Chen, Xiaohong, Favilukis, Jack and Ludvigson, Sydney C.

An estimation of economic models with recursive preferences

Article (Accepted version) (Refereed)

Original citation:
Chen, Xiaohong, Favilukis, Jack and Ludvigson, Sydney C. (2013) An estimation of economic models with recursive preferences. Quantitative Economics, 4 (1). pp. 39-83. ISSN 1759-7323
DOI: 10.3982/QE97

Quantitative Economics 4 (2013), 39–83 1759-7331/20130039

An estimation of economic models with recursive preferences

Xiaohong Chen
Yale University

Jack Favilukis
London School of Economics

Sydney C. Ludvigson
New York University and NBER

This paper presents estimates of key preference parameters of the Epstein and Zin (1989, 1991) and Weil (1989) 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 ranges from 17 to 60, with higher values for aggregate consumption than for stockholder consumption, while the estimated elasticity of intertemporal substitution is above 1. In addition, the estimated model-implied aggregate wealth return is found to be weakly correlated with the Center for Research in Security Prices value-weighted stock market return, suggesting that the return to human wealth is negatively correlated with the aggregate stock market return.

Keywords. Consumption based asset pricing, semiparametric estimation, limited stock market participation.

JEL classification. G12, E21.

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

Xiaohong Chen: xiaohong.chen@yale.edu
Jack Favilukis: j.favilukis@lse.ac.uk
Sydney C. Ludvigson: sydney.ludvigson@nyu.edu
We acknowledge financial support from the National Science Foundation (Chen and Ludvigson) and from the Alfred P. Sloan Foundation (Ludvigson). We are grateful to Orazio Attanasio, Richard Blundell, Darrel Duffie, Lars Hansen, Monika Piazzesi, Annette Vissing-Jorgensen, Gianluca Violante, Motohiro Yogo, Stanley Zin, and to seminar participants at NYU, UC Berkeley, the University of Chicago, the June 2007 SED Annual Meetings, the July 2007 NBER Summer Institute Methods and Applications for Dynamic, Stochastic General Equilibrium Models Workshop, and the September 2007 CMU Conference in Honor of Hansen–Singleton's 1982 paper 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.

DOI: 10.3982/QE97

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 with regard to attitudes toward 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 of 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 semiparametric 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 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

[1]See, for example, Attanasio and Weber (1989), Campbell (1993, 1996), Tallarini (2000), Campbell and Viceira (2001), Bansal and Yaron (2004), Colacito and Croce (2004), Bansal, Dittmar, and Kiku (2009), Campbell and Vuolteenaho (2004), Gomes and Michaelides (2005), Krueger and Kubler (2006), Hansen, Heaton, and Li (2008), Kiku (2005), Malloy, Moskowitz, and Vissing-Jorgensen (2009), Campanale, Castro, and Clementi (2006), Croce (2012), Bansal, Dittmar, and Lundblad (2005), Croce, Lettau, and Ludvigson (2012), Hansen and Sargent (2006), and Piazzesi and Schneider (2006).

[2]Epstein and Zin (1991) used an aggregate stock market return to proxy for the aggregate wealth return. Campbell (1996) assumed that the aggregate wealth return is a portfolio weighted average of a human capital return and a financial return, and obtained an estimable expression for an approximate log-linear 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) followed Campbell's approach to estimate the model using household-level consumption data.

the potential importance of human capital and other unobservable 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 log-linear vector time-series process, the continuation value has an analytical solution and is a function of observable consumption data (e.g., Hansen, Heaton, and Li (2008)). Alternatively, if consumption and asset returns are assumed to be jointly log-normally distributed and homoskedastic (e.g., Attanasio and Weber (1989)) 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 over time.[3] Moreover, even first-order expansions of the EZW model around an EIS of unity may not capture the magnitude of variability of the consumption–wealth ratio (Hansen et al. (2007)). Second, although aggregate consumption growth itself appears to be well described by a log-normal process, empirical evidence suggests that the joint distribution of consumption and asset returns exhibits significant departures from log-normality (Lettau and Ludvigson (2009)). Third, Kocherlakota (1990) pointed out that joint log-normality 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 technique that allows us to conduct estimation and evaluation 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 using 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.

[3]Lettau and Ludvigson (2001a) argued 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 found 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 (2007) and Lustig, Van Nieuwerburgh, and Verdelhan (2007).

The continuation value is still an unknown function of the relevant state variables, however; 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 inference 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 such as generalized method of moments (GMM), with plugged-in estimated continuation value function. Motivated by the arguments of Hansen and Jagannathan (1997), our approach 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 1982–2002 and since household-level consumption data are known to contain significant measurement error, we follow Malloy, Moskowitz, and Vissing-Jorgensen (2009) 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 to 60, with higher values for the representative agent version of the model than the representative stockholder version. The estimated elasticity of intertemporal substitution is above 1, and differs considerably from the inverse of the coefficient of relative risk aversion. This estimate is of particular interest because the value of the EIS has important consequences for the asset pricing implications of models with EZW recursive utility. For example, if consumption growth is normally distributed, it is straightforward to show that the price–consumption ratio implied by EZW recursive utility is increasing in expected consumption growth only if the EIS is greater than 1. In addition, when relative risk aversion exceeds unity, the price–consumption ratio will be decreasing in the volatility of consumption growth only if the EIS exceeds 1.

We find that the estimated aggregate wealth return is weakly correlated with the Center for Research in Security Prices (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 (2008), 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 Capital Asset Pricing Model (CAPM) model, but not as well 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 (2007) and Bansal, Kiku, and Yaron (2007) estimated 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.

The rest of this paper is organized as follows. The next section describes the model we estimate. Section 3 discusses our main idea, which is to estimate the latent continuation value function nonparametrically using observable data. Section 4 describes the empirical procedure; Section 5 describes the data. Empirical results are discussed in Section 6. Section 7 investigates the implications of our estimates for the return to aggregate wealth and the return to human wealth. Section 8 concludes. The Appendix to this paper is provided in a supplementary file on the journal website, http://qeconomics.org/supp/97/supplement.pdf.

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 $t = 0, 1, \ldots$. 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

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

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

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

$$U_t = E\left[\sum_{j=0}^{\infty} \beta^j \frac{C_{t+j}^{1-\theta}}{1-\theta} \Big| \mathcal{F}_t\right], \tag{3}$$

where $U_t \equiv V_t^{1-\theta}/(1-\beta)$.

As in Hansen, Heaton, and Li (2008), the utility function can be rescaled and expressed as a function of stationary variables:

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

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

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

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

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

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

where $R_{w,t+1}$ is the return to aggregate wealth, where aggregate wealth 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 log-linear 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) assumed 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.

Instead, 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 $E_t[M_{t+1}R_{i,t+1}] = 1$ 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

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

Since the expected product of any traded asset return with M_{t+1} equals 1, 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 $(\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}/\mathcal{R}_t(\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}))^{\rho-\theta}$.

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, $\mathcal{R}_t(\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t})$. 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 can be plugged into (7). The resulting expression, for any observed sequence of traded asset returns $\{R_{i,t+1}\}_{i=1}^N$, takes the form

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

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

By estimating the fully nonlinear Euler equations (8), we obviate the need to linearize the model or to place parametric restrictions on preference parameters β, θ, and ρ. We also use a distribution-free estimation procedure, thereby obviating the need to place tight restrictions on the law of motion for, or joint distribution of, consumption and asset return data. Finally, the moment restrictions (8) make no reference to $R_{w,t+1}$; thus we obviate the need to find an observable proxy for the unobservable aggregate wealth return. Of course, the continuation value–consumption ratio $\frac{V_{t+1}}{C_{t+1}}$ is itself a latent variable. In the next section we show how it can be estimated nonparametrically from observable data, as a function of state variables.

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

This section discusses the main idea of our study, which is to nonparametrically estimate the latent component $\frac{V_{t+1}}{C_{t+1}}$ of the added risk factor $(\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}/\mathcal{R}_t(\frac{V_{t+1}}{C_{t+1}}\frac{C_{t+1}}{C_t}))^{\rho-\theta}$ in the EZW stochastic discount factor. To do so, we proceed in two steps. First, because $\frac{V_{t+1}}{C_{t+1}}$ is a function of state variables that govern the evolution of the distribution of consumption growth, we begin with assumptions on the dynamic behavior of consumption growth that allow 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 (2008). Here we assume that consumption growth is a function of a hidden univariate first-order Markov process x_t, a specification that encompasses a range of stationary, dynamic models for consumption growth. Second, because the state variable x_t is latent, it must be replaced in empirical work with either an estimate, $\widehat{x}_t$, or with other variables that subsume the information in $\widehat{x}_t$. We discuss this in the next subsections.

3.1 *The dynamics of consumption growth*

Let lowercase letters denote log variables, for example, $\ln(C_{t+1}) \equiv c_{t+1}$. We assume that consumption growth is a linear function of a hidden first-order univariate Markov process x_t that summarizes information about future consumption growth

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

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

where $\boldsymbol{\epsilon}_{t+1}$ is a (2×1) independent and identically distributed (i.i.d.) vector with mean zero and identity covariance matrix I, and where $\mathbf{C}$ 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 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)).

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, the econometric methodology, discussed in the next section, leaves the law of motion of the consumption process unspecified.

3.2 *Forming an estimate of the latent* x_t

The state variable x_t that is taken as the input of the unknown function $\frac{V_t}{C_t}$ is unobservable to the econometrician and must be inferred from observable data. One way to do this is to filter the consumption data so as to obtain an estimate of x_t. Given (9)–(10), optimal forecasts of future consumption growth are formed from an estimate of the hidden factor x_t, obtained by filtering the observable consumption data. Given the linearity of the system (9)–(10), the Kalman filter is a natural filtering algorithm. Applying the Kalman filter to (9)–(10), the dynamic system converges asymptotically to a time-invariant innovations representation taking the form

$$\Delta c_{t+1} = \mu + H\widehat{x}_t + \varepsilon_{t+1}, \tag{11}$$

$$\widehat{x}_{t+1} = \phi\widehat{x}_t + K\varepsilon_{t+1}, \tag{12}$$

where the scalar variable $\varepsilon_{t+1} \equiv \Delta c_{t+1} - \Delta\widehat{c}_{t+1} = H(x_t - \widehat{x}_t) + \mathbf{C}\boldsymbol{\epsilon}_{t+1}$, $\widehat{x}_t$ denotes a linear least squares projection of x_t onto $\Delta c_t, \Delta c_{t-1}, \ldots, \Delta c_{-\infty}$ and K is a scalar "Kalman gain" defined recursively from the Kalman updating equations as a function of the primitive parameters of the dynamic system (9) and (10). The Appendix gives the precise recursive function that defines K. Unlike the dynamic system (9)–(10), the representation (11)–(12) is a function of an observable (from filtered consumption data) state variable, $\widehat{x}_t$. The econometrician could, therefore, replace the latent state variable x_t as the argument over which $\frac{V_t}{C_t}$ is defined with the observable Kalman filter estimate $\widehat{x}_t$, implying $\frac{V_t}{C_t} = f(\widehat{x}_t)$ for some function f.

Rather than using $\widehat{x}_t$ directly in our estimation—a cumbersome approach that would require embedding the Kalman filter algorithm into our outer semiparametric estimation procedure—we assume that $\frac{V_t}{C_t}$ is an invertible function $f(\widehat{x}_t)$. As shown in the Appendix, under this assumption and given (9)–(12), the information contained in $\widehat{x}_t$ is fully summarized by two other variables: the lagged continuation value ratio $\frac{V_{t-1}}{C_{t-1}}$ and current consumption growth $\frac{C_t}{C_{t-1}}$. Thus, rather than modeling $\frac{V_t}{C_t}$ as an unknown function $f(\widehat{x}_t)$, we work with an equivalent specification in which $\frac{V_t}{C_t}$ is modeled as an unknown function $F:\mathbb{R}^2 \to \mathbb{R}$ of $\frac{V_{t-1}}{C_{t-1}}$ and $\frac{C_t}{C_{t-1}}$:

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

The Appendix also shows that the function $F(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}})$ may display negative serial dependence under a variety of plausible parameter-value combinations that govern the dynamic system (9)–(10), implying $\frac{\partial F}{\partial(V_{t-1}/C_{t-1})} < 0$. For example, if $f'(\widehat{x}_t) > 0$, then $\frac{\partial F}{\partial(V_{t-1}/C_{t-1})} < 0$ if ϕ is not too large, and/or if the innovations in (9) and (10) are positively correlated. As we show below, all of our estimated functions $\frac{V_t}{C_t} = F(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}})$ display such negative serial dependence.

An alternative motivation for the specification (13) may be obtained if consumption dynamics evolve as

$$\frac{C_{t+1}}{C_t} = h(X_{t+1}, X_t), \tag{14}$$

where $\{X_t\}$ is a first-order hidden, stationary Markov process characterizing the time t information set $\mathcal{F}_t$. In a recent paper, Hansen and Scheinkman (2012) established the existence and uniqueness of a solution of the form

$$\frac{V_t}{C_t} = f(X_t) \tag{15}$$

to the recursive continuation utility forward equation (4), under the assumption (14). If the latent state variable X_t is a scalar and the function $f(\cdot)$ is one-to-one, then we obtain

$$\frac{C_{t+1}}{C_t} = h\left(f^{-1}\left(\frac{V_{t+1}}{C_{t+1}}\right), f^{-1}\left(\frac{V_t}{C_t}\right)\right).$$

If further, $h(\cdot, \cdot)$ is one-to-one in its first argument, then we obtain our specification (13):

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

Note that (14) is more general than the specification (9) plus (10) in that it allows for general nonlinearities in consumption growth as a function of the first-order Markov process, but it is less general in that it does not allow consumption dynamics to additionally depend on an independent shock ε_{t+1}.

To summarize, the asset pricing model we entertain in this paper consists of the conditional moment restrictions (8), subject to the specification (13). Without placing tight parametric restrictions on the model, the continuation value ratio is an unknown function $\frac{V_t}{C_t} = F(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}})$. We, therefore, estimate $\frac{V_t}{C_t}$ nonparametrically, as described below. Our overall model is semiparametric in the sense that it contains both finite-dimensional parameters (β, θ, ρ) and infinite-dimensional unknown parameters in the unknown function $F(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}})$.

3.3 *Information structure*

It is important to emphasize that the procedure just described when consumption dynamics evolve according to (11) and (12) recovers the information in the Kalman filter estimate $\widehat{x}_t$ of x_t. This is *not* the same as recovering the information contained in x_t, which from the econometrician's perspective is latent. It follows that, in this case, we cannot recover $\frac{V_t}{C_t} = f(x_t)$ with some function $F(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}})$; we can only recover $f(\widehat{x}_t)$, where $\widehat{x}_t$ is the Kalman filter estimate, with some function $F(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}})$.

The Kalman filter estimate $\widehat{x}_t$ of x_t uses information contained only in the history of consumption growth and, in particular, it does not use information in asset prices. Might there be additional information about future consumption growth in asset prices? The answer to this question depends not only on whether (9)–(10) is a good description of the dynamics of consumption growth, but also on what information the representative agent in the asset pricing model we seek to evaluate actually has about x_t. Suppose the true data generating process for consumption is given by (9)–(10) but the representative agent—whose behavior determines asset prices—cannot observe the latent variable x_t or the separate innovations in (9) and (10). The agent could employ historical consumption data to form an estimate $\widehat{x}_t$ of x_t to be used in making the optimal consumption and portfolio decisions that determine equilibrium asset prices. The representative agent's continuation value function would then be a function of $\widehat{x}_t$, implying that once $\widehat{x}_t$ is included as an argument over which the function is defined, asset price information (also a function of $\widehat{x}_t$) would be redundant. On the other hand, if the true data generating process is (9)–(10) but the representative agent can observe x_t while the econometrician cannot, asset prices as equilibrium outcomes could contain additional information about future consumption growth that is not contained in $\widehat{x}_t$. Thus, our approach is justified when we assume both that (9)–(10) is good description of the dynamics of consumption growth and that agents in the model, like econometricians, cannot observe x_t. Croce, Lettau, and Ludvigson (2012) investigated the equilibrium asset pricing implications of this sort of "incomplete information," whereby investors must form an estimate $\widehat{x}_t$ of x_t based on information in the history of consumption growth when making optimal decisions. Since x_t is, in fact, a latent conditional moment, we view this information structure as more plausible than one in which agents are presumed to directly observe x_t.

But even if we allowed for reasons that the econometrician might benefit from using asset price information (e.g., the price–dividend ratio) in place of, or in addition to, the information in $\widehat{x}_t$ (e.g., optimizing agents really can observe x_t, so asset prices reveal the information in x_t), there would be a difficulty with specifying $\frac{V_t}{C_t}$ to be a function of such information in terms of the interpretation of results: By doing so, we would in effect specify a stochastic discount factor that is a function of the very return data that the model is being asked to explain. While there is nothing invalid about this approach (conditional on the assumption that agents can directly observe x_t), estimates obtained this way would tell us nothing about whether the empirical consumption dynamics alone—which are exogenous inputs into the asset pricing model—are consistent with what would be required to explain the return behavior observed. This situation would muddle the interpretation of results. For example, if an EZW model with the value function defined over asset price data performed well, this could be because a varient of the model in which agents directly observe x_t really is true, or it could be because the consumption-based model is fundamentally wrong and the approach merely delivers a back-door means of explaining asset returns with other asset returns. Moreover, while such an empirical model for the SDF might provide a good *description* of asset returns, it cannot provide a satisfactory *explanation* for asset return behavior in terms of primitive macroeconomic risk. For these reasons, we focus on evaluating the extent to which the EZW asset pricing model can explain asset return data, without reference to return data as part of the stochastic discount factor that explains returns.

4. Empirical implementation

This section presents the details of our empirical procedure. 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 let $F : \mathbb{R}^2 \to \mathbb{R}$ 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 $i = 1, \ldots, N$, denote

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

where $\mathbf{z}_{t+1}$ is a vector containing all the strictly stationary observations, including consumption growth rate and return data. We let $F_o(\cdot; \boldsymbol{\delta})$ denote the minimizer of

$$\inf_{F \in \mathcal{V}} E\left[\sum_{i=1}^{N} \left(E\{\gamma_i(\mathbf{z}_{t+1}, \boldsymbol{\delta}, F) | \mathcal{F}_t\}\right)^2\right] \tag{16}$$

and let $\boldsymbol{\delta}_o \equiv (\beta_o, \rho_o, \theta_o)' \in \mathcal{D}$ denote the minimizer of

$$\min_{\boldsymbol{\delta} \in \mathcal{D}} E\left[\sum_{i=1}^{N} \left(E\{\gamma_i(\mathbf{z}_{t+1}, \boldsymbol{\delta}, F_o(\cdot; \boldsymbol{\delta})) | \mathcal{F}_t\}\right)^2\right]. \tag{17}$$

Let $F_o \equiv F_o(\mathbf{z}_t; \boldsymbol{\delta}_o) \equiv F_o(\cdot; \boldsymbol{\delta}_o) \in \mathcal{V}$. We say that the model consisting of (8) plus (13) is correctly specified if

$$E\{\gamma_i(\mathbf{z}_{t+1}, \boldsymbol{\delta}_o, F_o(\cdot, \boldsymbol{\delta}_o)) | \mathcal{F}_t\} = 0, \quad i = 1, \ldots, N. \tag{18}$$

Equation (18) implies that the N-vector of conditional means $E\{\gamma(\cdot)|\mathcal{F}_t\}$ should be zero in every time period, t. It follows that the true values $F_o(\cdot; \boldsymbol{\delta})$ and $\boldsymbol{\delta}_o$ should be those that minimize the squared distance from zero (quadratic norm) of the conditional means for each t. But since we have more time periods $t = 1, \ldots, T$ than parameters to be estimated, we weight each time period equally, as indicated by the unconditional expectation operator in (16)–(17).

The general estimation methodology is based on estimation of the conditional moment restrictions (18), 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 was previously emphasized by Hansen and Jagannathan (1997). As Hansen and Jagannathan stressed, all models are approximations of reality and, therefore, potentially are misspecified. The estimation procedure used here explicitly takes this possibility into account in the empirical implementation. 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 principle, include more lags, and misspecification of the arguments of the CES utility function, which could, in principle, include a broader measure of durable consumption or leisure. More generally, when we conduct model comparisons 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 subset of $\mathcal{F}_t$.[4] Equation (18) implies

$$E\{\gamma_i(\mathbf{z}_{t+1}, \boldsymbol{\delta}_o, F_o(\cdot, \boldsymbol{\delta}_o))|\mathbf{w}_t\} = 0, \quad i = 1, \ldots, N. \tag{19}$$

Denote

$$\begin{aligned} &m(\mathbf{w}_t, \boldsymbol{\delta}, F) \equiv E\{\gamma(\mathbf{z}_{t+1}, \boldsymbol{\delta}, F)|\mathbf{w}_t\}, \\ &\gamma(\mathbf{z}_{t+1}, \boldsymbol{\delta}, F) = (\gamma_1(\mathbf{z}_{t+1}, \boldsymbol{\delta}, F), \ldots, \gamma_N(\mathbf{z}_{t+1}, \boldsymbol{\delta}, F))'. \end{aligned} \tag{20}$$

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

$$\inf_{F \in \mathcal{V}} E[m(\mathbf{w}_t, \boldsymbol{\delta}, F)' m(\mathbf{w}_t, \boldsymbol{\delta}, F)]. \tag{21}$$

It is clear that $F_o(\mathbf{z}_t, \boldsymbol{\delta}_o) = F^*(\mathbf{z}_t, \boldsymbol{\delta}_o)$ when the model (19) is correctly specified. We say the model (19) is *misspecified* if

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

We estimate the possibly misspecified model (19) using a *profile* semiparametric minimum distance procedure, which consists of two steps; see, for example, Newey (1994), Chen, Linton, and van Keilegom (2003), and Chen (2007). In the first step, for any

[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 $E_t[M_{t+1}R_{i,t+1} - 1]$, which includes individual asset returns, is likely to be a highly nonlinear function of the state variables. In addition, one of 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 $\mathbf{w}_t$.

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. Notice that the estimation procedure itself leaves the law of motion of the data unspecified.[5]

4.1 *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 profile sieve minimum distance (SMD) estimator, described below. In practice, this is achieved by applying the SMD estimator at each point in a three-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 (21). The procedure has two essential parts. First, we replace the conditional expectation $m(\mathbf{w}_t, \boldsymbol{\delta}, F)$ 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 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 $\frac{V_t}{C_t} = F^*(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}}; \boldsymbol{\delta})$, with the initial value for $\frac{V_t}{C_t}$ at time $t = 0$, denoted $\frac{V_0}{C_0}$, taken as an unknown scalar parameter to be estimated. Second, the unknown function $F^*(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}}; \boldsymbol{\delta})$ 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 $\{a_0, a_1, \ldots, a_{K_T}\}$ depend on $\boldsymbol{\delta}$, but the sieve basis functions $\{B_j(\cdot, \cdot) : j = 1, \ldots, K_T\}$ have known functional forms that are independent of $\boldsymbol{\delta}$; see the Appendix for a discussion of the sieve basis functions $B_j(\cdot, \cdot)$. To provide a nonparametric estimate of the unknown function $F^*(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}}; \boldsymbol{\delta})$, K_T must grow with the sample size to insure consistency of the method.[6] We are not interested in the sieve parameters

[5] The estimation procedure requires stationary ergodic observations but does not restrict to linear time series specifications or specific parametric laws of motion of the data.

[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 larger 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 below the results for the more parsimonious specification using $K_T = 9$.

$(a_0, a_1, \ldots, a_{K_T})'$ per se, but 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 $\frac{V_0}{C_0}$, $\{a_j\}_{j=1}^{K_T}$, and $\{B_j(\cdot)\}_{j=1}^{K_T}$, and data on consumption $\{\frac{C_t}{C_{t-1}}\}_{t=1}^{T}$, the function F_{K_T} is used to generate a sequence $\{\frac{V_i}{C_i}\}_{i=1}^{T}$ that can be taken as data to be used in the estimation of (21).

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 $\{p_{0j}(\mathbf{w}_t), j = 1, 2, \ldots, J_T\}$ be a sequence of known basis functions (including a constant function) that map from $\mathbb{R}^{d_w}$ into $\mathbb{R}$. Denote $p^{J_T}(\cdot) \equiv (p_{01}(\cdot), \ldots, p_{0J_T}(\cdot))'$ and the $T \times J_T$ matrix $\mathbf{P} \equiv (p^{J_T}(\mathbf{w}_1), \ldots, p^{J_T}(\mathbf{w}_T))'$. Then

$$\widehat{m}(\mathbf{w}, \boldsymbol{\delta}, F) = \left(\sum_{t=1}^{T} \gamma(\mathbf{z}_{t+1}, \boldsymbol{\delta}, F) p^{J_T}(\mathbf{w}_t)' \left(\mathbf{P}'\mathbf{P}\right)^{-1}\right) p^{J_T}(\mathbf{w}) \tag{22}$$

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 (21):

$$\widehat{F}(\cdot, \boldsymbol{\delta}) = \underset{F_{K_T}}{\arg\min} \frac{1}{T} \sum_{t=1}^{T} \widehat{m}(\mathbf{w}_t, \boldsymbol{\delta}, F_{K_T})' \widehat{m}(\mathbf{w}_t, \boldsymbol{\delta}, F_{K_T}). \tag{23}$$

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

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

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 $F^*(\cdot, \boldsymbol{\delta})$.

4.2 *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

$$\widehat{\boldsymbol{\delta}} = \arg\min_{\boldsymbol{\delta}\in\mathcal{D}} Q_T(\boldsymbol{\delta}), \tag{24}$$

$$Q_T(\boldsymbol{\delta}) = \big[\mathbf{g}_T\big(\boldsymbol{\delta}, \widehat{F}(\cdot,\boldsymbol{\delta}); \mathbf{y}^T\big)\big]'\mathbf{W}\big[\mathbf{g}_T\big(\boldsymbol{\delta}, \widehat{F}(\cdot,\boldsymbol{\delta}); \mathbf{y}^T\big)\big], \tag{25}$$

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

$$\mathbf{g}_T\big(\boldsymbol{\delta}, \widehat{F}(\cdot,\boldsymbol{\delta}); \mathbf{y}^T\big) \equiv \frac{1}{T}\sum_{t=1}^{T} \gamma\big(\mathbf{z}_{t+1}, \boldsymbol{\delta}, \widehat{F}(\cdot,\boldsymbol{\delta})\big) \otimes \mathbf{x}_t \tag{26}$$

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

$$E\big\{\gamma_i\big(\mathbf{z}_{t+1}, \boldsymbol{\delta}_o, F^*(\cdot,\boldsymbol{\delta}_o)\big) \otimes \mathbf{x}_t\big\} = 0, \quad i = 1, \dots, N, \tag{27}$$

where $\mathbf{x}_t$ is any chosen measurable function of $\mathbf{w}_t$.

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 $\widehat{F}(\cdot,\widehat{\boldsymbol{\delta}})$.

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 $i, j = 1, \dots, N$).

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 principle, all the parameters of the model (including the finite-dimensional preference parameters) could be estimated in one step by minimizing the sample SMD criterion:

$$\min_{\boldsymbol{\delta}\in\mathcal{D}, F_{K_T}} \frac{1}{T}\sum_{t=1}^{T} \widehat{m}(\mathbf{w}_t, \boldsymbol{\delta}, F_{K_T})'\widehat{m}(\mathbf{w}_t, \boldsymbol{\delta}, F_{K_T}). \tag{28}$$

It is important to clarify why the two-step profile procedure employed here is superior to the one-step procedure in (28) for our application. First, we want estimates of standard preference parameters such as risk aversion and the EIS (those contained in $\boldsymbol{\delta}$) 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. Note that the estimator of $\boldsymbol{\delta}$ in the two procedures differs not only because the procedures employ different weighting matrices; they also use different information sets. In the two-step profile procedure, the first step (which is required to estimate the unknown function $F(\,)$) is done using *conditional* moment restrictions, which corresponds to infinitely many unconditional moment restrictions. (Of course, this correspondence holds in econometric theory; we must approximate with finitely many restrictions in implementation.) The second step, which is used only to estimate the finite dimensional parameters $\boldsymbol{\delta}$, can be implemented using finitely many unconditional moments, as in GMM. As a consequence, with the two-step procedure, we are free to choose those finitely many unconditional moment restrictions so that the finite-dimensional preference parameters, such as risk aversion and the EIS, reflect values required to match the unconditional moments commonly emphasized in the asset pricing literature (e.g., in Bansal and Yaron (2004) and others). We are not free to make this choice if the procedure is done in a single step, since in that case the finite-dimensional parameter estimates are forced to be those that match the very same conditional moment restrictions required to identify the unknown function. (The unknown function *cannot* be identified from unconditional moment restrictions.)

A second reason that the two-step procedure is important is that both the weighting scheme inherent in the SMD procedure (28) 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 may 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. While this change in the effective set of test assets is necessary to estimate the unknown function $F(\,)$, it is unnecessary to consistently estimate the finite-dimensional parameters $\boldsymbol{\delta}$. We can estimate the finite-dimensional parameters $\boldsymbol{\delta}$ on the original set of test assets by again breaking the procedure up into two steps and 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 1.

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.

4.3 *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) studied the effects of time aggregation in a consumption-based asset pricing model with habit formation and concluded, 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, their 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 ($\mathbf{w}_t$) in the first-step profile estimation of $F^*(\cdot, \boldsymbol{\delta})$. 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).

5. 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 section 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, 1993)). These assets include the 3-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 seven 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 real 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 1980–2002, and since household-level consumption data are known to contain significant measurement error, we follow Malloy, Moskowitz, and Vissing-Jorgensen (2009) and generate a longer time series of data by constructing consumption-mimicking factors for aggregate stockholder consumption growth. The CEX interviews households 3 months apart and households are asked to report consumption for the previous 3 months. Thus, while each household is interviewed 3 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 (2009), 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 (1982–2002), the construction of mimicking factors allows a longer time series of data to be constructed. The procedure follows Malloy, Moskowitz, and Vissing-Jorgensen (2009). 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

TABLE 1. First-stage estimated weights in the stockholder consumption model: $\Delta c_t^{\mathrm{SH}} = \gamma_0 + \gamma_1 \Delta \ln(IP_t) + \gamma_2 \Delta \ln(SV_t) + \varepsilon_t$.[a]

	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	

[a]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.

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, although 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 (comparable R^2 values can be found in Malloy, Moskowitz, and Vissing-Jorgensen (2009)).

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 3 consecutive months corresponding to the 3 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 (2009) for further details on the CEX data and the construction of mimicking factors.

The empirical procedure also requires computation of instruments to estimate the conditional moment functions $\widehat{m}(\mathbf{w}_t, \boldsymbol{\delta}, \widehat{F}(\cdot, \boldsymbol{\delta}))$. These instruments, $p^{J_T}(\mathbf{w}_t)$, 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 3 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 3-month Treasury-bill rate (see Campbell (1991), Hodrick (1992), Lettau and Ludvigson (2001a)). We denote the relative bill rate as RREL

[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.

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, and is denoted $\widehat{cay}_t$.[10] Lettau and Ludvigson (2004) found 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 $\mathbf{w}_t = [\widehat{cay}_t, \mathrm{RREL}_t, \mathrm{SPEX}_t, \frac{C_t}{C_{t-1}}]'$. 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 pairwise cross products of each variable in $\mathbf{w}_t$, or 15 instruments in total.

6. Empirical results

6.1 *Parameter estimates*

The shape of our estimated continuation value ratio function $\frac{V_t}{C_t} = F(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}})$ can be illustrated by plotting $\widehat{F}(\cdot, \widehat{\boldsymbol{\delta}})$ 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, especially in Figure 2, from the finding that the shape depends on where in the domain space the function is evaluated. Notice that the serial dependence of $\widehat{F}$ is negative in both figures. Such a pattern is possible in the linear state space model

[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) reported that this variable also forecasts returns on portfolios sorted by size and book–market ratios.

[10]See Lettau and Ludvigson (2001a, 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 preestimation 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 1 to each variable, so that they have roughly the same units as unscaled returns.

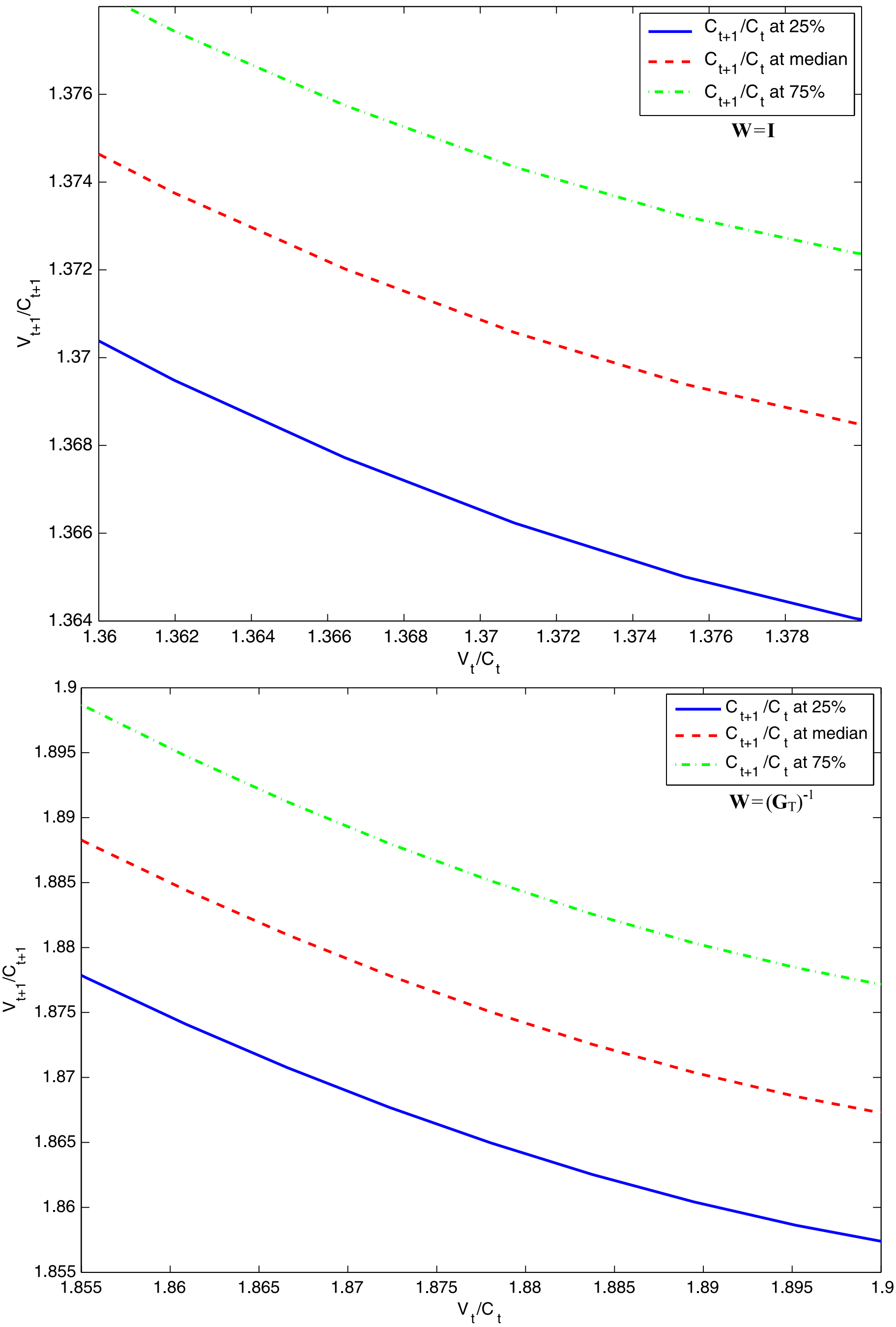

FIGURE 1. Plots of the estimated continuation value–consumption ratio against lagged values of the continuation value with consumption growth held alternately at the 25th, 50th, and 75th percentiles in the sample. Consumption is measured as aggregate consumption; the **W**s indicate the weighting matrix used in the second-step estimation. The sample is 1952:Q1–2005Q1.

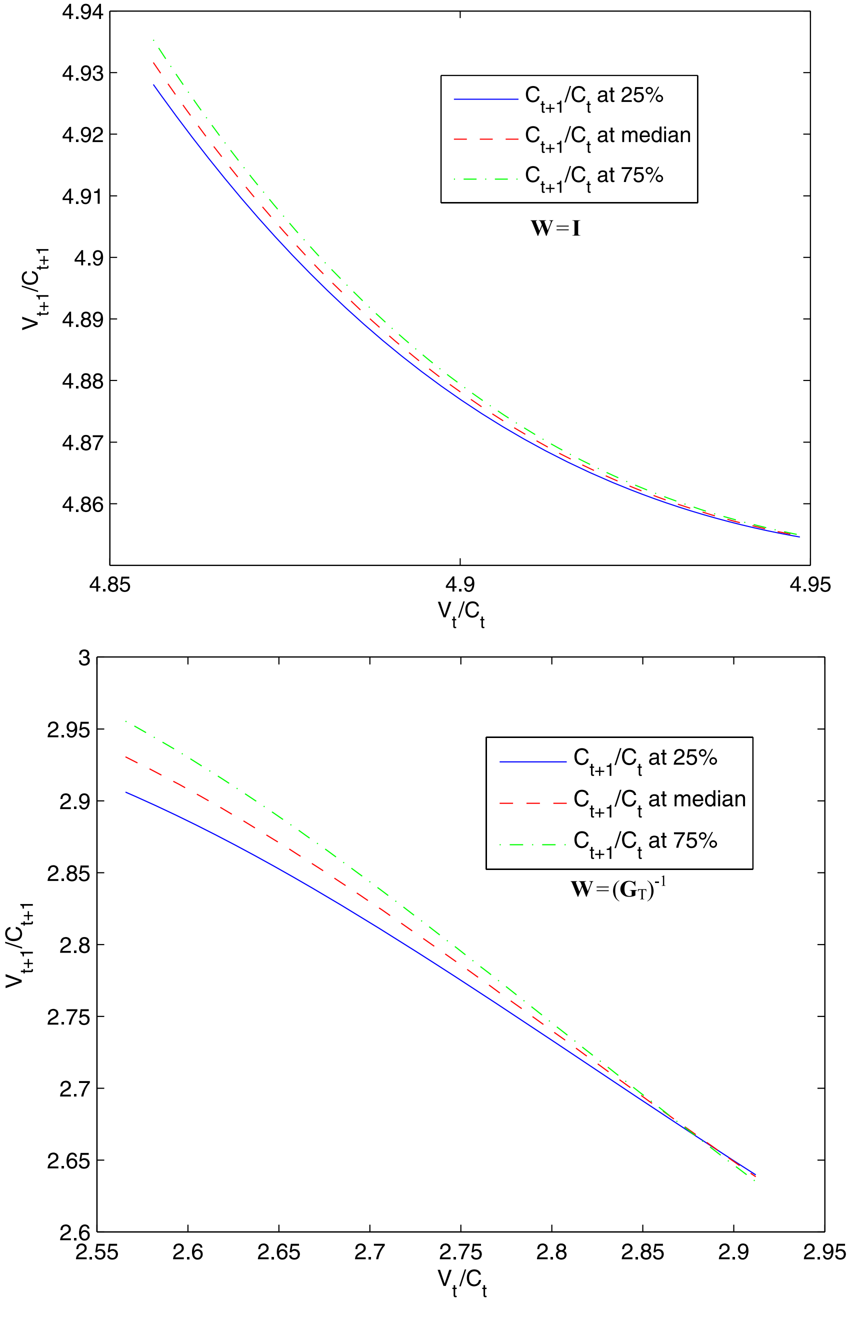

FIGURE 2. Plots of the estimated continuation value–consumption ratio against lagged values of the continuation value with consumption growth held alternately at the 25th, 50th, and 75th percentiles in the sample. Consumption is measured as stockholder consumption; the **W**s indicate the weighting matrix used in the second-step estimation. The sample is 1952:Q1–2005Q1.

Table 2. Value function statistics and preference parameter estimates.[a]

Value Function Statistics			
	Mean	Std	AC
Agg cons, $\mathbf{W}=\mathbf{I}$	1.37	0.011	0.53
Agg cons, $\mathbf{W}=\mathbf{G}_T^{-1}$	1.87	0.019	0.58
SH cons, $\mathbf{W}=\mathbf{I}$	4.89	0.025	−0.24
SH cons, $\mathbf{W}=\mathbf{G}_T^{-1}$	2.77	0.047	−0.29
Preference Parameter Estimates			
Second-step estimation	β	θ	ρ
Aggregate Consumption			
$\mathbf{W}=\mathbf{I}$	0.990	57.5	0.60
	(0.985, 0.996)	(27.5, 129)	(0.24, 0.99)
$\mathbf{W}=\mathbf{G}_T^{-1}$	0.999	60	0.50
	(0.994, 0.9999)	(42, 144)	(0.20, 0.75)
Stockholder Consumption			
$\mathbf{W}=\mathbf{I}$	0.994	20.00	0.90
	(0.993, 0.9995)	(0.25, 40)	(0.38, 1.24)
$\mathbf{W}=\mathbf{G}_T^{-1}$	0.998	17.0	0.68
	(0.992, 0.9999)	(1, 43.3)	(0.23, 1.01)

[a]The top panel presents statistics (mean, standard deviation (Std), autocorrelation (AC)) of the estimated value function. The bottom panel reports second-step estimates of preference parameters, with 95% confidence intervals in parentheses. 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 $\mathbf{W}=\mathbf{G}_T^{-1}$, where in both cases $\mathbf{x}_t = \mathbf{1}_N$, an $N \times 1$ vector of 1. The sample is 1952:Q1–2005:Q1.

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.

The top panel of Table 2 presents statistics of the estimated continuation value–consumption ratio function for cases estimated using aggregate (Agg cons) or stockholder (SH cons) consumption, and using one of two weighting matrices employed in the second step ($\mathbf{W}=\mathbf{I}$ or $\mathbf{W}=\mathbf{G}_T^{-1}$). These statistics are calculated by reading in the historical data as arguments to the estimated function $\frac{V_t}{C_t} = F(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}})$ and then computing statistics for the resulting time series on V_t/C_t. Not surprisingly given Figures 1 and 2, the mean of the estimated value function is greater than 1, more so for estimates using stockholder consumption growth. But the panel also shows that the function estimated on stockholder consumption growth is more volatile than that estimated on aggregate consumption growth; when $\mathbf{W}=\mathbf{I}$, V_t/C_t is about $3\frac{1}{3}$ times more volatile when estimated on stockholder consumption growth than when estimated on aggregate consumption growth. This plays a role in the lower risk aversion estimates discussed below. Finally, the last column of the top panel of Table 2 reports the autocorrelation statistics. Note that these are based on a linear univariate relation between V_t/C_t and V_{t-1}/C_{t-1}; thus they do not control for the influence of contemporaneous consumption growth, the second argument of the function $\frac{V_t}{C_t} = F(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}})$. For this reason, the function is positively autocorrelated in a univariate sense when computed using aggregate consumption growth, even though Figure 1 shows that, conditional on consumption growth, the function is negatively autocorrelated. This occurs because consumption growth is positively related to $\frac{V_t}{C_t} = F(\frac{V_{t-1}}{C_{t-1}}, \frac{C_t}{C_{t-1}})$ and is itself positively autocorrelated in aggregate data, implying that the univariate autoregressive coefficient is "biased up." The same bias is not present for estimates of the value–consumption ratio using stockholder consumption because stockholder consumption growth is not positively autocorrelated.

Table 2 presents estimates of the model's preference parameters $\boldsymbol{\delta} = (\beta, \rho, \theta)'$. The subjective time–discount factor, β, is close to 1 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 to 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 (2009), who focused on the special case of the EZW utility model in which the EIS $1/\rho$ is unity. 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 1, indicating that the EIS is above 1 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 1.47 in the representative stockholder version of the model. The estimates for this parameter are in line with those reported in Bansal, Gallant, and Tauchen (2007), who estimated a model of long-run consumption risk with EZW utility. In theoretical work, Bansal and Yaron (2004) 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. Lettau and Ludvigson (2009) emphasized the large empirical Euler equation errors generated by the standard power utility, representative agent asset pricing model when confronted with stock market data. Consistent with these findings, we find that the estimated Euler equation errors in this study are larger and considerably different from zero when the EIS is restricted to equal the inverse of the coefficient of relative risk aversion compared to when these parameters are left unrestricted.

Recall that the mean value of the continuation value–consumption ratio is higher using stockholder consumption data than it is using aggregate consumption data (Table 2, top panel). The preference parameter estimates for each case help explain these different mean values for V_t/C_t depending on whether the estimation is carried out using aggregate consumption data or stockholder consumption data. To understand how, consider a simple example of an EZW asset pricing model that can be solved analytically: $\Delta \ln C_{t+1} \sim$ i.i.d. $N(\mu, \sigma^2)$. Under this assumption, the Euler equations can be solved analytically for V_t/C_t, which is a constant equal to

$$V/C = \frac{\Omega}{1-\Omega},$$
$$\Omega = \beta \exp\left[(1-\rho)\mu + \frac{(1-\theta)(1-\rho)}{2}\sigma^2\right].$$

It is straightforward to show that V/C is increasing in β, decreasing in θ if $\rho < 1$ (the case we estimate), and increasing in ρ when θ is sufficiently greater than 1. Comparing estimates with the same weighting matrix (i.e., $\mathbf{W} = \mathbf{I}$ or $\mathbf{W} = \mathbf{G}^{-1}$), we see that those using stockholder consumption have higher β, lower θ, and higher ρ than do those using aggregate consumption, helping to explain why estimates using stockholder consumption data produce higher mean values of V/C than do those using aggregate consumption data. Of course, the data in our study do not necessarily conform to the distributional assumptions of this simple example. Nevertheless, plausible departures from these assumptions are likely to lead to (numerical) solutions for V_t/C_t that generate the same qualitative relationships between the mean of V_t/C_t and the EZW preference parameter values.

6.2 *Model misspecification and standard errors*

The estimation procedure used here allows for model misspecification, in the sense that the moment conditions are allowed to not hold with equality. In this event, the parameters estimated are pseudo-true parameters. The implementation itself is affected by the allowance for misspecification in the computation of standard errors. In the class of semiparametric models considered here, Ai and Chen (2007) proved that when the model is misspecified, as long as the pseudo-true parameter values are unique and are in the interior of the parameter space, the estimator is still root-T asymptotically normally distributed, centered at the pseudo-true parameter values, except that the asymptotic variance now includes extra terms that would be zero under correct specification. Due to the complication of the asymptotic variance expressions under misspecification, we compute block bootstrap estimates of the finite sample distributions of $\widehat{\boldsymbol{\delta}}$.

In the bootstrap, the sieve parameters $\frac{V_0}{C_0}$ and $\{a_j\}_{j=1}^{K_T}$, 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] The procedure is highly numerically intensive and takes several

[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 showed that the asymptotically optimal block length for estimating a symmetrical distribution function is $l \propto T^{1/5}$; also see Horowitz (2003).

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 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 1, or an EIS above 1.

6.2.1 *Cyclical properties of estimated pricing kernel* Figures 3–5 give a visual impression of the cyclical properties of the estimated EZW pricing kernel. For these figures, we focus on the properties of the estimated EZW model using aggregate consumption where the weighting matrix $\mathbf{W} = \mathbf{I}$ is employed in the second stage estimation. The estimated pricing kernel, M_{t+1}, is the product of two pieces, $M_{1,t+1}$ and $M_{2,t+1}$, denoted separately in the graphs:

$$M_{t+1} = \underbrace{\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\rho}}_{M_{1,t+1}} \underbrace{\left(\frac{\dfrac{V_{t+1}}{C_{t+1}}\dfrac{C_{t+1}}{C_t}}{\mathcal{R}_t\left(\dfrac{V_{t+1}}{C_{t+1}}\dfrac{C_{t+1}}{C_t}\right)}\right)^{\rho-\theta}}_{M_{2,t+1}}.$$

The first piece corresponds to the part of the pricing kernel that is present in the standard, constant relative risk aversion, power utility model that arises as a special case when $\rho = \theta$. The second piece is an additional multiplicative piece that is present more generally when $\rho \neq \theta$ and is attributable to the recursive preference structure of the EZW utility function.

Figure 3 plots the estimated pricing kernel M_{t+1} over time, along with real gross domestic product (GDP) growth (top panel). Both series are five-quarter moving averages. The middle and bottom panels plot the estimated values of $M_{1,t+1}$ and $M_{2,t+1}$ separately, over time. The pricing kernel M_{t+1} has a clear countercyclical component, rising in recessions and falling in booms. Its correlation with real GDP growth is -0.26 over our sample. Both $M_{1,t+1}$ and $M_{2,t+1}$ contribute to this negative correlation, but since $M_{1,t+1}$ is much less volatile than $M_{2,t+1}$, the overall correlation is close to that with just $M_{2,t+1}$.

The cyclical properties of the pricing kernel are of interest because they determine the cyclical properties of risk premia. Figures 4 and 5 plot an estimate of the risk premium (and its components) over time for the aggregate stock market implied by our estimate of M_{t+1}, computed as a five-quarter moving average of

$$\text{risk premium} = \frac{-\operatorname{Cov}(M_{t+1}, R_{\mathrm{CRSP},t+1} - R_{f,t+1})}{E(M_{t+1})},$$

where $R_{\mathrm{CRSP},t+1}$ denotes the return on the CRSP value-weighted stock market index and $R_{f,t+1}$ denotes the 3-month Treasury-bill rate. To give a rough idea of how the two components of the pricing kernel contribute to its dynamic behavior, some plots also exhibit the properties of $M_{1,t+1}$ and $M_{2,t+1}$ separately. In viewing these plots, the reader should

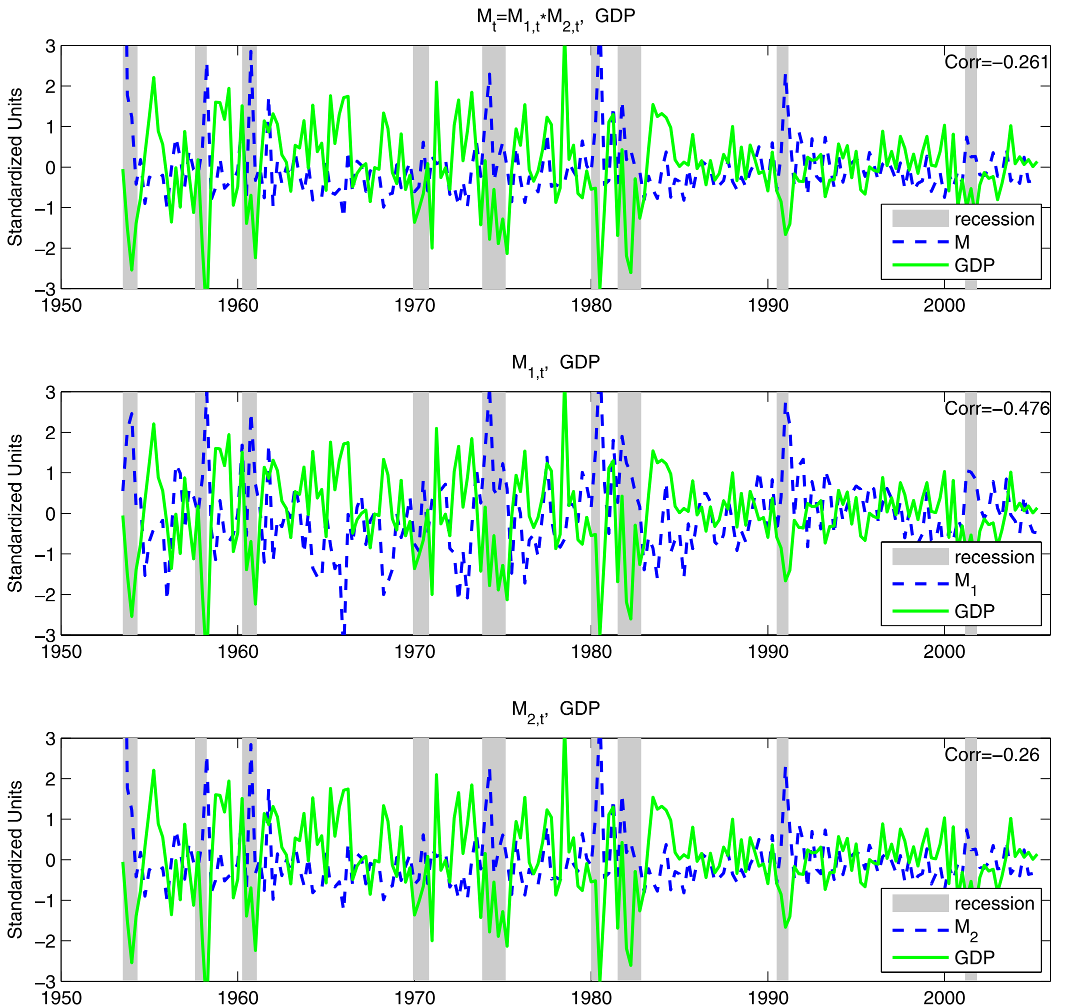

FIGURE 3. Cyclical properties of estimated EZW pricing kernel. The top panel plots the estimated pricing kernel, $M_t = M_{1,t} \cdot M_{2,t}$, as the product of two components, $M_{1,t}$ and $M_{2,t}$, along with real GDP growth over time. $M_{1,t}$ corresponds to the conventional CRRA piece, $M_{1,t+1} = \beta(C_{t+1}/C_t)^{-\rho}$; $M_{2,t}$ corresponds to the multiplicative piece added by EZW preferences, $M_{2,t+1} = (\frac{(V_{t+1}/C_{t+1})(C_{t+1}/C_t)}{\Re((V_{t+1}/C_{t+1})(C_{t+1}/C_t))})^{\rho-\theta}$. Corr indicates the correlation between the pricing kernel or one of its components and GDP growth. Shaded areas denote a recession as designated by the National Bureau of Economic Research. The SDF plotted is estimated using aggregate consumption, with $\mathbf{W} = \mathbf{I}$ as the weighting matrix in the second-step estimation. The sample is 1952:Q1–2005Q1.

keep in mind that the two components are likely to be correlated; thus the plots do not display orthogonal movements in $M_{1,t+1}$ and $M_{2,t+1}$.

Several aspects of Figures 4 and 5 are noteworthy. First, Figure 4 shows that the stock market risk premium has a marked countercyclical component: it rises in recessions and falls in expansions, and has a correlation of -0.16 with a five-quarter moving average of

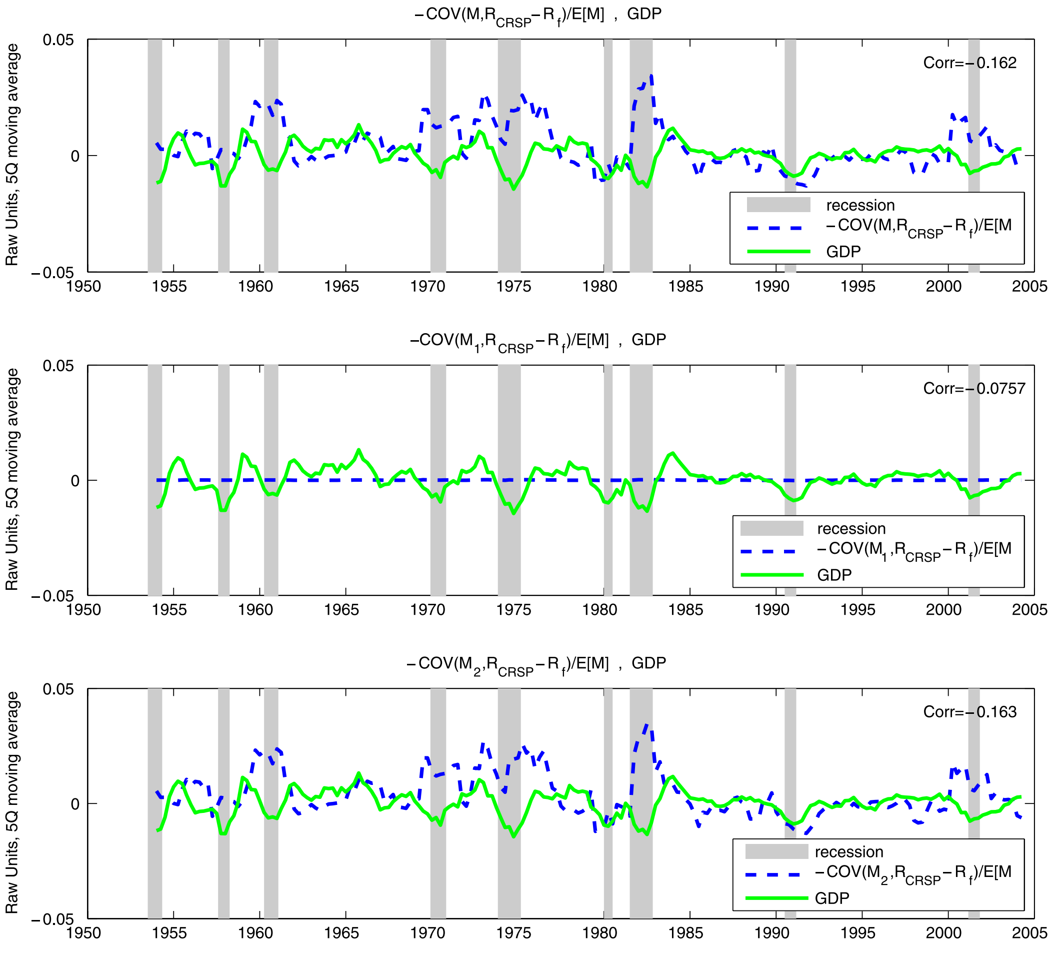

FIGURE 4. Cyclical properties of the market risk premium implied by EZW estimation. The top panel plots rolling, five-quarter estimates of the risk premium for the aggregate stock market, computed as the covariance of $M_t = M_{1,t} \cdot M_{2,t}$ with the CRSP excess stock market return, $R_{\text{CRSP},t} - R_{f,t}$, divided by the mean of M_t. Also plotted is real GDP growth over time. Corr indicates the correlation between the risk premium and GDP growth. $M_{1,t}$ corresponds to the conventional CRRA piece, $M_{1,t+1} = \beta(C_{t+1}/C_t)^{-\rho}$; $M_{2,t}$ corresponds to the multiplicative piece added by EZW preferences, $M_{2,t+1} = (\frac{(V_{t+1}/C_{t+1})(C_{t+1}/C_t)}{\Re((V_{t+1}/C_{t+1})(C_{t+1}/C_t))})^{\rho-\theta}$. Shaded areas denote a recession as designated by the National Bureau of Economic Research. The SDF plotted is estimated using aggregate consumption, with $\mathbf{W} = \mathbf{I}$ as the weighting matrix in the second-step estimation. The sample is 1952:Q1–2005Q1.

real GDP growth. Second, the next two panels show the (negative of the) covariance between $M_{1,t+1}$ and $R_{\text{CRSP},t+1} - R_{f,t+1}$ (middle panel) and the (negative of the) covariance between $M_{2,t+1}$ and $R_{\text{CRSP},t+1} - R_{f,t+1}$ (bottom panel). The covariance with $M_{2,t+1}$ is much larger than that with $M_{1,t+1}$ because the former has a much larger standard deviation. (Given our parameter estimates, the variable in parentheses of $M_{2,t+1}$is raised to a large number in absolute value.) However, both components of the pricing kernel

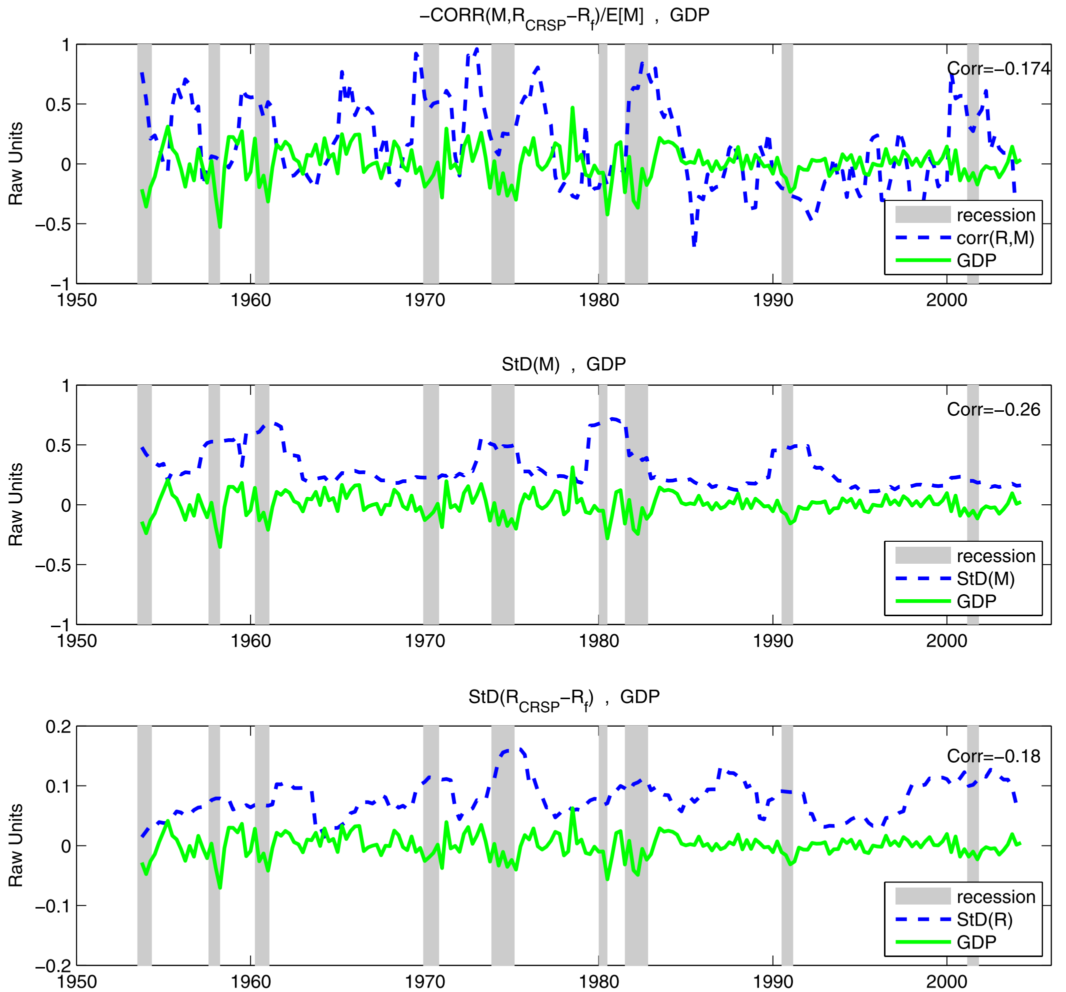

FIGURE 5. Cyclical properties of the components of the market risk premium implied by EZW estimation. The estimated pricing kernel is $M_t = M_{1,t} \cdot M_{2,t}$. The top panel plots rolling, five-quarter estimates of $\text{Corr}(M_t, R_{\text{CRSP},t} - R_{f,t})/E(M_t)$, along with real GDP growth over time. The bottom subpanels plot rolling, five-quarter estimates of the standard deviations of $M_{1,t}$ and $M_{2,t}$. $M_{1,t}$ corresponds to the conventional CRRA piece, $M_{1,t+1} = \beta(C_{t+1}/C_t)^{-\rho}$; $M_{2,t}$ corresponds to the multiplicative piece added by EZW preferences, $M_{2,t+1} = (\frac{(V_{t+1}/C_{t+1})(C_{t+1}/C_t)}{\Re((V_{t+1}/C_{t+1})(C_{t+1}/C_t))})^{\rho-\theta}$. Corr indicates the correlation between the dashed line in each subplot and GDP growth. Shaded areas denote a recession as designated by the National Bureau of Economic Research. The SDF plotted is estimated using aggregate consumption, with $\mathbf{W} = \mathbf{I}$ as the weighting matrix in the second-step estimation. The sample is 1952:Q1–2005Q1.

display a countercyclical correlation with the excess stock market return, rising in recessions and falling in expansions.

Third, the countercyclicality of $-\text{Cov}(M_{t+1}, R_{\text{CRSP},t+1} - R_{f,t+1})/E(M_{t+1})$ is attributable to countercyclicality in the correlation, $-\text{Corr}(M_{t+1}, R_{\text{CRSP},t+1} - R_{f,t+1})/E(M_{t+1})$, but also to countercyclical heteroskedasticity in the pricing kernel and in excess returns. Figure 5 plots the five-quarter moving average of $-\operatorname{Corr}(M_{t+1}, R_{\mathrm{CRSP},t+1} - R_{f,t+1})/E(M_{t+1})$ (top panel), of the standard deviation of M_{t+1}, $\operatorname{StD}(M_{t+1})$ (middle panel), and of the standard deviation of the excess return, $\operatorname{StD}(R_{\mathrm{CRSP},t+1} - R_{f,t+1})$. All three components rise sharply in recessions and fall in booms. The correlation component has a correlation of -0.17 with real GDP growth, but the standard deviation of the pricing kernel is even more countercyclical, having a correlation with real GDP growth of -0.26. The correlation between the standard deviation of excess returns and real GDP growth is -0.18.

6.3 *Model comparison*

In this section we address the question of how well the EZW recursive utility model explains asset pricing data relative to competing specifications. We use the methodology provided by Hansen and Jagannathan (1997), which allows all stochastic discount factor models to be treated as misspecified proxies for the true unknown SDF.

Hansen and Jagannathan suggested 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 $\mathbf{g}_T^{\mathrm{HJ}}(b) \equiv \{g_T(b)\}'\mathbf{G}_T^{-1}g_T(b)$, where $g_T(b) = (g_{1T}(b), \ldots, g_{NT}(b))'$ 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 $i = 1, \ldots, N$) 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)th element of $\mathbf{G}_T$ is $\frac{1}{T}\sum_{t=1}^{T} R_{i,t}R_{j,t}$ for $i, j = 1, \ldots, N$). The measure of model misspecification is then the square root of this minimized quadratic form, $d_T \equiv \sqrt{\mathbf{g}_T^{\mathrm{HJ}}(\widehat{b})}$, 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, $d_T \equiv \sqrt{\mathbf{g}_T^{\mathrm{HJ}}(\widehat{b})}$, as the *Hansen–Jagannathan (HJ) distance*.

We also compute a conditional version of the distance metric that incorporates conditioning information Z_t. In this case, $g_T(b) = \frac{1}{T}\sum_{t=1}^{T}[(M_{t+1}(b)R_{t+1} - 1_N) \otimes Z_t]$ 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) 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

$$M_{t+1}(b) = b_0 + \sum_{i=1}^{k} b_i F_{i,t+1}, \tag{29}$$

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 developed 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 that is 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, 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) showed 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. 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 a 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 SDFs. To do so, we choose the unknown parameters $b = (b_0, b_1, \ldots, b_k)'$ of

[13] SMB is the difference between the returns on small and big stock portfolios with the same weighted-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 webpage (see the Appendix).

the linear models to minimize the squared HJ distance for that model, subject to the constraint that the SDF proxy must 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 (19) 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 (23). Note that this places the EZW model (19) 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 Akaike information criterion (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.[15]

Table 3 reports the measure of specification error given by the HJ distance (HJ Dist), $d_T \equiv \sqrt{\mathbf{g}_T^{\mathrm{HJ}}(\widehat{b})}$ 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 the scaled CCAPM: the HJ distance is smaller when models are compared on the basis of either the unconditional or conditional HJ distance, regardless of which measure of consumption is used.[16] Third, when the representative stockholder version of the model is estimated,

[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.

[15]The adjusted criterion function (with AIC penalty) is

$$\sqrt{d_T^2 + \frac{\#\mathrm{param}}{T}},$$

where #param refers to the number of free parameters b chosen to minimize the Hansen–Jagannathan distance.

[16]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

Table 3. Specification errors for alternative models: HJ distance.[a]

Model (1)	Unconditional HJ Dist (2)	Conditional HJ Dist (3)
Aggregate Consumption		
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		
Recursive	0.463	0.605
CRRA utility	0.517	0.627
Fama–French	0.363	0.515
Scaled CCAPM	0.490	0.620

[a]The table reports the Hansen–Jagannathan distance metric

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

where b are parameter values associated with the model listed in column 1. In column 2, $g_T(b) \equiv \frac{1}{T}\sum_{t=1}^{T}[\{M_t(b)\}^+ R_t - 1_N]$ 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, $g_T(b) \equiv \frac{1}{T}\sum_{t=1}^{T}[(\{M_{t+1}(b)\}^+ R_{t+1} - 1_N) \otimes Z_t]$ 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.

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.[17]

Several authors have focused on the cross-sectional implications of EZW preferences when the EIS, ρ^{-1}, is restricted to unity (e.g., Hansen, Heaton, and Li (2008), Malloy, Moskowitz, and Vissing-Jorgensen (2009)). The Appendix presents results when we re-

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.

[17]Following Hansen and Jagannathan (1997), HJ^+ is computed numerically as

$$\text{HJ}^+ = \left\{ \max_{\lambda \in \mathbb{R}^N} (1/T) \sum_{t=1}^{T} \left[(M_{t+1})^2 - \left(M_{t+1} - \lambda' R_{t+1}\right)^{+2} - 2\lambda' \right] \right\}^{1/2},$$

where $(M_{t+1} - \lambda' R_{t+1})^+ = \max\{M_{t+1} - \lambda' R_{t+1}, 0\}$.

TABLE 4. Specification errors for alternative models: HJ$^+$ distance.[a]

Model (1)	Unconditional HJ$^+$ Dist (2)	Conditional HJ$^+$ Dist (3)
Aggregate Consumption		
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		
Recursive	0.463	0.605
CRRA utility	0.517	0.627
Fama–French	0.338	0.506
Scaled CCAPM	0.467	0.661

[a]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.

peat our estimation with $\rho = 1$ fixed. We find qualitatively similar results in an estimation of the representative stockholder version of the model.

7. 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 $\{\widehat{\boldsymbol{\delta}}, \widehat{F}(\cdot, \widehat{\boldsymbol{\delta}})\}$, with its theoretical representation based on consumption growth and the return to aggregate wealth (6):

$$\beta\left(\frac{C_{t+1}}{C_t}\right)^{-\rho}\left(\frac{\dfrac{V_{t+1}}{C_{t+1}}\dfrac{C_{t+1}}{C_t}}{\mathcal{R}_t\left(\dfrac{V_{t+1}}{C_{t+1}}\dfrac{C_{t+1}}{C_t}\right)}\right)^{\rho-\theta}$$

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

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 assumed 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 started with the relationship

$$R_{w,t+1} = (1 - \nu_t)R_{a,t+1} + \nu_t R_{y,t+1}, \tag{30}$$

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 (30) is that the wealth shares may, in principle, vary over time. Campbell dealt with this by linearizing (30) 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 can 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 (30) 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, $R_{y,t+1}$, can be inferred.

The exercise in this section is similar in spirit to the investigation of Lustig and Van Nieuwerburgh (2008). These authors, following Campbell (1996), investigated a log-linear version of the EZW recursive utility model under the assumption that asset returns and consumption are jointly log-normal and homoskedastic. With these assumptions, the authors backed out the human wealth return from observable aggregate consumption data and found 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 log-normal 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 (2008), 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, and for the $\mathbf{W} = \mathbf{G}_T^{-1}$ case in Table 6. For comparison, summary statistics on the CRSP value-weighted return, $R_{\mathrm{CRSP},t+1}$, 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 $R_{\mathrm{CRSP},t+1}$ is 0.165. Second, in the representative agent model, the mean of $R_{w,t+1}$ is less than the mean of $R_{\mathrm{CRSP},t+1}$, 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 $R_{\mathrm{CRSP},t+1}$, and given that the mean of $R_{\mathrm{CRSP},t+1}$ is 0.084, the

TABLE 5. Summary statistics for return to aggregate wealth and human wealth, $\mathbf{W} = \mathbf{I}$.[a]

Model-Implied Aggregate Wealth Return				
	Representative Agent		Representative Stockholder	
	$R_{w,t}$	$R_{\mathrm{CRSP},t}$	$R_{w,t}$	$R_{\mathrm{CRSP},t}$
Panel A: Correlation Matrix				
$R_{w,t}$	1.00	0.171	1.00	−0.049
$R_{\mathrm{CRSP},t}$		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

Model-Implied Human Wealth Return, $\nu = 0.333$				
	Representative Agent		Representative Stockholder	
	$R_{y,t}$	$R_{\mathrm{CRSP},t}$	$R_{y,t}$	$R_{\mathrm{CRSP},t}$
Panel A: Correlation Matrix				
$R_{y,t}$	1.00	−0.996	1.00	−0.953
$R_{\mathrm{CRSP},t}$		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		Representative Stockholder	
	$R_{y,t}$	$R_{\mathrm{CRSP},t}$	$R_{y,t}$	$R_{\mathrm{CRSP},t}$
Panel A: Correlation Matrix				
$R_{y,t}$	1.00	−0.982	1.00	−0.847
$R_{\mathrm{CRSP},t}$		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

[a] 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, $R_{\mathrm{CRSP},t}$. 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 1. The sample is 1952:Q1–2005:Q1.

mean of the human wealth return can be quite small if, as in the representative agent model, the mean of aggregate wealth return 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 0.023), the gross return on human wealth can even be less than 1, so that the simple net return is negative. Third,

TABLE 6. Summary statistics for return to aggregate wealth and human wealth, $\mathbf{W} = \mathbf{G}_T^{-1}$.[a]

Model-Implied Aggregate Wealth Return

	Representative Agent		Representative Stockholder	
	$R_{w,t}$	$R_{\mathrm{CRSP},t}$	$R_{w,t}$	$R_{\mathrm{CRSP},t}$
Panel A: Correlation Matrix				
$R_{w,t}$	1.00	0.18	1.00	0.004
$R_{\mathrm{CRSP},t}$		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

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

	Representative Agent		Representative Stockholder	
	$R_{y,t}$	$R_{\mathrm{CRSP},t}$	$R_{y,t}$	$R_{\mathrm{CRSP},t}$
Panel A: Correlation Matrix				
$R_{y,t}$	1.00	−0.994	1.00	−0.921
$R_{\mathrm{CRSP},t}$		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		Representative Stockholder	
	$R_{y,t}$	$R_{\mathrm{CRSP},t}$	$R_{y,t}$	$R_{\mathrm{CRSP},t}$
Panel A: Correlation Matrix				
$R_{y,t}$	1.00	−0.975	1.00	−0.764
$R_{\mathrm{CRSP},t}$		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

[a]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, $R_{\mathrm{CRSP},t}$. 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.

the return to human wealth is a weighted average (where the weights exceed 1 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 $R_{\mathrm{CRSP},t+1}$ ranges 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 (2008) for the cases where the EIS exceeds 1.

The Appendix of this paper presents additional results from an investigation of the implications of the findings above for forecastability of the multihorizon excess return to the aggregate wealth portfolio, $R_{w,t+h}$, using the log aggregate wealth–consumption ratio $\ln W_t - \ln C_t$ as a predictor variable.

8. Conclusion

In this paper we undertake a semiparametric econometric evaluation of the Epstein–Zin–Weil recursive utility model, a framework upon which a large and growing body of theoretical work in 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 the 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 to 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 log-linear version of the model where the EIS is restricted to unity, which was studied by Malloy, Moskowitz, and Vissing-Jorgensen (2009). On the other hand, we find that the estimated elasticity of intertemporal substitution is typically above 1, 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 log-linear version of the model studied by Lustig and Van Nieuwerburgh (2008).

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 principle, 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 so as to fit the equity return data.

A possible objection to our estimation approach concerns the applicability of the model to microeconomic data. Suppose we take the model of preferences we have estimated as literally true at the individual level. There is no general aggregation result stating that these same preferences hold for a representative agent, that is, for the average consumption of some set of heterogeneous households. In this case, the resulting parameter estimates on average consumption data may be biased estimates of the preference parameters applicable to an individual. Attanasio and Weber (1993) emphasized this point in documenting that estimates of the EIS are typically lower for aggregate data than they are for average cohort data. A second possible objection concerns the use of average stockholder consumption data when stock market participation rates have fluctuated over the sample. If different individuals move in and out of the stock market, the average consumption growth of stockholders may not correspond to that of any single stockholder or even to the growth rate of the average consumption of individuals who remained stockholders between t and $t+1$ (Attanasio, Banks, and Tanner (2002)).

If the null hypothesis is that the preferences we have estimated are an accurate representation of the true preferences of individuals, these considerations point to important areas for future research using household-level data. Preference heterogeneity across households (including possibly the nonparametric part of the utility function), and possible nonclassical measurement error in household-level data are important challenges that would need to be addressed in the context of nonlinear estimation with an unknown function. But the applicability to microeconomic data is not the primary concern of the present paper. Our goal, challenging enough, is to take the representative agent specifications that have been routinely employed in the large and growing asset pricing literature on EZW preferences and provide some empirical content to the parameter values of the utility function as well as provide formal statistical tests of the model's ability to fit the data relative to competing specifications. The representative agent preferences used in this literature could take the same form as those of individual agents or they could result from aggregation of heterogeneous agents with quite different preferences. An important aspect of this approach is that the model of the stochastic discount factor need not be correctly specified, thereby permitting estimation under misspecification. Misspecification could arise for a number of reasons, including lack of complete aggregation when markets are incomplete or mismeasurement of stockholder consumption over time. If the model is misspecified, the methodology here will not allow us to uncover the true preference parameter estimates, but it does allow us to estimate the pseudo-true parameters (those that best fit the data) of the representative agent approximating specification and to assess the magnitude of misspecification relative to competing specifications. An important area of future research is to investigate how the magnitude of specification error in the representative agent versions of the model compares to that when these same preferences are applied to individual level data.

References

Ai, C. and X. Chen (2003), "Efficient estimation of models with conditional moment restrictions containing unknown functions." *Econometrica*, 71, 1795–1843. [52]

Ai, C. and X. Chen (2007), "Estimation of possibly misspecified semiparametric conditional moment restriction models with different conditioning variables." *Journal of Econometrics*, 141, 5–43. [52, 64]

Attanasio, O. P., J. Banks, and S. Tanner (2002), "Asset holding and consumption volatility." *Journal of Political Economy*, 110 (4), 771–792. [78]

Attanasio, O. and G. Weber (1989), "Intertemporal substitution, risk aversion and the Euler equation for consumption." *Economic Journal*, 99, 59–73. [40, 41]

Attanasio, O. P. and G. Weber (1993), "Consumption growth, the interest rate and aggregation." *Review of Economic Studies*, 60 (3), 631–649. [78]

Bansal, R., R. F. Dittmar, and D. Kiku (2009), "Cointegration and consumption risk in equity returns." *Review of Financial Studies*, 22, 1343–1375. [40]

Bansal, R., R. F. Dittmar, and C. T. Lundblad (2005), "Consumption, dividends, and the cross-section of equity returns." *Journal of Finance*, 60 (4), 1639–1672. [40]

Bansal, R., A. R. Gallant, and G. Tauchen (2007), "Rational pessimism, rational exuberance, and asset pricing models." *Review of Economic Studies*, 74, 1005–1033. [43, 63]

Bansal, R., D. Kiku, and A. Yaron (2007), "Risks for the long-run: Estimation and inference." Unpublished paper, Fuqua School, Duke University. [43]

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. [40, 46, 55, 63]

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. [40]

Campbell, J. Y. (1991), "A variance decomposition for stock returns." *Economic Journal*, 101, 157–179. [58]

Campbell, J. Y. (1993), "Intertemporal asset pricing without consumption data." *American Economic Review*, 83, 487–512. [40]

Campbell, J. Y. (1996), "Understanding risk and return." *Journal of Political Economy*, 104 (2), 298–345. [40, 42, 45, 73, 74]

Campbell, J. Y. (2003), "Consumption-based asset pricing." In *Handbook of the Economics of Finance*, Vol. I-B (G. Constantinides, M. Harris, and R. Stulz, eds.), 803–887, Elsevier Science B.V., Amsterdam. [41]

Campbell, J. Y. and L. Viceira (2001), *Strategic Asset Allocation: Portfolio Choice for Long-Term Investors*. Oxford University Press, London, UK. [40]

Campbell, J. Y. and T. Vuolteenaho (2004), "Bad beta, good beta." *American Economic Review*, 94 (5), 1249–1275. [40]

Chen, X. (2007), "Large sample sieve estimation of semi-nonparameteric models." In *Handbook of Econometrics*, Vol. 6B (J. Heckman and E. Leamer, eds.), North-Holland, Amsterdam, The Netherlands. [51]

Chen, X., O. Linton, and I. van Keilegom (2003), "Estimation of semiparametric models when the criterion function is not smooth." *Econometrica*, 71, 1591–1608. [51]

Cochrane, J. H. (2001), *Asset Pricing*. Princeton University Press, Princeton, NJ. [59]

Colacito, R. and M. M. Croce (2004), "Risks for the long run and the real exchange rate." Unpublished paper, New York University. [40]

Croce, M. M. (2012), "Welfare costs for the long-run." Unpublished paper, Kenan-Flagler Business School. [40]

Croce, M. M., M. Lettau, and S. C. Ludvigson (2012), "Investor information, long-run risk and the term structure of equity." Unpublished paper, New York University. [40, 49]

Epstein, L. and S. Zin (1989), "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework." *Econometrica*, 57, 937–969. [39, 40, 41, 44]

Epstein, L. and S. Zin (1991), "Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis." *Journal of Political Economy*, 99, 263–286. [39, 40, 41, 44, 45]

Fama, E. F. and K. R. French (1992), "The cross-section of expected returns." *Journal of Finance*, 47, 427–465. [56]

Fama, E. F. and K. R. French (1993), "Common risk factors in the returns on stocks and bonds." *Journal of Financial Economics*, 33, 3–56. [43, 56, 70]

Fama, E. F. and K. R. French (1996), "Multifactor explanations of asset pricing anomalies." *Journal of Finance*, 51, 55–84. [70]

Gabaix, X. and D. Laibson (2002), "The 6-D bias and the equity premium puzzle." In *NBER Macroeconomics Annual: 2002* (B. Bernanke and K. Rogoff, eds.), 257–312, MIT Press, Cambridge, MA. [56]

Gomes, F. and A. Michaelides (2005), "Optimal life-cycle asset allocation: Understanding the empirical evidence." *Journal of Finance*, 60 (2), 869–904. [40]

Grenander, U. (1981), *Abstract Inference*. Wiley, New York. [42]

Hall, P., J. L. Horowitz, and B. Y. Jing (1995), "On blocking rules for the bootstrap with dependent data." *Biometrika*, 82, 561–574. [64]

Hansen, L. P. (1982), "Large sample properties of generalized methods of moments estimators." *Econometrica*, 50, 1029–1054. [54]

Hansen, L. P., J. Heaton, and N. Li (2008), "Consumption strikes back?: Measuring long-run risk." *Journal of Political Economy*, 116 (2), 260–302. [40, 41, 44, 46, 72]

Hansen, L. P., J. Heaton, N. Roussanov, and J. Lee (2007), "Intertemporal substitution and risk aversion." In *Handbook of Econometrics*, Vol. 6B (J. Heckman and E. Leamer, eds.), North-Holland, Amsterdam, The Netherlands. [41]

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. [78]

Hansen, L. P. and R. Jagannathan (1997), "Assessing specification errors in stochastic discount factor models." *Journal of Finance*, 52, 557–590. [42, 51, 55, 69, 72]

Hansen, L. P. and T. J. Sargent (2006), "Fragile beliefs and the price of model uncertainty." Unpublished Paper, New York University. [40]

Hansen, L. P. and J. A. Scheinkman (2012), "Recursive utility in a Markov environment with stochastic growth." *Proceedings of the National Academy of Sciences*, 109 (30), 11967–11972. [48]

Hansen, L. P. and K. Singleton (1982), "Generalized instrumental variables estimation of nonlinear rational expectations models." *Econometrica*, 50 (5), 1269–1286. [40]

Hansen, L. P. and K. Singleton (1983), "Stochastic consumption, risk aversion, and the temporal behavior of asset returns." *Journal of Political Economy*, 91, 249–265. [40]

Heaton, J. (1993), "The interaction between time-nonseparable preferences and time aggregation." *Econometrica*, 61 (2), 353–385. [55]

Hodrick, R. (1992), "Dividend yields and expected stock returns: Alternative procedures for inference and measurement." *Review of Financial Studies*, 5, 357–386. [58]

Horowitz, J. L. (2003), "The bootstrap." In *Handbook of Econometrics*, Vol. 5 (J. J. Heckman and E. Leamer, eds.), Elsevier Science B.V., Amsterdam. [64]

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. [56]

Kandel, S. and R. F. Stambaugh (1995), "Portfolio inefficiency and the cross-section of expected returns." *Journal of Finance*, 50, 157–184. [55, 69]

Kiku, D. (2005), "Is the value premium a puzzle?" Unpublished paper, Duke University. [40]

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. [41, 78]

Krueger, D. and F. Kubler (2006), "Pareto improving social security reform when financial markets are incomplete!?" *American Economic Review*, 96 (3), 737–755. [40]

Lettau, M. and S. C. Ludvigson (2001a), "Consumption, aggregate wealth and expected stock returns." *Journal of Finance*, 56 (3), 815–849. [41, 58, 59]

Lettau, M. and S. C. Ludvigson (2001b), "Resurrecting the (C)CAPM: A cross-sectional test when risk premia are time-varying." *Journal of Political Economy*, 109 (6), 1238–1287. [43, 59, 69, 70, 71]

Lettau, M. and S. C. Ludvigson (2004), "Understanding trend and cycle in asset values: Reevaluating the wealth effect on consumption." *American Economic Review*, 94 (1), 276–299. [59]

Lettau, M. and S. C. Ludvigson (2009), "Euler equation errors." *Review of Economic Dynamics*, 12 (2), 255–283. [41, 63]

Lustig, H. and S. Van Nieuwerburgh (2008), "The returns on human capital: Good news on Wall Street is bad news on Main Street." *Review of Financial Studies*, 21, 2097–2137. [43, 74, 77, 78]

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. [41]

Malloy, C. J., T. J. Moskowitz, and A. Vissing-Jorgensen (2009), "Long-run stockholder consumption risk and asset returns." *Journal of Finance*, 64, 2427–2479. [40, 42, 57, 58, 63, 72, 77]

Newey, W. K. (1994), "The asymptotic variance of semiparametric estimators." *Econometrica*, 62, 1349–1382. [51]

Newey, W. K. and J. Powell (2003), "Instrumental variable estimation of nonparametric models." *Econometrica*, 71, 1565–1578. [52]

Newey, W. K. and K. D. West (1987), "A simple, positive semidefinite, heteroskedasticity and autocorrelation consistent covariance matrix." *Econometrica*, 55, 703–708. [58]

Piazzesi, M. and M. Schneider (2006), "Equilibrium yield curves." In *NBER Macroeconomics Annual: 2006* (M. W. D. Acemoglu and K. Rogoff, eds.), MIT Press, Cambridge, MA. [40]

Restoy, F. and P. Weil (1998), "Approximate equilibrium asset prices." Working Paper 6611, NBER. [41]

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), 518–529. [58]

Tallarini, T. D. (2000), "Risk-sensitive real business cycle models." *Journal of Monetary Economics*, 45, 507–532. [40]

Vissing-Jorgensen, A. (2002), "Limited asset market participation and intertemporal substitution." *Journal of Political Economy*, 110 (4), 825–853. [42, 57, 58]

Vissing-Jorgensen, A. and O. P. Attanasio (2003), "Stock market participation, intertemporal substitution and risk-aversion." *American Economic Review, AEA Papers and Proceedings*, 93 (2), 383–391. [40]

Weil, P. (1989), "The equity premium puzzle and the risk-free rate puzzle." *Journal of Monetary Economics*, 24 (3), 401–421. [39, 40, 41]

Submitted July, 2010. Final version accepted October, 2012.