Appendix to: "An Estimation of Economic Models with Recursive Preferences"*

Xiaohong Chen — Yale
Jack Favilukis — LSE
Sydney C. Ludvigson — NYU and NBER

First draft: January 27, 2005
This draft: October 3, 2012

*Chen: Department of Economics Yale University, Box 208281, New Haven, CT 06520, Email: xiaohong.chen@yale.edu, Tel: (203) 432-5852; http://cowles.econ.yale.edu/faculty/chen.htm. Favilukis: Department of Finance, London School of Economics, Houghton Street, London WC2A 2AE; Email: j.favilukis@lse.ac.uk, http://pages.stern.nyu.edu/~jfaviluk. Ludvigson: Department of Economics, New York University, 19 W. 4th Street, 6th Floor, New York, NY 10012; Email: sydney.ludvigson@nyu.edu; Tel: (212) 998-8927; http://www.econ.nyu.edu/user/ludvigsons/. We acknowledge financial support from the National Science Foundation (Chen and Ludvigson) and from the Alfred P. Sloan Foundation (Ludvigson). Any errors or omissions are the responsibility of the authors, and do not necessarily reflect the views of the National Science Foundation.

Appendix to: "An Estimation of Economic Models with Recursive Preferences"

Abstract

This document is the on-line Appendix to accompany "An Estimation of Economic Models with Recursive Preferences," by Jack Favilukis, Sydney C. Ludvigson, and Xiaohong Chen. The appendix consist of several parts: Section 1 describes the data. Section 2 discusses how the unknown continuation value function is approximated, including discussion of the arguments of $\frac{V_t}{C_t}$, and the choice of sieve function to approximate $F(\cdot)$. Section 3 provides details of the two-step semiparametric estimation procedure, including the implementation of the SMD estimator as an instance of GMM. Section 4 presents additional results from the estimation not reported in the paper.

1 Data Description

The sources and description of each data series we use are listed below.

AGGREGATE CONSUMPTION

Aggregate consumption is measured as expenditures on nondurables and services, excluding shoes and clothing. The quarterly data are seasonally adjusted at annual rates, in billions of chain- weighted 2000 dollars. The components are chain-weighted together, and this series is scaled up so that the sample mean matches the sample mean of total personal consumption expenditures. Our source is the U.S. Department of Commerce, Bureau of Economic Analysis.

STOCKHOLDER CONSUMPTION

The definition of stockholder status, the consumption measure, and the sample selection follow Vissing-Jorgensen (2002). Consumption is measured as nondurables and services expenditures. Details on this construction can be found in Appendix A of Malloy, Moskowitz, and Vissing-Jorgensen (2009). We use their "simple" measure of stockholders, based on responses to the survey indicating positive holdings of "stocks, bonds, mutual funds and other such securities." Nominal consumption values are deflated by the BLS deflator for nondurables for urban households. Our source is the Consumer Expenditure Survey.

POPULATION

A measure of population is created by dividing real total disposable income by real per capita disposable income. Consumption, wealth, labor income, and dividends are in per capita terms. Our source is the Bureau of Economic Analysis.

PRICE DEFLATOR

Real asset returns are deflated by the implicit chain-type price deflator (2000=100) given for the consumption measure described above. Our source is the U.S. Department of Commerce, Bureau of Economic Analysis.

MONTHLY INDUSTRIAL PRODUCTION INDEX

Industrial production is measured as the seasonally adjusted total industrial production index (2002=100). Our source is the Board of Governors of the Federal Reserve System.

MONTHLY SERVICES EXPENDITURES

Measured as personal consumption expenditures on services, billions of dollars; months seasonally adjusted at annual rates. Nominal consumption is deflated by the implicit price deflator for services expenditures. Our source is the Bureau of Economic Analysis.

ASSET RETURNS

- 3-Month Treasury Bill Rate: secondary market, averages of business days, discount basis percent; Source: H.15 Release – Federal Reserve Board of Governors.

- 6 size/book-market returns: Six portfolios, monthly returns from July 1926-December 2001. The portfolios, which are constructed at the end of each June, are the intersections of 2 portfolios formed on size (market equity, ME) and 3 portfolios formed on the ratio of book equity to market equity (BE/ME). The size breakpoint for year t is the median NYSE market equity at the end of June of year t. BE/ME for June of year t is the book equity for the last fiscal year end in t-1 divided by ME for December of t-1. The BE/ME breakpoints are the 30th and 70th NYSE percentiles. Source: Kenneth French's homepage, http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

PROXY FOR LOG CONSUMPTION-WEALTH RATIO, $\widehat{cay}$

The proxy for the log consumption-wealth ratio is computed as described in Lettau and Ludvigson (2001).

RELATIVE BILL RATE, $RREL$

The relative bill rate is the 3-month treasury bill yield less its four-quarter moving average. Our source is the Board of Governors of the Federal Reserve System.

LOG EXCESS RETURNS ON S&P 500 INDEX: $SPEX$

SPEX is the log difference in the Standard and Poor 500 stock market index, less the log 3-month treasury bill yield. Our source is the Board of Governors of the Federal Reserve System.

R_m, SMB, HML

The Fama/French benchmark factors, Rm, SMB, and HML, are constructed from six size/book-to-market benchmark portfolios that do not include hold ranges and do not incur transaction costs. Rm, the return on the market, is the value-weighted return on all NYSE, AMEX, and NASDAQ stocks. Source: Kenneth French's homepage,
http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

2 Approximation to Continuation Value Function $F()$

The arguments of $F(\cdot)$*.* If the Markov structure is linear, we give assumptions under which $\frac{V_t}{C_t} = F\left(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}}\right)$. The system in the text is repeated here for exposition:

$$c_{t+1} - c_t = \mu + Hx_t + \mathbf{C}\boldsymbol{\epsilon}_{t+1} \tag{1}$$

$$x_{t+1} = \phi x_t + \mathbf{D}\boldsymbol{\epsilon}_{t+1}, \tag{2}$$

where $\boldsymbol{\epsilon}_{t+1}$, $\mathbf{C}$, and $\mathbf{D}$ are 2×1 vectors defined:

$$\boldsymbol{\epsilon}_{t+1} \equiv \begin{bmatrix} \epsilon_{c,t+1} \\ \epsilon_{x,t+1} \end{bmatrix}$$

$$\begin{aligned} \mathbf{C} &\equiv [\sigma_c + \phi_x \sigma_x] \\ \mathbf{D} &\equiv [\phi_c \sigma_c + \sigma_x]. \end{aligned}$$

First, note that the state variable x_t is latent to the econometrician. We therefore form an estimate of x_t by applying the Kalman filter to the system (1) and (2). The Kalman filter implies that the dynamic system (1) and (2) converges asymptotically to time-invariant innovations representation taking the form

$$\Delta c_{t+1} = \mu + H\widehat{x}_t + \varepsilon_{t+1} \tag{3}$$

$$\widehat{x}_{t+1} = \phi \widehat{x}_t + K\varepsilon_{t+1}, \tag{4}$$

where the scalar variable $\varepsilon_{t+1} \equiv \Delta c_{t+1} - \Delta \widehat{c}_{t+1} = H(x_t - \widehat{x}_t) + \mathbf{C}\boldsymbol{\epsilon}_{t+1}$, $\widehat{x}_t$ denotes a linear least squares projection of x_t onto $\Delta c_t, \Delta c_{t-1}, ...\Delta c_{-\infty}$, and

$$\begin{aligned} K &\equiv (\mathbf{D}\mathbf{C}' + \phi P H)(H' P H + \mathbf{C}\mathbf{C}')^{-1} \\ &= (\mathbf{D}\mathbf{C}' + \phi P H)\left(H^2 P + \mathbf{C}\mathbf{C}'\right)^{-1} \end{aligned} \tag{5}$$

(because H is scalar here) where P solves

$$P = (\phi - KH)^2 P + (\mathbf{D} - K\mathbf{C})(\mathbf{D} - K\mathbf{C})'.$$

(See Hansen and Sargent (2007).) The representation (3)-(4) shows that the observable state variable $\widehat{x}_t$ replaces the latent state variable x_t as the argument of the function over which $\frac{V_t}{C_t}$ is defined. Note that, if either of the parameters ϕ_x or ϕ_c is different from zero, the innovation in the observation equation (1) will be correlated with the innovation in the state equation (2).

Assume $\frac{V_t}{C_t}$ is an invertible function $f\left(\widehat{x}_t\right)$ with $f'\left(\widehat{x}_t\right) > 0$. Then,

$$\widehat{x}_t = f^{-1}\left(\frac{V_t}{C_t}\right)$$

is also an increasing function. From (4) we have

$$\frac{V_t}{C_t} = f\left(\widehat{x}_t\right) = f\left(\phi\widehat{x}_{t-1} + K\varepsilon_t\right). \tag{6}$$

By inverting (3), we obtain

$$\begin{aligned} \varepsilon_t &= \Delta c_t - \mu - H\widehat{x}_{t-1} \\ &= \Delta c_t - \mu - Hf^{-1}\left(\frac{V_{t-1}}{C_{t-1}}\right). \end{aligned} \tag{7}$$

Plugging (7) into (6), we have

$$\begin{aligned} \frac{V_t}{C_t} &= f\left(\phi f^{-1}\left(\frac{V_{t-1}}{C_{t-1}}\right) + K\left[\Delta c_t - \mu - Hf^{-1}\left(\frac{V_{t-1}}{C_{t-1}}\right)\right]\right) \\ &= f\left(\left[\phi - KH\right] f^{-1}\left(\frac{V_{t-1}}{C_{t-1}}\right) + K\left[\Delta c_t - \mu\right]\right) \\ &\equiv F\left(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}}\right). \end{aligned} \tag{8}$$

where $F : \mathbb{R}^2 \rightarrow \mathbb{R}$.

Two aspects of this application of the Kalman filter bear noting. First, the linear Markov structure implies that F is a monotonic function of $\frac{C_t}{C_{t-1}}$. Second, the function $\frac{V_t}{C_t}$ may display negative serial dependence under a variety of circumstances, for example if the innovations in (1) and (2) are positively correlated, or if the autocorrelation in x_t is low. To see this, notice from equation (8) we can see that $\frac{V_t}{C_t}$ will display negative serial dependence if $\phi < KH$. From (5) we have

$$\begin{aligned} K &\equiv \left(\mathbf{DC}' + \phi PH\right)\left(H^2P + \mathbf{CC}'\right)^{-1} \\ &= \left(\phi_c\sigma_c^2 + \phi_x\sigma_x^2 + \phi PH\right)\left(H^2P + \sigma_c^2 + \phi_x^2\sigma_x^2\right)^{-1} => \\ KH &= \phi\left[\frac{PH^2}{H^2P + \sigma_c^2 + \phi_x^2\sigma_x^2}\right] + \left[\frac{H\phi_c\sigma_c^2 + H\phi_x\sigma_x^2}{H^2P + \sigma_c^2 + \phi_x^2\sigma_x^2}\right]. \end{aligned}$$

The first term in square brackets, multiplying ϕ, is less than one in absolute value, so that if the second term in square brackets is zero (which will be true if for example $\phi_x = \phi_c = 0$),

$\phi > KH$ and $\frac{V_t}{C_t}$ will display positive serial dependence. But more generally, the second term can be sufficiently positive such that $\phi < KH$ and $\frac{V_t}{C_t}$ exhibits negative serial dependence. To determine the circumstances under which $\phi < KH$ requires using a numerical algorithm to solve recursively for K and P under some values for the primitive parameters of Markov system (1)-(2), and then computing the resulting value for $\phi - KH$. A range of cases can be found for which $\phi - KH < 0$, including those when ϕ is not too large, and when ϕ_x or $\phi_y \neq 0$.

The assumptions embedded in this example are meant to be illustrative: more general nonlinear state space models and distributional assumptions are likely to produce more complicated dynamic relationships between $\frac{V_t}{C_t}$ and its own lagged value, as well as consumption growth.

B-spline Approximation of $F(\cdot)$. We use cubic B-splines to approximate the unknown continuation value-consumption ratio function because unlike other basis functions (e.g., polynomials) they are shape-preserving (Chui (1992)). The multivariate sieve functions $\{B_j(\cdot,\cdot) : j = 1, ..., K_T\}$ are implemented as a tensor product cubic B-spline taking the form:

$$F(z,c) = \alpha_0 + \sum_{i=1}^{K_{1T}} \sum_{j=1}^{K_{2T}} a_{ij} B_m\left(z - i + \frac{m}{2}\right) B_m\left(\frac{c}{\Delta_2} + \varsigma - j\right), \tag{9}$$

where $z \equiv \frac{V_t}{C_t}$, $c \equiv \frac{C_{t+1}}{C_t}$, $B_m(.)$ is a B-spline of degree m, and a_{ij} are parameters to be estimated. The term $\frac{m}{2}$ recenter the function, which insures that the function is shape-preserving (preserving nonnegativity, monotonicity and convexity of the unknown function to be approximated). For consumption growth the parameters Δ_2 and ς are set to guarantee that the support of B_m stays within the bounds $[0.97, 1.04]$ since this is the range for which we observe variation in gross consumption growth data. This insures that as j goes from 1 to K_{2T}, B_m is always evaluated only over the support $[0.97, 1.04]$. Δ_2 fixes the support of the spline. By shifting i and j, the spline is moved on the real line.

We use a cardinal B-spline given by

$$B_m(y) = \frac{1}{(m-1)!} \sum_{k=0}^{m} (-1)^k \binom{m}{k} \left[\max(0, y-k)\right]^{m-1}, \quad \text{with } \binom{m}{k} \equiv \frac{m!}{(m-k)!k!}.$$

The order of the spline, m, for our application is set to 3. For the dimensionality of the B-spline sieve, we set $K_{1T} = K_{2T} = 3$. Because asymptotic theory only provides guidance about the *rate* at which $K_{1T} \cdot K_{2T} + 1$ must increase with the sample size T, other considerations must be used to judge how best to set this dimensionality. The bigger are K_{1T} and K_{2T}, the greater is the number of parameters that must be estimated, therefore the dimensionality of the sieve is naturally limited by the size of our data set. With $K_{1T} = K_{2T} = 3$, the dimension of the total unknown parameter vector, $(\boldsymbol{\delta},F)' = \left(\beta, \rho, \theta, a_0, a_{11}, ..., a_{K_{1T}K_{2T}}, \frac{V_0}{C_0}\right)'$, is 14, estimated using a sample of size $T = 213$. In practice, we obtained very similar results setting $K_{1T} = K_{2T} = 4$.

3 Semiparametric Two-Step Estimation Procedure

We use $\mathcal{D} \equiv [\underline{\beta}, \overline{\beta}] \times [\underline{\theta}, \overline{\theta}] \times [\underline{\rho}, \overline{\rho}]$ to denote the compact parameter space for the finite-dimensional unknown parameters $\boldsymbol{\delta} = (\beta, \theta, \rho)'$, and $\mathcal{V}$ denotes the function space for the infinite dimensional unknown function $F()$. In the application we assume that $\mathcal{V}$ is a Holder ball:

$$\mathcal{V} \equiv \{g : (0, \infty) \times (0, \infty) \to (0, \infty) : \|g\|_{\Lambda^s} \leq const. < \infty\}, \quad \text{for some } s > 1, \tag{10}$$

here the norm $\|g\|_{\Lambda^s}$ is defined as

$$||g||_{\Lambda^s} \equiv \sup_{x,y} |g(x, y)| + \max_{a_1+a_2=[s]} \sup_{(x,y)\neq(\overline{x},\overline{y})} \frac{|\partial_x^{a_1} \partial_y^{a_2} g(x, y) - \partial_x^{a_1} \partial_y^{a_2} g(\overline{x}, \overline{y})|}{\sqrt{(x - \overline{x})^2 + (y - \overline{y})^2}^{s-[s]}} < \infty,$$

where $[s]$ denotes the largest non-negative integer such that $[s] < s$, and (a_1, a_2) is any pair of non-negative integers such that $a_1 + a_2 = [s]$.

For any candidate value $\boldsymbol{\delta} = (\beta, \theta, \rho)' \in \mathcal{D}$, we define

$$F^* (\cdot; \boldsymbol{\delta}) \equiv \arg \inf_{F \in \mathcal{V}} E\{m(\mathbf{w}_t, \boldsymbol{\delta}, F)' m(\mathbf{w}_t, \boldsymbol{\delta}, F)\},$$

where $m(\mathbf{w}_t, \boldsymbol{\delta}, F)' \equiv E\{\gamma(\mathbf{z}_{t+1}, \boldsymbol{\delta}, F)|\mathbf{w}_t\} = (m_1(\mathbf{w}_t, \boldsymbol{\delta}, F), ..., m_N(\mathbf{w}_t, \boldsymbol{\delta}, F))$ and $m_i(\mathbf{w}_t, \boldsymbol{\delta}, F) \equiv E\{\gamma_i(\mathbf{z}_{t+1}, \boldsymbol{\delta}, F)|\mathbf{w}_t\}$ for $i = 1, ..., N$. Next we define the pseudo true value $\boldsymbol{\delta}^* = (\beta^*, \theta^*, \rho^*)' \in \mathcal{D}$ as

$$\boldsymbol{\delta}^*_{\mathbf{W}} \equiv \arg \min_{\boldsymbol{\delta} \in \mathcal{D}} \left[E\left\{\gamma(\mathbf{z}_{t+1}, \boldsymbol{\delta}, F^* (\cdot, \boldsymbol{\delta})) \otimes \mathbf{x}_t\right\}\right]' \mathbf{W} \left[E\left\{\gamma(\mathbf{z}_{t+1}, \boldsymbol{\delta}, F^* (\cdot, \boldsymbol{\delta})) \otimes \mathbf{x}_t\right\}\right],$$

where $\mathbf{W}$ is some positive definite weighting matrix and $\mathbf{x}_t$ is any chosen measurable function of $\mathbf{w}_t$.

We say the model is correctly specified if

$$E\left\{\gamma_i(\mathbf{z}_{t+1}, \boldsymbol{\delta}_o, F^* (\cdot, \boldsymbol{\delta}_o)) \otimes \mathbf{x}_t\right\} = 0, \qquad i = 1, ..., N. \tag{11}$$

When the model is correctly specified, we have $\boldsymbol{\delta}^*_{\mathbf{W}} = \boldsymbol{\delta}_o$ and $F^* (\cdot, \boldsymbol{\delta}_o) = F_o$, and these true parameter values $\boldsymbol{\delta}_o, F^* (\cdot, \boldsymbol{\delta}_o)$ do not depend on the choice of the weighting matrix

$\mathbf{W}$. However, when the model could be misspecified, then the pseudo true values $\boldsymbol{\delta}^*_{\mathbf{W}}$ and $F^*(\cdot, \boldsymbol{\delta}^*_{\mathbf{W}})$ typically will depend on the weighting matrix $\mathbf{W}$.

Two-step Semiparametric Estimation Procedure. In **Step One**, for any candidate value $\boldsymbol{\delta} = (\beta, \theta, \rho)' \in \mathcal{D}$, we estimate $F^*(\cdot; \boldsymbol{\delta})$ by the sieve minimum distance (SMD) estimator $\widehat{F}_T(\cdot; \boldsymbol{\delta})$:

$$\widehat{F}_T(\cdot, \boldsymbol{\delta}) = \arg\min_{F_T \in \mathcal{V}_T} \frac{1}{T} \sum_{t=1}^{T} \widehat{m}(\mathbf{w}_t, \boldsymbol{\delta}, F_T)' \widehat{m}(\mathbf{w}_t, \boldsymbol{\delta}, F_T), \tag{12}$$

where $\widehat{m}(\mathbf{w}_t, \boldsymbol{\delta}, F)' = (\widehat{m}_1(\mathbf{w}_t, \boldsymbol{\delta}, F), ..., \widehat{m}_N(\mathbf{w}_t, \boldsymbol{\delta}, F))$ is some nonparametric estimate of $m(\mathbf{w}_t, \boldsymbol{\delta}, F)$, and $\mathcal{V}_T$ is a sieve space that approximates $\mathcal{V}$. In the application we let $\mathcal{V}_T$ be the tensor product B-spline (9) sieve space, which becomes dense in $\mathcal{V}$ as sample size $T \to \infty$.

In **Step Two**, we estimate $\boldsymbol{\delta}^*_{\mathbf{W}}$ by minimizing a sample GMM objective function:

$$\widehat{\boldsymbol{\delta}}_{\mathbf{W}} = \arg\min_{\boldsymbol{\delta} \in \mathcal{D}} \left[\mathbf{g}_T(\boldsymbol{\delta}, \widehat{F}_T(\cdot, \boldsymbol{\delta}); \mathbf{y}^T)\right]' \mathbf{W}_T \left[\mathbf{g}_T((\boldsymbol{\delta}, \widehat{F}_T(\cdot, \boldsymbol{\delta}); \mathbf{y}^T)\right], \tag{13}$$

where $\mathbf{y}^T = \left(\mathbf{z}'_{T+1}, ...\mathbf{z}'_2, \mathbf{x}'_T, ...\mathbf{x}'_1\right)'$ denotes the vector containing all observations in the sample of size T, and $\mathbf{W}_T$ is a positive, semi-definite possibly random weighting matrix that converges to $\mathbf{W}$, also,

$$\mathbf{g}_T(\boldsymbol{\delta}, \widehat{F}_T(\cdot, \boldsymbol{\delta}); \mathbf{y}^T) = \frac{1}{T} \sum_{t=1}^{T} \gamma(\mathbf{z}_{t+1}, \boldsymbol{\delta}, \widehat{F}_T(\cdot, \boldsymbol{\delta})) \otimes \mathbf{x}_t \tag{14}$$

are the sample moment conditions.

We have considered two kinds of GMM estimation of $\boldsymbol{\delta}^*_{\mathbf{W}}$ in Step Two: (i) GMM estimation of $\boldsymbol{\delta}^*_{\mathbf{W}}$ using $\mathbf{x}_t = 1_N$ as the instruments and $\mathbf{W}_T = \mathbf{G}_T^{-1}$ as the weighting matrix, where the $(i, j)th$ element of $\mathbf{G}_T$ is $\frac{1}{T} \sum_{t=1}^{T} R_{i,t} R_{j,t}$ for $i, j = 1, ..., N$. This leads to the GMM estimate using HJ criterion. (ii) GMM estimation of $\boldsymbol{\delta}^*_{\mathbf{W}}$ using $\mathbf{x}_t = 1_N$ as the instruments and $\mathbf{W}_T = \mathbf{I}$ as the weighting matrix, where $\mathbf{I}$ is the $N \times N$ identity matrix.

The SMD procedure in Step One has been proposed respectively in Newey and Powell (2003) for nonparametric IV regression, and in Ai and Chen (2003) for semi/nonparametric conditional moment restriction models. The SMD procedure needs a nonparametric estimator $\widehat{m}(\mathbf{w}_t, \boldsymbol{\delta}, F)$ for $m(\mathbf{w}_t, \boldsymbol{\delta}, F)$. There are many nonparametric procedures such as kernel, local linear regression, nearest neighbor and various sieve methods that can be used to estimate $m_i(\mathbf{w}_t, \boldsymbol{\delta}, F)$, $i = 1, ..., N$. In our application we consider the sieve Least Squares (LS) estimator. For each fixed $(\mathbf{w}_t, \boldsymbol{\delta}, F)$, we approximate $m_i(\mathbf{w}_t, \boldsymbol{\delta}, F)$ by

$$m_i(\mathbf{w}_t, \boldsymbol{\delta}, F) \approx \sum_{j=1}^{J_T} a_j(\boldsymbol{\delta}, F) p_{0j}(\mathbf{w}_t), \quad i = 1, ..., N,$$

where p_{0j} some known fixed basis functions, and $J_T \to \infty$ slowly as $T \to \infty$. We then estimate the sieve coefficients $\{a_j\}$ simply by OLS regression:

$$\min_{\{a_j\}} \frac{1}{T} \sum_{t=1}^{T} [\gamma_i(\mathbf{z}_{t+1}, \boldsymbol{\delta}, F) - \sum_{j=1}^{J_T} a_j(\boldsymbol{\delta}, F) p_{0j}(\mathbf{w}_t)]'[\gamma_i(\mathbf{z}_{t+1}, \boldsymbol{\delta}, F) - \sum_{j=1}^{J_T} a_j(\boldsymbol{\delta}, F) p_{0j}(\mathbf{w}_t)]$$

and the resulting estimator is denoted as: $\widehat{m}_i(\mathbf{w}, \boldsymbol{\delta}, F) = \sum_{j=1}^{J_T} \widehat{a}_j(\boldsymbol{\delta}, F) p_{0j}(\mathbf{w}_t)$. In the following we denote: $p^{J_T}(\mathbf{w}) = (p_{01}(\mathbf{w}), ..., p_{0J_T}(\mathbf{w}))'$ and $\mathbf{P} = (p^{J_T}(\mathbf{w}_1), ..., p^{J_T}(\mathbf{w}_T))'$, then:

$$\widehat{m}_i(\mathbf{w}, \boldsymbol{\delta}, F) = \sum_{t=1}^{T} \gamma_i(\mathbf{z}_{t+1}, \boldsymbol{\delta}, F) p^{J_T}(\mathbf{w}_t)'(\mathbf{P}'\mathbf{P})^{-1} p^{J_T}(\mathbf{w}), \quad i = 1, ..., N. \tag{15}$$

Many known sieve bases could be used as $\{p_{0j}\}$. In our application we take the power series and Fourier series as the $p^{J_T}(\mathbf{w})$. The empirical findings are not sensitive to the different choice of sieve bases, and we only report the results based on power series due to the length of the paper.

GMM Implementation of SMD Estimation. When the nonparametric estimator $\widehat{m}_i(\mathbf{w}, \boldsymbol{\delta}, F)$ is the linear sieve estimator (15), the first step SMD estimation of $F^*(\cdot; \boldsymbol{\delta})$ can be alternatively implemented via the following GMM criterion (16):

$$\widehat{F}_T(\cdot, \boldsymbol{\delta}) = \arg\min_{F_T \in \mathcal{V}_T} \left[\mathbf{g}_T(\boldsymbol{\delta}, F_T; \mathbf{y}^T)\right]' \left\{\mathbf{I}_N \otimes (\mathbf{P}'\mathbf{P})^{-1}\right\} \left[\mathbf{g}_T(\boldsymbol{\delta}, F_T; \mathbf{y}^T)\right], \tag{16}$$

where $\mathbf{y}^T = \left(\mathbf{z}'_{T+1}, ...\mathbf{z}'_2, \mathbf{w}'_T, ...\mathbf{w}'_1\right)'$ denotes the vector containing all observations in the sample of size T and

$$\mathbf{g}_T(\boldsymbol{\delta}, F_T; \mathbf{y}^T) = \frac{1}{T} \sum_{t=1}^{T} \gamma(\mathbf{z}_{t+1}, \boldsymbol{\delta}, F_T) \otimes p^{J_T}(\mathbf{w}_t) \tag{17}$$

are the sample moment conditions associated with the $NJ_T \times 1$ -vector of population unconditional moment conditions: $E\left\{\gamma_i(\mathbf{z}_{t+1}, \boldsymbol{\delta}, F^*(\cdot, \boldsymbol{\delta})) p_{0j}(\mathbf{w}_t)\right\}$, $i = 1, ..., N$, $j = 1, ..., J_T$.

4 Additional Empirical Results

4.1 Fixing the EIS $= 1$

Several authors have focused on the cross-sectional implications of EZW preferences when the EIS, ρ^{-1}, is restricted to unity (e.g., Hansen, Heaton, and Li (2008), Malloy, Moskowitz, and Vissing-Jorgensen (2009)). Malloy et. al., conjecture that risk-aversion estimates identified from a cross-section of returns are unlikely to be greatly affected by the value of the EIS. To investigate this possibility in our setting, we repeated our estimation fixing $\rho = 1$. The results are presented in Table A.1.

The results are somewhat sensitive to the weighting matrix used in the second step estimation. For example, in an estimation of the representative agent version of the model with $\rho = 1$ and $\mathbf{W} = \mathbf{I}_N$, the relative risk aversion coefficient θ is estimated to be 20, much lower than the value of almost 60 reached when ρ is freely estimated (Table 2). But when $\mathbf{W}=\mathbf{G}_T^{-1}$, the coefficient of relative risk aversion θ is estimated to be 60, precisely the same value obtained when ρ is left unrestricted. In addition, the HJ distance is about the same when $\rho = 1$, equal to 0.448 compared to 0.451 when ρ is unrestricted (the HJ distance is slightly smaller when $\rho = 1$ because, when ρ is fixed, one fewer parameter is estimated, reducing the AIC penalty). Thus, the results using $\mathbf{W}=\mathbf{G}_T^{-1}$ are largely supportive of the conjecture of Malloy, Moskowitz, and Vissing-Jorgensen (2009). We note, however, that if the model with $\rho = 1$ is misspecified, parameter estimates can be sensitive to the objective function minimized, as we find here.

We find qualitatively similar results in an estimation of the representative stockholder version of the model. In this case, when $\rho = 1$ and $\mathbf{W} = \mathbf{I}_N$, the relative risk aversion coefficient θ is estimated to be 20, the same value obtained when ρ is left unrestricted. This is not surprising because the unrestricted value of ρ is already quite close to unity, equal to 0.9. On the other hand, when $\mathbf{W}=\mathbf{G}_T^{-1}$, θ is estimated to be 10, considerably smaller than the value of 17 estimated when ρ is unrestricted with a point estimate of 0.68. But the HJ distance is 0.469 when $\rho = 1$, only slightly larger than the value of 0.463 found when ρ is unrestricted. We conclude that the model's cross-sectional performance, as measured by the HJ distance, is not sensitive to fixing the EIS at unity.

4.2 Forecasting the Aggregate Wealth Portfolio $R_{w,t+h}$ with $\ln W_t - \ln C_t$

We investigate the implications of the findings above for forecastability of the multi-horizon excess return to the aggregate wealth portfolio, R_w. To do so, consider the accumulation equation for aggregate wealth, written[1]

$$W_{t+1} = R_{w,t+1}(W_t - C_t). \tag{18}$$

[1]Labor income does not appear explicitly in this equation because of the assumption that the market value of tradable human capital is included in aggregate wealth.

If the consumption-aggregate wealth ratio is stationary, the budget constraint may be approximated by taking a first-order Taylor expansion of the equation following Campbell and Mankiw (1989). The resulting approximation gives

$$\Delta \ln W_{t+1} \approx k + \ln R_{w,t+1} + (1 - 1/\rho_w)(\ln C_t - \ln W_t), \tag{19}$$

where $\rho_w \equiv 1 - \exp\left(\overline{c - w}\right)$, and k is a constant that plays no role in the forecasting analysis. Solving this difference equation forward, imposing the condition that $\lim_{i \to \infty} \rho_w^i (\log C_{t+i} - \log W_{t+i}) = 0$ and taking expectations, the log wealth-consumption ratio may be written (ignoring constants):

$$\ln W_t - \ln C_t = \sum_{i=1}^{\infty} \rho_w^i (\Delta \ln C_{t+i} - \ln R_{w,t+i}). \tag{20}$$

Under rational expectations, the expression above should hold *ex-ante* as well as *ex-post*:

$$\ln W_t - \ln C_t = E_t \sum_{i=1}^{\infty} \rho_w^i (\Delta \ln C_{t+i} - \ln R_{w,t+i}). \tag{21}$$

where E_t is the expectation operator conditional on information available at time t. Equation (**??**) implies that the log consumption-wealth ratio forecasts future returns to aggregate wealth, or future consumption growth, or some combination of the two. Multiply both sides of (**??**) by $(\ln W_t - \ln C_t) - E(\ln W_t - \ln C_t)$ and take unconditional expectations to obtain

$$\begin{aligned} \operatorname{Var}(\ln W_t - \ln C_t) &= \operatorname{Cov}\left((\ln W_t - \ln C_t), \sum_{j=0}^{\infty} \rho_w^j \Delta \ln C_{t+1+j} \right) \\ &\quad - \operatorname{Cov}\left((\ln W_t - \ln C_t), \sum_{j=0}^{\infty} \rho_w^j \ln R_{w,t+1+j} \right). \end{aligned} \tag{22}$$

The above expression says that the variance of the log wealth-consumption ratio must be attributable to its covariance with future consumption growth minus its covariance with future returns to aggregate wealth.

It is straightforward to investigate the implications of the estimated V/C functions for the forecasting identities above using the equilibrium relation between the wealth-consumption ratio and the continuation value-consumption ratio implied by EZW preferences:

$$\frac{W_t}{C_t} = \frac{1}{(1-\beta)} \left(\frac{V_t}{C_t} \right)^{1-\rho}.$$

We use the relation above, along with the estimated continuation value-consumption ratio functions, to investigate how the variance of $\ln C_t - \ln W_t$ is related to future returns and future consumption growth.

Table A.2 presents forecasts of $\ln R_{w,t+H}$ and $\Delta \ln C_{t+H}$ by $\ln W_t - \ln C_t$ for various horizons H. Also reported are empirical counterparts to the variance decomposition (**??**) computed by truncating the infinite sums at 20 quarters and multiplying by 100 to express as a percent:

$$\frac{100 \times \operatorname{Cov}\left(\ln W_t - \ln C_t, \sum_{j=0}^{20} \rho_w^j \Delta \ln C_{t+1+j}\right)}{\operatorname{Var}\left(\ln W_t - \ln C_t\right)}$$
$$\frac{-100 \times \operatorname{Cov}\left(\ln W_t - \ln C_t, \sum_{j=0}^{20} \rho_w^j \ln R_{w,t+1+j}\right)}{\operatorname{Var}\left(\ln W_t - \ln C_t\right)}.$$

The table shows that $\ln W_t - \ln C_t$ forecasts future returns to aggregate wealth, but is weakly related to consumption growth over long horizons. A high wealth-consumption ratio is indicative of lower future returns to aggregate wealth, and not of higher future consumption growth. Moreover, the covariance of $\ln W_t - \ln C_t$ with future returns accounts for 107% of the variance of $\ln W_t - \ln C_t$.[2] These results are reminiscent of the behavior of the log price-dividend ratio for the aggregate stock market.[3]

[2] A high wealth-consumption ratio forecasts a slightly *lower* discounted value of future consumption growth, so covariance with future returns has to account for more than 100% of the variance of $\ln W_t - \ln C_t$.

[3] For a recent review of this evidence see Cochrane (2011), Ludvigson (2012).

Table A.1

Preference Parameter Estimates, EIS=1

2nd Step Estimation	β	θ	HJ Dist
	Aggregate Consumption		
$\mathbf{W}=\mathbf{I}$	0.985	20	—
$\mathbf{W}=\mathbf{G}_T^{-1}$	0.985	60	0.448
	Stockholder Consumption		
$\mathbf{W}=\mathbf{I}$	0.990	20.00	—
$\mathbf{W}=\mathbf{G}_T^{-1}$	0.999	10.0	0.469

Notes: The table reports second-step estimates of preference parameters, when the EIS $=\rho^{-1}$ is fixed at one. β is the subjective time discount factor,and θ is the coefficient of relative risk aversion. Second-step estimates are obtained by minimizing the GMM criterion with either $\mathbf{W}=\mathbf{I}$ or with $W=G_T^{-1}$, where in both cases $x_t = 1_N$, an $N \times 1$ vector of ones. The sample is 1952:Q1-2005:Q1.

Table A.2

Dynamics of Estimated Aggregate Wealth-Consumption Ratio

Horizon H (quarters)	$\ln R_{W,t\to t+H} = a + b\,[\ln W_t - \ln C_t]$			$\Delta \ln C_{t\to t+H} = a + b\,[\ln W_t - \ln C_t]$		
	b	t-stat	$\overline{R}^2$	b	t-stat	$\overline{R}^2$
2	-0.0003	-7.2706	0.1223	0.0001	3.0292	0.0180
4	-0.0007	-8.4251	0.2070	0.0000	0.2397	0.0002
8	-0.0008	-8.8560	0.1260	-0.0001	-1.0239	0.0021
16	-0.0007	-6.1469	0.0430	0.0001	0.6498	0.0005

Variance decomposition of $\ln W_t - \ln C_t$	
R_W	C
106.87	-5.26

Notes: The top panel reports results from regressions of the estimated aggregate wealth return R_W (left) and aggregate consumption growth (right) from t to $t+H$ on the estimated log aggregate wealth-consumption ratio. The bottom panel reports the results of a variance decomposition of the estimated log wealth-consumption ratio. The column labeled "R_W" in the variance decomposition denotes $\frac{-100\times \mathrm{Cov}\left(\ln W_t - \ln C_t, \sum_{j=0}^{20} \rho_w^j \ln R_{w,t+1+j}\right)}{\mathrm{Var}(\ln W_t - \ln C_t)}$. The column lableled "$C$" in the variance decomposition stands for $\frac{100\times \mathrm{Cov}\left(\ln W_t - \ln C_t, \sum_{j=0}^{20} \rho_w^j \Delta \ln C_{t+1+j}\right)}{\mathrm{Var}(\ln W_t - \ln C_t)}$. The results reported are for estimates on aggregate consumption, $W = G_T^{-1}$. The sample is 1952:Q1-2005:Q1.

References

Ai, C., and X. Chen (2003): "Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, 71, 1795–1843.

Campbell, J. Y., and G. Mankiw (1989): "Consumption, Income, and Interest Rates: Reinterpreting the Time Series Evidence," in *NBER Macroeconomics Annual: 1989*, ed. by O. Blanchard, and S. Fischer, pp. 185–216. MIT Press, Cambridge, MA.

Chui, C. (1992): *An Introduction to Wavelets.* Academic Press, Inc., San Diego, CA.

Cochrane, J. H. (2011): "Discount Rates," *Journal of Finance,* forthcoming, American Finance Association Presidential Speech.

Hansen, L. P., J. Heaton, and N. Li (2008): "Consumption Strikes Back?: Measuring Long-run Risk," *Journal of Political Economy*, 116(2), 260–302.

Hansen, L. P., and T. J. Sargent (2007): *Robustness.* Princeton University Press forthcoming, Princeton, NJ.

Lettau, M., and S. C. Ludvigson (2001): "Consumption, Aggregate Wealth and Expected Stock Returns," *Journal of Finance*, 56(3), 815–849.

Ludvigson, S. C. (2012): "Advances in Consumption-Based Asset Pricing: Empirical Tests," in *Handbook of the Economics of Finance Vol. II,* forthcoming, ed. by G. Constantinides, M. Harris, and R. Stulz. Elsevier Science B.V., North Holland, Amsterdam.

Malloy, C. J., T. J. Moskowitz, and A. Vissing-Jorgensen (2009): "Long-run Stockholder Consumption Risk and Asset Returns," *Journal of Finance*, 64, 2427–2479.

Newey, W. K., and J. Powell (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1557–1569.

Vissing-Jorgensen, A. (2002): "Limited Asset Market Participation and Intertemporal Substitution," *Journal of Political Economy*, 110(4), 825–853.