ISSN 0956-8549-603

An Estimation of Economic Models with Recursive Preferences

By

Xiaohong Chen
Jack Favilukis
Sydney C. Ludvigson

DISCUSSION PAPER NO 603

DISCUSSION PAPER SERIES

November 2007

**Xiaohong Chen is Professor of Economics at Yale University. She obtained her PhD in economics PhD at the University of California, San Diego. Her research interests include semi-nonparametric econometrics estimation and testing, non-linear time series and stochastic process modelling and non-parametric adaptive learning. Jack Favilukis is a Lecturer in the Finance Department at The London School of Economics. He received his PhD in finance from the Stern School of Business at New York University. His areas of interests are consumption based asset pricing, macro-finance, and incomplete markets. Sydney C. Ludvigson is the William R. Berkley Associate Professor of Economics at New York University. She is also a Research Associate in the Asset Pricing group of the National Bureau of Economic Research. She received her Ph.D. and M.A. from Princeton University, and undergraduate degree from University of California, Los Angeles. She is a current and past recipient of the Alfred P. Sloan research fellowship and of several National Science Foundation grants. Ludvigson's recent research focuses on the relationship between financial market volatility, risk premia, and macroeconomic fundamentals, and on the empirical estimation and testing of leading asset pricing models.

An Estimation of Economic Models with Recursive Preferences*

Xiaohong Chen
Yale

Jack Favilukis
LSE

Sydney C. Ludvigson
NYU and NBER

Preliminary and Incomplete
Comments Welcome
First draft: January 27, 2005
This draft: November 2, 2007

---

*Chen: Department of Economics Yale University, Box 208281, New Haven, CT 06520, Email: xiaohong.chen@yale.edu, Tel: (203) 432-5852; http://homepages.nyu.edu/~xc6/. 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 and the C.V. Starr Center at NYU (Chen and Ludvigson) and from the Alfred P. Sloan Foundation (Ludvigson). We are grateful to Darrel Duffie, Lars Hansen, Monika Piazzesi, Annette Vissing-Jorgensen, Gianluca Violante, Motohiro Yogo, and to seminar participants at NYU, the June 2007 SED annual meetings and the July 2007 NBER Summer Institute Methods and Applications for Dynamic, Stochastic General Equilibrium Models Workshop for helpful comments. We also thank Annette Vissing-Jorgensen for help with the stockholder consumption data. Any errors or omissions are the responsibility of the authors, and do not necessarily reflect the views of the National Science Foundation.

An Estimation of Economic Models with Recursive Preferences

Abstract

This paper presents estimates of key preference parameters of the Epstein and Zin (1989, 1991) and Weil (1989) (EZW) recursive utility model, evaluates the model's ability to fit asset return data relative to other asset pricing models, and investigates the implications of such estimates for the unobservable aggregate wealth return. Our empirical results indicate that the estimated relative risk aversion parameter is high, ranging from 17-60, with higher values for aggregate consumption than for stockholder consumption, while the estimated elasticity of intertemporal substitution is above one. In addition, the estimated model-implied aggregate wealth return is found to be weakly correlated with the CRSP value-weighted stock market return, suggesting that the return to human wealth is negatively correlated with the aggregate stock market return. In quarterly data from 1952 to 2005, we find that an SMD estimated EZW recursive utility model can explain a cross-section of size and book-market sorted portfolio equity returns better than the standard consumption-based model based on power utility and better than the Lettau and Ludvigson (2001b) *cay*-scaled consumption CAPM model, but not as well as the Fama and French (1993) three-factor model with financial returns as risk factors.

JEL:

1 Introduction

A large and growing body of theoretical work in macroeconomics and finance models the preferences of economic agents using a recursive utility function of the type explored by Epstein and Zin (1989, 1991) and Weil (1989).[1] One reason for the growing interest in such preferences is that they provide a potentially important generalization of the standard power utility model first investigated in classic empirical studies by Hansen and Singleton (1982, 1983). The salient feature of this generalization is a greater degree of flexibility as regards attitudes towards risk and intertemporal substitution. Specifically, under the recursive representation, the coefficient of relative risk aversion need not equal the inverse of the elasticity of intertemporal substitution (EIS), as it must in time-separable expected utility models with constant relative risk aversion. This degree of flexibility is appealing in many applications because it is unclear why an individual's willingness to substitute consumption across random states of nature should be so tightly linked to her willingness to substitute consumption deterministically over time.

Despite the growing interest in recursive utility models, there has been a relatively small amount econometric work aimed at estimating the relevant preference parameters and assessing the model's fit with the data. As a consequence, theoretical models are often calibrated with little econometric guidance as to the value of key preference parameters, the extent to which the model explains the data relative to competing specifications, or the implications of the model's best-fitting specifications for other economic variables of interest, such as the return to the aggregate wealth portfolio or the return to human wealth. The purpose of this study is to help fill this gap in the literature by undertaking a formal econometric evaluation of the Epstein-Zin-Weil (EZW) recursive utility model.

The EZW recursive utility function is a constant elasticity of substitution (CES) aggregator over current consumption and the expected discounted utility of future consumption. This structure makes estimation of the general model difficult because the intertemporal marginal rate of substitution is a function of the unobservable continuation value of the future consumption plan. One approach to this problem, based on the insight of Epstein and

[1] See for example Campbell (1993); Campbell (1996); Tallarini (2000); Campbell and Viceira (2001) Bansal and Yaron (2004); Colacito and Croce (2004); Bansal, Dittmar, and Kiku (2005); Campbell and Voulteenaho (2005); Gomes and Michaelides (2005); Krueger and Kubler (2005); Hansen, Heaton, and Li (2005); Kiku (2005); Malloy, Moskowitz, and Vissing-Jorgensen (2005); Campanale, Castro, and Clementi (2006); Croce (2006); Bansal, Dittmar, and Lundblad (2006); Croce, Lettau, and Ludvigson (2006); Hansen and Sargent (2006); Piazzesi and Schneider (2006).

Zin (1989), is to exploit the relation between the continuation value and the return on the aggregate wealth portfolio. To the extent that the return on the aggregate wealth portfolio can be measured or proxied, the unobservable continuation value can be substituted out of the marginal rate of substitution and estimation can proceed using only observable variables (e.g., Epstein and Zin (1991), Campbell (1996), Vissing-Jorgensen and Attanasio (2003)).[2] Unfortunately, the aggregate wealth portfolio represents a claim to future consumption and is itself unobservable. Moreover, given the potential importance of human capital and other nontradable assets in aggregate wealth, its return may not be well proxied by observable asset market returns.

These difficulties can be overcome in specific cases of the EZW recursive utility model. For example, if the EIS is restricted to unity and consumption follows a loglinear time-series process, the continuation value has an analytical solution and is a function of observable consumption data (e.g., Hansen, Heaton, and Li (2005)). Alternatively, if consumption and asset returns are assumed to be jointly lognormally distributed and homoskedastic (or if a second-order linearization is applied to the Euler equation), the risk premium of any asset can be expressed as a function of covariances of the asset's return with current consumption growth and with news about future consumption growth (e.g., Restoy and Weil (1998), Campbell (2003)). In this case, the model's cross-sectional asset pricing implications can be evaluated using observable consumption data and a model for expectations of future consumption.

While the study of these specific cases has yielded a number of important insights, there are several reasons why it may be desirable to allow for more general representations of the model, free from tight parametric or distributional assumptions. First, an EIS of unity implies that the consumption-wealth ratio is constant, contradicting statistical evidence that it varies considerably over time.[3] Moreover, even first-order expansions of the EZW model

[2] Epstein and Zin (1991) use an aggregate stock market return to proxy for the aggregate wealth return. Campbell (1996) assumes that the aggregate wealth return is a portfolio weighted average of a human capital return and a financial return, and obtains an estimable expression for an approximate loglinear formulation of the model by assuming that expected returns on human wealth are equal to expected returns on financial wealth. Vissing-Jorgensen and Attanasio (2003) follow Campbell's approach to estimate the model using household level consumption data.

[3] Lettau and Ludvigson (2001a) argue that a cointegrating residual for log consumption, log asset wealth, and log labor income should be correlated with the unobservable log consumption-aggregate wealth ratio, and find evidence that this residual varies considerably over time and forecasts future stock market returns. See also recent evidence on the consumption-wealth ratio in Hansen, Heaton, Roussanov, and Lee (2006) and Lustig, Van Nieuwerburgh, and Verdelhan (2007).

around an EIS of unity may not capture the magnitude of variability of the consumption-wealth ratio (Hansen, Heaton, Roussanov, and Lee (2006)). Second, although aggregate consumption growth itself appears to be well described by a lognormal process, empirical evidence suggests that the joint distribution of consumption and asset returns exhibits significant departures from lognormality (Lettau and Ludvigson (2005)). Third, Kocherlakota (1990) points out that joint lognormality is inconsistent with an individual maximizing a utility function that satisfies the recursive representation used by Epstein and Zin (1989, 1991) and Weil (1989).

To overcome these issues, we employ a semiparametric estimation technique that allows us to conduct estimation and testing of the EZW recursive utility model without the need to find a proxy for the unobservable aggregate wealth return, without linearizing the model, and without placing tight parametric restrictions on either the law of motion or joint distribution of consumption and asset returns, or on the value of key preference parameters such as the EIS. We present estimates of all the preference parameters of the EZW model, evaluate the model's ability to fit asset return data relative to competing asset pricing models, and investigate the implications of such estimates for the unobservable aggregate wealth return and human wealth return.

To avoid having to find a proxy for the return on the aggregate wealth portfolio, we explicitly estimate the unobservable continuation value of the future consumption plan. By assuming that consumption growth falls within a general class of stationary, dynamic models, we may identify the state variables over which the continuation value is defined. However, without placing tight parametric restrictions on the model, the continuation value is still an unknown function of the relevant state variables. Thus, we estimate the continuation value function nonparametrically. The resulting empirical specification for investor utility is semiparametric in the sense that it contains both the finite dimensional unknown parameters that are part of the CES utility function (risk aversion, EIS, and subjective time-discount factor), as well as the infinite dimensional unknown continuation value function.

Estimation and testing are conducted by applying a profile Sieve Minimum Distance (SMD) procedure to a set of Euler equations corresponding to the EZW utility model we study. The SMD method is a distribution-free minimum distance procedure, where the conditional moments associated with the Euler equations are directly estimated nonparametrically as functions of conditioning variables. The "sieve" part of the SMD procedure requires that the unknown function embedded in the Euler equations (here the continuation value function) be approximated by a sequence of flexible parametric functions, with the number of parameters expanding as the sample size grows (Grenander (1981)). The unknown parameters of the marginal rate of substitution, including the sieve parameters of the continuation value function and the finite-dimensional parameters that are part of the CES utility function, may then be estimated using a profile two-step minimum distance estimator. In the first step, for arbitrarily fixed candidate finite dimensional parameter values, the sieve parameters are estimated by minimizing a weighted quadratic distance from zero of the nonparametrically estimated conditional moments. In the second step, consistent estimates of the finite dimensional parameters are obtained by solving a suitable sample minimum distance problem. Motivated by the arguments of Hansen and Jagannathan (1997), our asymptotic justification allows for possible model misspecification in the sense that the Euler equation may not hold exactly.

We estimate two versions of the model. The first is a representative agent formulation, in which the utility function is defined over per capita aggregate consumption. The second is a representative stockholder formulation, in which utility is defined over per capita consumption of stockholders. The definition of stockholder status, the consumption measure, and the sample selection follow Vissing-Jorgensen (2002), which uses the Consumer Expenditure Survey (CEX). Since CEX data are limited to the period 1982 to 2002, and since household-level consumption data are known to contain significant measurement error, we follow Malloy, Moskowitz, and Vissing-Jorgensen (2005) and generate a longer time-series of data by constructing consumption mimicking factors for aggregate stockholder consumption growth.

Once estimates of the continuation value function have been obtained, it is possible to investigate the model's implications for the aggregate wealth return. This return is in general unobservable but can be inferred from the model by equating the estimated marginal rate of substitution with its theoretical representation based on consumption growth and the return to aggregate wealth. If, in addition, we follow Campbell (1996) and assume that the return to aggregate wealth is a portfolio weighted average of the unobservable return to human wealth and the return to financial wealth, the estimated model also delivers implications for the return to human wealth.

Using quarterly data on consumption growth, assets returns and instruments, our empirical results indicate that the estimated relative risk aversion parameter is high, ranging from 17-60, with higher values for the representative agent version of the model than the representative stockholder version. The estimated elasticity of intertemporal substitution is typically above one, and differs considerably from the inverse of the coefficient of relative risk aversion. In addition, the estimated aggregate wealth return is found to be weakly correlated with the CRSP value-weighted stock market return and much less volatile, implying that the return to human capital is negatively correlated with the aggregate stock market return. This later finding is consistent with results in Lustig and Van Nieuwerburgh (2006), discussed further below. In data from 1952 to 2005, we find that an SMD estimated EZW recursive utility model can explain a cross-section of size and book-market sorted portfolio equity returns better than the time-separable, constant relative risk aversion power utility model and better than the Lettau and Ludvigson (2001b) *cay*-scaled consumption CAPM model, but not as well as purely empirical models based on financial factors such as the Fama and French (1993) three-factor model.

Our study is related to recent work estimating specific asset pricing models in which the EZW recursive utility function is embedded. Bansal, Gallant, and Tauchen (2004) and Bansal, Kiku, and Yaron (2006) estimate models of long-run consumption risk, where the data generating processes for consumption and dividend growth are explicitly modeled as linear functions of a small but very persistent long-run risk component and normally distributed shocks. These papers focus on the representative agent formulation of the model, in which utility is defined over per capita aggregate consumption. In such long-run risk models, the continuation value can be expressed as a function of innovations in the explicitly imposed driving processes for consumption and dividend growth, and inferred either by direct simulation or by specifying a vector autoregression to capture the predictable component. Our work differs from these studies in that our estimation procedure does not restrict the law of motion for consumption or dividend growth. As such, our estimates apply generally to the EZW recursive preference representation, not to specific asset pricing models of cash flow dynamics.

2 The Model

Let $\{\mathcal{F}_t\}_{t=0}^{\infty}$ denote the sequence of increasing conditioning information sets available to a representative agent at dates $t = 0, 1, ...$ Adapted to this sequence are consumption sequence $\{C_t\}_{t=0}^{\infty}$ and a corresponding sequence of continuation values $\{V_t\}_{t=0}^{\infty}$. The date t consumption C_t and continuation value V_t are in the date t information set $\mathcal{F}_t$ (but are typically not in the date $t-1$ information set $\mathcal{F}_{t-1}$). Sometimes we use $E_t[\cdot]$ to denote $E[\cdot|\mathcal{F}_t]$, the conditional expectation with respect to information set at date t.

The Epstein-Zin-Weil objective function is defined recursively by

$$V_t = \left[(1-\beta)\,C_t^{1-\rho} + \beta\{\mathcal{R}_t\,(V_{t+1})\}^{1-\rho}\right]^{\frac{1}{1-\rho}} \tag{1}$$

$$\mathcal{R}_t\,(V_{t+1}) = \left(E\left[V_{t+1}^{1-\theta}|\mathcal{F}_t\right]\right)^{\frac{1}{1-\theta}}, \tag{2}$$

where V_{t+1} is the continuation value of the future consumption plan. The parameter θ governs relative risk aversion and $1/\rho$ is the elasticity of intertemporal substitution over consumption (EIS). When $\theta = \rho$, the utility function can be solved forward to yield the familiar time-separable, constant relative risk aversion (CRRA) power utility model

$$V_t = \beta \frac{C_t^{1-\theta}}{1-\theta}. \tag{3}$$

As in Hansen, Heaton, and Li (2005), the utility function may be rescaled and expressed as a function of stationary variables:

$$\begin{aligned} \frac{V_t}{C_t} &= \left[(1-\beta) + \beta\left\{\mathcal{R}_t\left(\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}\right)\right\}^{1-\rho}\right]^{\frac{1}{1-\rho}} \\ &= \left[(1-\beta) + \beta\left\{E_t\left[\left(\frac{V_{t+1}}{C_{t+1}}\right)^{1-\theta}\left(\frac{C_{t+1}}{C_t}\right)^{1-\theta}\right]\right\}^{\frac{1-\rho}{1-\theta}}\right]^{\frac{1}{1-\rho}}. \end{aligned} \tag{4}$$

The intertemporal marginal rate of substitution (MRS) in consumption is given by

$$M_{t+1} = \beta\left(\frac{C_{t+1}}{C_t}\right)^{-\rho}\left(\frac{\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}}{\mathcal{R}_t\left(\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}\right)}\right)^{\rho-\theta}. \tag{5}$$

The MRS is a function of $\mathcal{R}_t\,(\cdot)$, the expected value of the continuation value-consumption ratio, $\frac{V_{t+1}}{C_{t+1}}$, referred to hereafter as the *continuation value ratio.*

Epstein and Zin (1989, 1991) show that the MRS can be expressed in an alternate form as

$$M_{t+1} = \left\{\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\rho}\right\}^{\frac{1-\theta}{1-\rho}}\left\{\frac{1}{R_{w,t+1}}\right\}^{\frac{\theta-\rho}{1-\rho}}, \tag{6}$$

where $R_{w,t+1}$ is the return to aggregate wealth, which represents a claim to future consumption. This return is in general unobservable, but some researchers have undertaken empirical work using an aggregate stock market return as a proxy, as in Epstein and Zin (1991). A difficulty with this approach is that $R_{w,t+1}$ may not be well proxied by observable asset market returns, especially if human wealth and other nontradable assets are quantitatively important fractions of aggregate wealth. Alternatively, approximate loglinear formulations of the model can be obtained by making specific assumptions regarding the relation between the return to human wealth and the return to some observable form of asset wealth. For example, Campbell (1996) assumes that expected returns on human wealth are equal to expected returns on financial wealth. Since the return to human wealth is unobservable, however, such assumptions are difficult to verify in the data. Consequently, we work with the formulation of the MRS given in (5), with its explicit dependence on the continuation value of the future consumption plan.

The first-order conditions for optimal consumption choice imply that $E_t\left[M_{t+1}R_{i,t+1}\right] = 1$, for any traded asset indexed by i, with a gross return at time $t+1$ of $R_{i,t+1}$. Using (5), the first-order conditions take the form

$$E_t\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\rho}\left(\frac{\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}}{\mathcal{R}_t\left(\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}\right)}\right)^{\rho-\theta}R_{i,t+1}-1\right]=0. \tag{7}$$

Since the expected product of any traded asset return with M_{t+1} equals one, the model implies that M_{t+1} is the stochastic discount factor (SDF), or pricing kernel, for valuing any traded asset return.

Equation (7) is a cross-sectional asset pricing model; it states that the risk premium on any traded asset return $R_{i,t+1}$ is determined in equilibrium by the covariance between returns and the stochastic discount factor M_{t+1}. Notice that, compared to the CRRA model where consumption growth is the single risk factor, the EZW model adds a second risk factor for explaining the cross-section of asset returns, given by the multiplicative term $\left(\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}/\mathcal{R}_t\left(\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}\right)\right)^{\rho-\theta}$.

The moment restrictions (7) are complicated by the fact that the conditional mean is taken over a highly nonlinear function of the conditionally expected value of discounted continuation utility, $\mathcal{R}_t\left(\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}\right)$. However, both the rescaled utility function (4) and the Euler equations (7) depend on $\mathcal{R}_t$. Thus, equation (4) can be solved for $\mathcal{R}_t$, and the solution plugged into (7). The resulting expression, for any observed sequence of traded asset returns $\{R_{i,t+1}\}_{i=1}^{N}$, takes the form

$$E_t\left[\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\rho}\left(\frac{\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}}{\left\{\frac{1}{\beta}\left[\left(\frac{V_t}{C_t}\right)^{1-\rho}-(1-\beta)\right]\right\}^{\frac{1}{1-\rho}}}\right)^{\rho-\theta}R_{i,t+1}-1\right]=0 \qquad i=1,...,N. \tag{8}$$

The moment restrictions (8) form the basis of our empirical investigation.

2.1 A nonparametric specification of $\frac{V_{t+1}}{C_{t+1}}$

To avoid having to find a proxy for the return on the aggregate wealth portfolio, we explicitly estimate the unobservable continuation value ratio $\frac{V_{t+1}}{C_{t+1}}$. To do so, we assume that consumption growth falls within a general class of stationary, dynamic models, thereby allowing us to identify the state variables over which the continuation value ratio is defined. Several examples of this approach are given in Hansen, Heaton, and Li (2005). Here, we assume that consumption growth is a possibly nonlinear function of a hidden first-order Markov process x_t that summarizes information about future consumption growth. Let lower case letters denote log variables, e.g., $\ln(C_{t+1}) \equiv c_{t+1}$. As a special case, consumption growth may be a linear function of a hidden first-order Markov process x_t

$$c_{t+1} - c_t = \mu + Hx_t + \mathbf{C}\boldsymbol{\epsilon}_{t+1}, \tag{9}$$

$$x_{t+1} = \phi x_t + \mathbf{D}\boldsymbol{\epsilon}_{t+1}, \tag{10}$$

where $\boldsymbol{\epsilon}_{t+1}$ is a (2×1) i.i.d. vector with mean zero and identity covariance matrix I and $\mathbf{C}$ and $\mathbf{D}$ are (1×2) vectors. Notice that this allows shocks in the observation equation (9) to have arbitrary correlation with those in the state equation (10). The specification (9)-(10) nests a number of stationary univariate representations for consumption growth, including a first-order autoregression, first-order moving average representation, a first-order autoregressive-moving average process, or $ARMA(1,1)$, and $i.i.d.$ The asset pricing literature on long-run consumption risk restricts to a special case of the above, where the innovations in (9) and (10) are uncorrelated and ϕ is close to unity (e.g., Bansal and Yaron (2004)).

More generally, we can allow consumption growth to be a potentially nonlinear function of a hidden Markov process x_t:

$$c_{t+1} - c_t = h(x_t) + \epsilon_{c,t+1} \tag{11}$$

$$x_{t+1} = \psi(x_t) + \epsilon_{x,t+1}, \tag{12}$$

where $h(x_t)$ and $\psi(x_t)$ are no longer necessarily linear functions of the state variable x_t, and $\epsilon_{c,t+1}$ and $\epsilon_{x,t+1}$ are i.i.d. random variables that may be correlated with one another.

In either case, given the first-order Markov structure, expected future consumption growth is summarized by the single state variable x_t, implying that x_t also summarizes the state space over which the function $\frac{V_t}{C_t}$ is defined. Notice that while we use the first-order Markov assumption as a motivation for specifying the state space over which continuation utility is defined, as discussed below, the econometric methodology itself leaves the law of motion of the consumption process unspecified.

There are two remaining complications that must be addressed before estimation can be undertaken. First, without placing tight parametric restrictions on the model, the continuation value ratio is an unknown function of the relevant state variables. Thus, we estimate $\frac{V_t}{C_t}$ nonparametrically. Second, the state variable x_t that is taken as the input of the unknown function is itself unobservable and must be inferred from consumption data. In the Appendix, we provide assumptions under which the first-order Markov structure in either (9)-(10) or (11)-(12) implies that the information contained in x_t is summarized by the lagged continuation value ratio $\frac{V_{t-1}}{C_{t-1}}$ and current consumption growth $\frac{C_t}{C_{t-1}}$. It follows that $\frac{V_t}{C_t}$ may be modeled as an unknown function F: $\mathbb{R}^2 \to \mathbb{R}$ of the lagged continuation value ratio and consumption growth:

$$\frac{V_t}{C_t} = F\left(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}}\right). \tag{13}$$

Observe that if the innovations in (9) and (10) are positively correlated, $\frac{V_t}{C_t}$ may display negative serial dependence, and we expect $F_1\left(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}}\right) < 0$, where $F_1(\cdot)$ denotes the partial derivative of F with respect to its first argument. In addition, although the linear specification (9)-(10) implies that F is a monotonic function of both arguments, if the stochastic process is nonlinear in x_t, as in (11)-(12), the function F can take on more general functional forms, potentially displaying nonmonotonicity in both its arguments.

To summarize, the asset pricing model we shall entertain in this paper consists of the conditional moment restrictions (8), subject to the nonparametric specification of (13). Our model is semiparametric in the sense that it contains both finite dimensional and infinite dimensional unknown parameters. Let $\boldsymbol{\delta} \equiv (\beta, \rho, \theta)'$ denote any vector of finite dimensional parameters in $\mathcal{D}$, a compact subset in $\mathbb{R}^3$, and F: $\mathbb{R}^2 \to \mathbb{R}$ denote any real-valued Lipschitz continuous functions in $\mathcal{V}$, a compact subset in the space of square integrable functions (with respect to some sigma-finite measure). For each $i = 1, ..., N$, denote

$$\gamma_i(\mathbf{z}_{t+1}, \boldsymbol{\delta}, F) \equiv \beta\left(\frac{C_{t+1}}{C_t}\right)^{-\rho}\left(\frac{F\left(\frac{V_t}{C_t}, \frac{C_{t+1}}{C_t}\right)\frac{C_{t+1}}{C_t}}{\left\{\frac{1}{\beta}\left[\left\{F\left(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}}\right)\right\}^{1-\rho} - (1-\beta)\right]\right\}^{\frac{1}{1-\rho}}}\right)^{\rho-\theta} R_{i,t+1} - 1,$$

where $\mathbf{z}_{t+1}$ is a vector containing all the strictly stationary observations, including consumption growth rate and return data. We define $\boldsymbol{\delta}_o \equiv (\beta_o, \rho_o, \theta_o)' \in \mathcal{D}$ and $F_o \equiv F_o(\mathbf{z}_t; \boldsymbol{\delta}_o) \equiv$

$F_o(\cdot;\boldsymbol{\delta}_o) \in \mathcal{V}$ as the solutions to

$$
\begin{aligned}
F_o(\cdot;\boldsymbol{\delta}) &= \arg\inf_{F\in\mathcal{V}} E\left[\sum_{i=1}^{N}\left(E\left\{\gamma_i(\mathbf{z}_{t+1},\boldsymbol{\delta},F)|\mathcal{F}_t\right\}\right)^2\right], \\
\boldsymbol{\delta}_o &= \arg\min_{\boldsymbol{\delta}\in\mathcal{D}} E\left[\sum_{i=1}^{N}\left(E\left\{\gamma_i(\mathbf{z}_{t+1},\boldsymbol{\delta},F_o(\cdot;\boldsymbol{\delta}))|\mathcal{F}_t\right\}\right)^2\right].
\end{aligned}
$$

We say that the model (8) and (13) is correctly specified if

$$E\left\{\gamma_i(\mathbf{z}_{t+1},\boldsymbol{\delta}_o,F_o(\cdot,\boldsymbol{\delta}_o))|\mathcal{F}_t\right\} = 0, \qquad i = 1,...,N. \tag{14}$$

3 Empirical Implementation

This section presents the details of our empirical procedure. The general methodology is based on estimation of the conditional moment restrictions (14), except that we allow for the possibility that the model could be misspecified.

The potential role of model misspecification in the evaluation of empirical asset pricing models has been previously emphasized by Hansen and Jagannathan (1997). As Hansen and Jagannathan stress, all models are approximations of reality and therefore potentially misspecified. The estimation procedure used here explicitly takes this possibility into account in its asymptotic justification. In the application of this paper, there are several possible reasons for misspecification, including possible misspecification of the arguments in the continuation value-consumption ratio function F, which could in principal include more lags, and misspecification of the arguments of the CES utility function, which could in principal include a broader measure of durable consumption or leisure. More generally, when we conduct model comparison in Section 5, we follow the advice of Hansen and Jagannathan (1997) and assume that all models are potentially misspecified.

Let $\mathbf{w}_t$ be a $d_w \times 1$ observable measurable function of $\mathcal{F}_t$ that does not contain a constant. Equation (14) implies

$$E\left\{\gamma_i(\mathbf{z}_{t+1},\boldsymbol{\delta}_o,F_o(\cdot,\boldsymbol{\delta}_o))|\mathbf{w}_t\right\} = 0, \qquad i = 1,...,N. \tag{15}$$

Denote[4]

$$m(\mathbf{w}_t,\boldsymbol{\delta},F) \equiv E\{\gamma(\mathbf{z}_{t+1},\boldsymbol{\delta},F)|\mathbf{w}_t\}, \gamma(\mathbf{z}_{t+1},\boldsymbol{\delta},F) = (\gamma_1(\mathbf{z}_{t+1},\boldsymbol{\delta},F),...,\gamma_N(\mathbf{z}_{t+1},\boldsymbol{\delta},F))'. \tag{16}$$

[4]If the model of consumption dynamics specified above were literally true, the state variables $\frac{V_{t-1}}{C_{t-1}}$ and $\frac{C_t}{C_{t-1}}$ (and all measurable transformations of these) are sufficient statistics for the agents information set $\mathcal{F}_t$. However, the fundamental asset pricing relation $E_t\left[M_{t+1}R_{i,t+1}-1\right]$, which includes individual asset

For any candidate value $\boldsymbol{\delta} \equiv (\beta, \rho, \theta)' \in \mathcal{D}$, we define $F^* \equiv F^*(\mathbf{z}_t, \boldsymbol{\delta}) \equiv F^*(\cdot, \boldsymbol{\delta}) \in \mathcal{V}$ as the solution to

$$F^*(\cdot, \boldsymbol{\delta}) \equiv \underset{F \in \mathcal{V}}{\arg\inf} E\left[m(\mathbf{w}_t, \boldsymbol{\delta}, F)' m(\mathbf{w}_t, \boldsymbol{\delta}, F)\right]. \tag{17}$$

It is clear that $F_o(\mathbf{z}_t, \boldsymbol{\delta}_o) = F^*(\mathbf{z}_t, \boldsymbol{\delta}_o)$ when the model (15) is correctly specified. We say the model (15) is *misspecified* if

$$\min_{\boldsymbol{\delta} \in \mathcal{D}} \inf_{F \in \mathcal{V}} E\left[m(\mathbf{w}_t, \boldsymbol{\delta}, F)' m(\mathbf{w}_t, \boldsymbol{\delta}, F)\right] = \min_{\boldsymbol{\delta} \in \mathcal{D}} E\left[m(\mathbf{w}_t, \boldsymbol{\delta}, F^*(\mathbf{z}_t, \boldsymbol{\delta}))' m(\mathbf{w}_t, \boldsymbol{\delta}, F^*(\mathbf{z}_t, \boldsymbol{\delta}))\right] > 0.$$

We estimate the possibly misspecified model (15) using a *profile* semiparametric minimum distance procedure, which consists of two steps; see e.g., Andrews (1994), Newey and McFadden (1994), Chen, Linton, and van Keilegom (2003) and Chen (2006). In the first step, for any candidate value $\boldsymbol{\delta} \equiv (\beta, \rho, \theta)' \in \mathcal{D}$, the unknown function $F^*(\cdot, \boldsymbol{\delta})$ is estimated using the sieve minimum distance (SMD) procedure developed in Newey and Powell (2003) and Ai and Chen (2003) (for correctly specified model) and Ai and Chen (2007) (for possibly misspecified model). In the second step, we estimate the finite dimensional parameters $\boldsymbol{\delta}$ by solving a suitable sample GMM problem. We show in the Appendix that, under the assumption of strictly stationary weakly dependent observations, the first-step SMD estimator of $F^*(\cdot, \boldsymbol{\delta})$ is consistent and converges at a rate $T^{1/4}$ under certain metric, uniformly over $\boldsymbol{\delta} \equiv (\beta, \rho, \theta)' \in \mathcal{D}$, where T is the sample size. The second-step GMM estimates of the finite-dimensional parameters $\boldsymbol{\delta} \equiv (\beta, \rho, \theta)'$ are $\sqrt{T}$ consistent and asymptotically normally distributed. Notice that the estimation procedure itself leaves the law of motion of the data unspecified.[5]

returns, is likely to be a highly nonlinear function of the state variables. In addition, one of the these state variables is the unknown function, $\frac{V_{t-1}}{C_{t-1}}$, and as such it embeds the unknown sieve parameters. These facts make the estimation procedure computationally intractable if the subset $\mathbf{w}_t$, over which the conditional mean $m(\mathbf{w}_t, \boldsymbol{\delta}, F)$ is taken, includes $\frac{V_{t-1}}{C_{t-1}}$. Fortunately, the procedure can be carried out on an observable measurable function $\mathbf{w}_t$ of $\mathcal{F}_t$, which need not contain $\frac{V_{t-1}}{C_{t-1}}$. A consistent estimate of the conditional mean $m(\mathbf{w}_t, \boldsymbol{\delta}, F)$ can be obtained using known basis functions of observed conditioning variables in $\mathbf{w}_t$. We take this approach here, using $\frac{C_t}{C_{t-1}}$ and several other observable conditioning variables as part of the econometrician's information $\mathbf{w}_t$.

[5]In the Appendix we provide asymptotic results on nonparametric consistency and parametric $\sqrt{T}-$asymptotic normality for possibly misspecified semiparametric conditional moment models, allowing for strictly stationary beta-mixing time series observations. Beta-mixing is one popular measure of temporal dependence for nonlinear time series that is satisfied by many widely used financial time series models including nonlinear ARCH, GARCH, stochastic volatility and diffusion models–see the Appendix for the formal definition. Thus, the estimation procedure requires stationary ergodic observations but does not restrict to linear time series specifications or specific parametric laws of motions of the data.

3.1 First-Step Profile SMD Estimation of $F^*(\cdot,\boldsymbol{\delta})$

For any candidate value $\boldsymbol{\delta} = (\beta, \rho, \theta)' \in \mathcal{D}$, an initial estimate of the unknown function $F^*(\cdot,\boldsymbol{\delta})$ is obtained using the sieve minimum distance (SMD) estimator, described below. In practice, this is achieved by applying the SMD estimator at each point in a 3-dimensional grid for $\boldsymbol{\delta} \in \mathcal{D}$. The idea behind the SMD estimator is to choose a flexible approximation to the value function $F^*(\cdot,\boldsymbol{\delta})$ to minimize the sample analog of the minimum distance criterion function (17). More precisely, this procedure itself has two essential parts. First, although the functional form of the conditional expectation function $m(\mathbf{w}_t, \boldsymbol{\delta}, F)$ defined in (16) is unknown, we may replace the conditional expectation itself with a consistent nonparametric estimator (to be specified later). Second, although the value function $F^*(\cdot,\boldsymbol{\delta})$ is an infinite-dimensional unknown function, we can approximate it by a sequence of finite-dimensional unknown parameters (sieves) $F_{K_T}(\cdot,\boldsymbol{\delta})$, where the approximation error decreases as the dimension K_T increases with the sample size T. For each $\boldsymbol{\delta} \in \mathcal{D}$, the function $F_{K_T}(\cdot,\boldsymbol{\delta})$ is estimated by minimizing a sample (weighted) quadratic norm of the nonparametrically estimated conditional expectation functions.

Estimation in the first profile SMD step is carried out by implementing the following algorithm. First, the ratio $\frac{V_t}{C_t}$ is treated as unknown function $\frac{V_t}{C_t} = F^*\left(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}}; \boldsymbol{\delta}\right)$, with the initial value for $\frac{V_t}{C_t}$ at time $t = 0$, denoted $\frac{V_0}{C_0}$, taken as a unknown scalar parameter to be estimated. Second, the unknown function $F^*\left(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}}; \boldsymbol{\delta}\right)$ is approximated by a bivariate sieve function

$$F^*\left(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}}; \boldsymbol{\delta}\right) \approx F_{K_T}(\cdot,\boldsymbol{\delta}) = a_0(\boldsymbol{\delta}) + \sum_{j=1}^{K_T} a_j(\boldsymbol{\delta}) B_j\left(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}}\right),$$

where the sieve coefficients $\{a_0, a_1, ..., a_{K_T}\}$ depend on $\boldsymbol{\delta}$, but the sieve basis functions $\{B_j(\cdot,\cdot) : j = 1, ..., K_T\}$ have known functional forms that are independent of $\boldsymbol{\delta}$; see the Appendix for examples of the sieve basis functions $B_j(\cdot,\cdot)$. To provide a nonparametric estimate of the true unknown function, K_T must grow with the sample size to insure consistency of the method.[6] We are not interested in the sieve parameters $(a_0, a_1, ..., a_{K_T})'$ *per se*, but

[6] Asymptotic theory only provides guidance about the *rate* at which K_T must increase with the sample size T. Thus, in practice, other considerations must be used to judge how best to set this dimensionality. The bigger is K_T, 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 $K_T = 9$, the dimension of the parameter vector, $\boldsymbol{\alpha}$ along with $\frac{V_0}{C_0}$, is 11, estimated using a sample of size $T = 213$. In practice, we obtained very similar results setting $K_T = 10$; thus we present the results for the more parsimonious specification using $K_T = 9$ below.

rather in the finite dimensional parameters $\boldsymbol{\delta}$, and in the dynamic behavior of the continuation value and the marginal rate of substitution, all of which depend on those parameters. For the empirical application below, we set $K_T = 9$ (see the Appendix for further discussion), leaving 10 sieve parameters to be estimated in F^*, plus the initial value $\frac{V_0}{C_0}$. The total number of parameters to be estimated, including the three finite dimensional parameters in $\boldsymbol{\delta}$, is therefore 14.

Given values $\frac{V_0}{C_0}$, $\{a_j\}_{j=1}^{K_T}$, $\{B_j(\cdot)\}_{j=1}^{K_T}$ and data on consumption $\left\{\frac{C_t}{C_{t-1}}\right\}_{t=1}^{T}$, the function F_{K_T} is used to generate a sequence $\left\{\frac{V_i}{C_i}\right\}_{i=1}^{T}$ that can be taken as data to be used in the estimation of (17).

Implementation of the profile SMD estimation requires a consistent estimate of the conditional mean function $m(\mathbf{w}_t, \boldsymbol{\delta}, F)$, which can be consistently estimated via a sieve least squares procedure. Let $\{p_{0j}(\mathbf{w}_t),\ j = 1, 2, ..., J_T\}$ be a sequence of known basis functions (including a constant function) that map from $\mathbb{R}^{d_w}$ into $\mathbb{R}$. Denote $p^{J_T}(\cdot) \equiv (p_{01}(\cdot), ..., p_{0J_T}(\cdot))'$ and the $T \times J_T$ matrix $\mathbf{P} \equiv \left(p^{J_T}(\mathbf{w}_1), ..., p^{J_T}(\mathbf{w}_T)\right)'$. Then

$$\widehat{m}(\mathbf{w}, \boldsymbol{\delta}, F) = \left(\sum_{t=1}^{T} \gamma(\mathbf{z}_{t+1}, \boldsymbol{\delta}, F) p^{J_T}(\mathbf{w}_t)'(\mathbf{P}'\mathbf{P})^{-1}\right) p^{J_T}(\mathbf{w}) \tag{18}$$

is a sieve least squares estimator of the conditional mean vector $m(\mathbf{w}, \boldsymbol{\delta}, F) = E\{\gamma(\mathbf{z}_{t+1}, \boldsymbol{\delta}, F)|\mathbf{w}_t = \mathbf{w}\}$. (Note that J_T must grow with the sample size to ensure that $m(\mathbf{w}_t, \boldsymbol{\delta}, F)$ is estimated consistently). We form the first-step profile SMD estimate $\widehat{F}(\cdot)$ for $F^*(\cdot)$ based on this estimate of the conditional mean vector and the sample analog of (17):

$$\widehat{F}(\cdot, \boldsymbol{\delta}) = \arg\min_{F_{K_T}} \frac{1}{T} \sum_{t=1}^{T} \widehat{m}(\mathbf{w}_t, \boldsymbol{\delta}, F_{K_T})' \widehat{m}(\mathbf{w}_t, \boldsymbol{\delta}, F_{K_T}). \tag{19}$$

See the Appendix for a detailed description of the profile SMD procedure.

As shown in the Appendix, an attractive feature of this estimator is that it can be implemented as an instance of GMM with a particular weighting matrix $\mathbf{W}$ given by

$$\mathbf{W} = \mathbf{I}_N \otimes (\mathbf{P}'\mathbf{P})^{-1}.$$

The procedure is equivalent to regressing each γ_i on the set of instruments $p^{J_T}(\cdot)$ and taking the fitted values from this regression as an estimate of the conditional mean, where the particular weighting matrix gives greater weight to moments that are more highly correlated with the instruments $p^{J_T}(\cdot)$. The weighting scheme can be understood intuitively by noting that variation in the conditional mean is what identifies the unknown function $F^*(\cdot, \boldsymbol{\delta})$.

3.2 Second-Step GMM Estimation of $\boldsymbol{\delta}$

Once an initial nonparametric estimate $\widehat{F}(\cdot, \boldsymbol{\delta})$ is obtained for $F^*(\cdot, \boldsymbol{\delta})$, we can estimate the finite dimensional parameters $\boldsymbol{\delta}_o$ consistently by solving a suitable sample minimum distance problem, for example by using a Generalized Method of Moments (GMM, Hansen (1982)) estimator. An advantage of this two-step approach is that the second-stage estimation need not be based on the sample SMD criterion

$$\min_{\boldsymbol{\delta} \in \mathcal{D}} \frac{1}{T} \sum_{t=1}^{T} \widehat{m}(\mathbf{w}_t, \boldsymbol{\delta}, \widehat{F}(\cdot, \boldsymbol{\delta}))' \widehat{m}(\mathbf{w}_t, \boldsymbol{\delta}, \widehat{F}(\cdot, \boldsymbol{\delta})),$$

which gives greater weight to moments that are more highly correlated with the instruments $p^{J_T}(\cdot)$. Such a weighting scheme is required to identify the unknown function $F^*(\cdot, \boldsymbol{\delta})$, but is not required for pinning down the finite dimensional preference parameters $\boldsymbol{\delta}_o$. We discuss this further below.

Notice that if the number of test asset returns $N \geq 3$, consistent estimation of $\boldsymbol{\delta} = (\beta, \rho, \theta)'$could in principal be based on the unconditional population moments implied by (15):

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

More generally, minimum distance estimation of $\boldsymbol{\delta}_o$ based on the moment conditions (15) could be conducted using any subset of the conditioning variables that make up the econometrician's information set $\mathbf{w}_t$, as long as the number of moment conditions is at least as large as the number of finite dimensional parameters to be estimated. Let the conditioning variables used in the second-step estimation of $\boldsymbol{\delta}_o$ be denoted $\mathbf{x}_t$, where $\mathbf{x}_t$ is a $d_x \times 1$ vector that could include a constant. We estimate $\boldsymbol{\delta}_o$ by minimizing a GMM objective function:

$$\widehat{\boldsymbol{\delta}} = \arg\min_{\boldsymbol{\delta} \in \mathcal{D}} Q_T(\boldsymbol{\delta}), \tag{20}$$

$$Q_T(\boldsymbol{\delta}) = \left[\mathbf{g}_T(\boldsymbol{\delta}, \widehat{F}_T(\cdot, \boldsymbol{\delta}); \mathbf{y}^T)\right]' \mathbf{W} \left[\mathbf{g}_T(\boldsymbol{\delta}, \widehat{F}_T(\cdot, \boldsymbol{\delta}); \mathbf{y}^T)\right], \tag{21}$$

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

$$\mathbf{g}_T(\boldsymbol{\delta}, \widehat{F}(\cdot, \boldsymbol{\delta}); \mathbf{y}^T) \equiv \frac{1}{T} \sum_{t=1}^{T} \gamma(\mathbf{z}_{t+1}, \boldsymbol{\delta}, \widehat{F}(\cdot, \boldsymbol{\delta})) \otimes \mathbf{x}_t \tag{22}$$

are the sample moment conditions associated with the $Nd_x \times 1$ -vector of population unconditional moment conditions:

$$E\left\{\gamma_i(\mathbf{z}_{t+1}, \boldsymbol{\delta}_o, F^*(\cdot, \boldsymbol{\delta}_o)) \otimes \mathbf{x}_t\right\} = 0, \qquad i = 1, ..., N. \tag{23}$$

Observe that $\widehat{F}(\cdot, \boldsymbol{\delta})$ is not held fixed in the second step, but instead depends on $\boldsymbol{\delta}$. Consequently, the second-step GMM estimation of $\boldsymbol{\delta}$ plays an important role in determining the final estimate of $F_o(\cdot)$, denoted $\widehat{F}\left(\cdot, \widehat{\boldsymbol{\delta}}\right)$.

In the empirical implementation, we use two different weighting matrices $\mathbf{W}$ to obtain the second-step GMM estimates of $\boldsymbol{\delta}$. The first is the identity weighting matrix $\mathbf{W} = \mathbf{I}$; the second is the inverse of the sample second moment matrix of the N asset returns upon which the model is evaluated, denoted $\mathbf{G}_T^{-1}$ (i.e., the $(i, j)th$ element of $\mathbf{G}_T$ is $\frac{1}{T}\sum_{t=1}^{T} R_{i,t}R_{j,t}$ for $i, j = 1, ..., N$.)

To understand the motivation behind using $\mathbf{W} = \mathbf{I}$ and $\mathbf{W} = \mathbf{G}_T^{-1}$ to weight the second-step GMM criterion function, it is useful to first observe that, in principal, all the parameters of the model (including the finite dimensional preference parameters), could be estimated in one step by minimizing the sample SMD criterion:

$$\min_{\boldsymbol{\delta} \in \mathcal{D}, F_{K_T}} \frac{1}{T} \sum_{t=1}^{T} \widehat{m}(\mathbf{w}_t, \boldsymbol{\delta}, F_{K_T})' \widehat{m}(\mathbf{w}_t, \boldsymbol{\delta}, F_{K_T}). \tag{24}$$

However, the two-step profile procedure employed here has several advantages for our empirical application. First, we want estimates of standard preference parameters such as risk aversion and the EIS to reflect values required to match unconditional moments commonly emphasized in the asset pricing literature, those associated with unconditional risk premia. This is not possible when estimates of $\boldsymbol{\delta}$ and $F()$ are obtained in one step, since the weighting scheme inherent in the SMD procedure (24) emphasizes conditional moments, placing greater weight on moments that are more highly correlated with the instruments. Second, both the weighting scheme inherent in the SMD procedure (24) and the use of instruments $p^{J_T}(\cdot)$ effectively change the set of test assets, implying that key preference parameters are estimated on linear combinations of the original portfolio returns. Such linear combinations often bear little relation to the original test asset returns upon which much of the asset pricing literature has focused. They may also imply implausible long and short positions in the original test assets and do not necessarily deliver a large spread in unconditional mean returns. These concerns can be alleviated by estimating the finite dimensional parameters in a second step, using the identity weighting matrix $\mathbf{W} = \mathbf{I}$ along with $\mathbf{x}_t = \mathbf{1}_N$, an $N \times 1$ vector of ones..

We also use $\mathbf{W} = \mathbf{G}_T^{-1}$ along with $\mathbf{x}_t = \mathbf{1}_N$. Parameter estimates computed in this way have the advantage that they are obtained by minimizing an objective function that is invariant to the initial choice of asset returns (Kandel and Stambaugh (1995)). In addition, the square root of the minimized GMM objective function has the appealing interpretation as the maximum pricing error per unit norm of any portfolio of the original test assets, and serves as a measure of model misspecification (Hansen and Jagannathan (1997)). We use this below to compare the performance of the estimated EZW model to that of competing asset pricing models.

3.3 Decision Interval of Household

We model the decision interval of the household at fixed horizons and measure consumption and returns over the same horizon. In reality, the decision interval of the household may differ from the data sampling interval. If the decision interval of the household is shorter than the data sampling interval, the consumption data are time aggregated. Heaton (1993) studies the effects of time aggregation in a consumption based asset pricing model with habit formation, and concludes, based on a first-order linear approximation of the Euler equation, that time aggregation can bias GMM parameter estimates of the habit coefficient. The extent to which time aggregation may influence parameter estimates in nonlinear Euler equation estimation is not generally known.

In practice, it is difficult or impossible to assess the extent to which time aggregation is likely to bias parameter estimates, for several reasons. First, the decision interval of the household is not directly observable. Time aggregation arises only if the decision interval of the household is shorter than the data sampling interval. Recently, several researchers have argued that the decision interval of the household may in fact be longer than the monthly, quarterly, or annual data sampling intervals typically employed in empirical work (Gabaix and Laibson (2002), Jagannathan and Wang (2007)). In this case, time aggregation is absent and has no influence on parameter estimates. Second, even if consumption data are time aggregated, its influence on parameter estimates is likely to depend on a number of factors that are difficult to evaluate in practice, such as the stochastic law of motion for consumption growth, and the degree to which the interval for household decisions falls short of the data sampling interval.

If time-aggregation is present, however, it may induce a spurious correlation between the estimated error terms over which conditional means are taken $(\gamma_i(\mathbf{z}_{t+1}, \boldsymbol{\delta}_o, F_o(\cdot, \boldsymbol{\delta}_o))$, above), and the information set at time t $(\mathbf{w}_t)$. Therefore, as a precaution, we conduct our empirical estimation using instruments at time t that do not admit the most recent lagged values of the variables (i.e., using two-period lagged instruments instead of one-period lagged instruments). The cost of doing so is that the two-period lagged instruments may not be as informative as the one-period lagged instruments; this cost is likely to be small, however, if the instruments are serially correlated, as are a number of those employed here (see the next section).

4 Data

A detailed description of the data and our sources is provided in the Appendix. Our aggregate data are quarterly, and span the period from the first quarter of 1952 to the first quarter of 2005.

The focus of this paper is on testing the model's theoretical restrictions for a cross-sections of asset returns. If the theory is correct, the cross-sectional asset pricing model (7) should be informative about the model's key preference parameters as well as about the unobservable continuation value function. Specifically, the first-order conditions for optimal consumption choice place tight restrictions both across assets and over time on equilibrium asset returns. Consequently, we study a cross-section of asset returns known to deliver a large spread in mean returns, which have been particularly challenging for classic asset pricing models to explain (Fama and French (1992) and Fama and French (1993)). These assets include the three-month Treasury bill rate and six value-weighted portfolios of common stock sorted into two size quantiles and three book value-market value quantiles, for a total of 7 asset returns. All stock return data are taken from Kenneth French's Dartmouth web page (URL provided in the appendix), created from stocks traded on the NYSE, AMEX and NASDAQ.

To estimate the representative agent formulation of the model, we use real, per-capita expenditures on nondurables and services as a measure of aggregate consumption. Since consumption is real, our estimation uses real asset returns, which are the nominal returns described above deflated by the implicit chain-type price deflator to measure consumption. We use quarterly consumption data because it is known to contain less measurement error than monthly consumption data.

We also construct a stockholder consumption measure to estimate the representative stockholder version of the model. The definition of stockholder status, the consumption measure, and the sample selection follow Vissing-Jorgensen (2002), which uses the Consumer Expenditure Survey (CEX). Since CEX data are limited to the period 1980 to 2002, and since household-level consumption data are known to contain significant measurement error, we follow Malloy, Moskowitz, and Vissing-Jorgensen (2005) and generate a longer time-series of data by constructing consumption mimicking factors for aggregate stockholder consumption growth. The CEX interviews households three months apart and households are asked to report consumption for the previous three months. Thus, while each household is interviewed three months apart, the interviews are spread out over the quarter implying that there will be households interviewed in each month of the sample. This permits the computation of quarterly consumption growth rates at a monthly frequency. As in Malloy, Moskowitz, and Vissing-Jorgensen (2005), we construct a time series of average consumption growth for stockholders from t to $t+1$ as

$$\frac{1}{H}\sum_{h=1}^{H}\frac{C_{t+1}^{h}}{C_{t}^{h}},$$

where C_{t+1}^{h} is the quarterly consumption of household h for quarter t and H is the number of stockholder households in quarter t. We use this average series to form a mimicking factor for stockholder consumption growth, by regressing it on aggregate variables (available at monthly frequency) and taking the fitted values as a measure of the mimicking factor for stockholder consumption growth.

Mimicking factors for stockholder consumption growth are formed for two reasons. First, the household level consumption data are known to be measured with considerable error, mostly driven by survey error. To the extent that measurement error is uncorrelated with aggregate variables, the mimicking factor will be free of the survey measurement error present in the household level consumption series. Second, since the CEX sample is short (1982 to 2002), the construction of mimicking factors allows a longer time-series of data to be constructed. The procedure follows Malloy, Moskowitz, and Vissing-Jorgensen (2005). We project the average consumption growth of stockholders on a set of instruments (available over a longer period) and use the estimated coefficients to construct a longer time-series of stockholder consumption growth, spanning the same sample as the aggregate consumption data. As instruments, we use two aggregate variables that display significant correlation with average stockholder consumption growth: the log difference of industrial production growth, $\Delta\ln(IP_t)$, and the log differences of real services expenditure growth, $\Delta\ln(SV_t)$. The regression is estimated using monthly data from July 1982 to February 2002, using the average CEX stockholder consumption growth rates. The fitted values from these regressions provide monthly observations on a mimicking factor for the quarterly consumption growth of stockholders. The results from this regression, with Newey and West (1987) t-statistics, are reported in Table 1. Average stockholder consumption growth is positively related to both the growth in industrial production, and to the growth in expenditures on services. Each variable has a statistically significant effect on average stockholder consumption growth, though the R^2 statistics are modest. The modest R^2 statistics are not surprising given the substantial amount of measurement error in household-level consumption data (for example, comparable R^2 values can be found in Malloy, Moskowitz, and Vissing-Jorgensen (2005)).

For the subsequent empirical analysis, we construct a quarterly measure of the stockholder consumption growth mimicking factor by matching the fitted values for quarterly consumption growth over the three consecutive months corresponding to the three months in a quarter (e.g., we use the observation on fitted consumption growth from March to January in a given year as a measure of first quarter consumption growth in that year). We refer the reader to Vissing-Jorgensen (2002) and Malloy, Moskowitz, and Vissing-Jorgensen (2005) for further details on the CEX data and the construction of mimicking factors.

The empirical procedure also requires computation of instruments, $p^{J_T}(\mathbf{w}_t)$, which are known basis functions (including a constant function) of conditioning variables, $\mathbf{w}_t$. We include lagged consumption growth in $\mathbf{w}_t$, as well as three variables that have been shown elsewhere to have significant forecasting power for excess stock returns and consumption growth in quarterly data.[7] Two variables that have been found to display forecasting power for excess stock returns at a quarterly frequency are the "relative T-bill rate" (which we measure as the three month Treasury-bill rate minus its 4-quarter moving average), and the lagged value of the excess return on the Standard & Poor 500 stock market index (S&P 500) over the three-month Treasury bill rate (see Campbell (1991), Hodrick (1992), Lettau and Ludvigson (2001a)). We denote the relative bill rate $RREL$ and the excess return on the S&P 500 index, $SPEX$.[8] We also use the proxy for the log consumption-wealth ratio studied in (Lettau and Ludvigson (2001a)) to forecast returns.[9] This proxy is measured as the cointegrating residual between log consumption, log asset wealth, and log labor income

[7] The importance of instrument relevance in a GMM setting (i.e., using instruments that are sufficiently correlated with the included endogenous variables) is now well understood. See Stock, Wright, and Yogo (2002) for a survey of this issue. No formal test of instrument relevance has been developed for estimation involving an unknown function. Thus we choose variables for $\mathbf{w}_t$ that are known to be strong predictors of asset returns and consumption growth in quarterly data.

[8] We focus on these variables rather than some others because, in samples that include recent data, they drive out many of the other popular forecasting variables for stock returns, such as an aggregate dividend-price ratio, earnings-price ratio, term spreads and default spreads (Lettau and Ludvigson (2001a)).

[9] This variable has strong forecasting power for stock returns over horizons ranging from one quarter to several years. Lettau and Ludvigson (2001b) report that this variable also forecasts returns on portfolios sorted by size and book-market ratios.

and is denoted $\widehat{cay}_t$.[10] Lettau and Ludvigson (2004) find that quarterly consumption growth is predictable by one lag of wealth growth, a variable that is highly correlated with $SPEX$, and results (not reported) confirm that it is also predictable by one lag of $SPEX$. Thus, we use $\mathbf{w}_t = \left[\widehat{cay}_t, RREL_t, SPEX_t, \frac{C_t}{C_{t-1}}\right]'$. We note that consumption growth–often thought to be nearly unforecastable–displays a fair amount of short-horizon predictability in the sample used here: a linear regression of consumption growth on the one-period lagged value $\mathbf{w}_t$ and a constant produces an $F-$statistic for the regression in excess of 12.[11]

Since the error term $\gamma_i(\mathbf{z}_{t+1}, \boldsymbol{\delta}_o, F_o)$ is orthogonal to the information set $\mathbf{w}_t$, any measurable transformation of $\mathbf{w}_t$, $p^{J_T}(\mathbf{w}_t)$, can be used as valid instruments in the first-step estimation of F_o. We use power series as instruments, where the specification includes a constant, the linear terms, squared terms and pair-wise cross products of each variable in $\mathbf{w}_t$, or 15 instruments in total.

5 Empirical Results

5.1 Parameter Estimates

The shape of our estimated continuation value ratio function $\frac{V_t}{C_t} = F\left(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}}\right)$ can be illustrated by plotting $\widehat{F}\left(\cdot, \widehat{\boldsymbol{\delta}}\right)$ as a function of $\frac{V_{t-1}}{C_{t-1}}$, holding fixed current consumption growth, $\frac{C_t}{C_{t-1}}$. Figures 1 and 2 plot this relation for each estimation described above, using aggregate consumption (Figure 1) or the stockholder mimicking factor as a measure of stockholder consumption (Figure 2). For these plots, $\frac{V_{t-1}}{C_{t-1}}$ varies along the horizontal axis, with $\frac{C_t}{C_{t-1}}$ alternately held fixed at its median, 25th, and 75th percentile values in our sample.

We draw several conclusions from the figures. First, the estimated continuation value-consumption ratio function is nonlinear; this is evident from the curved shape of the functions and from the finding that the shape depends on where in the domain space the function is evaluated. In particular, for the representative agent version of the model (Figure 1), the serial dependence of $\widehat{F}$ depends on where in the domain space the function is evaluated. It is negative for low values of $\frac{V_{t-1}}{C_{t-1}}$ and positive for high values. Such a nonmonotone pattern is

[10] See Lettau and Ludvigson (2001a) and Lettau and Ludvigson (2004) for further discussion of this variable and its relation to the log consumption-wealth ratio. Note that standard errors do not need to be corrected for pre-estimation of the cointegrating parameters in $\widehat{cay}_t$, since cointegrating coefficients are "superconsistent," converging at a rate faster than the square root of the sample size.

[11] As recommended by Cochrane (2001), the conditioning variables in $\mathbf{w}_t$ are normalized by standardizing and adding one to each variable, so that they have roughly the same units as unscaled returns.

possible, for example, under the functionally non-linear state space model (11)-(12). Negative serial dependence can arise even in the linear state space model, if the innovation in the observation equation (9) is correlated with the innovation in the state equation (10). Second, the estimated continuation value ratio is increasing in current consumption growth, in both the representative agent (Figure 1) and representative stockholder (Figure 2) versions of the model. The estimated relation is, however, nonlinear in consumption growth, a finding that is especially evident in Figure 2. Third, in the representative stockholder version of the model (Figure 2), the serial dependence of $\widehat{F}$ is negative over most of the domain space.

The shapes of the estimated continuation value ratio functions imply that the functionally linear state space representation (9)-(10) commonly employed in asset pricing models may not provide a good description of these data. For example, as Hansen, Heaton, and Li (2005) show, if (9)-(10) holds and the EIS=1, $\log(\frac{V_{t-1}}{C_{t-1}})$ is linear in the state x_t. Since the log of $\widehat{F}\left(\cdot, \widehat{\boldsymbol{\delta}}\right)$ is clearly nonlinear, the findings suggest that a linear state space representation, in conjunction with an EIS=1, is unlikely to provide an accurate description of the data. In addition, the nonmonotonicity of $\widehat{F}\left(\cdot, \widehat{\boldsymbol{\delta}}\right)$ over its first argument is also inconsistent with the linear state space representation, though nonmonotonicities are possible with a functionally nonlinear state space representation as in (11)-(12).

Table 2 presents estimates of the model's preference parameters $\boldsymbol{\delta} = (\beta, \rho, \theta)'$. The subjective time-discount factor, β, is close to one in each estimation, with values between 0.99 and 0.999, depending on the measure of consumption and the weighting matrix employed in the second step ($\mathbf{W} = \mathbf{I}$ or $\mathbf{W}{=}\mathbf{G}_T^{-1}$). The estimated relative risk aversion parameter θ ranges from 17-60, with higher values for the representative agent version of the model than the representative stockholder version. For example, using aggregate consumption data, estimated risk aversion is around 60, regardless of which estimation is employed in the second step ($\mathbf{W} = \mathbf{I}$ or $\mathbf{W}{=}\mathbf{G}_T^{-1}$). By contrast, estimated risk aversion is either 20 or 17 when we use the stockholder mimicking factor as a measure of stockholder consumption. The finding that estimated risk aversion is higher for the model with aggregate consumption than for that with stockholder consumption is consistent with results in Malloy, Moskowitz, and Vissing-Jorgensen (2005), who focus on the special case of the EZW utility model in which the EIS, $1/\rho$, equals one. In this case, the pricing kernel simplifies to an expression that depends only on the expected present value of long horizon consumption growth.

The estimated value of ρ is less than one, indicating that the EIS is above one and considerably different from the inverse of the coefficient of relative risk aversion. The results are similar across estimations. The EIS is estimated to be between 1.667 and 2 in the representative agent version of the model, and between 1.11 and 2.22 in the representative stockholder version of the model. The estimates for this parameter are in line with those reported in Bansal, Gallant, and Tauchen (2004) who estimate a model of long-run consumption risk with EZW utility. In theoretical work, Bansal and Yaron (2004) have emphasized the importance of EZW preferences with an EIS >1, in conjunction with a persistent component of consumption growth, to explain the dynamics of aggregate stock market returns.

Under standard regularity conditions typically imposed in semiparametric models, the two-step estimator $\widehat{\boldsymbol{\delta}}$ is $\sqrt{T}$ asymptotically normally distributed even when the model (15) may be misspecified. However, the asymptotic variance-covariance matrix is of complicated form. We therefore compute block bootstrap estimates of their finite sample distributions, as suggested by Chen, Linton, and van Keilegom (2003). The sieve parameters $\frac{V_0}{C_0}$, $\{a_j\}_{j=1}^{K_T}$, the conditional mean $\widehat{m}(\mathbf{w}_t, \boldsymbol{\delta}, F)$, and the finite dimensional parameters $\boldsymbol{\delta} = (\beta, \rho, \theta)'$ are all estimated for each simulated realization.[12] Unfortunately, the procedure is highly numerically intensive, and takes several days to run on a workstation computer, thus limiting the number of bootstrap simulations that can be feasibly performed. We therefore conduct the two-step SMD estimation on 100 block bootstrap samples. The resulting confidence regions are wide, a finding that may in part be attributable to the imprecision in the bootstrapped confidence regions, itself a result of the small number of bootstrap iterations. Even with the large confidence regions, however, in the representative agent formulation of the model we can always reject the hypothesis that $\theta = \rho$. Moreover, the 95% confidence region for ρ is moderate and contains only values below one, or an EIS above one.

5.2 Model Comparison

How well does the EZW recursive utility model explain asset pricing data relative to competing specifications? To address this question, we use the methodology provided by Hansen and Jagannathan (1997), who develop a way to compare asset pricing models when all stochastic discount factor models are treated as misspecified proxies for the true unknown SDF, and the relevant question is which model contains the least specification error.

Hansen and Jagannathan suggest that we compare the pricing errors of various candidate SDF $M_t(b)$ models by choosing each model's parameters, b, to minimize the quadratic form

[12] The bootstrap sample is obtained by sampling blocks of the raw data randomly with replacement and laying them end-to-end in the order sampled. To choose the block length, we follow the recommendation of Hall, Horowitz, and Jing (1995) who show that the asymptotically optimal block length for estimating a symmetrical distribution function is $l \propto T^{1/5}$; also see Horowitz (2003).

$\mathbf{g}_T^{HJ}(b) \equiv \{g_T(b)\}'\mathbf{G}_T^{-1}g_T(b)$, where $g_T(b) = (g_{1T}(b), ..., g_{NT}(b))'$ is the vector of the sample average of pricing errors (i.e., $g_{iT}(b) = \frac{1}{T}\sum_{t=1}^{T} M_t(b)R_{i,t} - 1$ for $i = 1, ..., N$), and $\mathbf{G}_T$ is the sample second moment matrix of the N asset returns upon which the models are evaluated (i.e., the (i,j)-the element of $\mathbf{G}_T$ is $\frac{1}{T}\sum_{t=1}^{T} R_{i,t}R_{j,t}$ for $i, j = 1, ..., N$). The measure of model misspecification is then the square root of this minimized quadratic form, $d_T \equiv \sqrt{\mathbf{g}_T^{HJ}(\widehat{b})}$, which gives the maximum pricing error per unit norm on any portfolio of the N assets studied, and delivers a metric suitable for model comparison. It is also a measure of the distance between the candidate SDF proxy, and the set of all admissible stochastic discount factors (Hansen and Jagannathan (1997)). We refer to the square root of this minimized quadratic form, $d_T \equiv \sqrt{\mathbf{g}_T^{HJ}(\widehat{b})}$, as the *Hansen-Jagannathan distance*, or HJ distance for short.

We also compute a conditional version of the distance metric that incorporates conditioning information Z_t. In this case, $g_T(b) = \frac{1}{T}\sum_{t=1}^{T}\left[(M_{t+1}(b) R_{t+1} - 1_N) \otimes Z_t\right]$ and $\mathbf{G}_T \equiv \frac{1}{T}\sum_{t=1}^{T}(R_{t+1} \otimes Z_t)(R_{t+1} \otimes Z_t)'$. Because the number of test assets increases quickly with the dimension of Z_t, we use just a single instrument $Z_t = cay_t$. This instrument is useful because it has been shown elsewhere to contain significant predictive power for returns on the size and book-market sorted portfolios used in this empirical study (Lettau and Ludvigson (2001b)). We refer to the Hansen-Jagannathan distance metric that incorporates conditioning information as the *conditional HJ distance*, and likewise refer to the distance without conditioning information as the *unconditional HJ distance*.

An important advantage of this procedure is that the second moment matrix of returns delivers an objective function that is invariant to the initial choice of asset returns. The identity and other fixed weighting matrices do not share this property. Kandel and Stambaugh (1995) have suggested that asset pricing tests using these other fixed weighting matrices can be highly sensitive to the choice of test assets. Using the second moment matrix helps to avert this problem.

We compare the specification errors of the estimated EZW recursive utility model to those of the time-separable, constant relative risk aversion (CRRA) power utility model (3) and to two alternative asset pricing models that have been studied in the literature: the three-factor, portfolio-based asset pricing model of Fama and French (1993), and the approximately linear, conditional, or "scaled" consumption-based capital asset pricing model explored in Lettau and Ludvigson (2001b). These models are both linear stochastic discount factor models taking the form

$$M_{t+1}(b) = b_0 + \sum_{i=1}^{k} b_i F_{i,t+1}, \tag{25}$$

where $F_{i,t+1}$ are variable factors, and the coefficients b_0 and b_i are treated as free parameters to be estimated. Fama and French develop an empirical three-factor model ($k = 3$), with variable factors related to firm size (market capitalization), book equity-to-market equity, and the aggregate stock market. These factors are the "small-minus-big" (SMB_{t+1}) portfolio return, the "high-minus-low" (HML_{t+1}) portfolio return, and the market return, $R_{m,t+1}$, respectively.[13] The Fama-French pricing kernel is an empirical model not motivated from any specific economic model of preferences. It nevertheless serves as a benchmark because it has displayed unusual success in explaining the cross section of mean equity returns (Fama and French (1993), Fama and French (1996)). The model explored by Lettau and Ludvigson (2001b) can be interpreted as a "scaled" or conditional consumption CAPM ("scaled CCAPM" hereafter) and also has three variable factors ($k = 3$), $\widehat{cay}_t$, $\widehat{cay}_t \cdot \Delta \log C_{t+1}$, and $\Delta \log C_{t+1}$. Lettau and Ludvigson (2001b) show that such a model can be thought of as a linear approximation to any consumption-based CAPM (CCAPM) in which risk-premia vary over time.

To insure that the SDF proxies we explore preclude arbitrage opportunities over all assets in our sample (including derivative securities), the estimated SDF must always be positive. The SDF of the time-separable CRRA utility model and of the EZW recursive utility model is always positive, thus these models are arbitrage free. By contrast, the SDFs of the linear comparison models may often take on large negative values, and are therefore not arbitrage free. In order to avoid comparisons between models that are arbitrage free and those that are not, we restrict the parameters of the linear SDF to those that produce a positive SDF in every period. Although we cannot guarantee that the linear SDFs will always be positive out-of-sample, we can at minimum choose parameters so as to insure that they are positive in sample, and therefore suitable for pricing derivative claims in sample.

In practice, the set of parameters that deliver positive SDFs is not closed, so it is convenient to include limit points by choosing among parameters b that deliver nonnegative

[13] SMB is the difference between the returns on small and big stock portfolios with the same weight-average book-to-market equity. HML is the difference between returns on high and low book-to-market equity portfolios with the same weighted-average size. Further details on these variables can be found in Fama and French (1993). We follow Fama and French and use the CRSP value-weighted return as a proxy for the market portfolio, R_m. The data are taken from Kenneth French's Dartmouth web page (see the Appendix).

SDFs. To do so, we choose the unknown parameters $b = (b_0, b_1, ..., b_k)'$ of the linear models to minimize the squared HJ distance for that model, subject to the constraint that the SDF proxy be nonnegative in every period of our sample. In the computation of the HJ distance metric, this implies that we restrict $g_T(b) \equiv \frac{1}{T}\sum_{t=1}^{T}[\{M_{t+1}(b)\}^{+} R_{t+1} - 1_N]$ or $g_T(b) \equiv \frac{1}{T}\sum_{t=1}^{T}[(\{M_{t+1}(b)\}^{+} R_{t+1} - 1_N) \otimes Z_t]$, where $\{M_{t+1}(b)\}^{+} = \max\{0, M_{t+1}(b)\}$.

For the EZW recursive utility model, the SDF is always positive and the restriction is nonbinding. The HJ distance for the EZW model (15) is computed by using the parameter estimates obtained from the two-step procedure described in Section 3, for the case in which $\mathbf{W} = \mathbf{G}_T^{-1}$ in the second step GMM estimation of the finite-dimensional parameters $\boldsymbol{\delta} = (\beta, \rho, \theta)'$. Notice that this drastically restricts the number of parameters in the EZW model that are chosen to minimize the HJ distance. In particular, we choose only the finite-dimensional parameters $\boldsymbol{\delta} = (\beta, \rho, \theta)'$ of the EZW model to minimize the HJ distance–the parameters of the nonparametric $F()$ function are chosen to minimize the SMD criterion (19). Note that this places the EZW model (15) at a disadvantage because the sieve parameters of the unknown function $F()$ are not chosen to minimize the HJ criterion, which is the measure of model misspecification. By contrast all of the comparison models' parameters are chosen to minimize the HJ criterion.[14] To rank competing models, we apply an AIC penalty to the HJ criterion of each model, for the number of free parameters b chosen to minimize the HJ distance. The HJ distances for all models are reported in Table 3.

Table 3 reports the measure of specification error given by the HJ distance ("HJ Dist"), $d_T \equiv \sqrt{\mathbf{g}_T^{HJ}(\widehat{b})}$, for all the models discussed above. Several general patterns emerge from the results. First, for both the representative agent version of the model and the representative stockholder version of the model, the estimated EZW recursive utility model always displays smaller specification error than the time-separable CRRA model, but greater specification error than the Fama-French model. This is true regardless of whether the unconditional or conditional HJ distance is used to compare models. The unconditional HJ distance for the EZW recursive specification is 0.449, about 13 percent smaller than that of the time-separable CRRA model, but about 26 percent larger than the Fama-French model. When models are compared according to the conditional HJ distance, the distance metric for the recursive model is only 15 percent larger than that of the Fama-French model. Second, the EZW model performs better than than the scaled CCAPM: the HJ distance is smaller when models are compared on the basis of either the unconditional or conditional HJ distance, regardless

[14] Recall that the SMD minimization gives greater weight to moments that are more highly correlated with the instruments $p^{J_T}(\mathbf{w}_t)$, while the HJ minimization matches unconditional moments.

of which measure of consumption is used.[15] Third, when the representative stockholder version of the model is estimated, the recursive utility model performs better than every model except the Fama-French model according to both the conditional and unconditional distance metrics. These results are encouraging for the recursive utility framework, because they suggest that the model's ability to fit the data is in a comparable range with other models that have shown particular success in explaining the cross-section of expected stock returns.

Note that the HJ distances computed so as to insure that the SDF *proxies* are nonnegative, are in principle distinct from an alternative distance metric suggested by Hansen and Jagannathan (1997), denoted "HJ^+ Dist," which restricts the set of *admissible* stochastic discount factors to be nonnegative. In practice, however, the two distance metrics are quite similar. Estimates of "HJ^+ Dist" are reported in Table 4.

5.3 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 (2005), Malloy, Moskowitz, and Vissing-Jorgensen (2005)). 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 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

[15] The estimated HJ distances for the linear scaled CCAPM are larger than reported in previous work (e.g., Lettau and Ludvigson (2001b)) due to the restriction that the SDF proxy be positive. Although the scaled CCAPM does a good job of assigning the right prices to size and book-market sorted equity returns, its linearity implies that it can assign negative prices to some positive derivative payoffs on those assets. This is not surprising, since linear models–typically implemented as approximations of nonlinear models for use in specific applications–are not designed to price derivative claims.

conjecture of Malloy, Moskowitz, and Vissing-Jorgensen (2005). We note, however, that if the model with $\rho = 1$ is misspecified, parameter estimates can be quite 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.

5.4 The Return to Aggregate Wealth and Human Wealth

In this section, we investigate the estimated EZW recursive utility model's implications for the return to aggregate wealth, $R_{w,t+1}$, and the return to human wealth, denoted $R_{y,t+1}$ hereafter. The return to aggregate wealth represents a claim to future consumption and is in general unobservable. However, it can be inferred from our estimates of V_t/C_t by equating the marginal rate of substitution (5), evaluated at the estimated parameter values $\left\{\widehat{\boldsymbol{\delta}},\widehat{F}\left(\cdot,\widehat{\boldsymbol{\delta}}\right)\right\}$, with its theoretical representation based on consumption growth and the return to aggregate wealth (6):

$$\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\rho}\left(\frac{\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}}{\mathcal{R}_t\left(\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}\right)}\right)^{\rho-\theta}=\left\{\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\rho}\right\}^{\frac{1-\theta}{1-\rho}}\left\{\frac{1}{R_{w,t+1}}\right\}^{\frac{\theta-\rho}{1-\rho}}.$$

If, in addition, we explicitly model human wealth as part of the aggregate wealth portfolio, the framework also has implications for the return to human wealth, $R_{y,t}$. We do so by following Campbell (1996), who assumes that the return to aggregate wealth is a portfolio weighted average of the unobservable return to human wealth and the return to financial wealth. Specifically, Campbell starts with the relationship

$$R_{w,t+1} = (1-\nu_t)\,R_{a,t+1} + \nu_t R_{y,t+1}, \tag{26}$$

where ν_t is the ratio of human wealth to aggregate wealth, and $R_{a,t+1}$ is the gross simple return on nonhuman wealth (a refers to financial asset wealth). A difficulty with (26) is that the wealth shares may in principal vary over time. Campbell deals with this by linearizing (26) around the means of ν_t, the log return on nonhuman asset wealth, and the log return on human wealth, assuming that the means of the latter two are the same. Under these assumptions, an approximate expression for the log return on aggregate wealth may be obtained with constant portfolio shares. Unfortunately, this approximation assumes that the means of human and nonhuman wealth returns are the same. As a start, we instead adopt the crude assumption that portfolio shares in (26) are constant:

$$R_{w,t+1} = (1-\nu)\,R_{a,t+1} + \nu R_{y,t+1}.$$

Such an assumption is presumably a reasonable approximation if portfolio shares between human and nonhuman wealth are relatively stable over quarterly horizons. Given observations on $R_{w,t+1}$ from our estimation of the EZW recursive utility model, and given a value for ν, the return to human wealth, $R_{y,t+1}$, may be inferred.

The exercise in this section is similar in spirit to the investigation of Lustig and Van Nieuwerburgh (2006). These authors, following Campbell (1996), investigate a loglinear version of the EZW recursive utility model under the assumption that asset returns and consumption are jointly lognormal and homoskedastic. With these assumptions, the authors back out the human wealth return from observable aggregate consumption data, and find a strong negative correlation between the return to asset wealth and the return to human wealth. Our approach generalizes their exercise in that it provides an estimate of the fully nonlinear EZW model without requiring the assumption that asset returns and consumption are jointly lognormal and homoskedastic. An important question of this study is whether our approach leads to significantly different implications for both the aggregate wealth return and the human wealth return.

Tables 5 and 6 present summary statistics for our estimated aggregate wealth return, $R_{w,t+1}$ and human wealth return, $R_{y,t+1}$. Following Campbell (1996) and Lustig and Van Nieuwerburgh (2006), we use the CRSP value-weighted stock market return to measure $R_{a,t+1}$. The statistics for $R_{y,t+1}$ are presented for two different values of the share of human wealth in aggregate wealth: $\nu = 0.333$ and $\nu = 0.667$. There are two different sets of estimates, depending on whether $\mathbf{W} = \mathbf{I}$ or $\mathbf{W} = \mathbf{G}_T^{-1}$ in the second-step estimation of the EZW model. Summary statistics for the $\mathbf{W} = \mathbf{I}$ case are presented in Table 5, for the $\mathbf{W} = \mathbf{G}_T^{-1}$ case in Table 6. For comparison, summary statistics on the CRSP value-weighted return, $R_{CRSP,t+1}$ are also presented.

Several conclusions can be drawn from the results in Tables 5 and 6. First, the return to aggregate wealth is always considerably less volatile than the aggregate stock market return. For example, in Table 5, the annualized standard deviation of $R_{w,t+1}$ is 0.01 in the representative agent model and 0.036 in the representative stockholder model. By contrast, the annualized standard deviation of $R_{CRSP,t+1}$ is 0.165. Second, in the representative agent model, the mean of $R_{w,t+1}$ is less than the mean of $R_{CRSP,t+1}$, but is larger in the representative stockholder model. Since the mean of $R_{w,t+1}$ is a weighted average of the means of $R_{y,t+1}$ and $R_{CRSP,t+1}$, and given that the mean of $R_{CRSP,t+1}$ is 0.084, the mean of the human wealth return can be quite small if, as in the representative agent model, the mean of aggregate wealth is small. This is especially so when the share of human wealth takes on the smaller value of 0.333. Indeed, if the mean of aggregate wealth is sufficiently small (as it is in Table 6 where it equals 0.024), the gross return on human wealth can even be less than one, so that the simple net return is negative. Third, the return to human wealth is a weighted average (where the weights exceed one in absolute value) of the returns to aggregate wealth and the return to asset wealth. Thus, unless the correlation between the stock market return and the aggregate wealth return is sufficiently high, the return to human wealth can be quite volatile, especially when ν is small. This occurs in the representative stockholder versions of the model when $\nu = 0.333$.

Finally, the results show that the only way to reconcile a relatively stable aggregate wealth return with a volatile stock market return, is to have the correlation between the human wealth return and the stock market return be negative and large in absolute value. The correlation between $R_{y,t+1}$ and $R_{CRSP,t+1}$ range from -0.764 in Table 6 when $\nu = 0.667$, to -0.996 in Table 5 when $\nu = 0.333$. These numbers are strikingly close to those reported in Lustig and Van Nieuwerburgh (2006) for the cases where the EIS exceeds one. The finding reinforces their conclusion that "good news on Wall street is bad news on Main street." As Lustig and Van Nieuwerburgh (2006) point out, a negative correlation between human and financial wealth is inconsistent with the production functions typically employed in standard business cycle models, which imply a near perfect correlation between the two forms of wealth.

6 Conclusion

In this paper we undertake a formal econometric evaluation of the Epstein-Zin-Weil recursive utility model, a framework upon which a large and growing body of theoretical work macroeconomics and finance is based. We conduct estimation of the EZW model without employing an observable financial market return as a proxy for the unobservable aggregate wealth return, without linearizing the model, and without placing tight parametric restrictions on either the law of motion or joint distribution of consumption and asset returns, or on the value of key preference parameters such as the elasticity of intertemporal substitution. We present estimates of all the preference parameters of the EZW model, evaluate the model's ability to fit asset return data relative to competing asset pricing models, and investigate the implications of such estimates for the unobservable aggregate wealth return and human wealth return.

Using quarterly data on consumption growth, assets returns and instruments, we find evidence that the elasticity of intertemporal substitution in consumption differs considerably from the inverse of the coefficient of relative risk aversion, and that the EZW recursive utility model displays less model misspecification than the familiar time-separable CRRA power utility model. Taken together, these findings suggest that the consumption and asset return data we study are better explained by the recursive generalization of the standard CRRA model than by the special case of this model in which preferences are time-separable and the coefficient of relative risk aversion equals the inverse of the EIS.

Our results can be compared to those in the existing the literature. For example, we find that the estimated relative risk aversion parameter ranges from 17-60, with considerably higher values for the representative agent representation of the model than the representative stockholder representation. These findings echo those in the approximate loglinear version of the model where the EIS is restricted to unity, studied by Malloy, Moskowitz, and Vissing-Jorgensen (2005). On the other hand, we find that the estimated elasticity of intertemporal substitution is typically above one, regardless of which consumption measure is employed. Finally, the empirical estimates imply that the unobservable aggregate wealth return is weakly correlated with the CRSP value-weighted stock market return and only one-tenth to one-fifth as volatile. These findings suggest that the return to human wealth must be strongly negatively correlated with the aggregate stock market return, similar to results reported for an approximate loglinear version of the model studied by Lustig and Van Nieuwerburgh (2006).

As an asset pricing model, the EZW recursive utility framework includes an additional risk factor for explaining asset returns, above and beyond the single consumption growth risk factor found in the time-separable, CRRA power utility framework. The added risk factor in the EZW recursive utility model is a multiplicative term involving the continuation value of the future consumption plan relative to its conditional expected value today. This factor can in principal add volatility to the marginal rate of substitution in consumption, helping to explain the behavior of equity return data (Hansen and Jagannathan (1991)). One way this factor can be volatile is if the conditional mean of consumption growth varies over long horizons. The estimation procedure employed here allows us to assess the plausibility of this implication from the consumption and return data alone, without imposing restrictions on the data generating process for consumption. The results suggest that the additional risk factor in the EZW model has sufficient dynamics so as to provide a better description of the data than the CRRA power utility model, implying that the conditional mean of consumption growth is unlikely to be constant over time (Kocherlakota (1990)). At the same time, the added volatility coming from continuation utility is modest and must be magnified by a relatively high value for risk aversion in order to fit the equity return data.

7 Appendix

This appendices consist of several parts: Appendix 1 describes the data. Appendix 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 $F(\cdot)$. Appendix 3 provides details of the two-step semiparametric estimation procedure, including the implementation of the SMD estimator as an instance of GMM. Appendix 4 presents consistency and convergence rate of the first step profile SMD estimator of the unknown function $F^*(\cdot;\boldsymbol{\delta})$ when the model could be misspecified. Appendix 5 presents root-T asymptotic normality of the second step GMM estimator of $\boldsymbol{\delta}$ when the model could be misspecified.

Appendix 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 (2005). 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 (2001a).

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.

Appendix 2: Approximation to Continuation Value Function $F()$

The arguments of $F()$. If the Markov structure is linear, as in (9) and (10), we give assumptions under which $\frac{V_t}{C_t} = F\left(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}}\right)$. First note that the dynamic system (9) and (10) converges asymptotically to time-invariant innovations representation taking the form

$$\Delta c_{t+1} = \mu + H\widehat{x}_t + \varepsilon_{t+1} \tag{27}$$

$$\widehat{x}_{t+1} = \phi\widehat{x}_t + K\varepsilon_{t+1}, \tag{28}$$

where the scalar variable $\varepsilon_{t+1} \equiv \Delta c_{t+1} - \Delta\widehat{c}_{t+1} = H\left(x_t - \widehat{x}_t\right) + \mathbf{C}\boldsymbol{\epsilon}_{t+1}$, $\widehat{x}_t$ denotes a linear least squares projection of x_t onto $\Delta c_t, \Delta c_{t-1}, ...\Delta c_{-\infty}$, and $K \equiv \left(\mathbf{DC}' + \phi PH\right)\left(HPH + \mathbf{CC}'\right)^{-1}$, where P solves

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

(See Hansen and Sargent (2007).) The representation above shows that the state variable $\widehat{x}_t$ replaces x_t as the argument of the function over which $\frac{V_t}{C_t}$ is defined. Assume $\frac{V_t}{C_t}$ is an invertible function $f(\widehat{x}_t)$. Then,

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

>From (28) we have

$$\begin{aligned}\frac{V_t}{C_t} &= f(\widehat{x}) = g(\widehat{x}_{t-1}, \varepsilon_t) \\ &= g\left(f^{-1}\left(\frac{V_{t-1}}{C_{t-1}}\right), \varepsilon_t\right), \end{aligned} \tag{29}$$

for some function g. By inverting (27), we obtain

$$\begin{aligned}\varepsilon_t &= h(\exp(\Delta c_{t+1}), \widehat{x}_{t-1}) \\ &= h\left(\exp\left(\ln\left[\frac{C_t}{C_{t-1}}\right]\right), f^{-1}\left(\frac{V_{t-1}}{C_{t-1}}\right)\right), \end{aligned} \tag{30}$$

where $h\left(\frac{C_t}{C_{t-1}}, \widehat{x}_t\right) = \ln\left[\frac{C_t}{C_{t-1}}\right] - \mu - H\widehat{x}$. Plugging (30) into (29), we have $\frac{V_t}{C_t} = F\left(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}}\right)$, for $F : \mathbb{R}^2 \to \mathbb{R}$. Observe that if the innovations in (9) and (10) are positively correlated, $\frac{V_t}{C_t}$ may display negative serial dependence. The linear model implies that F is a monotonic function of $\frac{C_t}{C_{t-1}}$.

If the stochastic process for consumption growth is a nonlinear function of a hidden first-order Markov process x_t, the function F can take on more general functional forms, potentially displaying nonmonotonicity in both its arguments. For example, consider the functionally non-linear state space model:

$$c_{t+1} - c_t = h(x_t) + \epsilon_{c,t+1} \tag{31}$$

$$x_{t+1} = \psi(x_t) + \epsilon_{x,t+1}, \tag{32}$$

where $E(\epsilon_{c,t+1}) = E(\epsilon_{d,t+1}) = 0$, $\text{Var}(\epsilon_{j,t+1}) = \sigma_j$, $j = c, x$, $E(\epsilon_{c,t+1}\epsilon_{x,t+1}) = \sigma_{cx}$ and $h(x_t)$ and $\psi(x_t)$ are no longer necessarily linear functions of the state variable x_t. Harvey (1989) shows that, under the assumption that the innovations in (31)-(32) are Gaussian, an approximate innovations representation can be obtained by linearizing the model and then applying a modification of the usual Kalman filter to the resulting linearized representation of (31)-(32). Let $\widehat{\chi}_{t|t-1}$ denote the conditional mean of x_t. If the functions $h(x_t)$ and $\psi(x_t)$ are expanded in Taylor series around $\widehat{\chi}_{t|t-1}$, an innovations representation may be obtained which takes the form:

$$\widehat{\chi}_{t+1|t} = \psi\left(\widehat{\chi}_{t|t-1} + P_{t|t-1}H_{t+1}F_t^{-1}e_{t+1}\right) \tag{33}$$

$$\Delta c_{t+1} = h\left(\widehat{\chi}_{t|t-1}\right) + e_{t+1}, \tag{34}$$

where $e_{t+1} = \Delta c_{t+1} - h\left(\widehat{\chi}_{t|t-1}\right)$ and $P_{t|t-1}$solves a suitable recursion applied to the linearized state space representation corresponding to the dynamic system (31)-(32):

$$\begin{aligned} P_{t+1|t} &= \widetilde{\phi}^2 \left(P_{t|t-1} - \left\{ P_{t|t-1} \widetilde{H} + \sigma_{cx} \right\}^2 F_t^{-1} \right) + \sigma_x \\ F_t &= \widetilde{H}^2 P_{t+1|t} + 2 H_{t+1} \boldsymbol{\sigma}_{cx} + \sigma_c, \end{aligned}$$

where $\widetilde{\phi}$ and $\widetilde{H}$ are partial derivatives of ψ and h respectively. See Harvey (1989), Ch., 3.

Given invertibility, from either (27)-(28) or (33)-(34), we again have the implication that $\frac{V_t}{C_t} = F\left(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}}\right)$, for some $F : \mathbb{R}^2 \to \mathbb{R}$, but unlike the case for the linear Markov model, the function F may display nonmonotonicities as well as nonlinearities. 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 function B_j is implemented as a tensor product cubic B-spline taking the form:

$$F(z, c) = \alpha_0 + \sum_{i=1}^{K_{1T}} \sum_{j=1}^{K_{2T}} a_{ij} B_m\left(z - i + \frac{m}{2}\right) B_m\left(\frac{c}{\Delta_2} + \varsigma - j\right), \tag{35}$$

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

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

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 $K_{1T} = K_{2T} = 4$; thus we present the results for the more parsimonious specification using $K_{1T} = K_{2T} = 3$ below.

Appendix 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:

$$\mathcal{V} \equiv \left\{ g : (0, \infty) \times (0, \infty) \to (0, \infty) : \|g\|_{\Lambda^s} \leq const. < \infty \right\}, \quad \text{for some } s > 1, \tag{36}$$

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

$$E \left\{\gamma_i(\mathbf{z}_{t+1}, \boldsymbol{\delta}_o, F^* (\cdot, \boldsymbol{\delta}_o)) \otimes \mathbf{x}_t\right\} = 0, \qquad i = 1, ..., N. \tag{37}$$

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 $\widehat{F}_T(\cdot; \boldsymbol{\delta})$:

$$\widehat{F}_T(\cdot, \boldsymbol{\delta}) = \arg\min_{F_T \in \mathcal{V}_T} \frac{1}{T} \sum_{t=1}^{T} \widehat{m}(\mathbf{w}_t, \boldsymbol{\delta}, F_T)' \widehat{m}(\mathbf{w}_t, \boldsymbol{\delta}, F_T), \tag{38}$$

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 the tensor product B-spline (35) 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:

$$\widehat{\boldsymbol{\delta}}_{\mathbf{W}} = \arg\min_{\boldsymbol{\delta} \in \mathcal{D}} \left[\mathbf{g}_T(\boldsymbol{\delta}, \widehat{F}_T(\cdot, \boldsymbol{\delta}); \mathbf{y}^T)\right]' \mathbf{W}_T \left[\mathbf{g}_T((\boldsymbol{\delta}, \widehat{F}_T(\cdot, \boldsymbol{\delta}); \mathbf{y}^T)\right], \tag{39}$$

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,

$$\mathbf{g}_T(\boldsymbol{\delta}, \widehat{F}_T(\cdot, \boldsymbol{\delta}); \mathbf{y}^T) = \frac{1}{T} \sum_{t=1}^{T} \gamma(\mathbf{z}_{t+1}, \boldsymbol{\delta}, \widehat{F}_T(\cdot, \boldsymbol{\delta})) \otimes \mathbf{x}_t \tag{40}$$

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:

$$\widehat{m}_i(\mathbf{w}, \boldsymbol{\delta}, F) = \sum_{t=1}^{T} \gamma_i(\mathbf{z}_{t+1}, \boldsymbol{\delta}, F) p^{J_T}(\mathbf{w}_t)' (\mathbf{P}'\mathbf{P})^{-1} p^{J_T}(\mathbf{w}), \quad i = 1, ..., N. \tag{41}$$

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 (41), the first step SMD estimation of $F^*(\cdot; \boldsymbol{\delta})$ can be alternatively implemented via the following GMM criterion (42):

$$\widehat{F}_T(\cdot, \boldsymbol{\delta}) = \arg\min_{F_T \in \mathcal{V}_T} \left[\mathbf{g}_T(\boldsymbol{\delta}, F_T; \mathbf{y}^T)\right]' \left\{\mathbf{I}_N \otimes (\mathbf{P}'\mathbf{P})^{-1}\right\} \left[\mathbf{g}_T(\boldsymbol{\delta}, F_T; \mathbf{y}^T)\right], \tag{42}$$

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

$$\mathbf{g}_T(\boldsymbol{\delta}, F_T; \mathbf{y}^T) = \frac{1}{T} \sum_{t=1}^{T} \gamma(\mathbf{z}_{t+1}, \boldsymbol{\delta}, F_T) \otimes p^{J_T}(\mathbf{w}_t) \tag{43}$$

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

Appendix 4: Convergence Rate of First Step SMD Estimator $\widehat{F}_T(\cdot, \boldsymbol{\delta})$

To be completed.

Appendix 5: Root-T Asymptotic Normality of Second Step GMM Estimator $\widehat{\boldsymbol{\delta}}$

To be completed.

References

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

——— (2007): "Estimation of Possibly Misspecified Semiparametric Conditional Moment Restriction Models with Different Conditioning Variables," *Journal of Econometrics,* forthcoming.

Andrews, D. (1994): "Asymptotics for SEmi-parametric Econometric Models via Stochastic Equicontinuity," *Econometrica*, 62, 43–72.

Bansal, R., R. F. Dittmar, and D. Kiku (2005): "Long Run Risks and Equity Returns," Unpublished paper, Fuqua School, Duke University.

Bansal, R., R. F. Dittmar, and C. T. Lundblad (2006): "Consumption, Dividends, and the Cross-Section of Equity Returns," *Journal of Finance,* forthcoming.

Bansal, R., A. R. Gallant, and G. Tauchen (2004): "Rational Pessimism, Rational Exuberance, adn Markets for Macro Risks," Unpublished paper, Fuqua School, Duke University.

Bansal, R., D. Kiku, and A. Yaron (2006): "Risks for the Long-Run: Estimation and Inference," Unpublished paper, Fuqua School, Duke University.

Bansal, R., and A. Yaron (2004): "Risks for the Long-Run: A Potential Resolution of Asset Pricing Puzzles," *Journal of Finance*, 59(4), 1481–1509.

Campanale, C., R. Castro, and G. L. Clementi (2006): "Asset Pricing in a General Equilibrium Production Economy with Chew-Dekel Risk Preferences," Unpublished paper, New York University.

Campbell, J. Y. (1991): "A Variance Decomposition for Stock Returns," *Economic Journal*, 101, 157–179.

——— (1993): "Intertemporal Asset Pricing Without Consumption Data," *American Economic Review*, 83, 487–512.

——— (1996): "Understanding Risk and Return," *Journal of Political Economy*, 104(2), 298–345.

——— (2003): "Consumption-Based Asset Pricing," in *Handbook of the Economics of Finance Vol. I-B*, ed. by G. Constantinides, M. Harris, and R. Stulz, pp. 803–887. North-Holland, Amsterdam.

Campbell, J. Y., and L. Viceira (2001): *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors.* Oxford Univeristy Press, London, UK.

Campbell, J. Y., and T. Voulteenaho (2005): "Good Beta, Bad Beta," *American Economic Review*, 94(5), 1249–1275.

Chen, X. (2006): "Large Sample Sieve Estimation of Semi-Nonparameteric Models," in *Handbook of Econometrics,* forthcoming, ed. by J. Heckman, and E. Leamer, vol. 6. North Holland, Amsterdam, The Netherlands.

Chen, X., O. Linton, and I. van Keilegom (2003): "Estmiation of Semiparametric Models when the Criterion Function is Not Smooth," *Econometrica*, 71, 1591–1608.

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

Cochrane, J. H. (2001): *Asset Pricing*. Princeton University Press, Princeton, NJ.

Colacito, R., and M. M. Croce (2004): "Risks for the Long Run and the Real Exchange Rate," Unpublished paper, New York University.

Croce, M. M. (2006): "Welfare Costs and Long Run Consumption Risk in a Production Economy," Unpublished paper, New York University.

Croce, M. M., M. Lettau, and S. C. Ludvigson (2006): "Investor Information, Long-Run Risk and the Duration of Risky Cash Flows," Unpublished paper, New York University.

Epstein, L., and S. Zin (1989): "Substitution Risk Aversion and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework," *Econometrica*, 57, 937–968.

——— (1991): "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Investigation," *Journal of Political Economy*, 99, 555–576.

Fama, E. F., and K. R. French (1992): "The Cross-Section of Expected Returns," *Journal of Finance*, 47, 427–465.

——— (1993): "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3–56.

——— (1996): "Multifactor Explanations of Asset Pricing Anomalies," *Journal of Finance*, 51, 55–84.

Gabaix, X., and D. Laibson (2002): "The 6-D Bias and the Equity Premium Puzzle," in *NBER Macroeconomics Annual: 2002*, ed. by B. Bernanke, and K. Rogoff, pp. 257–312. MIT Press, Cambridge, MA.

Gomes, F., and A. Michaelides (2005): "Optimal Life-Cycle Asset Allocation: Understanding the Empirical Evidence," *Journal of Finance*, 60(2), 869–904.

GRENANDER, U. (1981): *Abstract Inference*. Wiley Series, New York, NY.

HALL, P., J. L. HOROWITZ, AND B. Y. JING (1995): "On Blocking Rules for the Bootstrap with Dependent Data," *Biometrika*, 82, 561–574.

HANSEN, L. P. (1982): "Large Sample Properties of Generalized Methods of Moments Estimators," *Econometrica*, 50, 1029–54.

HANSEN, L. P., J. HEATON, AND N. LI (2005): "Consumption Strikes Back?: Measuring Long-run Risk," Unpublished Paper, University of Chicago.

HANSEN, L. P., J. HEATON, N. ROUSSANOV, AND J. LEE (2006): "Intertemporal Substitution and Risk Aversion," Unpublished Paper, University of Chicago.

HANSEN, L. P., AND R. JAGANNATHAN (1991): "Restrictions on Intertemporal Marginal Rates of Substitution Implied by Asset Returns," *Journal of Political Economy*, 99, 225–262.

——— (1997): "Assessing Specific Errors in Stochastic Discount Factor Models," *Journal of Finance*, 52, 557–590.

HANSEN, L. P., AND T. J. SARGENT (2006): "Fragile Beliefs and the Price of Model Uncertainty," Unpublished Paper, New York University.

——— (2007): *Robustness*. Princeton University Press forthcoming, Princeton, NJ.

HANSEN, L. P., AND K. SINGLETON (1982): "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50(5), 1269–86.

——— (1983): "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Returns," *Journal of Political Economy*, 91, 249–65.

HARVEY, A. C. (1989): *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge, United Kingdon.

HEATON, J. (1993): "The Interaction Between Time-Nonseparable Preferences and Time Aggregation," *Econometrica*, 61(2), 353–85.

HODRICK, R. (1992): "Dividend Yields and Expected Stock Returns: Alternative Procedures for Inference and Measurement," *Review of Financial Studies*, 5, 357–386.

HOROWITZ, J. L. (2003): "The Bootstrap," in *Handbook of Econometrics*, ed. by J. J. Heckman, and E. Leamer, vol. 5. Elsevier Science B.V., North Holland.

Jagannathan, R., and Y. Wang (2007): "Calendar Cycles, Infrequent Decisions and the Cross Section of Stock Returns," Unpublished paper, Kellog School of Management, Northwestern University.

Kandel, S., and R. F. Stambaugh (1995): "Portfolio Inefficiency and the Cross-Section of Expected Returns," *Journal of Finance*, 50, 157–184.

Kiku, D. (2005): "Is the Value Premium a Puzzle?," Unpublished paper, Duke University.

Kocherlakota, N. R. (1990): "Disentangling the Coefficient of Relative Risk Aversion from the Elasticity of Intertemporal Substitution: An Irrelevance Result," *Journal of Finance*, 45(1), 175–190.

Krueger, D., and F. Kubler (2005): "Pareto Improving Social Security Reform When Financial Markets are Incomplete!?," *American Economic Review*, 96(3), 737–755.

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

——— (2001b): "Resurrecting the (C)CAPM: A Cross-Sectional Test When Risk Premia are Time-Varying," *Journal of Political Economy*, 109(6), 1238–1287.

——— (2004): "Understanding Trend and Cycle in Asset Values: Reevaluating the Wealth Effect on Consumption," *American Economic Review*, 94(1), 276–299.

——— (2005): "Euler Equation Errors," Unpublished paper, New York University.

Lustig, H., and S. Van Nieuwerburgh (2006): "The Returns on Human Capital: Good News on Wall Street is Bad News on Main Street," *Review of Financial Studies,* forthcoming.

Lustig, H., S. Van Nieuwerburgh, and A. Verdelhan (2007): "The Consumption-Wealth Ratio: A Litmus Test for Consumption-based Asset Pricing Models," Unpublished paper, NYU Stern School.

Malloy, C. J., T. J. Moskowitz, and A. Vissing-Jorgensen (2005): "Long-run Stockholder Consumption Risk and Asset Returns," Unpublished manuscript, University of Chicago, GSB.

Newey, W. K., and D. L. McFadden (1994): "Large Sample Estimation and Hypothesis Testing," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. 4. Elsevier Science B.V., North Holland, Amsterdam.

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

Newey, W. K., and K. D. West (1987): "A Simple, Positive Semidefinite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Piazzesi, M., and M. Schneider (2006): "Equilibrium Yield Curves," in *NBER Macroeconomics Annual: 2006*, ed. by M. W. Daron Acemoglu, Kenneth Rogoff. MIT Press, Cambridge.

Restoy, F., and P. Weil (1998): "Approximate Equilibrium Asset Prices," NBER Working paper No. 6611.

Stock, J. H., J. H. Wright, and M. Yogo (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics*, 20(4), 518–528.

Tallarini, T. D. (2000): "Risk-Sensitive Real Business Cycle Models," *Journal of Monetary Economics*, 45, 507–532.

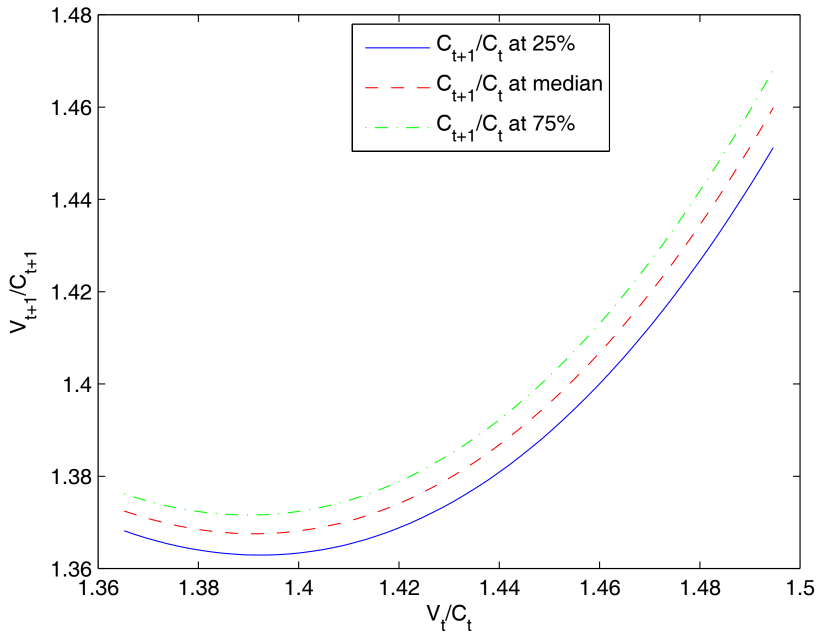
Vissing-Jorgensen, A. (2002): "Limited Asset Market Participation and Intertemporal Substitution," *Journal of Political Economy*, 110(4), 825–853.

Vissing-Jorgensen, A., and O. P. Attanasio (2003): "Stock Market Participation, Intetemporal Substitution and Risk-Aversion," *American Economic Review, AEA Papers and Proceedings*, 93(2), 383–391.

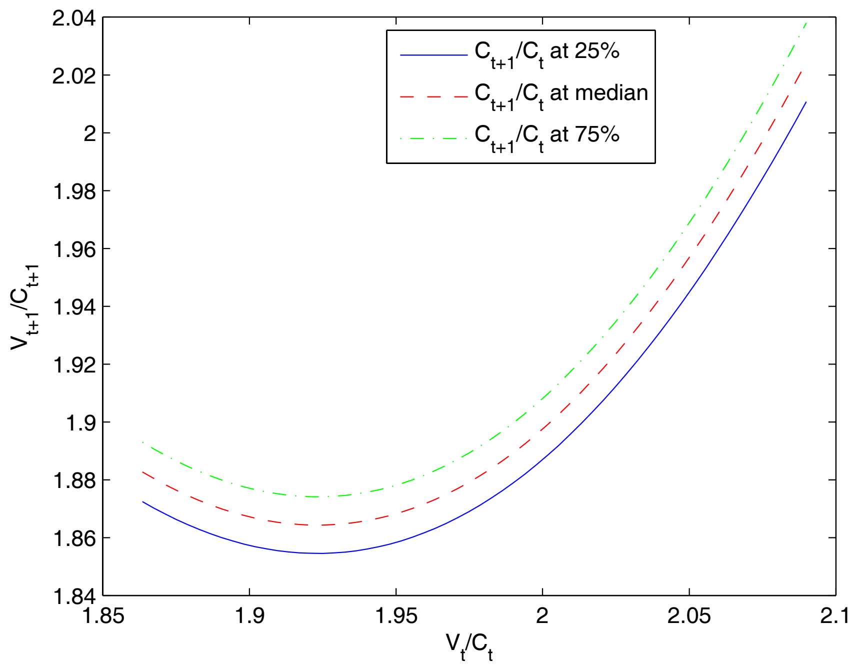
Weil, P. (1989): "The Equity Premium Puzzle and the Risk-Free Rate Puzzle," *Journal of Monetary Economics*, 24(3), 401–421.

Figure 1

Estimated Continuation Value-Consumption Ratio, Aggregate Consumption, **W**=**I**



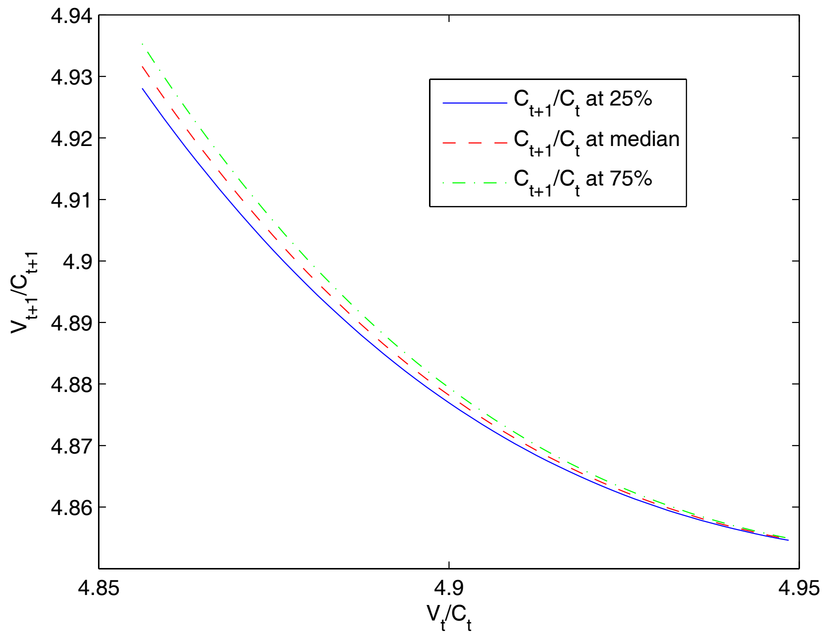
Estimated Continuation Value-Consumption Ratio, Aggregate Consumption, **W**=($\mathbf{G}_T$)$^{-1}$



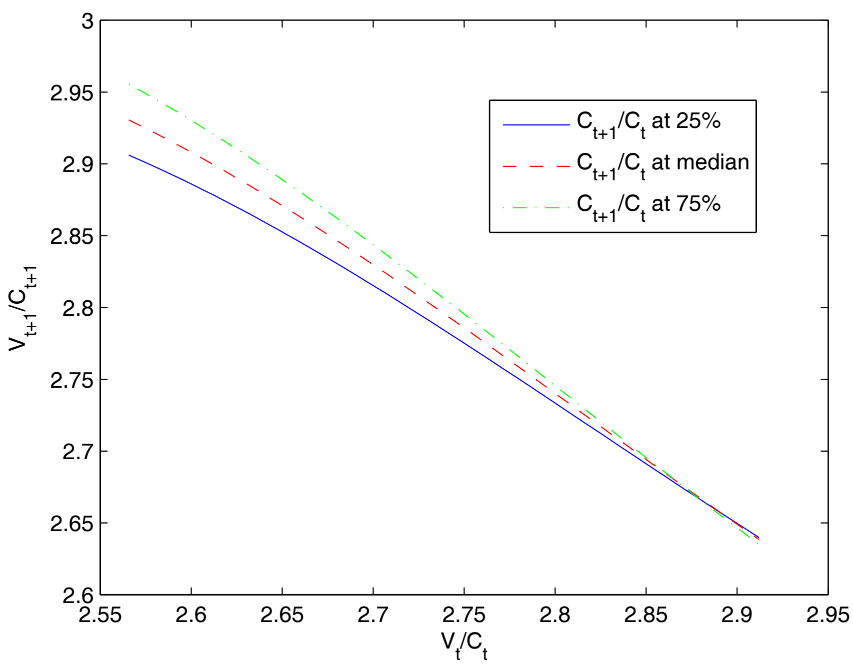
Notes: The figure plots the estimated continuation value-consumption ratio against lagged values of the continuation value with consumption growth held alternately held at the 25th, 50th and 75th percentiles in the sample. Consumption is measured as aggregate consumption, "**W**=" indicates the weighting matrix used in second-step estimation. The sample is 1952:Q1-2005Q1.

Figure 2

Estimated Continuation Value-Consumption Ratio, Stockholder Consumption, **W**=**I**



Estimated Continuation Value-Consumption Ratio, Stockholder Consumption, **W**=(**G**$_T$)$^{-1}$



Notes: The figure plots the estimated continuation value-consumption ratio against lagged values of the continuation value with consumption growth held alternately held at the 25th, 50th and 75th percentiles in the sample. Consumption is measured as stockholder consumption, "**W**=" indicates the weighting matrix used in second-step estimation. The sample is 1952:Q1-2005Q1.

Table 1

First-Stage Estimates of Weights for Stockholder Consumption

Model: $\Delta c_t^{SH} = \gamma_0 + \gamma_1 \Delta \ln(IP_t) + \gamma_2 \Delta \ln(SV_t) + \varepsilon_t$

	Est.	(t-$stat$)
γ_0	0.007	(1.447)
γ_1	0.833	(6.780)
γ_2	1.992	(2.204)
R^2	0.075	

Notes: The table reports the results from regressing stockholder consumption growth on the log difference of industrial production growth, $\Delta \ln(IP_t)$, and the log differences of real services expenditure growth, $\Delta \ln(SV_t)$. Point estimates are reported, along with Newey and West (1987) corrected t-statistics in parentheses. The sample period is 1982:M7-2002:M2.

Table 2

Preference Parameter Estimates

2nd Step Estimation	β (95% CI)	θ (95% CI)	ρ (95% CI)
	Aggregate Consumption		
$\mathbf{W} = \mathbf{I}$	0.990	57.5	0.60
	(.985, .996)	(27.5, 129)	(.24, .99)
$\mathbf{W} = \mathbf{G}_T^{-1}$	0.999	60	0.50
	(.994, .9999)	(42,144)	(.20, .75)
	Stockholder Consumption		
$\mathbf{W} = \mathbf{I}$	0.994	20.00	0.90
	(.993, .9995)	(.25, 40)	(.38, 1.24)
$\mathbf{W} = \mathbf{G}_T^{-1}$	0.998	17.0	0.68
	(.992, .9999)	(1, 43.3)	(.23, 1.01)

Notes: The table reports second-step estimates of preference parameters, with 95% confidence intervals in parenthesis. β is the subjective time discount factor, θ is the coefficient of relative risk aversion, and ρ is the inverse of the elasticity of intertemporal substitution. Second-step estimates are obtained by minimizing the GMM criterion with either $\mathbf{W} = \mathbf{I}$ or with $\mathbf{W} = \mathbf{G}_T^{-1}$, where in both cases $\mathbf{x}_t = \mathbf{1}_N$, an $N \times 1$ vector of ones. The sample is 1952:Q1-2005:Q1.

Table 3

Specification Errors for Alternative Models: HJ Distance

	Unconditional	Conditional
Aggregate Consumption		
Model	HJ Dist	HJ Dist
(1)	(2)	(3)
Recursive	0.451	0.591
CRRA Utility	0.514	0.627
Fama-French	0.363	0.515
Scaled CCAPM	0.456	0.625
Stockholder Consumption		
Model	HJ Dist	HJ Dist
(1)	(2)	(3)
Recursive	0.463	0.605
CRRA Utility	0.517	0.627
Fama-French	0.363	0.515
Scaled CCAPM	0.490	0.620

Notes: The table reports the Hansen-Jagannathan distance metric

$$HJ\ Dist_T(b) = \min_b \sqrt{g_T(b)' \, \mathbf{G}_T^{-1} g_T(b)},$$

where b are parameter values associated with the model listed in column 1. In column 2, $g_T(b) \equiv \frac{1}{T}\sum_{t=1}^{T}\left[\{M_t(b)\}^{+}R_t - 1_N\right]$, and $\mathbf{G}_T \equiv \frac{1}{T}\sum_{t=1}^{T} R_t R_t'$, where $M_t(b)$ is the stochastic discount factor associated with the model listed in column 1 and $\{M_t(b)\}^{+} = \max\{0, M_t(b)\}$. In column 3, $g_T(b) \equiv \frac{1}{T}\sum_{t=1}^{T}\left[(\{M_{t+1}(b)\}^{+}R_{t+1} - 1_N) \otimes Z_t\right]$ and $\mathbf{G}_T \equiv \frac{1}{T}\sum_{t=1}^{T}(R_{t+1} \otimes Z_{t+1})(R_{t+1} \otimes Z_t)'$ with $Z_t = cay_t$. The sample is 1952:Q1-2005:Q1.

Table 4

Specification Errors for Alternative Models: HJ^+ Distance

	Unconditional	Conditional
Aggregate Consumption		
Model (1)	HJ^+ Dist (2)	HJ^+ Dist (3)
Recursive	0.451	0.591
CRRA Utility	0.514	0.627
Fama-French	0.341	0.519
Scaled CCAPM	0.464	0.643
Stockholder Consumption		
Model (1)	HJ^+ Dist (2)	HJ^+ Dist (3)
Recursive	0.463	0.605
CRRA Utility	0.517	0.627
Fama-French	0.338	0.506
Scaled CCAPM	0.467	0.661

Notes: For each model in column 1, "HJ^+ Dist" is the distance between the model proxy and the family of admissible nonnegative stochastic discount factors. The sample is 1952:Q1-2005:Q1.

Table 5

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 $\mathbf{W} = \mathbf{G}_T^{-1}$, where in both cases $\mathbf{x}_t$=$\mathbf{1}_N$, an $N \times 1$ vector of ones. The sample is 1952:Q1-2005:Q1.

Table 6

Summary Statistics for Return to Aggregate Wealth, Human Wealth, $\mathbf{W} = \mathbf{I}$

Model-Implied Aggregate Wealth Return

	Representative Agent		Rep Stockholder	
	$R_{w,t}$	$R_{CRSP,t}$	$R_{w,t}$	$R_{CRSP,t}$
Panel A: Correlation Matrix				
$R_{w,t}$	1.00	0.171	1.00	-0.049
$R_{CRSP,t}$		1.00		1.00
Panel B: Univariate Summary Statistics				
Mean	0.057	0.084	0.109	0.084
Standard deviation	0.010	0.165	0.036	0.165
Autocorrelation	0.234	0.055	-0.08	0.055

Notes: See next page.

Table 6, continued

Model-Implied Human Wealth Return, $\nu = 0.333$

	Representative Agent		Rep Stockholder	
	$R_{y,t}$	$R_{CRSP,t}$	$R_{y,t}$	$R_{CRSP,t}$
Panel A: Correlation Matrix				
$R_{y,t}$	1.00	-0.996	1.00	-0.953
$R_{CRSP,t}$		1.00		1.00
Panel B: Univariate Summary Statistics				
Mean	0.003	0.084	0.160	0.084
Standard deviation	0.327	0.165	0.353	0.165
Autocorrelation	0.044	0.055	0.042	0.055

Model-Implied Human Wealth Return, $\nu = 0.667$

	Representative Agent		Rep Stockholder	
	$R_{y,t}$	$R_{CRSP,t}$	$R_{y,t}$	$R_{CRSP,t}$
Panel A: Correlation Matrix				
$R_{y,t}$	1.00	-0.982	1.00	-0.847
$R_{CRSP,t}$		1.00		1.00
Panel B: Univariate Summary Statistics				
Mean	0.043	0.084	0.121	0.084
Standard deviation	0.082	0.165	0.101	0.165
Autocorrelation	0.036	0.055	0.016	0.055

Notes: The table reports summary statistics for the return to the aggregate wealth portfolio, $R_{w,t}$, and the return to human wealth, $R_{y,t}$, implied by the estimates of the model, and for the CRSP value-weighted stock market return, $R_{CRSP,t}$. The parameter ν is the steady state fraction of human wealth in aggregate wealth. Means and standard deviations are annualized. Results for the model-implied returns are based on second-step estimates obtained by minimizing the GMM criterion with $\mathbf{W} = \mathbf{I}$ and $\mathbf{x}_t = \mathbf{1}_N$, an $N \times 1$ vector of ones. The sample is 1952:Q1-2005:Q1.

Table 7

Summary Statistics for Return to Aggregate Wealth, Human Wealth, $\mathbf{W} = \mathbf{G}_T^{-1}$

Model-Implied Aggregate Wealth Return

	Representative Agent		Rep Stockholder	
	$R_{w,t}$	$R_{CRSP,t}$	$R_{w,t}$	$R_{CRSP,t}$
	Panel A: Correlation Matrix			
$R_{w,t}$	1.00	0.18	1.00	0.004
$R_{CRSP,t}$		1.00		1.00
	Panel B: Univariate Summary Statistics			
Mean	0.023	0.084	0.092	0.084
Standard deviation	0.012	0.165	0.046	0.165
Autocorrelation	0.055	0.055	-0.434	0.055

Notes: See next page.

Table 7, continued

Model-Implied Human Wealth Return, $\nu = 0.333$

	Representative Agent		Rep Stockholder	
	$R_{y,t}$	$R_{CRSP,t}$	$R_{y,t}$	$R_{CRSP,t}$
Panel A: Correlation Matrix				
$R_{y,t}$	1.00	-0.994	1.00	-0.921
$R_{CRSP,t}$		1.00		1.00
Panel B: Univariate Summary Statistics				
Mean	-0.093	0.084	0.110	0.084
Standard deviation	0.326	0.165	0.359	0.165
Autocorrelation	0.043	0.055	0.013	0.055

Model-Implied Human Wealth Return, $\nu = 0.667$

	Representative Agent		Rep Stockholder	
	$R_{y,t}$	$R_{CRSP,t}$	$R_{y,t}$	$R_{CRSP,t}$
Panel A: Correlation Matrix				
$R_{y,t}$	1.00	-0.975	1.00	-0.764
$R_{CRSP,t}$		1.00		1.00
Panel B: Univariate Summary Statistics				
Mean	-0.007	0.084	0.097	0.084
Standard deviation	0.081	0.165	0.108	0.165
Autocorrelation	0.032	0.055	-0.103	0.055

Notes: The table reports summary statistics for the return to the aggregate wealth portfolio, $R_{w,t}$, and the return to human wealth, $R_{y,t}$, implied by the estimates of the model, and for the CRSP value-weighted stock market return, $R_{CRSP,t}$. The parameter ν is the steady state fraction of human wealth in aggregate wealth. Means and standard deviations are annualized statistics from quarterly data. Results for the model-implied returns are based on second-step GMM estimation using the $\mathbf{W} = \mathbf{G}_T^{-1}$ and $\mathbf{x}_t = \mathbf{1}_N$. The sample is 1952:Q1-2005:Q1.