beta_t \\ z_t \end{pmatrix} \\
&\text{Evolution equation } S_{t+1} = \begin{pmatrix} 1 & 0 \\ 0 & \alpha \end{pmatrix} S_t + \begin{pmatrix} 0 \\ v_t \end{pmatrix} \\
&\text{Observations } y_t = \begin{pmatrix} t & 1 \end{pmatrix} S_t + g_t + \epsilon_t
\end{aligned}
$$

Let $P_{t|t-1}$ and $P_{t|t}$ denote the prior covariance matrices of the observation S_t before and after the income realization y_t is observed respectively. Similarly, let $\hat{S}_{t|t-1}$, $\hat{S}_{t|t}$, $\hat{\beta}_{t|t-1}$, $\hat{\beta}_{t|t}$, $\hat{z}_{t|t-1}$, and $\hat{z}_{t|t}$ be the corresponding estimates of these quantities. The Kalman filter framework then gives the following rule for updating beliefs:

$$
\begin{aligned}
K_t &= P_{t|t-1}\begin{pmatrix} t \\ 1 \end{pmatrix}\left(\begin{pmatrix} t & 1 \end{pmatrix} P_{t|t-1}\begin{pmatrix} t \\ 1 \end{pmatrix} + \sigma_u^2\right)^{-1} \\
\hat{S}_{t|t} &= \hat{S}_{t|t-1} + K_t(y_t - \hat{\beta}_{t|t-1}t - \hat{z}_{t|t-1} - g_t - \omega_i) \\
P_{t|t} &= \left(I - K_t\begin{pmatrix} t & 1 \end{pmatrix}\right) P_{t|t-1}
\end{aligned}
\tag{2}
$$

Given a set of beliefs and having observed the realization y_t, the beliefs about the next period are given by:

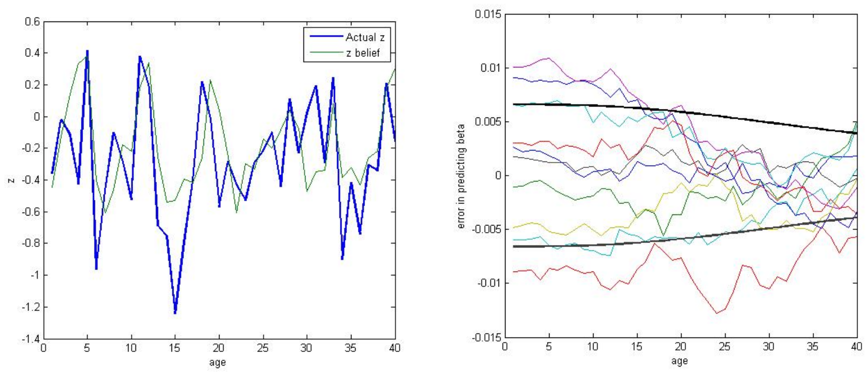
$$
\begin{aligned}
\hat{S}_{t+1|t} &= \begin{pmatrix} 1 & 0 \\ 0 & \alpha \end{pmatrix} \hat{S}_{t|t} = \begin{pmatrix} \hat{\beta}_{t|t} \\ \alpha \hat{z}_{t|t} \end{pmatrix} \\
P_{t+1|t} &= \begin{pmatrix} 1 & 0 \\ 0 & \alpha \end{pmatrix} P_{t|t} \begin{pmatrix} 1 & 0 \\ 0 & \alpha \end{pmatrix} + \begin{pmatrix} 0 \\ \sigma_u^2 \end{pmatrix} \\
y_{t+1} &\sim N\left(\hat{\beta}_{t+1|t}(t+1) + \hat{z}_{t+1|t} + g_{t+1} + \omega_i, \begin{pmatrix} t & 1 \end{pmatrix} P_{t+1|t} \begin{pmatrix} t \\ 1 \end{pmatrix} + \sigma_u^2\right)
\end{aligned}
\tag{3}
$$

Note that uncertainty and learning about the income process affect behavior an individual agent through the distribution of y_t – that is, through the expectation term in the Euler equation **??** below. Also note that having observed this period's income realization, one is able to specify the distribution of the next period's income using only the values of the state variable, $\hat{\beta}_t$ and $\hat{z}_t$. This means that the value function can be described as a function of these two variables without explicit reference to y_t.

Not all individuals have the same β_i, and it is this, along with the persistence of the shocks to z_{it}, that creates substantial increasing income heterogeneity within cohorts over time. Individuals are assigned a β_i and ω_i at the beginning of their life, which are drawn from Normal distributions with mean zero and standard deviations σ_β and σ_ω respectively. Their initial priors about β_i and z_{i1} are Normal, with the β_i standard deviation some fraction of the actual population variance:

$$P_{1|0} = \begin{pmatrix} (1-\lambda)\sigma_\beta^2 & 0 \\ 0 & \sigma_v^2 \end{pmatrix}$$

The parameter λ can then be used to control the degree to which income uncertainty about β_i early in life plays a role in decision making. The means of the initial prior for β_i were set to the true β_i plus a mean zero random error with variance $(1-\lambda)\sigma_\beta^2$. The mean of the prior for z_{i1} is zero.



(a) A Typical Individual Predicting z_t

(b) Ten Typical Individuals Learning β_i (thick lines are one standard deviation of the prior)

Figure 1: Sample Numerical Results

2.2 Lifecycle Model

Now that we have an income process defined, we can turn to the lifecycle modeling. The individual solves the optimization problem:

$$V_t(X_t, \hat{\beta}_{t|t}, \hat{z}_{t|t}) =$$

$$\max_{C_t} \left\{ v_t u(C_t) + \delta E_t \left(V_{t+1}(X_{t+1}, \hat{\beta}_{t+1|t+1}, \hat{z}_{t+1|t+1}) \right) \right\} \qquad \text{for } t \leq T$$

Subject to

$$X_{t+1} = R(X_t - C_t) + Y_{t+1}$$

$$X_{t+1} \geq 0$$

The expectation is taken with respect to the present beliefs about the income process, which affects the future values Y_{t+1}, $\hat{z}_{t+1|t+1}$, and $\hat{\beta}_{t+1|t+1}$. The variable v_t is known a time-dependent taste shifter, assumed here to be a function of family size alone. Again, as explained above, the individuals decision depends only on the assets on hand and the distribution of the future state variables. Since the distribution of Y_{t+1} can be fully characterized by $\hat{\beta}_{t|t}$ and $\hat{z}_{t|t}$, $\hat{\beta}_{t+1|t}$ and $\hat{z}_{t+1|t}$ can be derived from $\hat{\beta}_{t|t}$ and $\hat{z}_{t|t}$ alone, and since a particular realization of Y_{t+1} combined with $\hat{\beta}_{t+1|t}$ and $\hat{z}_{t+1|t}$ gives a value for $\hat{\beta}_{t+1|t+1}$ and $\hat{z}_{t+1|t+1}$, only X_t, $\hat{\beta}_{t|t}$, and $\hat{z}_{t|t}$ need to be included as state variables. These value functions need to be solved on a three dimensional state grid for each time step, as described in section **??**. Throughout, we will assume CRRA utility with parameter ρ.

Rather than try to simulate retirement in detail, we assume, like G&P, that the individual retires at time $T+1$, at which point they have the value function

$$V_{T+1}(X_{T+1}, z_T, \beta) = kv_{T+1}\left(X_{T+1} + \gamma P_T\right)^{1-\rho}$$
$$x_{T+1} = R(X_T - C_T)$$
$$P_T = \left(G_T Z_T \exp\left[\beta(T+1) + \omega\right]\right)$$

Here, P_T represents the true permanent income shock in period T. Unlike in Gourinchas and Parker, the individual does not know exactly what part of their observed income is permanent and what is transitory (though by the end of their life they may have an accurate estimate). Consequently, the expectation does not disappear in the value function for time T, as it it taken over the individual's priors.

Since the retirement value function is fully specified, we can use backwards induction to solve the consumption problem for each period given a particular income belief at that period. Let

$$f_t(X_t) = f_t(X_t, \hat{\beta}_{t|t}, \hat{z}_{t|t}) = C_t$$

For simplicity we suppress the other arguments, it being understood that the consumption problem is solved for a particular belief about the income process. Since the utility function is concave and differentiable, we can use the first-order conditions to solve for the consumption rule. Neglecting the liquidity constraint for the moment, the decision rule for the last period must satisfy:

$$f_T(X_T)^{-\rho} = \delta R\kappa(1-\rho)\frac{v_{T+1}}{v_T}E_T\left[(R(X_T - f_T(X_T)) + \gamma P_T))^{-\rho}\right]$$

Since the left hand side is decreasing in f_T, the right hand side is increasing in f_T, and as f_T approaches zero the left side exceeds the right, there is a unique solution to this problem, which we will call $\tilde{c}_T = f_T(x_T)$. Note that the expectation is taken with respect to the income beliefs – since it is not known with precision at time T what the value of z_T and β are in that period, it is not known exactly what proportion of their income will be available for retirement, and it is over these prior beliefs that the expectation is taken.

Once a consumption rule is derived for a period $t+1$, it can be derived for the preceding period t using the first order conditions and the Envelope Theorem in a similar manner:

$$f_t(X_t)^{-\rho} = \delta R \frac{v_{t+1}}{v_t} E_t[(f_{t+1}(R(X_t - f_t(X_t)) + Y_{t+1}, \hat{\beta}_{t+1|t+1}, \hat{z}_{t+1|t+1})^{-\rho}] \tag{4}$$

Here the prime denotes the derivative with respect to the first argument. Since different beliefs about the income process result in different consumption rules, this must be taken into account when evaluating the expectation. As before, this equation yields a unique solution, $\tilde{c}_t = f_t(x_t)$. As usual, the non-borrowing constraint results in the consumption rule:

$$C_t = min\left\{X_t, \tilde{C}_t\right\}$$
$$1 \leq t \leq T$$

Note that the final decision rule is homogeneous, that is, for any A:

$$AC_T = Af_T(X_T, \hat{z}_t, \hat{\beta}) = f_T(AX_T, \hat{z}_t + \ln A, \hat{\beta})$$

This follows from standard arguments as well as the observation that

$$\begin{aligned}\pi(AZ_T = \eta | A\hat{Z}_T = \zeta) &= \pi(z_T = \ln\eta - \ln A | \hat{z}_T = \ln\zeta - \ln A)\\ &= \pi(z_T = \ln\eta | \hat{z}_T = \ln\zeta)\\ &= \pi(Z_T = \eta | \hat{Z}_T = \zeta)\end{aligned}$$

By π we denote the prior of Z_T, and the second line follows from the fact that $\ln z_t | \hat{z}_t$ is distributed normally with mean $\hat{z}_t$ and variance independent of $\hat{z}_t$. Since the last period's decision rule is homogeneous, all the other decision rules are homogeneous as well by induction. Consequently, the problem is independent of multiplicative scale, and can be solved for every individual treating $\omega_i = 0$ in the income generating process and then scaling the consumption and money variables by $\exp(\omega_i)$ afterwards.

2.3 Numerical Solution

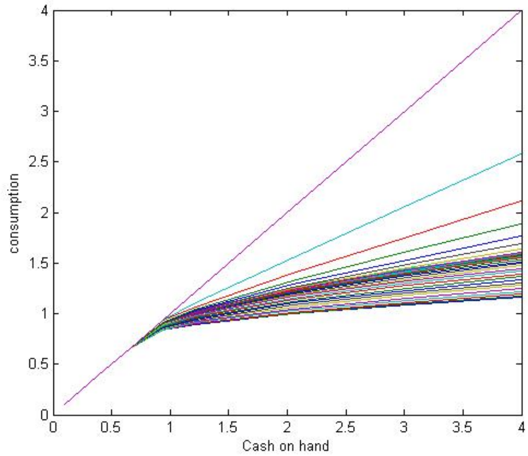
Since the value function has three state variables, we must numerically solve the consumption problem on a three-dimensional grid for every time step. We used Matlab for this purpose. As described above, we first solve the last period's consumption function, and then proceed with backwards induction. (Note that the covariance matrix P does not depend on the data and can be generated entirely in advance.) Given a consumption rule for the period $t+1$, for each point in the grid, we perform the following steps:

1. Given $\hat{\beta}_{t|t}$, $\hat{z}_{t|t}$, use (**??**) generate a grid of possible y_{t+1} with their corresponding probabilities. We used a grid of ten points spaced over three standard deviations of y_{t+1}.
2. Using (**??**), find $\hat{\beta}_{t+1|t}$ $\hat{z}_{t+1|t}$
3. For each possible y_{t+1}, use $\hat{\beta}_{t+1|t}$ and $\hat{z}_{t+1|t}$ to find the corresponding $\hat{\beta}_{t+1|t+1}$ and $\hat{z}_{t+1|t+1}$.
4. These, with an interpolation / extrapolation routine, can then be used to estimate the expectation in (**??**).
5. Use these expectations to solve numerically for the consumption rule.

Since the problem is three dimensional and we are solving for 40 time steps, there are a large number of grid points at which a decision rule must be calculated. Moreover, to solve the nonlinear equation, the interpolation / extrapolation routine must be executed many times for each grid point, and the built-in Matlab routine was found to be too slow to solve the problem in a reasonable amount of time. (Depending on the grid size, using the built-in interpolation routine, Matlab took about fifteen minutes to solve for the decision rules.) Consequently, we wrote a compiled routine in C that performs tri-linear interpolation and linear extrapolation on a three-dimensional grid which sped up the solution of the problem an order of magnitude. Extrapolation was performed by tracing a line from the extrapolation point to the nearest point of the cubic grid on which the function was defined, using interpolation to evaluate the function on the edge point and a small distance inside the grid, using the difference to estimate a directional derivative, and then using this derivative to extrapolate linearly to the extrapolation point. This routine was found to perform reasonably well small distances from the grid with a known function. However, as shown in section **??** (for example, figure **??**) , there may exist significant nonlinearities in the consumption rules, and so a more robust extrapolation routine might be a improvement to our method.

Once the decision rule has been calculated, a large number (we generally used 50 000) of incomes and true β_i are simulated according to (**??**), and, using these, their income beliefs are updated according to (**??**). (See section **??** for discussion of ω_i – suffice to say here that when we generate the moments, we treat ω_i as zero.) We also assume that initial cash-on-hand is distributed lognormally with mean μ_X and variance σ_X^2. These incomes and beliefs are used chronologically to find consumption and cash-on-hand for each period given the consumption rule. The result is a collection of consumption profiles whose moments can be constructed and compared with the CEX data as described in section **??**. Note that in order to calculate derivatives of individual consumption data with respect to income-generating parameters, necessary for estimation of the variance shown in equation (**??**), the same random seed was used for every income generation.

Figure 2: An Example of Decision Rules for Different Ages (here, everything is consumed at retirement)



2.4 Numerical Minimization

The MSM method described in section **??** requires minimization of an objective function, each iteration of which requires the above lifecycle model to be solved. The solution described in section **??** requires about fifty seconds to generate consumption rules and simulate lives even on a grid of size only 8x5x5 (in the X, β, and z dimensions, respectively). Because we are minimizing the criterion with respect to a relatively large number of variables, there are likely to be local minima (as described below), and discretization can mean apparent discontinuities (resulting in poor derivative estimates when using finite difference methods), standard quasi-Newton approaches were found to be inappropriate. Consequently, we adopted a two-tiered approach. First, the state space was explored randomly to give an idea of a starting point for a numerical optimizer. Then that starting point was put into a derivative-free simplex algorithm which then search the neighborhood of that point for a local minimum. Throughout, the state space was bounded by a hyper cube describing reasonable or economically meaningful values of the parameters. For example, ρ was bounded below by 0.2. This was because the numerical routine had a tendency to drive parameters to infinity or to economically meaningless values, at which the precise minimimum of the objective function is not very interesting for our purposes. If the parameters were interior, then the variance in (**??**) could be estimated, though none of our estimates were found to be interior points.

This entire process took many hours of devoted computer time for each experiment, meaning we were not provided with as much opportunity to explore as we would have liked. Preliminary investigation showed that the results in section **??** may not be robust to solving the model on a finer grid size. Unfortunately, we were not able to solve the problem on a grid any finer than 8x5x5 with the resources available to us. A good avenue for further research would be to apply a more robust search algorithm, such as a simulated annealing or genetic algorithm, on a more powerful computer than is available at the LSE.

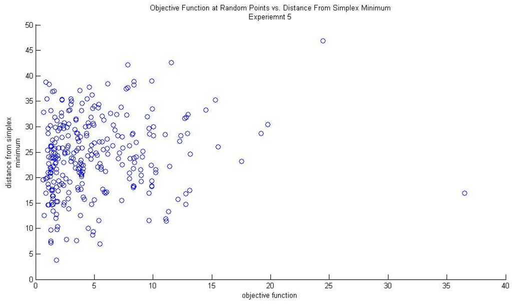
The random search provides some evidence for the existence of local minima, since very different sets of parameter values seem to perform equally well. To illustrate this, consider the following data, which is draws of parameter values from a random distribution and their associated criterion from section **??**. A measure of the "distance" between two parameter sets could be:

$$d_{ij} = \frac{(\beta_i - \beta_j)^2}{\sigma_\beta^2} + \frac{(\rho_i - \rho_j)^2}{\sigma_\rho^2} + ...$$

Here, σ_x represents the standard deviation of the parameter x within the entire sample. (We use this to correct for the different scales of the parameters). One can graph this measure of distance from the best parameter set of the random sample as a function of the goodness of fit.

The fact that parameter sets with similar, low criterion values are no closer to one another than those with worse criterion values suggests that there may be local minima. In light of this, the estimates presented in section **??** are best taken with a grain of salt, as, without more time or computing power, it was not possible to make a convincing case that the estimates we found were the genuine global minima.

Figure 3: Evidence For Local Minima



3 CEX Data

3.1 Regression

We took all the data for this project from the National Bureau of Economic Research (NBER) family- and member-level extracts of raw data from the CEX. These data are freely available on the web at `www.nber.org/data/ces_cbo.` In these extracts, for each quarter of each year from 1980 through 2003, the NBER reports the equivalent of yearly income and consumption data for family units which began the CEX survey in that year and quarter. It also reports demographic information about each member of each family unit. Because of a change of sampling frame, data were unavailable for the third and fourth quarters of 1985 and 1995.

Since our model does not treat non durables, we defined consumption as follows (for detailed descriptions of the variable names, see the NBER file Cexfam.pdf available at the above website):

$$\begin{aligned} cons = & food + nonfood + clothing + perscare + \\ & books + pubs + recsport + othrec + gambling + charity + \\ & gasoline + tolls + masstran + othtrans + airfare \end{aligned}$$

We also used income data from the CEX to estimate g_t, the expected income growth rate. Our definition of consumable income is:

$$cinc = income - (expenditure - cons) - socsec - pension - ssi - taxplus$$

We subtracted durable expenditure (which was taken as the difference between total expenditure and consumable expenditure as defined above) to get a measure of the money available for the sort of consumption we are modeling. We are assuming that durable and non-durable felicity is separable. Having extracted this data, we followed the data-cleaning methodology of G&P closely, so for details, the reader is referred to that paper. The consumption data in real terms using the seasonally adjusted Gross National Product Implicit Price Deflater, downloaded from the St. Louis Federal Reserve. The four quarterly deflaters were averaged into a single yearly deflater. We obtained yearly unemployment data from the US Bureau of Labor Statistics "Employment and Earnings" dataset.

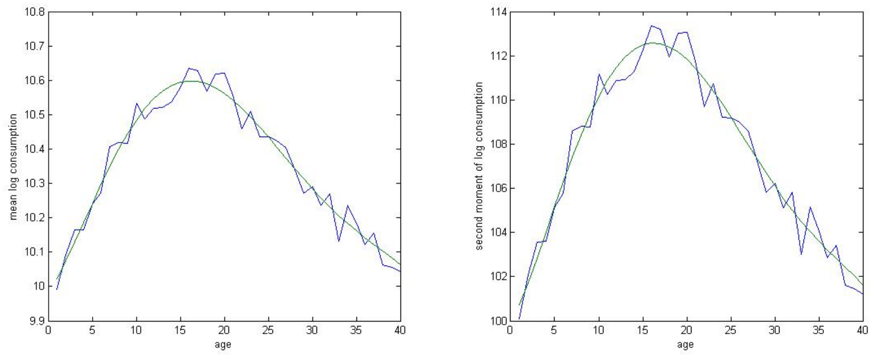
Note that because, unlike G&P, we use an exlplict liquidity constraint, we must assume without further justification that consumption is a multiplicative combination of family effects, cohort effects, age effects, time effects (which are assumed to be captured in the national unemployment rate for that year), retirement effects, and classical error:

$$\begin{aligned}\ln C_{it} = c_{it} =& \pi_1 \times \ln(famsize)_{it} + \pi_2 \times age_{it} + \\ & \pi_3 \times cohort_{it} + \pi_4 \times unemployment_{it} + \pi_5 \times ret_{it} + \epsilon_{it}\end{aligned}$$

Having run this regression and measured estimates $\hat{\pi}_i$, we then removed idiosyncratic cohort, family size, time, and retirement effects by evaluating at within-age averages (except for the retirement, which is simply removed):

$$\begin{aligned}\hat{c}_{it} =& \hat{\pi}_1 \times \overline{\log(famsize)}_t + \hat{\pi}_2 \times age_{it} + \\ & \pi_4 \times \overline{unemployment}_t + \hat{\epsilon}_{it}\end{aligned}$$

Having calculated this, the average and second moment were calculated for each age and smoothed with a fifth-order polynomial. Averages were weighted with the *adjwt* variable.



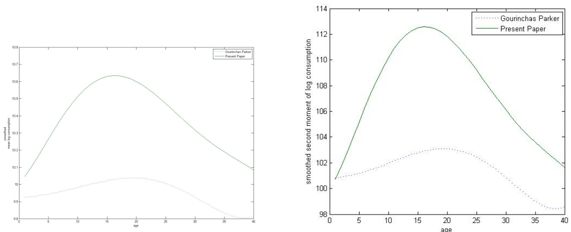
(a) Mean Consumption From CEX Data (Smoothed and Un-smoothed)
(b) Second Moment of Consumption From CEX Data (Smoothed and Un-smoothed)

Figure 4: Sample Numerical Results

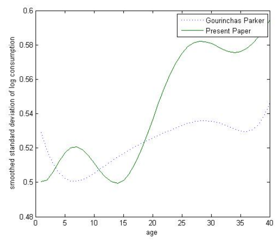
The parameters g_t and v_t were calculated as in G&P.

3.2 Differences Between Our Results and Other Data Sets

Unfortunately, we were unable to acquire detailed data from G&P for comparison. However, we were able to get a version of the above $\hat{c}_{it}$ (though the provider was not sure exactly what stage of the regression it was from). Performing the same averaging and smoothing as was performed on our data allows us to compare our target moments:



(a) Mean
(b) Second Moments



(c) Standard Deviation

Figure 5: Our Estimates vs. Gourinchas and Parker

Note that the graphs in figure **??** are in logs (which we targeted with MSM), whereas the figures in G&P's paper are in levels. It is clear that there is substantially more time variation in our moments than in Gourinchas and Parker, though it is hard to say exactly why that might be without more detailed access to their data set. Creating a more dramatic hump shape is probably a more difficult task for the lifecycle model. Further exploration of the discrepancy between our data sets is warranted.

More significantly, our measure of consumption variance is very different from the measure in Deaton and Paxson, on which Gourinchas's estimates are based. Whereas Deaton and Paxson[**?**] find an increase in consumption standard deviation from the age of 25 to the age of 65 of approximately 0.2 (based on their Figure 4), we find an increase of only about 0.1. This is almost certainly due to the fact that our data controls for family size whereas Deaton and Paxson do not. If family size is left out of the regression, we measure an increase in standard deviation of about 0.18, which is more consistent with their estimates. This has significant implications for our results, particularly in Experiment Three and Experiment Four, as described in section **??**.

4 Estimation

4.1 Five Experiments

According to the model developed in section **??**, the simulated moments depend on the following variables:

Parameters estimated only in a first stage:

v_t The taste shifters

g_t The expected income growth

σ_X The standard deviation of the log initial cash endowment

μ_X The mean of the log initial cash endowment

R The real interest rate

Parameters estimated only by MSM:

λ The proportion of the initial β variance known at the start of the life

γ The proportion of the last period's permanent income available at retirement

κ The proportionality constant in the retirement value function

Parameters that could be estimated either with MSM or in a first stage:

α The AR coefficient of the persistent shock

σ_u The standard deviation of the persistent shock

σ_v The standard deviation of the transitory shock

σ_β The standard deviation of the population distribution of β

ρ The CRRA coefficient

δ The discounting factor

We estimated the parameters v_t and g_t as described in section **??**. The remainder of the parameters from the first section were set equal to the estimates of G&P [**?**]:

σ_X=1.784

μ_X=-2.794

R=1.0344

We always treated the parameters in the second category as free variables in the MSM estimation. This is largely because they lack intuitive economic meaning outside of the context of matching consumption profiles. It is true that Guvenen [**?**] estimated λ from a different set of consumption data, given the parameters of the income distribution, and that G&P estimated parameters for another rather arbitrary retirement consumption rule. However, since this estimation is qualitatively different from those settings, and there is no real economic intuition to suggest that these parameters should be the same across different models, we will not use those previous estimates.

The third category of parameters could be estimated either from MSM in this context or taken from previous studies. We performed five different experiments with these: taking G&P's estimation of δ and ρ as given, taking Guvenen's estimation of σ_u, σ_v, α, and σ_β as given, using the HIP and RIP parameters from Guvenen [**?**], and estimating all of the parameters simultaneously with MSM.

Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5
α=?	α=?	α=0.988	α=0.821	α=?
σ_u^2=?	σ_u^2=?	σ_u^2=0.061	σ_u^2=0.047	σ_u^2=?
σ_v^2=?	σ_v^2=?	σ_v^2=0.015	σ_v^2=0.029	σ_v^2=?
σ_β^2=?	σ_β^2=?	σ_β^2=0	σ_β^2=0.00038	σ_β^2=?
ρ=0.5140	ρ=2	ρ=?	ρ=?	ρ=?
δ=0.9598	δ=0.965	δ=?	δ=?	δ=?

G&P estimate ρ and δ to match the moments of consumption, the precise solution of which was certainly dependent on their measurement of σ_u, σ_v, and α. In this sense, the justification for Experiment One is a little unclear. However, these parameter estimates are intended to have economic meaning that extends beyond the model – that is, they are supposed to be a reasonable measurement of "true" rates of future discounting and risk aversion. If this is the case, then it is reasonable to ask whether or not the details of the income process (and our above ad-hoc parameters) can alone match the empirical data if these are taken as given. It should be noted that there is another important difference between G&P 's model and this: even absent income process heterogeneity, agents here do not observe their AR shock perfectly, whereas in G&P they do.

Experiment two is similar to experiment one, except for δ and ρ we use the values assumed in Guvenen[**?**]. This is comparable to the results in his paper, except instead of matching panel income data we match pseudo-panel consumption data. It is interesting to see how these two estimates differ.

Experiments three and four attempt to estimate ρ and δ from the income parameters. The estimates in Guvenen[**?**] are made from a different data set (the PSID) and certainly do not depend on any estimates of δ or ρ. If these are indeed the "true" income-generating parameters, then it is reasonable to expect that by changing other, unrelated parameters of consumption behavior, one should be able match other sets of empirical data. Experiment three uses the values from the RIP (restricted income profile) without β heterogeneity. Experiment four uses Guvenen's HIP (heterogeneous income profile) estimates. This is directly analogous to the first-stage process used in G&P [**?**] to estimate the income parameters.

Experiment five is simply a measure of what the data suggests for all these parameters if one is only trying to account for the shape of the mean and standard deviation of consumption. The accuracy of the fit in experiment three is also a measure of the descriptive power of the model.

4.2 Method of Simulated Moments

We used the Method of Simulated Moments (MSM) to match the first two (non central) moments of the data. We define

$$\mu_t(\psi) = E[c_{it}|\psi]$$
$$\sigma_t(\psi)^2 = E[c_{it}^2|\psi]$$

where ψ represents the vector of model parameters and, as before, lower case denotes logs, that is, $\ln(C_{it}) = c_{it}$. This is an unconditional expectation, meaning that it is taken over the state of all individuals in a cohort at a particular age. We assume that the first $2 + \eta$ moments of this unconditional distribution exist for some $\eta > 0$ so that a law of large numbers applies.

Assuming that observations from the CEX are draws from this distribution plus mean zero, uniformly integrable (after being squared) observation errors, uncorrelated with the observed consumption, we have that

$$\hat{\mu}_t = \frac{1}{n_t}\sum_{i=1}^{n_t}(\hat{c}_{it} + \epsilon_{it}) = \frac{1}{n_t}\sum_{i=1}^{n_t} c_{it} + o_p(1) \rightarrow \mu_t(\psi_0)$$
$$\hat{\sigma}_t^2 = \frac{1}{n_t}\sum_{i=1}^{n_t}(\hat{c}_{it} + \epsilon_{it})^2 = \frac{1}{n_t}\sum_{i=1}^{n_t} c_{it}^2 + \frac{1}{n_t}\sum_{i=1}^{n_t} \epsilon_{it}^2 + o_p(1) \rightarrow \sigma_t(\psi_0)^2 + \sigma_u^2$$

Here, n_t is the number of individuals observed at time t and ψ_0 is the true value of the parameters. Note that our estimate of the second moment from the CEX data will be biased upwards by the variance of the observation error. We will assume that this error is small relative to the variance in consumption itself. It would be desirable to also estimate the differences rather than levels of the variance, and this would be a valuable future task.

Although we don't know the exact distribution of S_{it}, given a candidate set of parameters, $\tilde{\psi}$, we can use the numerical model described above to simulate a large number of lifecycles and their corresponding consumptions,

$\tilde{c}_{it}$, and calculate their sample moments:

$$\tilde{\mu}_t = \frac{1}{N}\sum_{i=1}^{N} \tilde{c}_{it} =\longrightarrow \mu_t(\tilde{\psi})$$
$$\tilde{\sigma}_t^2 = \frac{1}{N}\sum_{i=1}^{N} \tilde{c}_{it}^2 \to \sigma_t(\tilde{\psi})^2$$

Here, N is the number of individuals simulated. Using the observation from section **??** that the model is homogeneous and unaffected by scale, we then used treated ω_i as the same across all individuals and set is so that the mean in the first period of the simulated and CEX data were the same.

Given these definitions, ψ was found to minimize

$$\text{criterion} = \begin{pmatrix} (\tilde{\mu}-\hat{\mu})' & (\tilde{\sigma}^2-\hat{\sigma}^2)' \end{pmatrix} W \begin{pmatrix} (\tilde{\mu}-\hat{\mu}) \\ (\tilde{\sigma}^2-\hat{\sigma}^2) \end{pmatrix}$$

Under regularity conditions discussed in Pakes and Pollard [?], this estimator is consistent with covariance matrix

$$\begin{aligned} &(G_\theta' W G_\theta)^{-1} G_\theta' W \Omega^{-1} W G_\theta (G_\theta' W G_\theta)^{-1} \\ &\text{where} \\ &G_\theta' = E\left[\frac{\partial}{\partial \theta'}\zeta_i\right] \\ &\Omega = E\left[\zeta_i \zeta_i'\right] \\ &\zeta_i = \begin{pmatrix} \tilde{c}_{i1}-\hat{c}_{i1} \\ \vdots \\ \tilde{c}_{iT}-\hat{c}_{iT} \\ \tilde{c}_{i1}^2-\hat{c}_{i1}^2 \\ \vdots \\ \tilde{c}_{iT}^2-\hat{c}_{iT}^2 \end{pmatrix} \end{aligned} \tag{5}$$

Each of these terms were estimated with sample analogues using first differencing to estimate the derivatives.

For a weighting matrix, we used the following:

$$W = \begin{pmatrix} (\max(\hat{\mu}) - \min(\hat{\mu}))^{-2} I_{40} & 0 \\ 0 & (\max(\hat{\sigma}^2) - \min(\hat{\sigma}^2))^{-2} I_{40} \end{pmatrix}$$

Here, I_{40} denotes the forty-by-forty identity matrix. Our rationale for this was that the scales of the consumption mean and variance were quite different, and we didn't want that to cause the model to preferentially match one set of moments over the other. We did not attempt to generate a more efficient weighting matrix, nor to incorporate the first-stage errors into the covariance matrix. These may be fruitful avenues for further research.

5 Results

The results of the experiments are as follows:

Variable	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5
δ	**0.9598**	**0.965**	*1.000*	0.968	0.972
ρ	**0.514**	**2.00**	1.278	*0.2*	0.451
σ_β	*0.000*	0.012	**0.000**	**0.019**	0.011
σ^2_β	*0.000*	1.475E-04	**0.000**	**3.61E-04**	1.210E-04
λ	*0.000*	*0.000*	0.273	0.011	0.221
γ	*0.000*	*0.000*	*0.000*	14.851	4.097
α	0.561	0.161	**0.988**	**0.821**	0.617
σ_u	0.405	0.483	**0.247**	**0.216**	*0.001*
σ^2_u	0.164	0.233	**0.0610**	**0.047**	*0.000*
σ_v	0.600	*0.001*	**0.122**	**0.170**	0.512
σ^2_v	0.360	*0.000*	**0.015**	**0.029**	0.262
k	7.452	*50.0*	39.224	*0.500*	9.292
criterion	0.174	0.438	0.128	4.820	0.013

The variances are redundant, but are included for comparing with other papers. Boldfaced numbers were fixed throughout that experiment, and numbers in italics were on the boundary of the sample space. We did not generally calculate covariance matrices when variables were on the boundary, as the standard asymptotic theory is not valid in that case. However, for interest, we did calculate the (inappropriate) estimator for Experiment Five.

These were minimized from the following starting values:

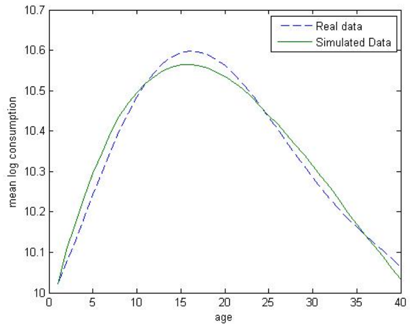
Variable	Exp. 1	Exp. 2	Exp. 3	Exp. 4	Exp. 5
δ_0	0.960	0.965	0.968	0.960	0.968
ρ_0	0.009	2.000	0.200	15.455	0.482
$\sigma_{\beta 0}$	0.030	0.008	0.019	0.019	0.010
λ_0	0.500	0.104	0.011	0.000	0.750
γ_0	18.615	15.000	14.851	24.554	5.451
α_0	0.462	0.391	0.821	0.821	0.467
σ_{u0}	0.128	0.517	0.217	0.217	0.491
σ_{v0}	0.190	0.009	0.170	0.170	0.380
k_0	2.000	10.000	0.500	6.314	10.122

Note that since the first and second non-central moments have such a similar shape and weight, it should not be surprising that many of the fits looks similar for both graphs. However, the variance depends on the relationship between these two quantities in a way that may not be immediately obvious upon inspection. Consequently, we report the implied variance for each of the estimates although it was not itself fit explicitly. It is worth noting that the fit does not appear to improve if the variance itself, rather than the non central moment, is matched directly in the optimization.

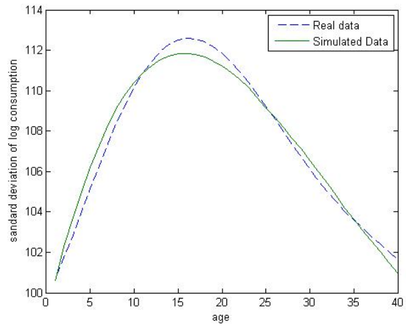
In general, the estimates are not realistic. Consequently, we reiterate that this is best understood as an exploration of the expressive power of the model rather than an attempt to discern the "true" underlying parameters of the American consumer. Even if estimation were the goal, every minimization had at least one variable on the boundary of the sample space, making standard asymptotic theory invalid and depriving us of covariance estimates. More accurate estimation would above all require more computing power and a data-driven approach to the retirement function.

5.1 Experiments One and Two

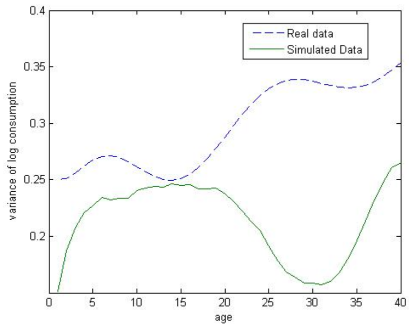
Experiments one and two are interesting to compare and contrast. The main difference between then was that the risk aversion factor was larger in the second case. Recall that the first corresponds to estimates obtained by G&P in [?] and the second to values assumed by Guvenen [?]. As can be seen in figures **??** and **??**, the values assumed by Guvenen were not capable of matching the downwards turn in consumption at the end of the life. However, the resulting variance more closely resembles the shape found in Carroll et. al [?], which is sensible since Guvenen was trying to match this variance without considering consumption. As observed in section ??, however, the consumption variance from Carrol et. al [?] does not control for family size effects, and so overstates the increase in variance relative to our set. Again, this is consistent with the below results, and results in a worse overall fit for Experiment Two.



(a) First Moments

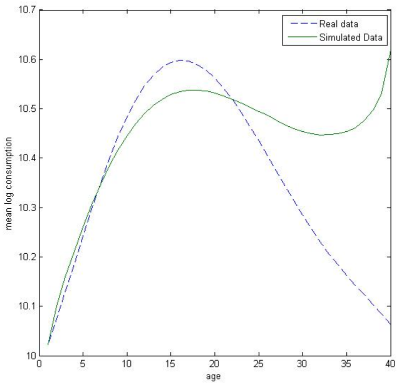


(b) Second Moments

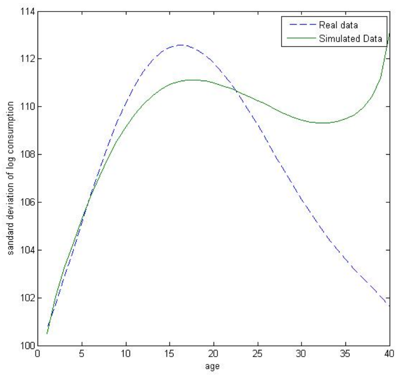


(c) Variance

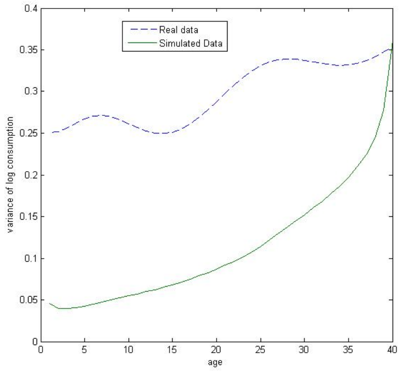
Figure 6: Experiment One Fit



(a) First Moments

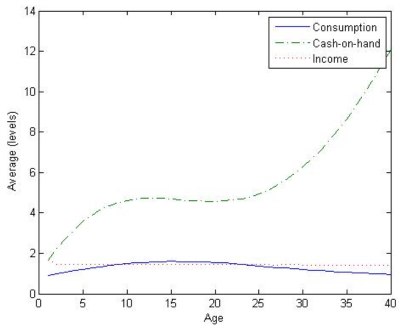


(b) Second Moments

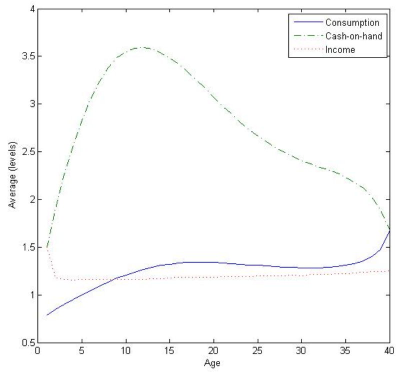


(c) Variance

Figure 7: Experiment Two Fit



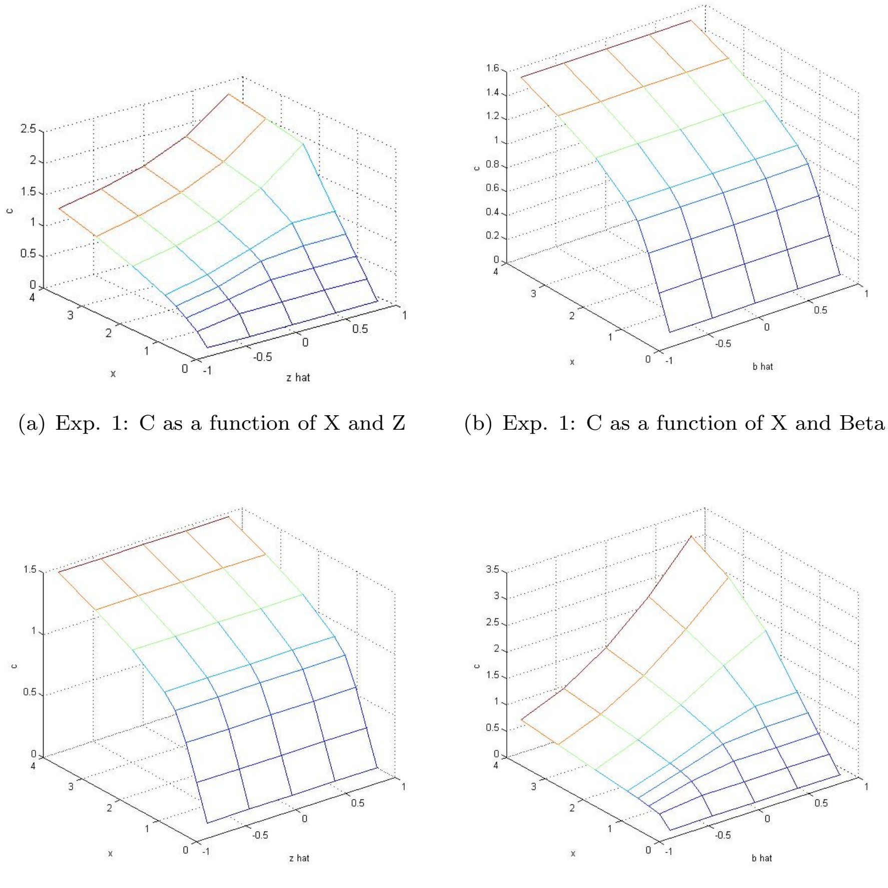
(a) Experiment One



(b) Experiment Two

Figure 8: Comparison of Average Lifecycle Behaviour

It is also interesting to observe that Experiment One results in an unrealistic cash-on-hand behavior although it generates a better fit overall. This is driven by the very different kinds of uncertainty faced by individuals in Experiment One and Experiment Two. Note that in Experiment One, uncertainty is dominated by the persistent shocks, which have a higher AR coefficient and (unrealistically) large variances. Permanent shocks are not necessarily learned with increasing accuracy over time, resulting in substantial uncertainty for the agent. However, in Experiment Two, the permanent shocks were eliminated altogether, and $\hat{\beta}_i$ was learned fairly well over the course of the life. These facts are evident in the decision rules for the two experiments (figure ??), as well as the time-varying histogram of the error of $\hat{\beta}_i$ prediction in Experiment Two (figure ??).



(a) Exp. 1: C as a function of X and Z

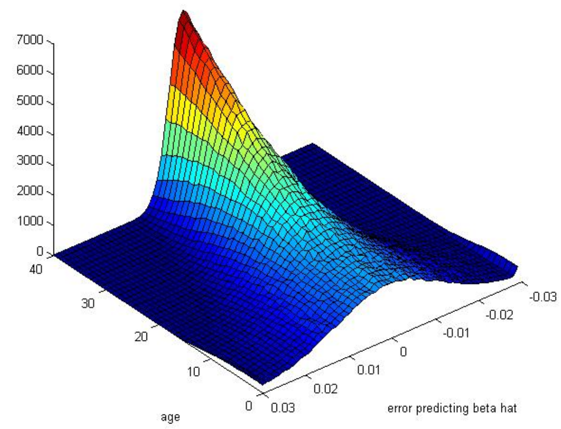
(b) Exp. 1: C as a function of X and Beta

(c) Exp. 2: C as a function of X and Z

(d) Exp. 2: C as a function of X and Beta

Figure 9: Two Different Sources of Uncertainty (at age=20)

Figure 10: Exp. 2: Time-varying Histogram of Error Predicting Beta

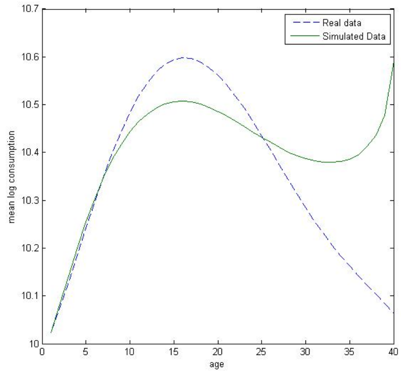


It would be interesting to perform further study with a more extensive survey of starting points to see whether these results are a robust effect of increasing ρ or simply an artifact of different local attractors. At least it illustrates how having a model with two different mechanisms for the same qualitative effect can give very different predictions even for indirectly related parameters (such as the retirement function in this case).

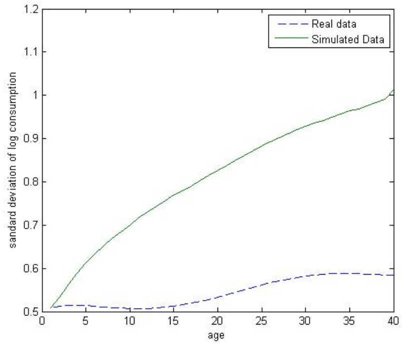
5.2 Experiments Three and Four

Experiments Three and Four treat Guvenen's estimates from [**?**] as a first-stage income process estimation and proceed to estimate the retirement and preference parameters in the style of G&P. Pseudo cross-sectional consumption data is not likely to reveal very much information about the details of an income process, so it is natural to estimate income parameters in a first stage and estimate preferences, which are directly related to consumption, in the second. As such, this is the most revealing test of the ability of the income estimates of [**?**] to accommodate the desired consumption behavior. Experiment Three corresponds to his RIP (restricted income process) estimates without heterogeneity, and Experiment Four to the HIP (heterogeneous income process) estimates.

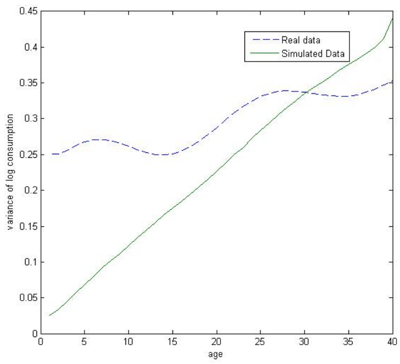
The results are shown in figures **??** and **??**. Note that Experiment Four is the worst fit of all the models by far. It produces too much consumption and too much heterogeneity. As observed above, it happens to be optimal for the model to deviate from the first and second moments the same amount, as they have very similar shapes and are re-scaled by the weighting matrix. Consequently, this should not be interpreted to mean that the HIP profile cannot match the mean consumption, but rather that it would have to do so at the expense of failing to match the second moment. Given the differences between our measure of variance and Guvenen's, as discussed in section ??, this is perhaps not too surprising.



(a) First Moments

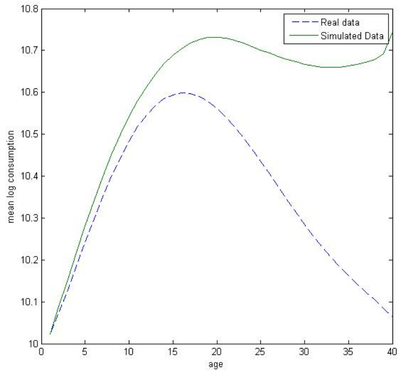


(b) Second Moments

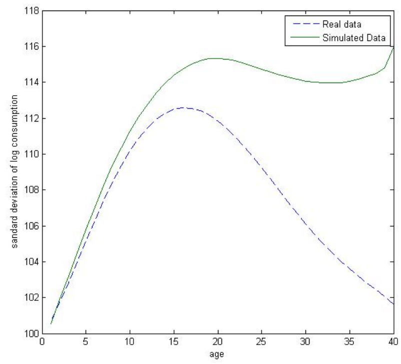


(c) Variance

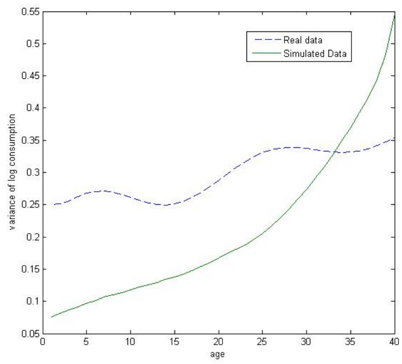
Figure 11: Experiment Three Fit



(a) First Moments

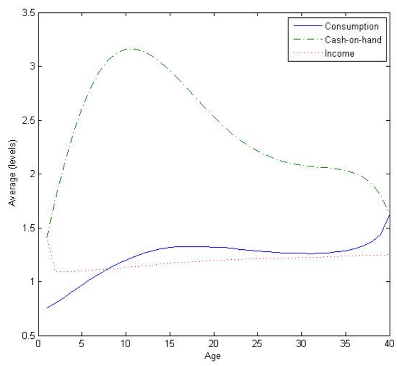


(b) Second Moments

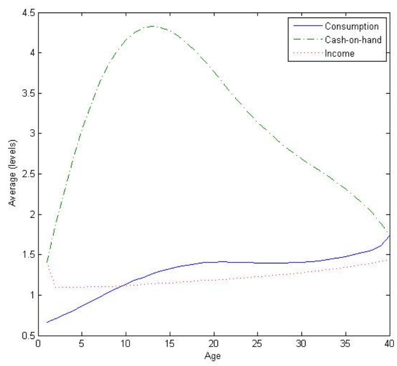


(c) Variance

Figure 12: Experiment Four Fit



(a) Experiment Three



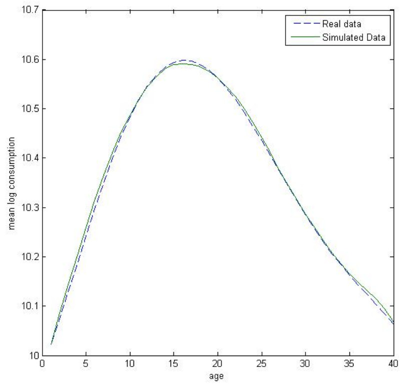
(b) Experiment Four

Figure 13: Comparison of Average Lifecycle Behaviour

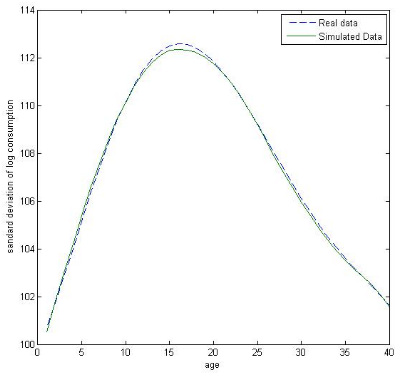
As seen in figure **??**, both models generate an intuitive saving and dissaving pattern. However, both did so at the expense of realistic values for δ or ρ.

5.3 Experiment Five

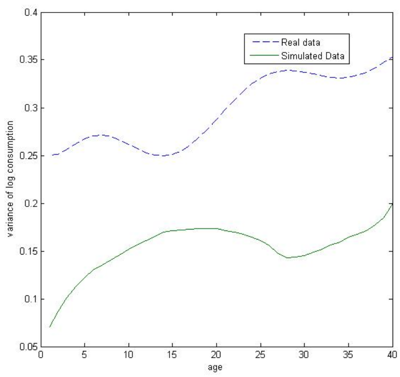
Interestingly, experiment five, which treated all the parameters as free, resulted in the most “realistic” parameter values. The only counterintuitive parameter estimate was the elimination of the transitory shock by setting σ_u^2 to zero. As expected, the fit is also the best among all the models as well. Figure ?? shows the fit, and figure ?? shows the time-varying histogram of consumption.



(a) First Moments



(b) Second Moments



(c) Variance

Figure 14: Experiment Five Fit

Figure 15: Exp. 5: Average Lifecycle Behaviour

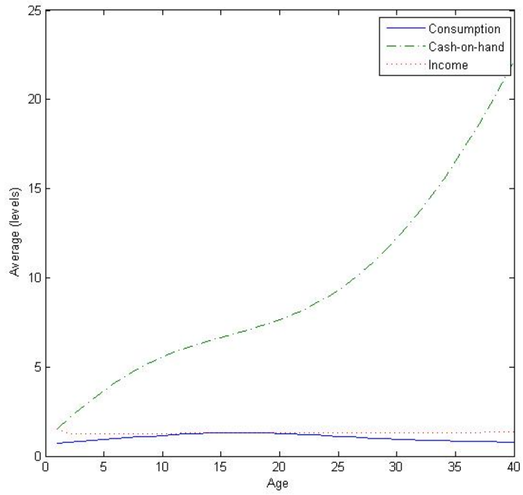
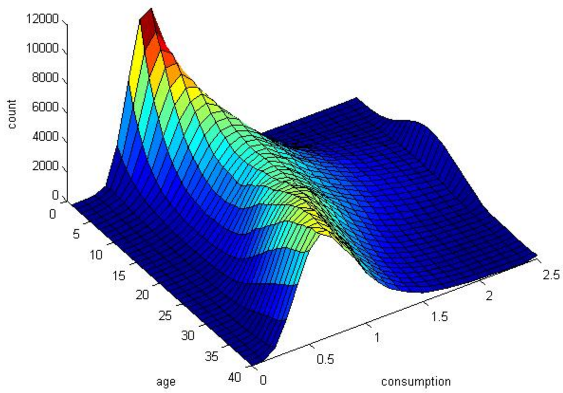
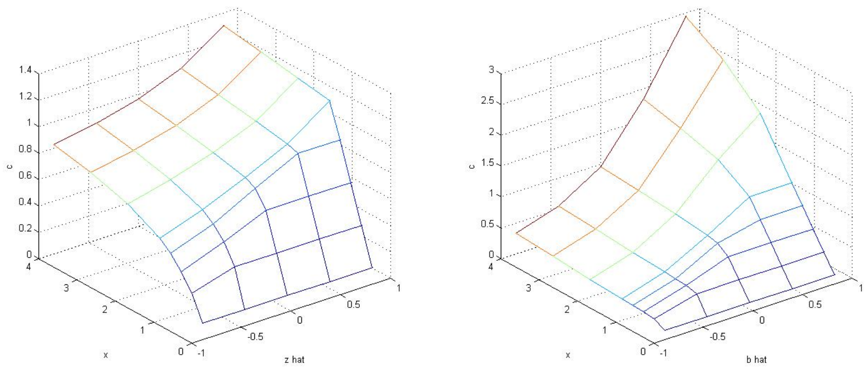


Figure 16: Exp. 5: Time-varying Histogram of Consumption



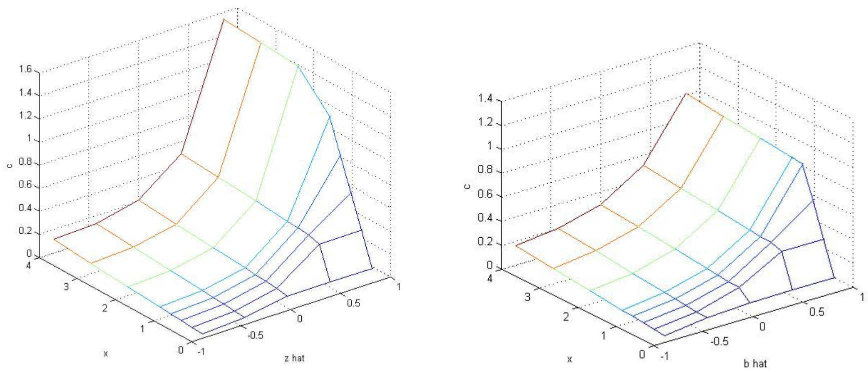
Note that, like Experiment One, it predicts a non-realistic growth in cash-on-hand (figure **??**). This case is more complex than Experiment One, however, since both $\hat{z}_t$ and $\hat{\beta}_t$ influence decisions (figure **??**). This experiment also has a meaningful retirement decision which depends on the individual's state, since the retirement parameters were more realistic then in other experiments (figure **??**). In the other experiments, the final consumption rule was to consume everything. Note that since the AR shock variance is so high, even though the transitory shock is absent, β_i is learned very slowly, since it is dominated by the permanent shock. As the same time, the AR shock is observed relatively accurately (figure **??**).



(a) C as a function of X and Z

(b) C as a function of X and Beta

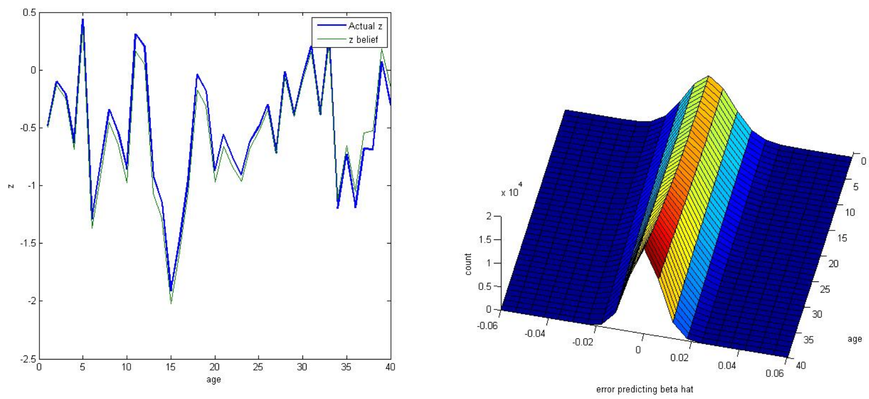
Figure 17: Exp. 5: Consumption Decision Rules (at age=20)



(a) C as a function of X and Z

(b) C as a function of X and Beta

Figure 18: Exp. 5: Consumption Decision Rules (at age=40)



(a) Typical Ability to Observe Z

(b) Time-varying Histogram of Error Predicting Beta

Figure 19: Exp. 5: Ability to Observe Income Components

Although the asymptotic covariance matrix is not appropriate here since σ_u is on the boundary, it is still interesting to see what equation (**??**) gives if σ_u is excluded. The diagonal of the covariance matrix is

$$
\begin{aligned}
var(\hat{\delta}) &= 0.0017 \\
var(\hat{\rho}) &= 0.2959 \\
var(\hat{\sigma}_\beta) &= 0.0003 \\
var(\hat{\lambda}) &= 213.1423 \\
var(\hat{\gamma}) &= 5738.2305 \\
var(\hat{\alpha}) &= 1.9413 \\
var(\hat{\sigma}_v) &= 3.0329 \\
var(\hat{k}) &= 283.0091
\end{aligned}
$$

Note that, as G&P found, the retirement function is the most poorly identified. A cursory examination of the results from various experiments suggests, intuitively, that γ and k may have similar effects, since when one is large, the other is generally on its lower boundary. This could cause the poor identification here. Interestingly, though the dispersion in β_i does not have an astronomical estimated variance (though the variance is still large relative to the scale of the parameter), the amount of initial uncertainty (λ) does. In this case, this may well be a result of the slow learning of β_i as shown in figure **??**. The other income process parameters are very poorly identified relative to their scale. However, δ and ρ have quite low variances.

6 Conclusion

We have solved a consumption model that incorporates heterogeneous income profiles with Bayesian learning and attempted to match them to the first and second non-central moments of consumption data from the CEX under different parameter restrictions. Although the numerical solution of the model was straightforward, it was by its nature extremely time-consuming to minimize numerically. Furthermore, there is some evidence of local minima. Preliminary experiments showed that the results above were not robust to increasing the fineness of the grid, but improving the model in that regard was not possible because of time and computing power constraints. If this model is to be taken further, it must be done with more computing power than we had available to us.

In general, the model was found to have a good ability to minimize the objective function, especially when the parameters were unconstrained. However, the time evolution of the actual consumption variance, which is the parameter of intuitive interest, depends very subtly on the non-central moments, and in no case did the change in variance over time look like a very satisfactory fit. It should be noted that the variance derived from the CEX after accounting for family size gives a much smaller increase in consumption variance over time than similar measures that do not control for family size, and that this may account for the poor performance of experiments Three and Four.

In every case, the model either predicted counterintuitive parameter values (such as in Experiments Two, Three, and Four), or unrealistic predictions about savings behavior (as in Experiments One and Five). In every case, at least one parameter was found to lie on the boundary of the parameter space, prohibiting estimation of the asymptotic covariance matrix. However, when the (incorrect) covariance matrix for an interior point was calculated for the Experiment Five, it was found that some parameters – particularly the retirement parameters, but also the income process – had very high variances. This militates for using first-stage estimators when attacking this kind of problem, as the expressive power of the model works against identification by overfitting. On the other hand, first-stage estimates from PSID data, such as those providing the baselines for Experiments Three and Four, were found to give poor fits to the data. Again, this is probably as a result of the first-stage estimates failing to control for the effect of family size on within-cohort variance.