NBER WORKING PAPER SERIES

LAND OF ADDICTS? AN EMPIRICAL INVESTIGATION OF HABIT-BASED ASSET PRICING BEHAVIOR

Xiaohong Chen
Sydney C. Ludvigson

Working Paper 10503
http://www.nber.org/papers/w10503

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2004

We acknowledge financial support from the National Science Foundation (Chen), the Alfred P. Sloan Foundation (Ludvigson), and the C.V. Starr Center at NYU. We thank Kenneth French for providing the portfolio return data, and are grateful to John Campbell, David Chapman, John Cochrane, Yuichi Kitamura, Per Krusell, Lars Hansen, Oliver Linton, Anthony Lynch, Monika Piazzesi, Masao Ogaki, B. Ravikumar, Martin Schneider, Jim Stock, George Tauchen, Zhenyu Wang and seminar participants at Carnegie Mellon GSIA, Columbia University, Norwegian School of Economics and Business Administration, Norwegian School of Management, Notre Dame, NYU, University of Chicago, University of Maryland, University of Pennsylvania, 2004 CIREQ-CIRANO Financial Econometrics Conference at Montreal, 2004 NBER Spring Asset Pricing meeting, 2003 NBER Summer Institute, 2004 American Finance Association Meetings, and 2004 Econometric Society winter meetings for useful comments. Jinyong Kim and Artem Voronov provided excellent research assistance. Any errors or omissions are the responsibility of the authors. The views expressed herein are those of the author(s) and not necessarily those of the National Bureau of Economic Research.

Land of Addicts? An Empirical Investigation of Habit-Based Asset Pricing Behavior
Xiaohong Chen and Sydney C. Ludvigson
NBER Working Paper No. 10503
May 2004
JEL No. G1

ABSTRACT

This paper studies the ability of a general class of habit-based asset pricing models to match the conditional moment restrictions implied by asset pricing theory. We treat the functional form of the habit as unknown, and to estimate it along with the rest of the model's finite dimensional parameters. Using quarterly data on consumption growth, assets returns and instruments, our empirical results indicate that the estimated habit function is nonlinear, the habit formation is better described as internal rather than external, and the estimated time-preference parameter and the power utility parameter are sensible. In addition, the estimated habit function generates a positive stochastic discount factor (SDF) proxy and performs well in explaining cross-sectional stock return data . We find that an internal habit SDF proxy can explain a cross-section of size and book-market sorted portfolio equity returns better than (i) the Fama and French (1993) three-factor model, (ii) Lettau and Ludvigson (2001) scaled consumption CAPM model, (iii) an external habit SDF proxy, (iv) the classic CAPM, and (v) the classic consumption CAPM.

Xiaohong Chen
Department of Economics
New York University
269 Mercer Street
7th Floor
New York, NY 10003
xiaohong.chen@nyu.edu

Sydney C. Ludvigson
Department of Economics
New York University
269 Mercer Street
7th Floor
New York, NY 10003
and NBER
sydney.ludvigson@nyu.edu

1 Introduction

A prominent explanation of aggregate stock market behavior says that assets are priced as if there were a representative investor whose utility is a power function of the difference between aggregate consumption and a habit level.[1] In these theories, the habit function is the central part of the asset pricing model. With few exceptions (discussed below), these models have been empirically evaluated by engaging in calibration and simulation exercises, in which a chosen set of moments computed from model-simulated data are informally compared to those computed from historical data. Although such calibration exercises are undoubtedly important as an initial step, a complete assessment of the habit-based paradigm requires moving beyond calibration, to formal estimation, hypothesis testing, and model comparison.

Formal estimation, testing, and model comparison present some significant challenges. Among the most important is that theory does not provide precise guidelines about the functional form of the habit. Consider the range of habit-based asset pricing models cited in footnote 1. All models assume that the habit stock is a function of past and (possibly) contemporaneous consumption. But there is substantial divergence across models in how the habit stock is specified to vary with aggregate consumption. Some work specifies the habit stock as a linear function of past consumption (e.g., Sundaresan (1989); Constantinides (1990); Heaton (1995); Jermann (1998); Boldrin, Christiano, and Fisher (2001)). More recent theoretical work often takes as a starting point the particular nonlinear habit specification that includes current consumption developed in Campbell and Cochrane (1999) (e.g., Campbell and Cochrane (2000); Li (2001); Wachter (2006); and Menzly, Santos, and Veronesi (2004)). Moreover, there is no theoretical reason why other forms of nonlinearities could not be entertained. These considerations imply that the functional form of the habit should be treated, not as a given, but as part and parcel of any empirical investigation.

This study econometrically evaluates a general class of habit-based asset pricing models, placing as few restrictions as possible on the specification of the habit and no parametric restrictions on the law of motion for consumption. The empirical model we explore follows the existing theoretical literature and presumes that investor utility is a power function of the difference between aggregate consumption and a habit level, but we allow the habit to be an unknown function of lagged and contemporaneous consumption. The resulting specification for investor utility is semiparametric in the sense that it contains both the finite dimensional set of unknown parameters that are part of the power function and time-preference, as well as the infinite dimensional unknown habit function that must be estimated nonparametrically.

[1] See Sundaresan (1989), Constantinides (1990), Ferson and Harvey (1992), Heaton (1995), Jermann (1998), Campbell and Cochrane (1999), Campbell and Cochrane (2000); Boldrin, Christiano, and Fisher (2001), Li (2001), Shore and White (2002); Dai (2003); Menzly, Santos, and Veronesi (2004); Wachter (2006). We discuss these papers further below.

We estimate all the unknown parameters of this semiparametric asset pricing model, and conduct statistical tests of hypotheses regarding the functional form of the unknown habit as well as statistical tests for whether an internal habit vs external habit specification better describes the data. We then compare the estimated habit model's ability to fit a cross-section of equity returns with that of other asset pricing models, both quantitatively and in formal statistical terms.

Estimation and testing are conducted by applying a Sieve Minimum Distance (SMD) procedure to a set of Euler equations corresponding to the habit-based framework we study. The SMD method is a distribution-free minimum distance method, 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 habit 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 stochastic discount factor, including the sieve parameters of the habit function and the finite-dimensional parameters that are part of the power function and time-preference, may then be estimated by minimizing a weighted quadratic distance from zero of the nonparametrically estimated conditional moments.

Using stationary quarterly data on consumption growth, assets returns and instruments, we apply the SMD procedure to directly estimate all the unknown parameters of interest in the Euler equations underlying the optimal consumption choice of an investor with access to N asset payoffs. The SMD procedure is especially appealing for this application because it can be implemented as an instance of Generalized Method of Moments (GMM, Hansen (1982)), an approach that is familiar from prior work in estimating fully parametric, consumption-based asset pricing models (e.g., Hansen and Singleton (1982)). In addition to being robust to misspecification of the functional form of the habit and the law of motion for the underlying fundamentals, the SMD procedure estimates the unknown habit function consistently at some nonparametric rate. The procedure also provides estimates of the finite dimensional parameters, such as the curvature of the power utility function and the subjective time-discount factor; these estimates converge at rate $\sqrt{T}$ (where T is the sample size) and are asymptotically normally distributed. See the Appendix for details.

Using this methodology, we empirically investigate a number of hypotheses about the specification of habit-based asset pricing models that have not been previously investigated. One hypothesis concerns whether the habit is better described as a linear or nonlinear function. We develop a statistical test of the hypothesis of linearity and find that the functional form of the habit is better described as nonlinear rather than linear.

A second hypothesis concerns the distinction between "internal" and "external" habit formation. About half of the theoretical papers cited in footnote 1 investigate models of internal habit formation, in which the habit is a function of the agent's own past consumption. The rest investi-

gate models of external habit formation, in which the habit depends on the consumption of some exterior reference group, typically per capita aggregate consumption. Abel (1990) calls external habit formation "catching up with the Joneses." Determining which form of habit formation is more empirically plausible is important because the two specifications can have dramatically different implications for optimal tax policy and welfare analysis (Ljungqvist and Uhlig (2000)), and for whether habit models can explain long-standing asset-allocation puzzles in the international finance literature (Shore and White (2002)). To address this issue, we derive a conditional moment restriction that nests the internal and external nonlinear habit function, under the assumption that both functions are specified over current and lagged consumption with the same finite lag length. Our empirical results indicate that the data are better described by internal habit formation than external habit formation.

The SMD approach also allows us to assess the quantitative importance of the habit in the power utility specification. Our empirical results suggest that the habit is a substantial fraction of current consumption–about 97 percent on average–echoing the specification of Campbell and Cochrane (1999) in which the steady-state habit-consumption ratio exceeds 94 percent. The SMD estimated habit function is concave and generates positive intertemporal marginal rate of substitution in consumption. The SMD estimated subjective time-discount factor is around 0.99 and the estimated power utility curvature parameter is about 0.80 for three different combinations of instruments and asset returns.

Finally, we undertake a statistical model comparison analysis. Because our habit-based asset pricing model makes some parametric assumptions that may not be fully accurate (e.g., it maintains the power utility specification), and because the SMD-estimated nonparametric habit function contains lagged consumption of only finite lag length, the implied Stochastic Discount Factor (SDF) should be best viewed as a proxy to the true unknown SDF. Thus, we evaluate the SMD-estimated habit model and several competing asset pricing models by employing the model comparison distance metrics recommended in Hansen and Jagannathan (1997) (the so-called HJ distance and the HJ^+ distance), where all the models are treated as SDF proxies to the unknown truth. In particular, the SMD-estimated internal habit model is compared to (i) the SMD-estimated external habit model, (ii) the three-factor asset pricing model of Fama and French (1993), (iii) the "scaled" consumption Capital Asset Pricing Model (CAPM) of Lettau and Ludvigson (2001b), (iv) the classic CAPM of Sharpe (1964) and Lintner (1965), and (v) the classic consumption CAPM of Breeden (1979) and Breeden and Litzenberger (1978). Doing so, we find that a SMD-estimated internal habit model can better explain a cross-section of size and book-market sorted equity returns, both economically and in a statistically significant way, than the other five competing models. These results are particularly encouraging for the internal habit specification, since the Fama and French (1993) three-factor model and the Lettau and Ludvigson (2001b) scaled consumption CAPM have

previously displayed relative success in explaining the cross-section of stock market portfolio returns.

The rest of this paper is organized as follows. In the next section we discuss related literature not discussed above. Section 3 presents the habit-based asset pricing framework to be estimated and tested. In Section 4 we explain the estimation technique and its implementation. Section 5 describes the data. Section 6 presents the results of estimation and hypothesis testing about linearity of the habit function. Section 7 presents the results of specification tests for internal versus external habit formation, and conducts statistical model comparison when all the competing models are treated as potentially misspecified. Section 8 concludes. In the Appendix we discuss the identification of model parameters, and the large sample properties of the SMD estimators, and the related test statistics.

2 Related Literature

Our work builds off and extends a preexisting empirical literature investigating the possible role of non-separabilities in preferences. Habit- and durability-based asset pricing models, where the habit is restricted to have a linear functional form, have been empirically investigated by Ferson and Constantinides (1991) and Heaton (1995). We discuss these papers further below. Dynan (2000) estimates a linear model of habit formation using household level data on food consumption. We add to this literature by relaxing the functional form of the habit and by testing internal vs external habit formation.

Our work also relates to a previous asset pricing literature that approximates the entire SDF nonparametrically. Gallant and Tauchen (1989) were among the first to nonparametrically estimate a consumption-based asset pricing model using the so-called SNP (semi-nonparametric) method. Their SNP procedure is a constrained sieve maximum likelihood method, where the conditional density function is approximated by a truncated Hermite polynomial sieve, and where all the unknown parameters, including the Hermite polynomial sieve coefficients and the parameters of the stochastic discount factor, are estimated by maximizing the approximate likelihood subject to the conditional moment restriction constraints implied by an asset pricing theory. Gallant, Hansen, and Tauchen (1990) employ the same SNP method to estimate the conditional density of a vector of monthly asset payoffs. Subsequently, Bansal and Viswanathan (1993) and Bansal, Hsieh, and Viswanathan (1993) apply semi-nonparametric techniques to estimate nonlinear arbitrage-pricing models, while Chapman (1997) approximates an asset pricing kernel using orthonormal polynomials in state variables implied by a real business cycle model.

Our asset pricing model is closest to that of Gallant and Tauchen (1989), but our work still differs from theirs and previous work in several respects. First, we introduce the SMD procedure as

a particularly useful approach in asset pricing studies of larger cross-sections of returns. The SNP procedure of Gallant and Tauchen (1989) requires nonparametric estimation by constrained sieve maximum likelihood of the entire joint conditional density of the consumption and return series, which involves nonlinear numerical optimization of a likelihood function whose dimensionality grows with the number of return series. For the application studied here, the SMD procedure is much easier to implement, for two reasons: (i) it does not require nonparametric estimation of the entire likelihood function; this is important because when the number of return series is large, the dimensionality of the joint density to be estimated becomes prohibitive. (ii) the SMD procedure works directly on the Euler equations implied by asset pricing theory and delivers closed-form expressions for the estimated conditional moments (see Section 4).

Second, while both the SMD and the SNP procedures have the advantage of avoiding any parametric specification of the law of motion of consumption growth, there is currently no large sample theory on asymptotic properties such as convergence rates and limiting distributions of the SNP estimator that could be applied to undertake statistical inference. Here we extend existing large sample theories on the SMD estimator for I.I.D. data (see e.g., Newey and Powell (2003), Ai and Chen (2003), Ai and Chen (2005)), to our asset pricing setting with stationary but not necessarily serially uncorrelated data.

Third, we investigate a number of interesting hypotheses specific to the habit-based theoretical framework (such as internal vs external habit formation) that have not been investigated elsewhere. To do so, we place more structure on the empirical asset pricing model than in previous work by embedding the unknown habit function in the familiar power-utility framework. By contrast, Gallant and Tauchen (1989) treat the entire period-by-period utility function as unknown and approximate it using a polynomial sieve.

Finally, we study the cross-sectional asset pricing properties of our estimated habit models using a large number (e.g., $N = 7, 17$ or 26) of return series (including the especially challenging book-market sorted portfolios of Fama and French 1992, 1993), and undertake an extensive statistical model comparison analysis. Gallant and Tauchen (1989) instead focus on the time series asset pricing properties of their SNP estimated model using a small number ($N = 2$) of asset return series, and do not undertake any model comparison.

The results in this paper may also be related to recent evidence on long-run consumption correlations with asset returns. For example, Parker and Julliard (2004) find that the covariance between stock returns and consumption growth over a horizon of three years explains a much larger fraction of the cross-sectional variation in size and book-market portfolio returns than does the covariance between stock returns and contemporaneous, one-period, consumption growth. Internal habit formation provides one possible explanation for these findings, since in this case agents take into account the affect of today's consumption on all future habit levels. This interpretation of

the data is supported by our empirical results. We find that the internal habit SDF performs as well or better than other popular models in the literature and better than a purely external habit SDF that omits forward-looking terms governing the affect of today's consumption on future habit levels. The findings in this paper may therefore provide a structural interpretation of the results in Parker and Julliard (2004).[2] We discuss this further below.

3 The Model

In this section we present a model of investor behavior in which utility is a power function of the difference between aggregate consumption and the habit. We do not consider models in which utility is a power function of the *ratio* of consumption to the habit stock, as in Abel (1990) and Abel (1999). Ratio models of external habit formation imply that relative risk-aversion is constant, hence they have difficulty accounting for the predictability of excess stock returns documented in the empirical asset pricing literature.[3] By contrast, difference models can generate time-variation in the equilibrium risk-premium because relative risk aversion varies countercyclically. Difference models are also far more common in the asset pricing literature; for example, the difference specification is used in all the habit-based asset pricing models referenced in footnote 1 above.

Throughout this paper we assume that identical agents maximize the utility function

$$U = E\sum_{t=0}^{\infty}\delta^t\frac{(C_t - X_t)^{1-\gamma} - 1}{1-\gamma}. \tag{1}$$

Here X_t is the level of the habit, and δ is the subjective time discount factor. X_t is assumed to be a function (known to the agent but unknown to the econometrician) of current and past consumption

$$X_t = f\left(C_t, C_{t-1}, ..., C_{t-L}\right),$$

such that $X_t < C_t, X_t \geq 0$. Note that we allow the habit to potentially depend on contemporaneous as well as past consumption, a modeling choice that is a feature of several habit models in the recent theoretical literature (e.g., Campbell and Cochrane (1999)).[4]

When the habit is internal, the agent takes into account the impact of today's consumption decisions on future habit levels. In this case the intertemporal marginal rate of substitution in

[2] Hansen, Heaton, and Li (2005) provide an alternative interpretation based on recursive preferences and long-run consumption risk.

[3] A large literature finds that excess stock returns are forecastable; see Shiller (1981), Fama and French (1988), Campbell and Shiller (1989), Campbell (1991), Harvey (1991) Hodrick (1992), Lamont (1998) Lettau and Ludvigson (2001a).

[4] In the conclusion we discuss a possible alternative specification (a significant extension of the empirical approach developed here), in which the habit is a function of an infinite number of lags and specified as a recursive functional of unknown form e.g., $X_t = r\left(C_t, C_{t-1}, X_{t-1}\right)$.

consumption is given by

$$M_{t+1} = \delta \frac{MU_{t+1}}{MU_t}, \tag{2}$$

where

$$MU_t = \frac{\partial U}{\partial C_t} = (C_t - X_t)^{-\gamma} - E_t \left[\sum_{j=0}^{L} \delta^j \left(C_{t+j} - X_{t+j} \right)^{-\gamma} \frac{\partial X_{t+j}}{\partial C_t} \right], \tag{3}$$

and where E_t is the expectation operator conditional on information available at time t. When the habit is external, agents maximize (1) but ignore the impact of today's consumption on tomorrow's habits, since the habit in this specification merely plays the role of an externality. In this case, only the first term on the right-hand-side of (3), $(C_t - X_t)^{-\gamma}$, is part of marginal utility. In equilibrium, identical individuals choose the same consumption, so that regardless of whether the habit is external or internal, individual consumption, C_t, is equal to aggregate consumption, C_t^a, which we denote C_t from now on.

Although we have motivated the time-nonseparable model above with the notion of habit-formation, the framework itself is suitable for studying more general forms of nonseparabilities. For example, if an increase in past consumption raises X_t, then X_t is interpreted as a habit because current consumption must now be higher to deliver the same utility. By contrast, if an increase in past consumption decreases X_t, then X_t is consistent with durability, since increases in the stock of durable goods add to current utility. In our empirical estimation we do not restrict the unknown X_t to impose either specification. Instead, we follow Gallant and Tauchen (1989), Ferson and Constantinides (1991) and Heaton (1995) and freely estimate the function, letting the data dictate the appropriate framework as pertains to habit-formation or durability. The results (discussed below) are that X_t increases with past consumption (at all lags), favoring the habit interpretation over the durability interpretation.

The asset pricing model comes from the first-order conditions for optimal consumption choice for an investor with access to N asset returns. The first-order conditions place restrictions on the joint distribution of the intertemporal marginal rate of substitution in consumption and asset returns and imply that, for any traded asset indexed by i, with a gross return at time $t+1$ of $R_{i,t+1}$, the following equation holds:

$$E_t \left[M_{t+1} R_{i,t+1} \right] = 1, \qquad i = 1, ..., N. \tag{4}$$

Equation (4) shows that the intertemporal marginal rate of substitution in consumption, M_t, is the stochastic discount factor (SDF), which in this setting depends on the unknown habit function. The resulting N equations yield a set of conditional moment restrictions containing a vector of unknown parameters, $(\delta, \gamma)'$, and a single unknown habit function $X_t = f\left(C_t, C_{t-1}, ..., C_{t-L}\right)$.

4 Empirical Implementation

This section presents the details of our estimation procedure. The model to be estimated in (4) is semiparametric in the sense that it contains both finite dimensional and infinite dimensional unknown parameters. Estimation in this setting may be undertaken using the SMD procedure first developed in Newey and Powell (2003) and Ai and Chen (2003) for independent and identically distributed (I.I.D.) data. In the Appendix we show that the SMD procedure and its large sample properties can be extended to stationary ergodic time series data.

4.1 Transformation of the model

Since consumption is trending over time, it is necessary to transform the model to use stationary observations on consumption growth. We address this problem by assuming that the unknown habit function $X_t = f(C_t, C_{t-1}, ..., C_{t-L})$ can be written as

$$X_t = C_t g\left(\frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t}\right), \tag{5}$$

where $g : \mathcal{R}^L \to \mathcal{R}$ is an unknown function of the gross growth rates of consumption, with domain space reduced by one dimension relative to f. Note that g now replaces f as the unknown function to be estimated along with (δ, γ) using the Euler equations (4) and the SMD procedure. As shown below, this assumption allows us to express the stochastic discount factor, M_{t+1}, as a function of gross growth rates in consumption, which are plausibly stationary. One way to motivate (5) is to presume that the original function $X_t = f(C_t, C_{t-1}, ..., C_{t-L})$ is homogeneous of degree one, which allows the function to be re-written as

$$X_t = C_t f\left(1, \frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t}\right), \tag{6}$$

and redefined as in (5). The homogeneous of degree one assumption is consistent with the habit models studied in the asset pricing literature cited above, including the nonlinear habit specification investigated in Campbell and Cochrane (1999).

When the habit stock is a homogeneous of degree one function of current and past consumption, marginal utility, MU_t, takes the form

$$\begin{aligned} MU_t &= C_t^{-\gamma}\left(1 - g\left(\frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t}\right)\right)^{-\gamma} \\ &\quad -C_t^{-\gamma} E_t\left[\sum_{j=0}^{L} \delta^j \left(\frac{C_{t+j}}{C_t}\right)^{-\gamma}\left(1 - g\left(\frac{C_{t+j-1}}{C_{t+j}}, ..., \frac{C_{t+j-L}}{C_{t+j}}\right)\right)^{-\gamma} \frac{\partial X_{t+j}}{\partial C_t}\right], \end{aligned} \tag{7}$$

where,

$$\frac{\partial X_{t+j}}{\partial C_t} = \begin{cases} g_j\left(\frac{C_{t+j-1}}{C_{t+j}}, ..., \frac{C_{t+j-L}}{C_{t+j}}\right) & \forall j \neq 0 \\ g\left(\frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t}\right) - \sum_{i=1}^{L} g_i\left(\frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t}\right)\frac{C_{t-i}}{C_t} & j = 0 \end{cases} \tag{8}$$

In the expression directly above, g_i denotes the derivative of g with respect to its $i-$th argument.

To obtain an estimable expression for the unknown parameters of interest $\boldsymbol{\alpha} = (\delta, \gamma, g)'$, the Euler equations (4) must be rearranged so that the conditional expectation $E_t(\cdot)$ appears only on the outside of the conditional moment restrictions. The Appendix presents several equivalent expressions of this form; here we present one. Denote the true values of the parameters with an "o" subscript: $\boldsymbol{\alpha}_o = (\delta_o, \gamma_o, g_o)'$. Combining (2), (7) and (4), and rearranging terms, we find a set of N conditional moment conditions:

$$E_t \left\{ \left(\delta_o \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma_o} F_{t+1} R_{i,t+1} - 1 \right) \Phi_{t+1} \right\} = 0, \quad i = 1, ..., N, \tag{9}$$

where

$$F_{t+1} \equiv \left(\begin{array}{c} \left(1 - g_o \left(\frac{C_t}{C_{t+1}}, ..., \frac{C_{t+1-L}}{C_{t+1}} \right) \right)^{-\gamma_o} \\ - \left[\sum_{j=0}^{L} \delta_o^j \left(\frac{C_{t+1+j}}{C_{t+1}} \right)^{-\gamma_o} \left(1 - g_o \left(\frac{C_{t+j}}{C_{t+1+j}}, ..., \frac{C_{t+j+1-L}}{C_{t+1+j}} \right) \right)^{-\gamma_o} \frac{\partial X_{t+1+j}}{\partial C_{t+1}} \right] \end{array} \right) / \Phi_{t+1},$$

$$\Phi_{t+1} \equiv \left(\begin{array}{c} \left(1 - g_o \left(\frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t} \right) \right)^{-\gamma_o} \\ - \left[\sum_{j=0}^{L} \delta_o^j \left(\frac{C_{t+j}}{C_t} \right)^{-\gamma_o} \left(1 - g_o \left(\frac{C_{t+j-1}}{C_{t+j}}, ..., \frac{C_{t+j-L}}{C_{t+j}} \right) \right)^{-\gamma_o} \frac{\partial X_{t+j}}{\partial C_t} \right] \end{array} \right).$$

We may write (9) more compactly as

$$E \left\{ \rho_i(\mathbf{z}_{t+1}, \delta_o, \gamma_o, g_o) | \mathbf{w}_t^* \right\} = 0, \qquad i = 1, ..., N, \tag{10}$$

where $\mathbf{z}_{t+1}$ is a vector containing all observations used to estimate the conditional moment (9) at time t, ρ_i is defined as

$$\rho_i(\mathbf{z}_{t+1}, \delta_o, \gamma_o, g_o) \equiv \left(\delta_o \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma_o} F_{t+1} R_{i,t+1} - 1 \right) \Phi_{t+1},$$

and the conditional expectation in (10) is taken with respect to agents' information set at time t, denoted $\mathbf{w}_t^*$.

Let $\mathbf{w}_t$ be a $d_w \times 1$ observable subset of $\mathbf{w}_t^*$ that does not contain a constant. Equation (10) implies

$$E \left\{ \rho_i(\mathbf{z}_{t+1}, \delta_o, \gamma_o, g_o) | \mathbf{w}_t \right\} = 0, \qquad i = 1, ..., N. \tag{11}$$

We assume that the econometricians' information set $\mathbf{w}_t$ is informative enough, such that *if* $E \left\{ \rho_i(\mathbf{z}_{t+1}, \delta, \gamma, g) | \mathbf{w}_t \right\} = 0$, $i = 1, ..., N$ *for any* $\delta > 0, \gamma > 0$, $0 \leq g < 1, g(0, ..., 0) = 0$, $g \neq$*constant, then* $(\delta, \gamma, g)' = (\delta_o, \gamma_o, g_o)'$. In the Appendix we discuss identification of $\boldsymbol{\alpha}_o = (\delta_o, \gamma_o, g_o)'$ based on the set of conditional moment restrictions (11) for the external habit model.

Given the theoretical restrictions implied by (11), the true parameter values must satisfy the minimum distance relation

$$\boldsymbol{\alpha}_o = (\delta_o, \gamma_o, g_o)' = \arg\min_{\boldsymbol{\alpha}} E \left[m(\mathbf{w}_t, \boldsymbol{\alpha})' m(\mathbf{w}_t, \boldsymbol{\alpha}) \right], \tag{12}$$

where $m(\mathbf{w}_t, \boldsymbol{\alpha}) = E\{\rho(\mathbf{z}_{t+1}, \boldsymbol{\alpha})|\mathbf{w}_t\}$, $\rho(\mathbf{z}_{t+1}, \boldsymbol{\alpha}) = (\rho_1(\mathbf{z}_{t+1}, \boldsymbol{\alpha}), ..., \rho_N(\mathbf{z}_{t+1}, \boldsymbol{\alpha}))'$ for any candidate value $\boldsymbol{\alpha} = (\delta, \gamma, g)'$.

4.2 SMD estimation

We are now in a position to describe the SMD estimation of $\boldsymbol{\alpha}_o = (\delta_o, \gamma_o, g_o)'$ using the conditional moment restrictions model (11). The idea behind the SMD procedure has three essential parts. First, although the functional form of the conditional mean $m(\mathbf{w}_t, \boldsymbol{\alpha}) = E\{\rho(\mathbf{z}_{t+1}, \boldsymbol{\alpha})|\mathbf{w}_t\}$ is unknown, we may replace it with a consistent nonparametric estimator $\widehat{m}(\mathbf{w}, \boldsymbol{\alpha})$ for any candidate parameter value $\boldsymbol{\alpha} = (\delta, \gamma, g)'$. Second, although the habit function g is an infinite-dimensional unknown parameter, we can approximate it by a sequence of finite-dimensional unknown parameters (sieves) denoted g_{K_T}, where the approximation error decreases as the dimension K_T increases with the sample size T. Third, the function g_{K_T} is estimated jointly with the finite-dimensional parameters $(\delta, \gamma)'$ by minimizing a (weighted) quadratic norm of estimated conditional expectation functions:

$$\widehat{\boldsymbol{\alpha}} = \arg \min_{\delta, \gamma, g_{K_T}} \frac{1}{T} \sum_{t=1}^{T} \widehat{m}(\mathbf{w}_t, \delta, \gamma, g_{K_T})' \widehat{m}(\mathbf{w}_t, \delta, \gamma, g_{K_T}). \tag{13}$$

Implementation of the SMD estimator requires a consistent estimate of the conditional mean $m(\mathbf{w}_t, \boldsymbol{\alpha})$. Let $\{p_{0j}(\mathbf{w}_t),\ j = 1, 2, ..., J_T\}$ be a sequence of known basis functions (including a constant function) that map from $\mathcal{R}^{d_w}$ into $\mathcal{R}$. Denote $p^{J_T}(\cdot) \equiv (p_{01}(\cdot), ..., p_{0J_T}(\cdot))'$ and the $T \times J_T$ matrix $\mathbf{P} \equiv \left(p^{J_T}(w_1), ..., p^{J_T}(w_T)\right)'$. Then

$$\widehat{m}(\mathbf{w}, \boldsymbol{\alpha}) = \left(\sum_{t=1}^{T} \rho(\mathbf{z}_{t+1}, \boldsymbol{\alpha}) p^{J_T}(\mathbf{w}_t)' (\mathbf{P}'\mathbf{P})^{-1} \right) p^{J_T}(\mathbf{w}) \tag{14}$$

is a sieve Least Squares estimator of the conditional mean vector $m(\mathbf{w}, \boldsymbol{\alpha}) = E\{\rho(\mathbf{z}_{t+1}, \boldsymbol{\alpha})|\mathbf{w}_t = \mathbf{w}\}$. (Note that J_T must grow with the sample size to ensure that $m(\mathbf{w}_t, \boldsymbol{\alpha})$ is estimated consistently).

We show in the Appendix that, under the assumption of stationary ergodic observations and some additional regularity conditions, the SMD estimate of g is consistent and converges at a rate faster than $T^{1/4}$, while the estimates of the finite-dimensional parameters $(\delta, \gamma)'$ are $\sqrt{T}$ consistent and asymptotically normally distributed; hence the standard errors for $\widehat{\delta}, \widehat{\gamma}$ can be computed. Unfortunately, a general asymptotic distribution theory for SMD estimators of unknown functions, such as the habit function g, has not been developed;[5] hence it is not possible to compute confidence intervals for habit function itself. Nevertheless, it is possible to conduct hypothesis tests about some aspects of the functional form of the habit, as discussed below.

[5] Some authors treat the sieve estimators of unknown functions as parametric ones and compute their standard errors by applying standard root-T asymptotic normality theory. Such practices ignore the uncertainty of the unknown functions and may in general lead to erroneous inference decisions.

We emphasize that the SMD method is likely to be particularly useful in asset pricing studies of larger cross-sections of returns. This is because the nonparametric estimate $\widehat{m}_i(\mathbf{w}, \boldsymbol{\alpha})$ of the conditional mean $m_i(\mathbf{w}, \boldsymbol{\alpha})$, $i = 1, ..., N$, has a closed-form expression (14). An attractive feature of this SMD estimator is that it is easy to implement as an instance of GMM with a particular weighting matrix. To see this, plug the sieve least squares estimator $\widehat{m}(\mathbf{w}, \boldsymbol{\alpha})$ into (13) to obtain:

$$\widehat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \left[\mathbf{g}(\boldsymbol{\alpha}; \mathbf{y}^T)\right]' \left\{\mathbf{I}_N \otimes \left(\mathbf{P}'\mathbf{P}\right)^{-1}\right\} \left[\mathbf{g}(\boldsymbol{\alpha}; \mathbf{y}^T)\right], \tag{15}$$

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

$$\mathbf{g}_T(\boldsymbol{\alpha}; \mathbf{y}^T) \equiv \frac{1}{T} \sum_{t=1}^{T} \rho(\mathbf{z}_{t+1}, \boldsymbol{\alpha}) p^{J_T}(\mathbf{w}_t) \tag{16}$$

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

$$E\left\{\rho_i(\mathbf{z}_{t+1}, \delta_o, \gamma_o, g_o) p_{0j}(\mathbf{w}_t)\right\} = 0, \qquad i = 1, ..., N, \quad j = 1, ..., J_T. \tag{17}$$

Thus the GMM criterion is

$$Q_T(\boldsymbol{\alpha}) = \left[\mathbf{g}_T(\boldsymbol{\alpha}; \mathbf{y}^T)\right]' \mathbf{W} \left[\mathbf{g}_T(\boldsymbol{\alpha}; \mathbf{y}^T)\right], \tag{18}$$

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

It is important to note that the SMD estimation (13) collapses to a case of GMM only when the specific weighting matrix $\mathbf{W} = \mathbf{I}_N \otimes (\mathbf{P}'\mathbf{P})^{-1}$ is employed. This procedure is equivalent to regressing each ρ_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 weighting matrix in (18) gives greater weight to moments that are more highly correlated with the instruments $p^{J_T}(\cdot)$. In this specific case, the GMM implementation preserves the desirable properties of the SMD estimator, which include parameter consistency, reasonable rates of convergence, and asymptotic normality of the finite dimensional parameters. We cannot, however, presume that these properties would be preserved in GMM estimation using an arbitrary weighting matrix $\mathbf{W}$. The weighting scheme can be understood intuitively by noting that variation in the conditional mean is what identifies the unknown function.

4.3 Implementation

Before we can estimate the model, several implementation issues must be addressed. First, it is clear that identification of the unknown habit function using (4) requires ruling out the case $X_t = KC_t$, where $K \in [0, 1)$ is a constant. This is not a problem since any model of habit formation naturally depends on past consumption. Second, estimation of $X_t = f(C_t, C_{t-1}, ..., C_{t-L})$ nonparametrically

is not practical if L is too large, generating a "curse of dimensionality."[6] Thus, we have to assume L is a reasonable finite number relative to the sample size. Such lag limitations are less restrictive than they might at first appear, since standard theoretical treatments of habit formation imply that more recent values of consumption have the greatest influence on the habit stock. Thus, the estimation procedure we propose may still do a good job of characterizing how the habit changes with consumption, by estimating the habit stock as a function the current and most recent lags of consumption.

Third, we need a sequence of sieves g_{K_T} to approximate the unknown function $g\left(\frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t}\right)$. We do so using an artificial neural network (ANN) sieve approximation, defined as

$$g_{K_T}\left(\frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t}\right) = \alpha_0 + \sum_{j=1}^{K_T} \alpha_j \psi\left(\gamma_{j,1}\frac{C_{t-1}}{C_t} + \cdots + \gamma_{j,L}\frac{C_{t-L}}{C_t} + \beta_j\right), \tag{19}$$

where $\psi(\cdot)$ an activation function, which can be any known function except a polynomial function of fixed finite degree (see Hornik, Stinchcombe, and White (1989)). A common choice for ψ is the logistic function, $\psi(x) = (1 + e^{-x})^{-1}$, a specification we use here. To provide a nonparametric estimate of the true unknown function, $g_o(\cdot)$, where "$(\cdot)$" denotes its generic argument, it is necessary to require K_T to grow with the sample size to ensure consistency of the method.[7] We use the ANN sieve for several reasons: (i) it allows the use of more lags in the unknown function g by delivering a relatively fast convergence rate when compared to other sieves (e.g., linear tensor product Fourier sieves or spline sieves) (Chen and White (1999)); (ii) the use of a nonlinear sieve such as (19) is often in practice better able than alternatives to allow for nonlinearities in the unknown function and nonseparabilities between elements of g; (iii) as long as consumption is strictly positive, the ANN sieve makes it possible to restrict coefficients in (19) so that the habit $X_t < C_t$, *for all possible shocks to consumption*, not just those observed in our sample. Imposing this restriction is straightforward because the logistic function $\psi(x) = (1 + e^{-x})^{-1}$ lies between zero and one, regardless of the values taken by its arguments. This insures that utility is always well defined, and avoids the danger that the model will break down out-of-sample. In previously specified habit-based asset pricing models, imposing this restriction is more difficult even when the habit is specified parametrically (for example, as a linear or polynomial function of past consumption.)

Forth, as with any nonlinear estimation procedure, it is necessary to require the parameter space

[6] A curse of dimensionality in this context refers to the situation in which, fixing the smoothness of the function f to be estimated, the rate of convergence of a nonparametric estimate approaches zero as the dimension of the domain of the function f approaches infinity.

[7] Hornik, Stinchcombe, and White (1989) show that any continuous function can be arbitrarily well approximated asymptotically using a neural network sieve. Chen and White (1999) provide convergence rates for a large class of single hidden layer feedforward artificial neural networks. Bansal and Viswanathan (1993) use a neural network to approximate the SDF of an nonlinear arbitrage pricing model. See Campbell, Lo, and MacKinlay (1997) for additional neural network sieve applications in finance.

to lie in a compact set. In practice, researchers use prior information to restrict the parameter space. Restriction of the parameter space is particularly important for our application, since sieve parameters which generate values for the ANN sieve logistic activation function $\psi(x) = (1 + e^{-x})^{-1}$ that lie in the tails of the function imply that the habit g is constant. Thus, we restrict the sieve parameters to a range that does not generate tail observations on $\psi(\cdot)$. We also restrict the rate of time-preference, $\delta \in (0, 1.2]$, and the curvature parameter $\gamma \in [.1, 100]$.

A final implementation issue concerns the sampling interval of our data relative to the decision interval of households. If consumption decisions occur more frequently than the data sampling interval, aggregate consumption data are time-aggregated. Heaton (1993) studies the interaction of time-aggregation and time-nonseparable preferences and concludes that it can influence the evidence in favor of habit formation, at least when habits are of the linear variety. Unfortunately, as Ferson and Constantinides (1991) and Heaton (1995) point out, it is not possible to model time-aggregation in a fully nonlinear framework using minimum distance estimation, which our procedure requires. To the extent that time-aggregation is a concern, this must be considered a limitation of our approach. Nevertheless, there are at least two reasons to think that time-aggregation may not unduly affect inference. First, Ferson and Constantinides (1991) note that estimates of the nonseparability parameter in Heaton (1993)—which uses a first-order linear approximation of the Euler equation but allows for time-aggregation—are similar to their own estimates generated from nonlinear GMM in which no time-aggregation is modeled. Second, Ferson and Constantinides also note that the effect of time-aggregation is to increase the order of the moving average process followed by the GMM error, which is straightforward to address using a nonparametric correction to the standard errors for serial correlation. Of course, the influence of time-aggregation may be more complex for nonlinear specifications and moreover with nonlinearities it is not clear in which direction time aggregation would influence the evidence on habit formation. We follow Ferson and Constantinides (1991) and at least partly account for these effects when computing the asymptotic standard errors for δ and γ, by using a higher order nonparametric correction for serial correlation in $\rho_i(\mathbf{z}_{t+1}, \delta_o, \gamma_o, g_o)$.

5 The Data

A detailed description of the data and our sources is provided in the Appendix. Our data are quarterly, and span the period from the fourth quarter of 1952 to the fourth quarter of 2001.

The focus of this paper is on testing and modeling cross-sections of asset returns, rather than one or two aggregate asset returns. As explained in the Appendix, exploiting the cross-section aids the empirical identification of the unknown habit function. We study three cross-sections of 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. The first group (Group 1) contains the three-month Treasury bill rate, 10 industry portfolios of common stocks based on 4-digit SIC codes, and six value-weighted portfolios of common stock sorted into two size (market equity) quantiles and three book value-market value quantiles. Thus Group 1 consists of 17 asset returns in total. The portfolios are created from all stocks traded on the NYSE, AMEX, and NASDAQ, as detailed on Kenneth French's web page. The second group of asset returns (Group 2), is comprised of the three-month Treasury bill rate and the six value-weighted portfolios of common stock sorted into two size quantiles and three book value-market value quantiles, for a total of 7 asset returns. The third group (Group 3) is comprised of the three-month Treasury bill rate plus 25 value-weighted returns for the intersections of 5 market equity quantiles and 5 book equity-market equity quantiles, or 26 asset returns in total (Group 3).

Our measure of consumption is real, per-capita expenditures on nondurables and services. 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 (1996=100) for our measure of consumption. We use quarterly consumption data because it is known to contain less measurement error than monthly consumption data.

The procedure requires computation of instruments, $p^{J_T}(\mathbf{w}_t)$, which are known basis functions (including a constant function) of conditioning variables, $\mathbf{w}_t$. We include lagged consumption growth in $\mathbf{w}_t$, as well as three variables that have been shown elsewhere to have significant forecasting power for excess stock returns and consumption growth in quarterly data.[8] Two variables that have been found to display forecasting power for excess stock returns at a quarterly frequency are the "relative T-bill rate" (which we measure as the three month Treasury-bill rate minus its 4-quarter moving average), and the lagged value of the excess return on the Standard & Poor 500 stock market index (S&P 500) over the three-month Treasury bill rate (see Campbell (1991), Hodrick (1992), Lettau and Ludvigson (2001a)). We denote the relative bill rate $RREL$ and the excess return on the S&P 500 index, $SPEX$.[9] We also use the proxy for the log consumption-wealth ratio studied in (Lettau and Ludvigson (2001a)) to forecast returns.[10] This proxy is measured as the cointegrating

[8] 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 we choose variables for $\mathbf{w}_t$ that are known to be strong predictors of asset returns and consumption growth in quarterly data.

[9] 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)).

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

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

Since the error term $\rho_i(\mathbf{z}_{t+1}, \delta_o, \gamma_o, g_o)$ 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. We use power series as instruments, investigating three different specifications. Each specification includes a constant (vector of ones). The first specification includes a constant, the linear terms plus the squared terms of each variable in $\mathbf{w}_t$, creating nine instruments; we use these basis functions when studying the asset returns in Group 1. The second set of instruments includes a constant, the linear terms, squared terms and pair-wise cross products of each variable in $\mathbf{w}_t$, or 15 instruments in total; we use these when studying the asset returns in Group 2. The third set of instruments utilizes just a constant and the linear terms of each variable, or five instruments in total; we use these when studying the asset returns in Group 3. This insures that the total number of moment conditions is about the same across estimations.[13]

6 Empirical Estimates

All the tables and figures are reported at the end of the Appendix. Table 1 and Figures 1-8 present the estimation results of the semiparametric habit model presented above, using the instruments and the three groups of assets described in the previous section.[14] The results reported below were very similar with $L = 4$ and $L = 3$. Thus, we opt for the more parsimonious specification, and in

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

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

[13] The number of total moment conditions is not uniquely determined by the estimation theory. The theory merely requires that there be more moment conditions than parameters to be estimated, $NJ_T \geq \dim(\boldsymbol{\alpha})$, and that the number of moments, NJ_T, increase with the sample size T, but at a slower rate than the sample size, so that $NJ_T/T \to 0$ and $NJ_T \to \infty$ as $T \to \infty$.

[14] The results reported here are based on $\mathbf{w}_t' = \left[\widehat{cay}_t, RREL_t, SPEX_t, \frac{C_t}{C_{t-1}}\right]$. Nevertheless, all the empirical results remain virtually unchanged when we use $[\widehat{cay}_t, RREL_t, SPEX_t]$ as $\mathbf{w}_t$. We have also tried $\mathbf{w}_t$, lagged $\mathbf{w}_t$ and lagged returns as instruments, again the empirical results change little with the additional lagged $\mathbf{w}_t$ and lagged returns.

all cases reported below set $L = 3$.[15] We emphasize that our use of three lags is a generalization of what has been done previously in the estimation of time-nonseparable asset pricing models, most of which have focused on specifications with $L = 1$ (e.g., Ferson and Constantinides (1991), Chapman (1997)) and/or $L = 2$ (e.g., Gallant and Tauchen (1989)).[16]

For the dimensionality of the ANN sieve, $g(x_1, ..., x_L) \approx \alpha_0 + \sum_{j=1}^{K_T} \alpha_j \psi(\Sigma_{l=1}^{L} \gamma_{j,l} x_l + \beta_j)$, we set $K_T = 3$. Because asymptotic theory only provides guidance about the *rate* at which $(L+2)K_T$ must increase with the sample size T, other considerations must be used to judge how best to set this dimensionality. The bigger is $(L + 2)K_T$, 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 $K_T = 3$, the dimension of the parameter vector, $\boldsymbol{\alpha} = (\delta, \gamma, g)'$, is 18, estimated using a sample of size $T = 200$. In practice, we obtained very similar results setting $K_T = 4$; thus we present the results for the more parsimonious specification using $K_T = 3$ below.

Table 1 displays estimates of the finite dimensional parameters, for three estimations. The the estimates of δ and γ are very similar across these three estimations. In each case, the subjective rate of time-preference is about 0.99, and the curvature parameter is between $\gamma = 0.76$ and $\gamma = 0.81$. The standard errors indicate that these variables are estimated precisely, but the estimates for γ are effectively estimates of unity, since the minimized value of the SMD criterion is very similar when γ is restricted to one. Boldrin, Christiano, and Fisher (2001) find that a business cycle model with linear habit formation and $\gamma = 1$ performs well in matching the mean equity premium and Sharpe ratio.

To get a sense of how important the habit is in the power utility specification, the top and bottom panels of Figure 1 plot the habit-consumption ratio and an estimate of the stochastic discount factor, respectively, over time, for the estimation on Group 1 assets, where the instruments are a constant, the linear and squared values of the elements of $\mathbf{w}_t$. The corresponding figures for the estimations on Group 2 and Group 3 assets are very similar and are therefore omitted to conserve space. The figure demonstrates significant evidence in favor of habit formation, conditional on the power utility framework. The habit is about 97 percent of current consumption on average, reminiscent

[15] Notice that, because of the forward-looking terms in (11), $\rho_i(\mathbf{z}_t, \boldsymbol{\alpha}_o)$ is not in the information set at time t. In particular, $E\left[\rho_i(\mathbf{z}_{t+s+1}, \boldsymbol{\alpha}_o)|\rho_i(\mathbf{z}_t, \boldsymbol{\alpha}_o)\right] \neq 0$ for $s \leq L$. Taken literally, therefore, the choice of $L = 3$ implies that the error term, $\{\rho_i(\mathbf{z}_{t+1}, \boldsymbol{\alpha}_o)\}$, is correlated with its three-period lagged value. If this were known to be the true lag structure of the error term, efficiency gains could be made by imposing this structure in estimation. The choice of lag length here, however, is largely dictated by our sample size. With no clear theoretical guidance on the value for L, possible efficiency gains are likely to be outweighed by the greater robustness afforded by foregoing efficient estimation in favor of consistent estimation. For this reason, we do not impose a third-order moving average (or other) structure on the error term, and instead simply apply a nonparameteric adjustment for higher-order serial correlation to the asymptotic standard errors of the finite dimensional parameters.

[16] Using stationary monthly data on consumption growth and two return series, Gallant and Tauchen (1989) experimented with 1 and 2 lags and found the appropriate lag length is 1 for their subutility approximation.

of the Campbell and Cochrane (1999) model, in which the steady state habit-consumption ratio is in excess of 94%. Since the procedure is free to estimate a zero habit, this evidence implies that habit formation significantly improves the model's ability to fit the data, and rejects the notion that preferences are well described as time-separable in the power utility framework. The habit-consumption ratio is stable, ranging only from 0.97 to 0.974.

The estimate of M_t is always positive, thus it is an arbitrage-free stochastic discount factor suitable for pricing derivative claims, as discussed in Hansen and Jagannathan (1997) and Wang and Zhang (2003).[17] The relative stability of the habit-consumption ratio translates into a relatively stable stochastic discount factor: the mean is slightly less than one (0.98), while the standard deviation is about 0.02. Although the standard deviation is not large, it is nonetheless significantly larger than the standard deviation of quarterly consumption growth, equal to 0.0045 in this sample. Still, it is evident that these specifications do not fit the unconditional volatility bounds for the stochastic discount factor implied by the work of Hansen and Jagannathan (1991) when matched to post-war data on aggregate stock returns. These bounds determine whether the model can match the mean equity premium and Sharpe ratio. This finding is not surprising, since the methodology used here must place very high weight on conditional moments (and therefore relatively little weight on unconditional moments) in order to nonparametrically estimate the unknown habit function with accuracy.[18]

We can check whether our estimates of the habit imply that the partial first derivatives, $\frac{\partial X_{t+i}}{\partial C_t}$, $i = 1, 2, 3$ are greater than zero, and decreasing in i. Such a structure is typical of linear habit models specified as a declining polynomial lag of past consumption. Of course, with a nonlinear habit, these partial derivatives are not constant, but vary over time given our estimated X function. Figure 2 plots the derivatives over time as a function of lagged consumption growth, again for the model estimated on Group 1 assets with a constant, the linear and squared terms of $\mathbf{w}_t$ as instruments. In each case, the partial derivative is positive everywhere; moreover, the partial derivative of the habit one-step ahead is everywhere greater than that two-steps, which in turn is everywhere greater than that three steps ahead. The estimated habit depends positively on lagged consumption, but this positive dependence decreases as the consumption becomes more distant.

The shape of our estimated habit function can be illustrated by plotting X_t as a function of lagged consumption, C_{t-1}, holding fixed current consumption, C_t, and the other lags of consumption, $C_{t-2}, ..., C_{t-L}$. Figures 3 through 5 plot this relation for each estimation described above. For these plots, one-period lagged consumption is allowed to vary, but C_t, $C_{t-2}, ..., C_{t-L}$ are alternately

[17] A related point is made by Chapman (1998) who notes that requiring MU_t itself to be positive places restrictions on habit models. The fitted marginal utility values are never negative in the sample of data used here.

[18] The finding is also similar to those in Ferson and Constantinides (1991), who concluded that habit persistence improves the fit of the standard consumption-based model largely through its influence on moments other than the mean equity premium and Sharpe ratio.

held fixed at their median, 25th, and 75th percentile values in our sample.

We can draw several conclusions from Figures 3-5. First, the estimated habit looks nonlinear; this is evident from the curved shape of the functions and from the finding that the shape depends on where in the domain space the function is evaluated. Second, the estimated habit is always increasing in past consumption. This suggests that the time-nonseparability in utility that we estimate is better interpreted as habit formation than durability in consumption, consistent with previous studies that use consumption growth measured over longer horizons (Gallant and Tauchen (1989), Ferson and Constantinides (1991), Heaton (1995)). Third, the estimated habit is increasing at a decreasing rate in past consumption. The next section presents a formal test of the linearity of the habit.

6.1 Is Habit Function Linear?

Figures 3 to 5 indicate that the estimated habit function looks nonlinear. In this section, we test for nonlinearity more formally. Whether habits are linear is an interesting empirical question because the functional form of the habit is often crucial in determining the asset pricing implications of the model. To address this question, we make use of the partial derivatives of our estimated habit function. Notice that linearity implies $\frac{\partial X_{t+i}}{\partial C_t}$ is constant for all $i \geq 0$, and in particular

$$\frac{\partial^2 X_{t+i}}{\partial C_t^2} = C_{t+i} \frac{\partial^2}{\partial C_t^2} g\left(\frac{C_{t+i-1}}{C_{t+i}}, ..., \frac{C_{t+i-L}}{C_{t+i}}\right) \equiv 0 \qquad \forall i \geq 1. \tag{20}$$

Figures 7-9 plot the values of (20) based on our estimates for the cases above for $i = 1$. Notice from (8) that $\partial^2 X_{t+i}/\partial C_t^2$ will take the form $(1/C_{t+i})\, g_{ii}\left(\frac{C_{t+i-1}}{C_{t+i}}, ..., \frac{C_{t+i-L}}{C_{t+i}}\right)$, where g_{ii} denotes the second partial derivative of g with respect to its $i-$th argument. Obviously $(1/C_{t+i})\, g_{ii}$ cannot be identically zero unless $g_{ii}\left(\frac{C_{t+i-1}}{C_{t+i}}, ..., \frac{C_{t+i-L}}{C_{t+i}}\right)$ is a zero function, since consumption is everywhere positive and finite. Therefore, in order to rid $(1/C_{t+i})\, g_{ii}$ of its dependence on the arbitrary units of C_t, we plot the "normalized" second derivative of X_{t+i}, which consists only of the term g_{ii}. To conserve space, we plot only the values $g_{11}\left(\frac{C_t}{C_{t+1}}, ..., \frac{C_{t+1-L}}{C_{t+1}}\right)$ corresponding to $\partial^2 X_{t+1}/\partial C_t^2$.

Figure 6 plots the time series of g_{11} for the model estimated on Group 1 assets with a constant, the linear and squared terms of $\mathbf{w}_t$ as instruments; Figure 7 plots the same for the model estimated on Group 2 assets and a constant, the linear, squared and cross terms of $\mathbf{w}_t$ as instruments; Figure 8 plots the same for the model estimated on Group 3 assets with a constant, the linear terms of $\mathbf{w}_t$ as instruments. In each figure, the second partial derivatives are everywhere negative; taken together with the estimates of the first partial derivatives (e.g., Figure 2), this implies that the habit is increasing in lagged consumption, but at a decreasing rate. Moreover, all three figures indicate that the second derivative of X_t is nonzero.

We could construct a consistent test of whether the second derivatives of the habit are identically zero functions. However, for simplicity, we provide a test based on *smooth functionals* of unknown

functions, as discussed in Chen and Shen (1998). Such smooth functionals converge at the standard parametric rate, $\sqrt{T}$, and have standard asymptotic distributions.

One such smooth functional is the unconditional mean of the second derivative

$$\mu \equiv E\left[\frac{\partial^2 X_{t+i}}{\partial C_t^2}\right]. \tag{21}$$

Clearly if (20) is everywhere identically equal to zero, its mean (21) must be zero, although the converse need not hold. Nevertheless, if we find that the mean (21) is statistically different from zero, we may conclude that (20) is not true, and the habit function is nonlinear. We present the results of such a test now, focusing, as discussed above, on the normalized second derivative corresponding to $\partial^2 X_{t+1}/\partial C_t^2$, which is $g_{11}\left(\frac{C_t}{C_{t+1}}, ..., \frac{C_{t+1-L}}{C_{t+1}}\right)$.

Let

$$\mu_g \equiv E\left[g_{11}\left(\frac{C_t}{C_{t+1}}, ..., \frac{C_{t+1-L}}{C_{t+1}}\right)\right],$$

and let $\widehat{g}()$ denote the SMD estimate of $g()$ and $\widehat{g}_{11}()$ be the second partial derivative of $\widehat{g}()$ with respect to its first argument. Then μ_g can be estimated by

$$\widehat{\mu}_g = \frac{1}{T}\sum_{t=L}^{T} \widehat{g}_{11}\left(\frac{C_t}{C_{t+1}}, ..., \frac{C_{t+1-L}}{C_{t+1}}\right).$$

By extending the theoretical result of Ai and Chen (2005) for I.I.D. data to stationary beta-mixing data, under some regularity conditions, we obtain that $\sqrt{T}(\widehat{\mu}_g - \mu_g)$ is asymptotically normally distributed with mean zero and variance $\sigma_{11}^2 > 0$. Let $\widehat{\sigma}_{11}^2$ be some consistent estimator of σ_{11}^2. We may use $\sqrt{T}\widehat{\mu}_g/\widehat{\sigma}_{11}$ as a test statistic, which has a standard normal limiting distribution under the null hypothesis $\mu_g = 0$.[19]

The mean of the second derivative is estimated very precisely with tight confidence intervals, indicating strong rejections of $\mu_g = 0$. For the estimation on Group 1 assets with a constant, and the linear and squared terms of $\mathbf{w}_t$ as instruments, the estimated 95 percent confidence region for μ_g is $[-0.052151, -0.051435]$. Since the estimated habit function is very similar across estimations, the confidence intervals for estimation on Group 2 assets using a constant, the linear, squared and cross-term of $\mathbf{w}_t$ as instruments, and on Group 3 assets with a constant and the linear terms of $\mathbf{w}_t$ as instruments, are almost identical to that reported above. In every case, zero is well outside the 95% confidence region. Thus, we reject the hypothesis that $\mu_g \equiv E[g_{11}]$ is zero, implying that (20) is not true and the habit function estimated is nonlinear.

[19] The asymptotic variance σ_{11}^2 takes a complicated form and the calculation of any consistent estimator $\widehat{\sigma}_{11}^2$ is complicated. For this reason, we use a bootstrap procedure to directly compute an empirical, 95% confidence region for μ_g. The bootstrap sample is obtained by sampling blocks of the raw data randomly with replacement and laying them end-to-end in the order sampled. To choose the block length, we follow the recommendation of Hall, Horowitz, and Jing (1995) who show that the asymptotically optimal block length for estimating a symmetrical distribution function is $l \propto T^{1/5}$; also see Horowitz (2003).We then conduct SMD estimation on 500 bootstrap samples so formed, which delivers 500 estimates of μ_g.

7 Specification and Model Comparison

7.1 Is Habit Formation Internal or External?

An interesting hypothesis concerns the distinction between "internal" and "external" habit formation. Much of our intuition about this distinction is based on simple linear models of habit formation. For example, Cochrane (2001) (Chapter 21, page 484) considers an example in which the habit is a distributed lag of past consumption. Under the additional assumptions of an I.I.D. endowment sequence and a constant risk-free rate R^f equal the inverse of the subjective rate of time preference, Cochrane (2001) shows that the asset pricing implications of internal and external habits are equivalent.

This equivalency breaks down if habits are nonlinear in past consumption or if the risk-free rate is not constant. With nonlinear habits, the partial derivatives $\frac{\partial X_{t+j}}{\partial C_t}$ in (3) vary with lagged consumption, implying that marginal utility under internal habit formation will not be proportional to that under external habit formation. The asset pricing implications of internal and external habit specifications will in general differ.

One way to assess which specification better describes the data is simply to compare the value of the minimized SMD criterion function when habit formation is restricted to be external (i.e., imposing the restriction that $MU_t = (C_t - X_t)^{-\gamma}$) with that of the internal habit cases estimated above. Doing so for each estimation described above, we find that the minimized SMD criterion is several orders of magnitude larger under external habit formation. For example, for estimation on Group 3 asset returns (using a constant and the linear terms of $\mathbf{w}_t$ as instruments), we find that the external habit model produces a minimized criterion equal to 1.2177e-04, compared to 1.3424e-07 for the internal habit case, or about 1000 times larger. The estimations using Group 1 and Group 2 assets produced similar results.

Such findings do not provide a statistical comparison of external and internal habit. Thus, we pursue an alternative way to address the issue by directly testing the Euler equation restrictions corresponding to the external habit formation:

$$E\left(\delta\left(\frac{C_{t+1}-X_{t+1}^{ex}}{C_t-X_t^{ex}}\right)^{-\gamma}R_{i,t+1}-1 \mid \mathbf{w}_t\right)=0, \qquad i=1,...,N. \tag{22}$$

In the appendix we show that the conditional moment restrictions for the internal habit asset pricing model can be written as:

$$E\left(\delta\left(\frac{C_{t+1}-X_{t+1}^{in}}{C_t-X_t^{in}}\right)^{-\gamma}R_{i,t+1}\widetilde{F}_{i,t+1}-1 \mid \mathbf{w}_t\right)=0, \qquad i=1,...,N, \tag{23}$$

where

$$\widetilde{F}_{i,t+1} \equiv 1-\sum_{j=0}^{L}\delta^j\left(\frac{C_{t+1+j}-X_{t+1+j}^{in}}{C_{t+1}-X_{t+1}^{in}}\right)^{-\gamma}\frac{\partial X_{t+1+j}^{in}}{\partial C_{t+1}}+\sum_{j=0}^{L}\delta^{j-1}\left(\frac{C_{t+j}-X_{t+j}^{in}}{C_{t+1}-X_{t+1}^{in}}\right)^{-\gamma}\frac{\partial X_{t+j}^{in}}{\partial C_t}\frac{1}{R_{i,t+1}}.$$

We assume that the internal and external habit functions are both homogeneous of degree one functions of current and lagged consumptions with the *same* finite lag length L. Under this assumption, the two habit functions may be distinguished. To see this, write the habit function generically as $X_t = C_t g\left(\frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t}\right)$. The external and the internal habit conditional moment restrictions are then nested into

$$E\left(\delta\left(\frac{C_{t+1}-X_{t+1}}{C_t-X_t}\right)^{-\gamma} R_{i,t+1}\widetilde{F}_{i,t+1}(\beta;\delta,\gamma,g) - 1 \mid \mathbf{w}_t\right) = 0, \qquad i = 1, ..., N, \tag{24}$$

with

$$\begin{aligned} & \widetilde{F}_{i,t+1}(\beta;\delta,\gamma,g) \qquad (25)\\ \equiv\ & 1 - \beta\sum_{j=0}^{L}\delta^j\left(\frac{C_{t+1+j}-X_{t+1+j}}{C_{t+1}-X_{t+1}}\right)^{-\gamma}\frac{\partial X_{t+1+j}}{\partial C_{t+1}} + \beta\sum_{j=0}^{L}\delta^{j-1}\left(\frac{C_{t+j}-X_{t+j}}{C_{t+1}-X_{t+1}}\right)^{-\gamma}\frac{\partial X_{t+j}}{\partial C_t}\frac{1}{R_{i,t+1}}. \end{aligned}$$

The external habit corresponds to (24) with $\beta = 0$; the internal habit corresponds to (24) with $\beta = 1$.

We estimate all the unknown parameters $(\delta, \gamma, \beta, g(\cdot))$ corresponding to the generalized conditional moment restrictions (24) by the SMD method employed before, using the same ANN sieve (19) to approximate the unknown $g(\cdot)$, the same three asset groups with the associated sets of instruments. The estimation procedure is therefore free to choose one set of parameter values $(\delta^{ex}, \gamma^{ex}, g^{ex}(\cdot))$ satisfying the external habit conditional moment restrictions (22) with $\beta = 0$, just as it is to choose another set of parameter values $(\delta^{in}, \gamma^{in}, g^{in}(\cdot))$ satisfying the internal habit conditional moment restrictions (23) with $\beta = 1$. It follows that a test of the hypothesis that $\beta = 0$ or $\beta = 1$ provides a test of the hypothesis of external versus internal habit. As before, the SMD estimates of $(\delta, \gamma, \beta, g(\cdot))$ converge in probability to their true values and the SMD estimates of finite-dimensional parameters δ, γ, β are root-T consistent and asymptotically normally distributed. Let $\widehat{\sigma}_\beta^2$ be a consistent estimator of the asymptotic variance of the SMD estimate of β. Using the result that $\sqrt{T}(\widehat{\beta} - \beta^*)/\widehat{\sigma}_\beta$ is asymptotically standard normal, we can then apply a sequential Wald test of the null hypothesis of $\beta = 0$ for external habit; and the null hypothesis of $\beta = 1$ for internal habit.

For all the three asset groups and the corresponding instruments, the SMD estimates of δ, γ and their standard errors are quite similar to those in Table 1;[20] the estimates of β are found to be extremely close to unity, ranging between 1.0272 (for Group 1), 1.0876 (for Group 2) and 1.0275 (for Group 3), with small standard errors. In all cases, the hypothesis of $\beta = 0$ (external habit formation) is strongly rejected. For example, using the estimation result for Group 2 assets with the constant, the linear, squared and cross terms of $\mathbf{w}_t$ as instruments, $\widehat{\beta} = 1.0876$ and $\widehat{\sigma}_\beta = 0.057$,

[20] For example, for Group 2 assets with the constant, the linear, squared and cross terms of $\mathbf{w}_t$ as instruments, the estimates of δ, γ are 0.9805 and 0.7991 with standard errors 0.018 and 0.014 respectively.

we can easily reject the null hypothesis of $\beta = 0$ but fail to reject the null hypothesis of $\beta = 1$. Thus we find that the stochastic discount factor is much better approximated by internal habit formation than external habit formation.

7.2 Model Comparison

How well habit-based models explain asset pricing data relative to other models that have been explored in the literature? We use the methodology provided by Hansen and Jagannathan (1997), who develop a way to compare asset pricing models when all Stochastic Discount Factor (SDF) models are treated as misspecified proxies for the true unknown SDF, and the relevant question is which model contains the least specification error.

Hansen and Jagannathan suggest that we compare the pricing errors of various candidate SDF models by choosing each model's parameters, $\boldsymbol{\theta}$, to minimize the quadratic form $\mathbf{g}_T^{HJ}(\boldsymbol{\theta}) \equiv \mathbf{w}_T'(\boldsymbol{\theta})\,\mathbf{G}_T^{-1}\mathbf{w}_T(\boldsymbol{\theta})$, where $\mathbf{w}_T(\boldsymbol{\theta}) = (w_{1T}(\boldsymbol{\theta}), ..., w_{NT}(\boldsymbol{\theta}))'$ is the vector of the sample average of pricing errors (i.e., $w_{iT}(\boldsymbol{\theta}) = \frac{1}{T}\sum_{t=1}^{T} M_t(\boldsymbol{\theta})R_{i,t} - 1$ for $i = 1, ..., N$), and $\mathbf{G}_T$ is the sample second moment matrix of the N asset returns upon which the models are evaluated (i.e., the (i,j)-the element of $\mathbf{G}_T$ is $\frac{1}{T}\sum_{t=1}^{T} R_{i,t}R_{j,t}$ for $i, j = 1, ..., N$). The measure of model misspecification is then the square root of this minimized quadratic form, $d_T \equiv \sqrt{\mathbf{g}_T^{HJ}(\widehat{\boldsymbol{\theta}})}$, 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. We refer to the square root of this minimized quadratic form, $d_T \equiv \sqrt{\mathbf{g}_T^{HJ}(\widehat{\boldsymbol{\theta}})}$, as the *Hansen-Jagannathan distance*, or HJ distance for short.

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

We compare the specification errors of the habit-based model to several alternative asset pricing models that have been studied in the literature. First, we compare the SMD-estimated habit model to two empirical asset pricing models that have displayed relative success in explaining the cross-section of stock market portfolio returns: 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

$$M_{t+1} = \theta_0 + \sum_{i=1}^{k} \theta_i F_{i,t+1}, \tag{26}$$

where $F_{i,t+1}$ are variable factors, and the coefficients θ_0 and θ_i are treated as free parameters to be estimated. Fama and French develop an empirical three-factor model ($k = 3$), with variable factors

related to firm size (market capitalization), book equity-to-market equity, and the aggregate stock market. These factors are the "small-minus-big" (SMB_{t+1}) portfolio return, the "high-minus-low" (HML_{t+1}) portfolio return, and the market return, $R_{m,t+1}$, respectively.[21] The Fama-French model has displayed unusual success in explaining the cross section of mean equity returns (Fama and French (1993), Fama and French (1996)). The model explored by Lettau and Ludvigson (2001b) can be interpreted as a "scaled" or conditional consumption CAPM ("scaled CCAPM" hereafter) and also has three variable factors ($k = 3$), $\widehat{cay}_t$, $\widehat{cay}_t \cdot \Delta \log C_{t+1}$, and $\Delta \log C_{t+1}$. Lettau and Ludvigson (2001b) show that such a model can be thought of as a linear approximation to any consumption-based CAPM (CCAPM) in which risk-premia vary over time.

We also compare specification errors of these models to those of a linearized version of the standard CCAPM (with consumption growth the single variable factor in (26), $k = 1$) (Breeden (1979); Breeden and Litzenberger (1978)), and to those of the classic CAPM, where the market return R_m is the single variable factor in (26), $k = 1$.

For all the non-habit asset pricing models, all the unknown parameters $\boldsymbol{\theta} = (\theta_0, \theta_1, ..., \theta_k)'$ are chosen to minimize the squared HJ distance, $\mathbf{g}_T^{HJ}(\boldsymbol{\theta})$, for that model. Because the fully parametric non-habit asset pricing models have fewer parameters than the nonparametric habit models, when computing the HJ distance for the habit models we deliberately restrict the number of parameters that are chosen to minimize the HJ distance. In particular, we choose only the finite-dimensional parameters $\boldsymbol{\theta} = (\delta, \gamma)'$ of our habit-based models to minimize the HJ distance–the parameters of the nonparametric habit function are restricted to be those that minimize the SMD criterion. The procedure is as follows. For any candidate $\boldsymbol{\theta} = (\delta, \gamma)'$ value in the habit-based asset pricing models, we first estimate the habit function by the SMD estimator $\widehat{g}(\cdot; \delta, \gamma) = \arg\min_{g_{K_T}} \frac{1}{T} \sum_{t=1}^{T} \widehat{m}(\mathbf{w}_t, \delta, \gamma, g_{K_T})' \widehat{m}(\mathbf{w}_t, \delta, \gamma, g_{K_T})$, where $\widehat{m}(\mathbf{w}_t, \delta, \gamma, g_{K_T})$ is given in (14), using, say, Group 2 assets (six size and book-market returns plus the T-bill rate). We then construct the SMD-estimated internal habit and external habit SDF proxies by replacing the unknown habit function $g_o(\cdot; \delta, \gamma)$ with the SMD-estimated $\widehat{g}(\cdot; \delta, \gamma)$. Finally, given $\widehat{g}(\cdot; \delta, \gamma)$, we choose the parameters $\boldsymbol{\theta} = (\delta, \gamma)'$ of the habit asset pricing models to minimize the corresponding HJ distances. Note that this places the habit models at a disadvantage because the habit parameters are not chosen to minimize the HJ criterion, the measure of model misspecification. By contrast all of the comparison models' parameters are chosen to minimize the HJ criterion.[22]

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

[22] 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.

We evaluate the specification errors of the competing asset pricing models using a time-series on two alternative sets of quarterly asset returns: (i) the six equity returns on portfolios double-sorted on size and book-to-market characteristics provided by Fama and French, ($N = 6$), and (ii) these six equity portfolio returns plus the three-month Treasury bill rate, ($N = 7$). We use equity returns on size and book-to-market sorted portfolios because Fama and French (1992) show that these two characteristics provide a "simple and powerful characterization" of the cross-section of average stock returns, and absorb the roles of leverage, earnings-to-price ratio and many other factors governing average stock return differentials.

Before discussing how each model fares according to specification error, we note that the estimates of δ generated from minimizing $\mathbf{g}_T^{HJ}(\boldsymbol{\theta})$ for the general internal habit model are similar to those estimated using the SMD procedure but generally smaller, equal to 0.70 when the model is evaluated on the equity portfolios alone, and 0.73 when the Treasury bill is included. Estimates of δ in the external habit case are similar to the internal habit case, equal to 0.9 when the Treasury bill is included and 0.6 when it is omitted. The estimates of the curvature parameter γ, when freely estimated to minimize the HJ distance, are substantially larger than those estimated using the SMD procedure. For the internal habit case, they equal to 26 when the model is evaluated on the equity portfolios alone, and 25 when the Treasury bill is included. For the external habit model, the resulting estimates of the curvature parameter γ are equal to 37 when the Treasury bill is included, and 62 when it is excluded. The larger values for γ are not surprising because the HJ procedure places emphasis on unconditional mean returns whereas the SMD procedure emphasizes conditional moments. Fitting unconditional moments requires a more volatile SDF (Hansen and Jagannathan (1991)), which can be generated by a higher value for γ.

Table 2 reports the measure of specification error given by the HJ distance ("HJ Dist"), $d_T \equiv \sqrt{\mathbf{g}_T^{HJ}(\widehat{\boldsymbol{\theta}})}$, for all the models discussed above. Regardless of whether the Treasury bill rate is included in the set of test assets, the smallest specification error is generated by the SMD-estimated internal habit model. The HJ distance for this model is equal to 0.18 on the set of stock returns and Treasury bill, and 0.17 on the set of size and book-market sorted equity returns alone. These numbers are substantially lower than those for all the other models; for example, when the T-bill is included, the next lowest pricing error is given by the Fama-French model, equal to 0.28. When the T-bill is excluded, the scaled CCAPM delivers the second lowest specification error, equal to 0.21, while the third lowest specification error is generated by the Fama-French model and the external habit model, which have almost identical values of the HJ distance, equal to 0.262 and 0.261, respectively. The scaled CCAPM performs much worse once the T-bill is included. Finally, the classic CCAPM and CAPM have errors substantially larger than those of most other models, regardless of the set of test assets upon which they are evaluated. The one exception is the external habit model which performs worst of all models when the T-bill is included, generating

a specification error of 0.43. These results are particularly encouraging for the internal habit framework.

Table 3 also reports an alternate distance metric, denoted "HJ^+ Dist," for all the models discussed above. This metric restricts the set of admissible stochastic discount factors to be positive. The SMD-estimated internal habit model continues to beat all the other models by a wide margin, according to this metric. The only models for which HJ^+ is substantially different from HJ, are the scaled CCAPM and the classic CCAPM. The distance metric for the scaled CCAPM increases by a factor of four in pricing the equity returns when the set of admissible stochastic discount factors is restricted to be positive. Although this model does a relatively 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 altogether surprising, since linear models–typically implemented as approximations of nonlinear models for use in specific applications–are not designed to price derivative claims.

7.2.1 Statistical Comparison of Competing Models

Table 2 reports the point estimates of the HJ distances for the six competing asset pricing models without taking into account the statistical uncertainty associated with those estimates. To our knowledge, no broadly applicable procedure has been developed for comparing HJ distances statistically.[23] Here we provide a new methodology for statistical model comparison of HJ distances based on the "reality check" method of White (2000). Given that the HJ distance is found to be smallest for the internal habit model, we make this our null hypothesis for model comparison.

Let $j = 1, 2, ..., 6$ index the six asset pricing models reported in Table 2, with $j = 1$ being the internal habit model. Let $E[\mathbf{w}_T^j(\boldsymbol{\theta}_j)]$ denote the vector of population average of pricing errors associated with model j and the candidate parameter value $\boldsymbol{\theta}_j$. Let

$$\mathbf{g}_j^{HJ}(\boldsymbol{\theta}_j) \equiv E[\mathbf{w}_T^j(\boldsymbol{\theta}_j)]'\{E[\mathbf{G}_T]\}^{-1}E[\mathbf{w}_T^j(\boldsymbol{\theta}_j)]$$

and $\boldsymbol{\theta}_j^* \equiv \arg\min \mathbf{g}_j^{HJ}(\boldsymbol{\theta}_j)$. Denote the population squared HJ distance associated with model j $d_j^2 \equiv \mathbf{g}_j^{HJ}(\boldsymbol{\theta}_j^*)$. The null hypothesis is:

$$H_0: \quad \max_{j=2,...,6}\{d_1^2 - d_j^2\} \leq 0,$$

meaning that, among the six competing models, model 1 (the internal habit model) has the smallest

[23] Wang and Zhang (2003) provide a way to compare HJ distance measures across models using Bayesian methods, under the assumption that the data follow linear, Gaussian processes. Their procedure is not directly applicable here, since our methodology requires no specification of the law of motion and permits the data to be stationary ergodic, allowing a wide variety of nonlinear time-series processes such as diffusion models, stochastic volatility, nonlinear ARCH, GARCH, Markov switching, and many more.

pricing error according to the squared HJ distance. The alternative hypothesis is:

$$H_1: \quad \max_{j=2,...,6}\{d_1^2 - d_j^2\} > 0,$$

meaning that there is at least one competing model has smaller pricing error in the squared HJ distance than model 1 (the internal habit model).

Let $d_{T,j}^2 \equiv \mathbf{g}_{T,j}^{HJ}(\widehat{\boldsymbol{\theta}}_j)$ denote the sample estimate of the squared HJ distance for model j. To apply White's reality check test, we employ the following test statistics, $\mathcal{T}^W$, based on (White (2000)) and Hansen's modified reality check test statistic, $\mathcal{T}^H$, based on (Hansen (2003)):[24]

$$\mathcal{T}^W \equiv \max_{j=2,...,6} \sqrt{T}\{d_{T,1}^2 - d_{T,j}^2\} \quad \text{and} \quad \mathcal{T}^H \equiv \max\left(\max_{j=2,...,6} \sqrt{T}\{d_{T,1}^2 - d_{T,j}^2\}\ ,\ 0\right).$$

Both tests have complicated null limiting distributions. Therefore, to implement the reality check test, we follow the suggestion of White (2000), Hansen (2003) and Corradi and Swanson (2003) and compute block bootstrap estimates of their finite sample distributions. Let White's original bootstrap test statistic and Hansen's modified bootstrap test statistic be denoted $\mathcal{T}^{W,b}$ and $\mathcal{T}^{H,b}$, respectively. By repeated sampling, we compute bootstrap estimates of the $p-values$

$$\widehat{p}_W \equiv \frac{1}{B}\sum_{b=1}^{B} 1\{\mathcal{T}^{W,b} > \mathcal{T}^W\}, \quad \widehat{p}_H \equiv \frac{1}{B}\sum_{b=1}^{B} 1\{\mathcal{T}^{H,b} > \mathcal{T}^H\},$$

where B is the number of bootstrap samples. To account for the first-stage estimation of the unknown habit function, in each bootstrap sample the parameters of the habit function are reestimated using the SMD procedure.

Since this is a one-sided test, at a 5% level of significance, the critical value is the value at the $95th$ percentile of the bootstrap test statistics. We do not reject the null if the historical value of the tests statistic ($\mathcal{T}^W$ or $\mathcal{T}^H$) is not unusually high, i.e., is not greater than the value at the $95th$ percentile. Therefore, the tests $\mathcal{T}^W$ and $\mathcal{T}^H$ reject the null hypothesis if $\widehat{p}_W$ or $\widehat{p}_H$ are close to zero, and do not reject the null if $\widehat{p}_W$ or $\widehat{p}_H$ are close to one. At a 5% level of significance, we reject the null if $\widehat{p}_W$ or $\widehat{p}_H$ are less than 0.05, but do not reject otherwise. The results of these tests in our data are as follows. For the six return case $\widehat{p}_W = 0.9731$ and $\widehat{p}_H = 0.9275$, while for the six return plus T-bill case $\widehat{p}_W = 0.9838$ and $\widehat{p}_H = 0.9358$. Thus, both tests imply that model 1 (the internal habit model) is the best according to the squared HJ distance measure.

[24] Using results in Hansen, Heaton, and Luttmer (1995) and Ai and Chen (2005), it can be shown, under mild regularity conditions,

$$\sqrt{T}\left(d_{T,1}^2 - d_{T,2}^2 - [d_1^2 - d_2^2], ..., d_{T,1}^2 - d_{T,6}^2 - [d_1^2 - d_6^2]\right)' \xrightarrow{D} (Z_2, ..., Z_6)'$$

$$(Z_2, ..., Z_6)' \sim \mathcal{N}(0, \Omega),$$

where Ω is a positive semi-definite matrix. This justifies our applications of White's reality check test (White (2000)), $\mathcal{T}^W$, and Hansen's modified reality check test (Hansen (2003)), $\mathcal{T}^H$.

7.2.2 Comparing Internal and External Habit Specifications

What explains the superior econometric performance of the internal habit model compared to the external habit specification? Certainly the SDFs of the two models behave differently over time: the internal habit model is more volatile and more autocorrelated (quarterly standard deviation equals to 2.8% and first-order autocorrelation equals to 0.21), than the external habit model (quarterly standard deviation equals to 1.05% and first-order autocorrelation equals to 0.04).[25] But such time-series properties are likely to be more important determinants of how well each specification rationalizes variation over time in returns rather than across assets.

Here, models are evaluated on their ability to explain cross-sectional moment restrictions. These moment restrictions imply that the unconditional mean excess return of any asset must be proportional to the covariance of the SDF with that asset's return.[26] Thus the relative performance of the two models can be intuitively understood by examining how the SDFs of each model covaries with the set of test asset returns. Table 4 shows sample estimates of these covariances, for the SDFs of the internal habit model and external habit model, with each of the six size and book-market sorted portfolio returns we use as test assets. For reference, the top panel of Table 4 shows the unconditional mean quarterly excess return on each portfolio over the three-month Treasury bill rate.

Notice that the average excess returns on these portfolios are large, ranging from 1.8 percent to 3.2 percent per quarter (top panel, Table 4). Explaining such a large mean excess returns requires the covariance of the SDF with the asset's return to be high. The bottom panel of Table 4 shows that an important reason for the superior performance of the internal habit model is that the absolute value of its covariance with each return is considerably larger (in many cases by an order of magnitude) than that of the external model.[27] The internal habit model also gets the cross-sectional patterns right. For example, the largest cross-sectional spread in returns occurs between the return on S1B3, the "value" portfolio in the smallest size category, and the return on S1B1, the "growth" portfolio in the same size category. The former has an average excess return of 3.2%, the latter just 1.8%. This difference captures the well-known value-premium in these data, which is especially pronounced for small capitalization firms. The internal habit model covaries more with S1B3 than it does with S1B1, thereby explaining the higher excess return on the former relative to

[25] These numbers are for the SDFs estimated when computing the HJ distances with δ and γ are reestimated to minimize $g_T^{HJ}(\boldsymbol{\theta})$.

[26] Start with $1 = E\left[M_t R_{j,t}\right]$ and rearrange to find

$$E\left(R_{j,t}\right) - 1/E\left(M_t\right) = \frac{-\mathrm{Cov}\left(M_t, R_{j,t}\right)}{E\left(M_t\right)}.$$

[27] We expect this covariance to be negative. Positive excess returns are associated with positive covariance with consumption growth, and therefore a negative covariance with marginal utility growth.

the latter. This is not the case for the external habit model, where these two covariances are about the same.

From (23) and (22), it is clear that the key difference between external and internal habit formation is that the internal habit specification contains forward looking terms in marginal utility not present in the external habit specification. Empirically, these forward-looking terms are modeled as projections onto instruments that have predictive power for the endogenous variables in the model. To evaluate the quantitative importance of these forward looking terms, we re-computed the HJ distance for the internal model using instead the ex-post values of the future terms in (3), rather than their ex-ante predicted values, as dictated by theory. In this case we find that the performance of the internal habit model is much closer to that of the external habit model. Thus, a key feature of the empirical success of the internal habit model compared to the external habit specification lies with the forward-looking nature of marginal utility. These findings may be related to recent work by Parker and Julliard (2004), who find that the covariance of size and book-market sorted asset returns with long-horizon consumption growth explains a large fraction of the cross-sectional variation in average returns on these portfolios. They refer to this as "ultimate consumption risk." To the extent that such long-horizon measures of consumption growth are correlated with the forward-looking terms in marginal utility arising from internal habit formation, the findings here provide a structural interpretation of why ultimate consumption risk better explains the cross-section of average returns than one-period consumption risk.

8 Conclusion

A large and growing body of theoretical work in financial economics derives its most salient asset pricing implications from the presence of habit formation in investor preferences. The importance of this work in recent literature calls out for formal empirical evaluation, but such an evaluation is complicated by the lack of theoretical agreement over the functional form of the habit. In this article, we empirically evaluate a general class of representative-agent habit-based asset pricing models. Rather than restricting the functional form of the habit, we pursue a semiparametric approach, treating the habit specification as unknown and estimating it along with the finite dimensional parameters of the power utility function.

This semiparametric approach allows us to empirically evaluate a number of hypotheses about habit-based asset pricing models that have not been previously evaluated. First, our results suggest that, conditional on the power utility framework, preferences are far from time separable: the estimated habit is found to be a quantitatively important part of the power utility specification and is a large fraction of current consumption. Second, we find that the habit specification is better described as a nonlinear function of current and past consumption, rather than as a linear function.

Several authors have argued that nonlinearities in the habit function are crucial for allowing the model to account for the joint behavior of aggregate consumption and asset returns (e.g., Campbell and Cochrane (1999)). Third, when the habits are estimated as nonlinear functions of current and lagged consumption with the same finite lag length, we strongly reject the hypothesis that habits are a pure externality governed by the consumption of everyone else in the economy; internal habit models based on own-consumption better describe the asset pricing data studied here.

Finally, we assess how well the habit-based paradigm explains asset pricing data compared to other asset pricing models. We find that an estimated internal habit model explains a cross-section of size and book-market sorted equity returns better than (i) the Fama and French (1993) three-factor model, (ii) the scaled consumption CAPM explored by Lettau and Ludvigson (2001b), (iii) a SMD-estimated external habit model, (iv) the classic CAPM, and (v) the classic consumption CAPM. The internal habit model performs better on statistical grounds as well as economic grounds.

In this paper we have restricted our analysis to habit specifications that depend on a finite number of consumption lags. Ideally, one would like to allow for an infinite number of lags by modeling the habit as a recursive function of past habits, e.g., $X_t = r(C_t, C_{t-1}, X_{t-1})$. The econometric difficulty with this type of model is that it involves an unknown recursive functional that has as one of its arguments an unknown function, both of which must be estimated nonparametrically. Any procedure that succeeds in identifying r separately from X is likely to be slow to converge. As such, the econometric results required to execute such an estimation have yet to be developed. We argue that the empirical work in this paper is a natural starting place for such an investigation, since the estimation of X_t nonparametrically is likely to comprise one step in the larger recursive procedure.

9 Appendix

This appendices consist of several parts: Appendix 1 describes the data. Appendix 2 presents alternative expressions of the conditional moment models, and also provides sufficient conditions for identification in the external habit case. Appendix 3 describes the general SMD procedure. Appendix 4 presents large sample properties of the SMD estimator. Appendix 5 provide limiting distributions of the test statistics for testing linear habit and testing internal vs external habit.

1. Data Description

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

CONSUMPTION

Consumption is measured as expenditures on nondurables and services, excluding shoes and clothing. The quarterly data are seasonally adjusted at annual rates, in billions of chain- weighted 1996 dollars. The components are chain-weighted together, and this series is scaled up so that the sample mean matches the sample mean of total personal consumption expenditures. Our source is the U.S. Department of Commerce, Bureau of Economic Analysis.

POPULATION

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

PRICE DEFLATOR

Real asset returns are deflated by the implicit chain-type price deflator (1996=100) given for the consumption measure described above. Our source is the U.S. Department of Commerce, Bureau of Economic Analysis.

ASSET RETURNS

- 3-Month Treasury Bill Rate: secondary market, averages of business days, discount basis percent; Source: H.15 Release – Federal Reserve Board of Governors.
- 25 size/book-market value weighted returns for NYSE, AMEX, NASDAQ; Returns were created using 200112 CRSP database. It contains value-weighted returns for the intersections of 5 market equity categories and 5 book equity-market equity categories. The portfolios are constructed at the end of June. Source: Kenneth French's homepage, http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

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

- 10 Industry Portfolios: The process assigns each NYSE, AMEX, and NASDAQ stock to an industry portfolio at the end of June of year t based on its four-digit SIC code at that time. Return data was created by CMPT_IND_RETS using the 200112 CRSP database. Returns are computed from July of t to June of t+1. Source: Kenneth French's homepage, http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html.

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

The proxy for the log consumption-wealth ratio is computed as described in Lettau and Ludvigson (2001a) using data from 1952:4-2001:4.

RELATIVE BILL RATE, $RREL$

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

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

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

R_m, SMB, HML

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

2. Conditional Moment Restrictions and Identification

Alternative expressions of the conditional moment restrictions:

$$E_t\left(M_{t+1}R_{i,t+1} - 1\right) = 0 \qquad i = 1, ..., N,$$

where

$$M_{t+1} = \delta_o \frac{MU_{t+1}}{MU_t},$$

where

$$\begin{aligned} MU_t &= \frac{\partial U}{\partial C_t} = (C_t - X_t)^{-\gamma_o} - E_t \left[\sum_{j=0}^{L} \delta_o^j \left(C_{t+j} - X_{t+j} \right)^{-\gamma_o} \frac{\partial X_{t+j}}{\partial C_t} \right] \\ &= (C_t - X_t)^{-\gamma_o} E_t \left\{ 1 - \sum_{j=0}^{L} \delta_o^j \left(\frac{C_{t+j} - X_{t+j}}{C_t - X_t} \right)^{-\gamma_o} \frac{\partial X_{t+j}}{\partial C_t} \right\}, \end{aligned}$$

where

$$\begin{aligned} X_t &= C_t f_o \left(1, \frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t} \right) = C_t g_o \left(\frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t} \right), \\ C_t - X_t &= C_t \left\{ 1 - g_o \left(\frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t} \right) \right\} \end{aligned}$$

hence

$$M_{t+1} = \delta_o \frac{MU_{t+1}}{MU_t} = \delta_o \left(\frac{C_{t+1} - X_{t+1}}{C_t - X_t} \right)^{-\gamma_o} \frac{E_{t+1} \left\{ 1 - \sum_{j=0}^{L} \delta_o^j \left(\frac{C_{t+1+j} - X_{t+1+j}}{C_{t+1} - X_{t+1}} \right)^{-\gamma_o} \frac{\partial X_{t+1+j}}{\partial C_{t+1}} \right\}}{E_t \left\{ 1 - \sum_{j=0}^{L} \delta_o^j \left(\frac{C_{t+j} - X_{t+j}}{C_t - X_t} \right)^{-\gamma_o} \frac{\partial X_{t+j}}{\partial C_t} \right\}}$$

and

$$E_t \left(\frac{E_{t+1} \left[\delta_o \left(\frac{C_{t+1} - X_{t+1}}{C_t - X_t} \right)^{-\gamma_o} \left\{ 1 - \sum_{j=0}^{L} \delta_o^j \left(\frac{C_{t+1+j} - X_{t+1+j}}{C_{t+1} - X_{t+1}} \right)^{-\gamma_o} \frac{\partial X_{t+1+j}}{\partial C_{t+1}} \right\} R_{i,t+1} \right]}{E_t \left\{ 1 - \sum_{j=0}^{L} \delta_o^j \left(\frac{C_{t+j} - X_{t+j}}{C_t - X_t} \right)^{-\gamma_o} \frac{\partial X_{t+j}}{\partial C_t} \right\}} - 1 \right) = 0.$$

This can be expressed three different ways:

$$E_t \left(\begin{array}{c} \left[\delta_o \left(\frac{C_{t+1} - X_{t+1}}{C_t - X_t} \right)^{-\gamma_o} \left\{ 1 - \sum_{j=0}^{L} \delta_o^j \left(\frac{C_{t+1+j} - X_{t+1+j}}{C_{t+1} - X_{t+1}} \right)^{-\gamma_o} \frac{\partial X_{t+1+j}}{\partial C_{t+1}} \right\} R_{i,t+1} \right] \\ - \left\{ 1 - \sum_{j=0}^{L} \delta_o^j \left(\frac{C_{t+j} - X_{t+j}}{C_t - X_t} \right)^{-\gamma_o} \frac{\partial X_{t+j}}{\partial C_t} \right\} \end{array} \right) = 0$$

$$E_t \left(\begin{array}{c} \delta_o \left(\frac{C_{t+1} - X_{t+1}}{C_t - X_t} \right)^{-\gamma_o} R_{i,t+1} - \sum_{j=0}^{L} \delta_o^{j+1} \left(\frac{C_{t+1+j} - X_{t+1+j}}{C_t - X_t} \right)^{-\gamma_o} \frac{\partial X_{t+1+j}}{\partial C_{t+1}} R_{i,t+1} \\ + \sum_{j=0}^{L} \delta_o^j \left(\frac{C_{t+j} - X_{t+j}}{C_t - X_t} \right)^{-\gamma_o} \frac{\partial X_{t+j}}{\partial C_t} - 1 \end{array} \right) = 0$$

$$E_t \left(\begin{array}{c} \frac{\partial X_t}{\partial C_t} + \delta_o \left(\frac{C_{t+1} - X_{t+1}}{C_t - X_t} \right)^{-\gamma_o} \left[R_{i,t+1} + \frac{\partial X_{t+1}}{\partial C_t} - \frac{\partial X_{t+1}}{\partial C_{t+1}} R_{i,t+1} \right] \\ + \sum_{j=2}^{L} \delta_o^j \left(\frac{C_{t+j} - X_{t+j}}{C_t - X_t} \right)^{-\gamma_o} \left[\frac{\partial X_{t+j}}{\partial C_t} - \frac{\partial X_{t+j}}{\partial C_{t+1}} R_{i,t+1} \right] \\ - \delta_o^{L+1} \left(\frac{C_{t+L+1} - X_{t+L+1}}{C_t - X_t} \right)^{-\gamma_o} \frac{\partial X_{t+L+1}}{\partial C_{t+1}} R_{i,t+1} - 1 \end{array} \right) = 0$$

Now if we specialize to the specification $X_t = a_o C_{t-1}$ with a unknown constant a_o, as in Ferson and Constantinides (1991) with no durable consumption, we have $\frac{\partial X_t}{\partial C_t} = 0$, $\frac{\partial X_{t+1}}{\partial C_t} = a_o$, $\frac{\partial X_{t+j}}{\partial C_t} = 0$ for all $j \geq 2$ and

$$E_t \left(\delta_o \left(\frac{C_{t+1} - X_{t+1}}{C_t - X_t} \right)^{-\gamma_o} [R_{i,t+1} + a_o] - \delta_o^2 \left(\frac{C_{t+2} - X_{t+2}}{C_t - X_t} \right)^{-\gamma_o} a_o R_{i,t+1} - 1 \right) = 0$$

which coincides with their expression.

Alternatively we can write the conditional moment restrictions as:

$$E_t \left(\delta_o \left(\frac{C_{t+1} - X_{t+1}}{C_t - X_t} \right)^{-\gamma_o} R_{i,t+1} \widetilde{F}_{i,t+1} - 1 \right) = 0, \quad i = 1, ..., N, \tag{27}$$

with

$$\widetilde{F}_{i,t+1} \equiv 1 - \sum_{j=0}^{L} \delta_o^j \left(\frac{C_{t+1+j} - X_{t+1+j}}{C_{t+1} - X_{t+1}} \right)^{-\gamma_o} \frac{\partial X_{t+1+j}}{\partial C_{t+1}} + \sum_{j=0}^{L} \delta_o^{j-1} \left(\frac{C_{t+j} - X_{t+j}}{C_{t+1} - X_{t+1}} \right)^{-\gamma_o} \frac{\partial X_{t+j}}{\partial C_t} \frac{1}{R_{i,t+1}}.$$

We note that $\widetilde{F}_{i,t+1} = 1$ for external habit.

Alternatively

$$E_t \left(\delta_o \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma_o} R_{i,t+1} F_{i,t+1} - 1 \right) = 0, \qquad i = 1, ..., N, \tag{28}$$

with

$$F_{i,t+1} \equiv \widetilde{F}_{i,t+1} \frac{\left(1 - g_o \left(\frac{C_t}{C_{t+1}}, ..., \frac{C_{t+1-L}}{C_{t+1}} \right) \right)^{-\gamma_o}}{\left(1 - g_o \left(\frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t} \right) \right)^{-\gamma_o}},$$

$$\begin{aligned} F_{i,t+1} \equiv & \frac{\left(\begin{array}{c} \left(1 - g_o \left(\frac{C_t}{C_{t+1}}, ..., \frac{C_{t+1-L}}{C_{t+1}} \right) \right)^{-\gamma_o} \\ - \left[\sum_{j=0}^{L} \delta_o^j \left(\frac{C_{t+1+j}}{C_{t+1}} \right)^{-\gamma_o} \left(1 - g_o \left(\frac{C_{t+j}}{C_{t+1+j}}, ..., \frac{C_{t+j+1-L}}{C_{t+1+j}} \right) \right)^{-\gamma_o} \frac{\partial X_{t+1+j}}{\partial C_{t+1}} \right] \end{array} \right)}{\left(1 - g_o \left(\frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t} \right) \right)^{-\gamma_o}} \\ & + \frac{\sum_{j=0}^{L} \delta^{j-1} \left(\frac{C_{t+j}}{C_{t+1}} \right)^{-\gamma_o} \left(1 - g_o \left(\frac{C_{t+j-1}}{C_{t+j}}, ..., \frac{C_{t+j-L}}{C_{t+j}} \right) \right)^{-\gamma_o} \frac{\partial X_{t+j}}{\partial C_t} \frac{1}{R_{i,t+1}}}{\left(1 - g_o \left(\frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t} \right) \right)^{-\gamma_o}}. \end{aligned}$$

Identification:

In this section, we provide sufficient conditions so that the conditional moment restrictions (11) identifies the parameters of interest $\boldsymbol{\alpha}_o = (\delta_o, \gamma_o, g_o)'$ in the external habit case, when $L = 1$. While we do not provide a formal proof of identification for the more complex internal habit case, the results for external habit are strongly suggestive for that case, since the presence of a time-varying component $\widetilde{F}_{i,t+1}$ under internal habit formation (a nonlinear function of $(\delta_o, \gamma_o, g_o)$) should

make identification easier than under external habit formation where $\widetilde{F}_{i,t+1} = 1$. Thus we study identification of the unknown true parameters $(\delta_o, \gamma_o, g_o)$ satisfying the conditional moment (28) under the special case of external habit with $L = 1$. We note that, in practice, all of the parameters of our model seem to be well identified, since (as shown above) the estimation results are remarkably similar across a variety specifications with different instruments and different asset returns.

The conditional moment restrictions for this case are:

$$E\left(h_o\left(\frac{C_t}{C_{t+1}}; \frac{C_{t-1}}{C_t}\right) R_{i,t+1} - 1 \mid \mathbf{w}_t\right) = 0, \quad i = 1, ..., N, \tag{29}$$

with $\mathbf{w}_{-c,t} \equiv [\widehat{cay}_t, RREL_t, SPEX_t]'$, $\mathbf{w}_t = \left[\mathbf{w}'_{-c,t}, \frac{C_{t-1}}{C_t}\right]'$ and

$$\begin{aligned} h_o\left(\frac{C_t}{C_{t+1}}; \frac{C_{t-1}}{C_t}\right) &= \delta_o \left(\frac{C_{t+1}}{C_t}\right)^{-\gamma_o} \frac{\left(1 - g_o\left(\frac{C_t}{C_{t+1}}\right)\right)^{-\gamma_o}}{\left(1 - g_o\left(\frac{C_{t-1}}{C_t}\right)\right)^{-\gamma_o}} \\ \text{with} \quad 0 &\leq g_o < 1, \ g_o \neq const., \ \gamma_o > 0, \ \delta_o > 0. \end{aligned} \tag{30}$$

We note that the conditional moment restriction (29) is treating the stochastic discount factor $M_{t+1} \equiv \delta \frac{MU_{t+1}}{MU_t}$ as a totally unknown function $M_{t+1} = h_o\left(\frac{C_t}{C_{t+1}}; \frac{C_{t-1}}{C_t}\right)$. We first provide sufficient conditions to identify the totally unknown $h_o()$ using the conditional moment restriction (29), and we then use the semiparametric specification (30) and the identified $h_o()$ function to identify $(\delta_o, \gamma_o, g_o)$.

The conditional moment restriction (29) is very similar to the equation (2.3) in Newey and Powell (1988). In the following, we denote $f_i(\frac{C_t}{C_{t+1}}, R_{i,t+1}|\mathbf{w}_t)$, $i = 1, ..., N$, as the conditional density of $(\frac{C_t}{C_{t+1}}, R_{i,t+1})$ given $\mathbf{w}_t$, and $f(\frac{C_t}{C_{t+1}}|\mathbf{w}_t)$ as the conditional density of $\frac{C_t}{C_{t+1}}$ given $\mathbf{w}_t = \left[\mathbf{w}'_{-c,t}, \frac{C_{t-1}}{C_t}\right]'$, where $\mathbf{w}_{-c,t} \equiv [\widehat{cay}_t, RREL_t, SPEX_t]'$. Following the result in Newey and Powell (1988), we have that the identification of $M_{t+1} = h_o\left(\frac{C_t}{C_{t+1}}; \frac{C_{t-1}}{C_t}\right)$ using the restriction (29) is equivalent to:

$$\begin{aligned} 0 &= \int\int \theta(y_{t+1}; y_t) R_{i,t+1} f_i(y_{t+1}, R_{i,t+1}|\mathbf{w}_t) dy_{t+1} dR_{i,t+1} \\ &= \int \theta(y_{t+1}; y_t) \left\{ \int \frac{R_{i,t+1}}{E[R_{i,t+1}|\mathbf{w}_{-c,t}, y_t]} f_i(y_{t+1}, R_{i,t+1}|\mathbf{w}_{-c,t}, y_t) dR_{i,t+1} \right\} dy_{t+1} \\ &= \int \theta(y_{t+1}; y_t) \left\{ \frac{E[R_{i,t+1}|y_{t+1}; \mathbf{w}_{-c,t}, y_t]}{E[R_{i,t+1}|\mathbf{w}_{-c,t}, y_t]} f(y_{t+1}|\mathbf{w}_{-c,t}, y_t) \right\} dy_{t+1} \end{aligned} \tag{31}$$

for all $i = 1, ..., N$, *implies* $\theta(y_{t+1}; y_t) = 0$ *almost surely.*

Thus, the identification condition for $M_{t+1} = h_o\left(\frac{C_t}{C_{t+1}}; \frac{C_{t-1}}{C_t}\right)$ becomes:

(*) *for each* $y_t \equiv \frac{C_{t-1}}{C_t}$, *and for all* $i = 1, ..., N$, *the "i-th return adjusted" conditional density,* $\frac{E[R_{i,t+1}|y_{t+1}; \mathbf{w}_{-c,t}, y_t]}{E[R_{i,t+1}|\mathbf{w}_{-c,t}, y_t]} f(y_{t+1}|\mathbf{w}_{-c,t}, y_t)$, *of* $y_{t+1} \equiv \frac{C_t}{C_{t+1}}$ *given* $\mathbf{w}_{-c,t}$ *is complete.*

See Newey and Powell (1988) for additional sufficient condition for the "completeness" in terms of exponential families.

Next we show that $(\delta_o, \gamma_o, g_o)$ is identified given the identified $h_o()$ function and the specification (30). In the following we let $g'()$ denote the derivative of g, $h'_k(\cdot;\cdot)$ denote the partial derivative of h with respect to its $k-$th element for $k = 1, 2$. Then the semiparametric specification (30) implies

$$\frac{h'_2\left(\frac{C_t}{C_{t+1}}; \frac{C_{t-1}}{C_t}\right)}{\gamma \times h\left(\frac{C_t}{C_{t+1}}; \frac{C_{t-1}}{C_t}\right)} = \frac{-g'\left(\frac{C_{t-1}}{C_t}\right)}{1 - g\left(\frac{C_{t-1}}{C_t}\right)} \quad \textit{almost surely}, \tag{32}$$

$$\frac{h'_1\left(\frac{C_t}{C_{t+1}}; \frac{C_{t-1}}{C_t}\right)}{\gamma \times h\left(\frac{C_t}{C_{t+1}}; \frac{C_{t-1}}{C_t}\right)} = \left(\frac{C_t}{C_{t+1}}\right)^{-1} + \frac{g'\left(\frac{C_t}{C_{t+1}}\right)}{1 - g\left(\frac{C_t}{C_{t+1}}\right)} \quad \textit{almost surely}. \tag{33}$$

Since $\frac{C_{t-1}}{C_t}$ and $\frac{C_t}{C_{t+1}}$ have the same distribution, we have

$$E\left\{\frac{g'\left(\frac{C_{t-1}}{C_t}\right)}{1 - g\left(\frac{C_{t-1}}{C_t}\right)}\right\} = E\left\{\frac{g'\left(\frac{C_t}{C_{t+1}}\right)}{1 - g\left(\frac{C_t}{C_{t+1}}\right)}\right\}.$$

After taking expectations on both sides of equations (32) and (33), we obtain:

$$E\left\{\frac{h'_2\left(\frac{C_t}{C_{t+1}}; \frac{C_{t-1}}{C_t}\right)}{\gamma \times h\left(\frac{C_t}{C_{t+1}}; \frac{C_{t-1}}{C_t}\right)}\right\} + E\left\{\frac{h'_1\left(\frac{C_t}{C_{t+1}}; \frac{C_{t-1}}{C_t}\right)}{\gamma \times h\left(\frac{C_t}{C_{t+1}}; \frac{C_{t-1}}{C_t}\right)}\right\} = E\left\{\left(\frac{C_t}{C_{t+1}}\right)^{-1}\right\},$$

hence γ is identified. We now take log on both sides of the equation (30):

$$\log h\left(\frac{C_t}{C_{t+1}}; \frac{C_{t-1}}{C_t}\right) = \log \delta - \gamma \log\left(\frac{C_{t+1}}{C_t}\right) - \gamma \log \frac{1 - g\left(\frac{C_t}{C_{t+1}}\right)}{1 - g\left(\frac{C_{t-1}}{C_t}\right)},$$

then take $E()$ on both sides, again since $\frac{C_{t-1}}{C_t}$ and $\frac{C_t}{C_{t+1}}$ have the same distribution, we have

$$E\left\{\log \frac{1 - g\left(\frac{C_t}{C_{t+1}}\right)}{1 - g\left(\frac{C_{t-1}}{C_t}\right)}\right\} = 0,$$

$$E \log h\left(\frac{C_t}{C_{t+1}}; \frac{C_{t-1}}{C_t}\right) + \gamma E \log\left(\frac{C_{t+1}}{C_t}\right) = \log \delta,$$

hence δ is identified.

Denote $\Upsilon(x) \equiv 1 - g(x)$, which should only take values in $(0, 1)$. Then equation (32) becomes:

$$(\log \Upsilon(x))' = \frac{h'_2\left(\frac{C_t}{C_{t+1}}; x\right)}{\gamma h\left(\frac{C_t}{C_{t+1}}; x\right)}$$

which can be solved for $\log \Upsilon(x)$ up to a scaling constant $\log \Upsilon(\overline{x})$ for a fixed $\overline{x}$ in the support of the distribution of $\frac{C_{t-1}}{C_t}$:

$$\log \Upsilon(x) - \log \Upsilon(\overline{x}) = \int_{\overline{x}}^{x} \frac{h'_2\left(\frac{C_t}{C_{t+1}}; y\right)}{\gamma h\left(\frac{C_t}{C_{t+1}}; y\right)} dy$$

hence $1-g(x)$ is identified up to a scaling constant $[1-g(\overline{x})]$:

$$1-g(x)=[1-g(\overline{x})]\exp\left\{\int_{\overline{x}}^{x}\frac{h_2'\left(\frac{C_t}{C_{t+1}};y\right)}{\gamma h\left(\frac{C_t}{C_{t+1}};y\right)}dy\right\}.$$

We maintain the assumptions (i) $X_t\geq 0$, and (ii) $X_t<C_t$ for C_t positive. It follows that if $C_t=0$, $X_t=0$. Hence, we have $g(0)=0$ and

$$1-g(x)=\exp\left\{\int_{0}^{x}\frac{h_2'\left(\frac{C_t}{C_{t+1}};y\right)}{\gamma h\left(\frac{C_t}{C_{t+1}};y\right)}dy\right\}.$$

3. Sieve Minimum Distance (SMD) Procedure

The sieve minimum distance (SMD) procedure has been proposed respectively in Newey and Powell (2003) for nonparametric IV regression, and in Ai and Chen (2003) for semiparametric conditional moment restrictions. Here we describe the SMD procedure in the estimation of $\boldsymbol{\alpha}_o=(\delta_o,\gamma_o,g_o)$ for the habit formation consumption-based asset pricing model (11). We assume that $\boldsymbol{\alpha}_o\in[\underline{\delta},\overline{\delta}]\times[\underline{\gamma},\overline{\gamma}]\times\mathcal{G}$, where $[\underline{\delta},\overline{\delta}]\times[\underline{\gamma},\overline{\gamma}]$ denotes the compact parameter space for the finite dimensional parameters (δ,γ), and $\mathcal{G}$ denotes the parameter space for the infinite dimensional unknown function g. In the application we assume $[\underline{\delta},\overline{\delta}]\times[\underline{\gamma},\overline{\gamma}]\subset(0,1.2]\times[0.1,100]$ for simplicity), and $g_o\in\mathcal{G}$ where

$$\mathcal{G}\equiv\left\{g\in L_2(\mathcal{X}):\int_{\mathcal{R}^L}|w||\widetilde{g}(w)|dw<\infty,\ 0\leq g<1\right\},$$

here $\mathcal{X}$ is a convex open bounded set in $\mathcal{R}^L$. This means $g\in\mathcal{G}$ if and only if it is square integrable and its Fourier transform $\widetilde{g}$ has finite first moment, where $\widetilde{g}(w)\equiv\int\exp(-iw'x)g(x)dx$ is the Fourier transform of g.

First we approximate a function $g\in\mathcal{G}$ by $g_T\in\mathcal{G}_T$, where $\mathcal{G}_T$ is the ANN sieve:

$$\mathcal{G}_T\equiv\left\{g\left(x_1,...,x_L\right)=\alpha_0+\sum_{j=1}^{K_T}\alpha_j\psi\left(\Sigma_{l=1}^{L}\gamma_{j,l}x_l+\beta_j\right),\ 0\leq g<1\right\},\tag{34}$$

which becomes dense in $\mathcal{G}$ as sample size $T\to\infty$. Then for arbitrarily fixed candidate value $\boldsymbol{\alpha}=(\delta,\gamma,g_T)\in[\underline{\delta},\overline{\delta}]\times[\underline{\gamma},\overline{\gamma}]\times\mathcal{G}_T$, we estimate the population conditional moment function:

$$m_i(\mathbf{w}_t,\boldsymbol{\alpha})\equiv E\left\{\rho_i(\mathbf{z}_{t+1},\delta,\gamma,g_T)|\mathbf{w}_t\right\},\qquad i=1,...,N$$

nonparametrically by $\widehat{m}_i(\mathbf{w}_t,\boldsymbol{\alpha})$ and denote $\widehat{m}(\mathbf{w}_t,\boldsymbol{\alpha})'=(\widehat{m}_1(\mathbf{w}_t,\boldsymbol{\alpha}),...,\widehat{m}_N(\mathbf{w}_t,\boldsymbol{\alpha}))$. Finally we estimate the δ,γ and the unknown ANN sieve coefficients jointly by a generalized version of minimal distance estimation procedure:

$$\min_{\boldsymbol{\alpha}\in[\underline{\delta},\overline{\delta}]\times[\underline{\gamma},\overline{\gamma}]\times\mathcal{G}_T}\frac{1}{T}\sum_{t=1}^{T}\widehat{m}(\mathbf{w}_t,\boldsymbol{\alpha})'\widehat{\Sigma}(\mathbf{w}_t)\widehat{m}(\mathbf{w}_t,\boldsymbol{\alpha}),\tag{35}$$

where $\widehat{\Sigma}(\mathbf{w}_t)$ is a positive definite weighting matrix that is used to take care of heteroskedasticity and serial dependence. We denote the resulting SMD estimator as $\widehat{\boldsymbol{\alpha}}_T = (\widehat{\delta}_T, \widehat{\gamma}_T, \widehat{g}_T) \in [\underline{\delta}, \overline{\delta}] \times [\underline{\gamma}, \overline{\gamma}] \times \mathcal{G}_T$.

There are many nonparametric procedures such as kernel, local linear regression, nearest neighbor and various sieve methods that can be used to estimate $m_i(\mathbf{w}_t, \boldsymbol{\alpha})$, $i = 1, ..., N$. In our application we again consider the sieve estimator. For each fixed $(\mathbf{w}_t, \boldsymbol{\alpha})$, we approximate $m_i(\mathbf{w}_t, \boldsymbol{\alpha})$ by

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

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

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

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

$$\widehat{m}_i(\mathbf{w}, \boldsymbol{\alpha}) = \sum_{t=1}^{T} \rho_i(\mathbf{z}_{t+1}, \boldsymbol{\alpha}) p^{J_T}(\mathbf{w}_t)' (\mathbf{P}'\mathbf{P})^{-1} p^{J_T}(\mathbf{w}), \quad i = 1, ..., N. \tag{36}$$

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

In general, the SMD criterion (35) can not be expressed as a GMM criterion. However, when the weighting matrix is the identity matrix $\widehat{\Sigma}(\mathbf{w}_t) = \mathbf{I}_N$ and when the nonparametric estimator $\widehat{m}_i(\mathbf{w}, \boldsymbol{\alpha})$ is the linear sieve estimator (36), the SMD criterion (35) becomes the GMM criterion (15).

4. Asymptotic Properties of the SMD Estimator $\widehat{\boldsymbol{\alpha}}_T = (\widehat{\delta}_T, \widehat{\gamma}_T, \widehat{g}_T)$

Beta-mixing:

We first introduce the concept of *beta-mixing* as a measure of temporal dependence for a time series. Let $\{\mathbf{y}_t = (\mathbf{z}_t', \widehat{cay}_t, RREL_t, SPEX_t)'\}_{t=-\infty}^{\infty}$ denote the vector time series. Let $\mathcal{I}_{-\infty}^{t}$ and $\mathcal{I}_{t+j}^{\infty}$ be sigma-fields generated respectively by $(\mathbf{y}_{-\infty}, \cdots, \mathbf{y}_t)$ and $(\mathbf{y}_{t+j}, \cdots, \mathbf{y}_{\infty})$. Define

$$\beta(j) \equiv \sup_t E \sup\{|P(B|\mathcal{I}_{-\infty}^{t}) - P(B)| : B \in \mathcal{I}_{t+j}^{\infty}\}.$$

$\{\mathbf{y}_t\}_{t=-\infty}^{\infty}$ is called *beta-mixing* if $\beta(j) \to 0$ as $j \to \infty$. For a stationary Markov process $\{Y_t\}$ with invariant measure F, the *beta*-mixing coefficients are also given by: $\beta(j) = \int \sup_{0 \le \phi \le 1} |E[\phi(Y_j)|Y_0 =$

$y] - E[\phi(Y_j)]|dF(y)$. Beta-mixing is one popular measure of temporal dependence for nonlinear time series–see the Appendix for the formal definition. It is satisfied by many widely used financial time series models including nonlinear ARX(p,q), ARCH, GARCH, stochastic volatility and diffusion models; see e.g., Doukhan (1994), Chen, Hansen, and Carrasco (2001) and Carrasco and Chen (2002). Thus, our SMD procedure requires stationary ergodic observations but does not restrict to linear time series specifications or specific parametric laws of motions of the data.

Consistency:

The consistency of the SMD estimator $\widehat{\boldsymbol{\alpha}}_T = (\widehat{\delta}_T, \widehat{\gamma}_T, \widehat{g}_T)$ can be easily obtained by applying Lemma A1 of Newey and Powell (2003), with their criterion function $\widehat{Q}(\boldsymbol{\alpha}) = \frac{1}{T}\sum_{t=1}^{T} \widehat{m}(\mathbf{w}_t, \boldsymbol{\alpha})'\widehat{m}(\mathbf{w}_t, \boldsymbol{\alpha})$, their $Q(\boldsymbol{\alpha}) = E\{m(\mathbf{w}_t, \boldsymbol{\alpha})'m(\mathbf{w}_t, \boldsymbol{\alpha})\}$, their parameter space Θ is our $[\underline{\delta}, \overline{\delta}] \times [\underline{\gamma}, \overline{\gamma}] \times \mathcal{G}$, and their sieve space $\widehat{\Theta}$ is our $[\underline{\delta}, \overline{\delta}] \times [\underline{\gamma}, \overline{\gamma}] \times \mathcal{G}_T$. Their assumption i) is satisfied with our identification result in Appendix 2. Note that our $\widehat{Q}(\boldsymbol{\alpha})$ and $Q(\boldsymbol{\alpha})$ are continuous in all the unknown parameters. To satisfy their assumption of compact parameter spaces Θ and $\widehat{\Theta}$, we can take the following function space $\mathcal{G}$ and the ANN sieve space $\mathcal{G}_T$:

$$\mathcal{G} \equiv \left\{ g \in L_2(\mathcal{X}) : \int_{\mathcal{R}^L} |w||\widetilde{g}(w)|dw \leq K < \infty,\ 0 \leq g \leq 0.999 \right\},$$

for some known big constant $K > 0$, and

$$\mathcal{G}_T \equiv \left\{ g \in \mathcal{G} : g(x_1, ..., x_L) = \alpha_0 + \sum_{j=1}^{K_T} \alpha_j \frac{\exp\left(\Sigma_{l=1}^{L} \gamma_{j,l} x_l + \beta_j\right)}{\exp\left(\Sigma_{l=1}^{L} \gamma_{j,l} x_l + \beta_j\right) + 1} \right\}.$$

Then by applying the ANN denseness result of Hornik, Stinchcombe, and White (1989), the assumption iii) of Lemma A1 in Newey and Powell (2003) is satisfied with the sup-norm:

$$\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_o\|_s = |\delta - \delta_o| + |\gamma - \gamma_o| + \sup_{x \in \mathcal{X}} |g(x) - g_o(x)|.$$

It remains to verify their uniform convergence assumption ii), which is

$$\sup_{[\underline{\delta}, \overline{\delta}] \times [\underline{\gamma}, \overline{\gamma}] \times \mathcal{G}} \left| \frac{1}{T} \sum_{t=1}^{T} \widehat{m}(\mathbf{w}_t, \boldsymbol{\alpha})'\widehat{m}(\mathbf{w}_t, \boldsymbol{\alpha}) - E\{m(\mathbf{w}_t, \boldsymbol{\alpha})'m(\mathbf{w}_t, \boldsymbol{\alpha})\} \right| = o_p(1). \tag{37}$$

This uniform convergence can be established either by applying Lemma A2 in Newey and Powell, or by showing the following two results hold:

$$(1) \quad \sup_{\mathbf{w}_t, \boldsymbol{\alpha}} |\widehat{m}(\mathbf{w}_t, \boldsymbol{\alpha}) - m(\mathbf{w}_t, \boldsymbol{\alpha})| = o_p(1);$$

and

$$(2) \quad \sup_{[\underline{\delta}, \overline{\delta}] \times [\underline{\gamma}, \overline{\gamma}] \times \mathcal{G}} \left| \frac{1}{T} \sum_{t=1}^{T} m(\mathbf{w}_t, \boldsymbol{\alpha})'m(\mathbf{w}_t, \boldsymbol{\alpha}) - E\{m(\mathbf{w}_t, \boldsymbol{\alpha})'m(\mathbf{w}_t, \boldsymbol{\alpha})\} \right| = o_p(1).$$

Result (1) can be established by modifying Lemma A.1 in Ai and Chen (2003) to allow for stationary beta-mixing data. In particular, we replace the Bernstein inequality for I.I.D. data in their proof by Lemma 1 in Chen and Shen (1998) for stationary beta-mixing data. Result (2) can be obtained by applying Lemma 1 in Chen and Shen (1998) (or any other uniform laws of large numbers) for stationary beta-mixing data.

Now by Lemma A1 in Newey and Powell (2003), we obtain: $\|\widehat{\boldsymbol{\alpha}}_T - \boldsymbol{\alpha}_o\|_s = o_p(1)$.

Convergence rate:

For any $\boldsymbol{\alpha} \in [\underline{\delta}, \overline{\delta}] \times [\underline{\gamma}, \overline{\gamma}] \times \mathcal{G}$, let $\{\boldsymbol{\alpha}(\tau) : \tau \in [0,1]\}$ be a continuous path in $[\underline{\delta}, \overline{\delta}] \times [\underline{\gamma}, \overline{\gamma}] \times \mathcal{G}$ such that $\boldsymbol{\alpha}(0) = \boldsymbol{\alpha}_o$ and $\boldsymbol{\alpha}(1) = \boldsymbol{\alpha}$. Suppose that for almost all $\mathbf{z}_{t+1}$, $\rho(\mathbf{z}_{t+1}, \boldsymbol{\alpha}(\tau))$ is continuously differentiable at $\tau = 0$. Denote the first pathwise derivative at the direction $[\boldsymbol{\alpha} - \boldsymbol{\alpha}_o]$ evaluated at $\boldsymbol{\alpha}_o$ by:

$$\frac{d\rho(\mathbf{z}_{t+1}, \boldsymbol{\alpha}_o)}{d\boldsymbol{\alpha}}[\boldsymbol{\alpha} - \boldsymbol{\alpha}_o] \equiv \frac{d(\mathbf{z}_{t+1}, \boldsymbol{\alpha}(\tau))}{d\tau}|_{\tau=0} \quad a.s.\ \mathbf{z}_{t+1}$$

and denote $\frac{dm(\mathbf{w}_t, \boldsymbol{\alpha}_o)}{d\boldsymbol{\alpha}}[\boldsymbol{\alpha} - \boldsymbol{\alpha}_o] \equiv E\{\frac{d\rho(\mathbf{z}_{t+1}, \boldsymbol{\alpha}_o)}{d\boldsymbol{\alpha}}[\boldsymbol{\alpha} - \boldsymbol{\alpha}_o] \mid \mathbf{w}_t\}$. For any $\boldsymbol{\alpha} \in [\underline{\delta}, \overline{\delta}] \times [\underline{\gamma}, \overline{\gamma}] \times \mathcal{G}$ we define the following pseudo metric:

$$||\boldsymbol{\alpha} - \boldsymbol{\alpha}_o|| \equiv \sqrt{E\left\{\{\frac{dm(\mathbf{w}_t, \boldsymbol{\alpha}_o)}{d\boldsymbol{\alpha}}[\boldsymbol{\alpha} - \boldsymbol{\alpha}_o]\}'\{\frac{dm(\mathbf{w}_t, \boldsymbol{\alpha}_o)}{d\boldsymbol{\alpha}}[\boldsymbol{\alpha} - \boldsymbol{\alpha}_o]\}\right\}}.$$

Then under assumptions similar to those for Theorem 3.1 in Ai and Chen (2003), we have: $||\widehat{\boldsymbol{\alpha}}_T - \boldsymbol{\alpha}_o|| = o_p(T^{-1/4})$.

This rate result can be proved by slightly modifying the proof of Theorem 3.1 in Ai and Chen (2003), that is, we simply replace several parts in their proof that rely on I.I.D. data by the corresponding ones for stationary beta-mixing data. In particular, their key Lemma A.1 can be established for stationary beta-mixing data by using Lemma 1 in Chen and Shen (1998).

Root-T asymptotic normality of $\widehat{\delta}$, $\widehat{\gamma}$:

Define $\omega^* = (\omega_\delta^*, \omega_\gamma^*)$ with

$$\omega_\delta^* = \arg\min_{\omega_\delta \in \mathcal{G}} E\left\{\left(\frac{dm(\mathbf{w}_t, \boldsymbol{\alpha}_o)}{d\delta} - \frac{dm(\mathbf{w}_t, \boldsymbol{\alpha}_o)}{dg}[\omega_\delta]\right)'\left(\frac{dm(\mathbf{w}_t, \boldsymbol{\alpha}_o)}{d\delta} - \frac{dm(\mathbf{w}_t, \boldsymbol{\alpha}_o)}{dg}[\omega_\delta]\right)\right\},$$

$$\omega_\gamma^* = \arg\min_{\omega_\gamma \in \mathcal{G}} E\left\{\left(\frac{dm(\mathbf{w}_t, \boldsymbol{\alpha}_o)}{d\gamma} - \frac{dm(\mathbf{w}_t, \boldsymbol{\alpha}_o)}{dg}[\omega_\gamma]\right)'\left(\frac{dm(\mathbf{w}_t, \boldsymbol{\alpha}_o)}{d\gamma} - \frac{dm(\mathbf{w}_t, \boldsymbol{\alpha}_o)}{dg}[\omega_\gamma]\right)\right\}.$$

Denote

$$D_{\omega^*}(\mathbf{w}_t) = \left(\frac{dm(\mathbf{w}_t, \boldsymbol{\alpha}_o)}{d\delta}, \frac{dm(\mathbf{w}_t, \boldsymbol{\alpha}_o)}{d\gamma}\right) - \left(\frac{dm(\mathbf{w}_t, \boldsymbol{\alpha}_o)}{dg}[\omega_\delta^*], \frac{dm(\mathbf{w}_t, \boldsymbol{\alpha}_o)}{dg}[\omega_\gamma^*]\right).$$

Assumption N. (i) $E[D_{\omega^*}(\mathbf{w}_t)'D_{\omega^*}(\mathbf{w}_t)]$ is positive-definite; (ii) $\delta_o \in (\underline{\delta}, \overline{\delta})$ and $\gamma_o \in (\underline{\gamma}, \overline{\gamma})$; (iii) $\Omega_o(\mathbf{w}) \equiv Var[\rho(\mathbf{z}_{t+1}, \boldsymbol{\alpha}_o)|\mathbf{w}_t = \mathbf{w}]$ is positive definite for all $\mathbf{w}$ in the support of $\mathbf{w}_t$; (iv) $||\widehat{\boldsymbol{\alpha}}_T - \boldsymbol{\alpha}_o|| = o_p(T^{-1/4})$.

Under Assumption N and other regularity conditions similar to those for Theorem 4.1 in Ai and Chen (2003), we obtain $\sqrt{T}\left(\widehat{\delta} - \delta_o, \widehat{\gamma} - \gamma_o\right)' \longrightarrow^D \mathcal{N}(0, V^{-1})$ with

$$V = E[D_{\omega^*}(\mathbf{w}_t)'D_{\omega^*}(\mathbf{w}_t)]\{E[D_{\omega^*}(\mathbf{w}_t)'\Omega_o(\mathbf{w}_t)D_{\omega^*}(\mathbf{w}_t)]\}^{-1}E\{D_{\omega^*}(\mathbf{w}_t)'D_{\omega^*}(\mathbf{w}_t)\}.$$

This result can be proved by slightly modifying the proof of Theorem 4.1 in Ai and Chen (2003), that is, we simply replace several parts in their proof that rely on I.I.D. data by the corresponding ones for stationary beta-mixing data.

5. Limiting Distributions of the Test Statistics

Root-T asymptotic normality of $\widehat{\beta}$ for testing internal vs. external habit

Recall that the pseudo-true value $\boldsymbol{\alpha}^* = (\delta^*, \gamma^*, \beta^*, g^*(\cdot))'$ solves the following conditional moment restrictions:

$$E\left(\delta^*\left(\frac{C_{t+1}}{C_t}\frac{1 - g^*\left(\frac{C_t}{C_{t+1}}, ..., \frac{C_{t+1-L}}{C_{t+1}}\right)}{1 - g^*\left(\frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t}\right)}\right)^{-\gamma^*} R_{i,t+1}\widetilde{F}_{i,t+1}(\beta^*; \delta^*, \gamma^*, g^*) - 1 \mid \mathbf{w}_t\right) = 0, \quad i = 1, ..., N,$$

with $\widetilde{F}_{i,t+1}(\beta; \delta, \gamma, g)$ defined in (25).

These unknown pseudo-true values $(\delta^*, \gamma^*, \beta^*, g^*(\cdot))$ can again be estimated by the SMD method using the same ANN sieve (19) to approximate the unknown $g^*(\cdot)$, the same three asset groups with the associated sets of instruments. All we need to do is to redefine $\boldsymbol{\alpha} = (\delta, \gamma, \beta, g(\cdot))'$ and $\rho(\mathbf{z}_{t+1}, \boldsymbol{\alpha}) = (\rho_1(\mathbf{z}_{t+1}, \boldsymbol{\alpha}), ..., \rho_N(\mathbf{z}_{t+1}, \boldsymbol{\alpha}))'$ with

$$\rho_i(\mathbf{z}_{t+1}, \boldsymbol{\alpha}) = \delta\left(\frac{C_{t+1}}{C_t}\frac{1 - g\left(\frac{C_t}{C_{t+1}}, ..., \frac{C_{t+1-L}}{C_{t+1}}\right)}{1 - g\left(\frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t}\right)}\right)^{-\gamma} R_{i,t+1}\widetilde{F}_{i,t+1}(\beta; \delta, \gamma, g) - 1$$

in the sieve LS estimation (14) of $m_i(\mathbf{w}_t, \boldsymbol{\alpha})$ and $m(\mathbf{w}_t, \boldsymbol{\alpha}) = (m_1(\mathbf{w}_t, \boldsymbol{\alpha}), ..., m_N(\mathbf{w}_t, \boldsymbol{\alpha}))'$.

Denote $\boldsymbol{\alpha}^* = (\delta^*, \gamma^*, \beta^*, g^*(\cdot))'$ and let $\widehat{\boldsymbol{\alpha}}_T = (\widehat{\delta}_T, \widehat{\gamma}_T, \widehat{\beta}_T, \widehat{g}_T)$ be the solution to

$$\min_{\boldsymbol{\alpha} \in [\underline{\delta}, \overline{\delta}] \times [\underline{\gamma}, \overline{\gamma}] \times [\underline{\beta}, \overline{\beta}] \times \mathcal{G}_T} \frac{1}{T}\sum_{t=1}^{T} \widehat{m}(\mathbf{w}_t, \boldsymbol{\alpha})'\widehat{m}(\mathbf{w}_t, \boldsymbol{\alpha}).$$

Let $\omega^* = (\omega^*_\delta, \omega^*_\gamma, \omega^*_\beta)$ with

$$\omega^*_\beta = \arg\min_{\omega_\beta \in \mathcal{G}} E\left\{\left(\frac{dm(\mathbf{w}_t, \boldsymbol{\alpha}^*)}{d\beta} - \frac{dm(\mathbf{w}_t, \boldsymbol{\alpha}^*)}{dg}[\omega_\beta]\right)'\left(\frac{dm(\mathbf{w}_t, \boldsymbol{\alpha}^*)}{d\beta} - \frac{dm(\mathbf{w}_t, \boldsymbol{\alpha}^*)}{dg}[\omega_\beta]\right)\right\},$$

and ω_δ^*, ω_γ^* defined similarly as those in Appendix 4 (above) except replacing $\boldsymbol{\alpha}_o$ by $\boldsymbol{\alpha}^*$. Denote $\frac{dm(\mathbf{w}_t,\boldsymbol{\alpha}^*)}{dg}[\omega^*] = (\frac{dm(\mathbf{w}_t,\boldsymbol{\alpha}^*)}{dg}[\omega_\delta^*], \frac{dm(\mathbf{w}_t,\boldsymbol{\alpha}^*)}{dg}[\omega_\gamma^*], \frac{dm(\mathbf{w}_t,\boldsymbol{\alpha}^*)}{dg}[\omega_\beta^*])$,

$$D_{\omega^*}(\mathbf{w}_t) = \left(\frac{dm(\mathbf{w}_t,\boldsymbol{\alpha}^*)}{d\delta}, \frac{dm(\mathbf{w}_t,\boldsymbol{\alpha}^*)}{d\gamma}, \frac{dm(\mathbf{w}_t,\boldsymbol{\alpha}^*)}{d\beta}\right) - \frac{dm(\mathbf{w}_t,\boldsymbol{\alpha}^*)}{dg}[\omega^*],$$

and $\Omega_*(\mathbf{w}_t) = Var[\rho(\mathbf{z}_{t+1},\boldsymbol{\alpha}^*)|\mathbf{w}_t]$. Finally let

$$V_* = E[D_{\omega^*}(\mathbf{w}_t)'D_{\omega^*}(\mathbf{w}_t)]\{E[D_{\omega^*}(\mathbf{w}_t)'\Omega_*(\mathbf{w}_t)D_{\omega^*}(\mathbf{w}_t)]\}^{-1}E\{D_{\omega^*}(\mathbf{w}_t)'D_{\omega^*}(\mathbf{w}_t)\}.$$

Then under assumptions similar to those suggested in Appendix 4, we obtain:

$$\sqrt{T}\left(\widehat{\delta}-\delta^*, \widehat{\gamma}-\gamma^*, \widehat{\beta}-\beta^*\right)' \longrightarrow^D \mathcal{N}(0, V_*^{-1}).$$

Root-T asymptotic normality of $\widehat{\mu}_g$ for testing linear habit

Recall that

$$\widehat{\mu}_g = \frac{1}{T}\sum_{t=L}^{T}\widehat{g}_{11}(\frac{C_t}{C_{t+1}}, ..., \frac{C_{t+1-L}}{C_{t+1}}),$$

where $\widehat{g}_{11}()$ is the second partial derivative of the SMD estimator $\widehat{g}()$ with respect to its first argument. Let $f()$ denote the true unknown probability density of $\left(\frac{C_t}{C_{t+1}}, ..., \frac{C_{t+1-L}}{C_{t+1}}\right)$ and $g_{11}^o()$ denote the second partial derivative of the true $g_o()$ with respect to its first argument. Then

$$\mu_g = \int g_{11}^o(z_1, z_2, ..., z_L)f(z_1, z_2, ..., z_L)dz.$$

Suppose that $f()$ is at least twice continuously differentiable with respect to its first argument, where $f_1()$ and $f_{11}()$ denote the first and second partial derivatives of $f()$ with respect to its first argument. Also assume that $f()$ and $f_1()$ go to zero smoothly as their first argument goes to the boundaries. Under some mild additional conditions and using a slight modification of Example 2.2 in Ai and Chen (2005), we obtain:

$$\begin{aligned}
&\sqrt{T}\left(\widehat{\mu}_g - \mu_g\right) \\
= \;& \frac{1}{\sqrt{T}}\sum_{t=L}^{T}\left(g_{11}^o(\frac{C_t}{C_{t+1}}, ..., \frac{C_{t+1-L}}{C_{t+1}}) - E\{g_{11}^o(\frac{C_t}{C_{t+1}}, ..., \frac{C_{t+1-L}}{C_{t+1}})\}\right) \\
& + E\left[\left\{\widehat{g}(\frac{C_t}{C_{t+1}}, ..., \frac{C_{t+1-L}}{C_{t+1}}) - g_o(\frac{C_t}{C_{t+1}}, ..., \frac{C_{t+1-L}}{C_{t+1}})\right\}\frac{f_{11}(\frac{C_t}{C_{t+1}}, ..., \frac{C_{t+1-L}}{C_{t+1}})}{f(\frac{C_t}{C_{t+1}}, ..., \frac{C_{t+1-L}}{C_{t+1}})}\right] + o_p(1),
\end{aligned}$$

and $\sqrt{T}\left(\widehat{\mu}_g - \mu_g\right) \longrightarrow^D \mathcal{N}(0, \sigma_{11}^2)$. Unfortunately, the limiting variance σ_{11}^2 is of a complicated form, which motivates our use of bootstrap approximation of the asymptotic distribution of $\sqrt{T}\left(\widehat{\mu}_g - \mu_g\right)$.

References

Abel, A. B. (1990): "Asset Prices Under Habit Formation and Catching Up With the Jones," *American Economic Review*, 80, 38–42.

——— (1999): "Risk Premia and Term Premia in General Equilibrium," *Journal of Monetary Economics*, 43, 3–33.

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

——— (2005): "Estimation of Possibly Misspecified Semiparametric Conditional Moment Restriction Models with Different Conditioning Variables," forthcoming in Journal of Econometrics.

Bansal, R., D. A. Hsieh, and S. Viswanathan (1993): "A New Approach to International Arbitrage Pricing," *The Journal of Finance*, 48, 1719–1747.

Bansal, R., and S. Viswanathan (1993): "No Arbitrage and Arbitrage Pricing: A New Approach," *The Journal of Finance*, 48(4), 1231–1262.

Boldrin, M., L. J. Christiano, and J. D. M. Fisher (2001): "Habit Persistence, Asset Returns and the Business Cycle," *American Economic Review*, 91(1), 149–166.

Breeden, D. (1979): "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," *Journal of Financial Economics*, 7, 265–296.

Breeden, D., and R. Litzenberger (1978): "Prices of State-Contingent Claims Implicit in Option Prices," *Journal of Business*, 51, 621–651.

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

Campbell, J. Y., and J. H. Cochrane (1999): "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," *Journal of Political Economy*, 107, 205–251.

——— (2000): "Explaining the Poor Performance of Consumption-Based Asset Pricing Models," *Journal of Finance*, 55(6), 2863–2878.

Campbell, J. Y., A. W. Lo, and C. MacKinlay (1997): *The Econometrics of Financial Markets*. Princeton University Press, Princeton, NJ.

Campbell, J. Y., and R. J. Shiller (1989): "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies*, 1(3), 195–228.

Carrasco, M., and X. Chen (2002): "Mixing and Moment Properties of Various GARCH and Stochastic Volatility Models," *Economic Theory*, 18, 17–39.

Chapman, D. A. (1997): "Approximating the Asset Pricing Kernel," *Journal of Finance*, 52(4), 1383–1410.

——— (1998): "Habit Formation and Aggregate Consumption," *Econometrica*, 66(5), 1223–1230.

Chen, X., L. P. Hansen, and M. Carrasco (2001): "Nonlinearity and Temporal Dependence," Unpublished paper, New York University.

Chen, X., and X. Shen (1998): "Sieve Extremum Estimates For Weakly Dependent Data," *Econometrica*, 66(2), 289–314.

Chen, X., and H. White (1999): "Improved Rates of Asymptotic Normality for Nonparametric Neural Network Estimators," *IEEE Information Theory*, 45, 682–691.

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

Constantinides, G. M. (1990): "Habit-formation: A Resolution of the Equity Premium Puzzle," *Journal of Political Economy*, 98, 519–543.

Corradi, V., and N. Swanson (2003): "A Test for Comparing Multiple Misspecified Conditional Distributions," Unpublished paper, Rutgers University.

Dai, Q. (2003): "Term Structure Dynamics in a Model with Stochastic Internal Habit," Unpublished paper, Stern School of Business, New York University.

Doukhan, P. (1994): *Mixing: Properites and Examples*. Springer-Verlag, New York.

Dynan, K. E. (2000): "Habit Formation in Consumer Preferences: Evidence from Panel Data," *American Economic Review*, 90, 391–406.

Fama, E. F., and K. R. French (1988): "Dividend Yields and Expected Stock Returns," *Journal of Financial Economics*, 22, 3–27.

——— (1992): "The Cross-Section of Expected Returns," *Journal of Finance*, 47, 427–465.

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

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

Ferson, W. E., and G. M. Constantinides (1991): "Habit Persistence and Durability in Aggregate Consumption," *Journal of Financial Economics*, 29, 199–240.

Ferson, W. E., and C. R. Harvey (1992): "Seasonality and Consumption-Based Asset Pricing," *Journal of Finance*, 47, 511–552.

Gallant, A. R., L. P. Hansen, and G. Tauchen (1990): "Using Conditional Moments of Asset Payoffs to infer the Volatility of Intertemporal Marginal Rates of Subsitution," *Journal of Econometrics*, 45(2), 141–179.

Gallant, A. R., and G. Tauchen (1989): "Seminonparametric Estimation of Conditionally Constrained Heterogeneous Processes: Asset Pricing Applications," *Econometrica*, 57(5), 1091–1120.

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

Hall, P., J. L. Horowitz, and B. Y. Jing (1995): "On Blocking Rules for the Bootstrap with Dependent Data," *Biometrika*, 82, 561–574.

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

Hansen, L. P., J. Heaton, and N. Li (2005): "Consumption Strikes Back?: Measuring Long-run Risk," Unpublished Paper, University of Chicago.

Hansen, L. P., J. Heaton, and E. G. Luttmer (1995): "Econometric Evaluation of Asset Pricing Models," *The Review of Financial Studies*, 8(2), 238–274.

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

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

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

Hansen, P. R. (2003): "A Test for Superior Predictive Ability," Unpublished paper, Brown University.

Harvey, C. R. (1991): "The World Price of Covariance Risk," *Journal of Finance*, 46, 111–117.

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

——— (1995): "An Empirical Investigation of Asset Pricing with Temporally Dependent Preference Specifications," *Econometrica*, 63, 681–717.

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

Hornik, K., M. Stinchcombe, and H. White (1989): "Multi-layer Feedforward Networks are Universal Approximators," *Neural Networks*, 2, 359–366.

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

Jermann, U. (1998): "Asset Pricing in Production Economies," *Journal of Monetary Economics*, 41(2), 257–275.

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

Lamont, O. (1998): "Earnings and Expected Returns," *Journal of Finance*, 53, 1563–87.

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

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

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

Li, Y. (2001): "Expected Returns and Habit Persistence," *Review of Financial Studies*, 14(3), 861–899.

Lintner, J. (1965): "Security Prices, Risk and Maximal Gains from Diversification," *Journal of Finance*, 20, 587–615.

Ljungqvist, L., and H. Uhlig (2000): "Tax Policy and Aggregate Demand Management Under Catching Up With the Jonses," *American Economic Review*, 90(3), 356–366.

Menzly, L., T. Santos, and P. Veronesi (2004): "Understanding Predictability," *Journal of Political Economy*, 112(1), 1–47.

Newey, W. K., and J. Powell (1988): "Instrumental Variable Estimation of Nonparametric Models," Working Paper Version of Published Paper, MIT.

——— (2003): "Instrumental Variable Estimation of Nonparametric Models," *Econometrica*, 71, 1557–1569.

Parker, J., and C. Julliard (2004): "Consumption Risk and the Cross-Section of Expected Returns," *Journal of Political Economy*, 113(1), 185–222.

Sharpe, W. (1964): "Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk," *Journal of Finance*, 19, 425–444.

Shiller, R. J. (1981): "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?," *American Economic Review*, 71, 421–436.

Shore, S. H., and J. S. White (2002): "External Habit Formation and the Home Bias Puzzle," Unpublished paper, Harvard University.

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

Sundaresan, S. (1989): "Intertemporally Dependent Preferences and the Volatility of Consumption and Wealth," *The Review of Financial Studies*, 2, 73–89.

Wachter, J. (2006): "A Consumption Based Model of the Term Structure of Interest Rates," *Journal of Financial Economics*, 79, 365–399.

Wang, Z., and X. Zhang (2003): "The Role of Artibtrage in the Empirical Evaluation of Asset Pricing Models," Unpublished paper, Cornell University.

White, H. (2000): "A Reality Check for Data Snooping," *Econometrica*, 68(5), 1097–1126.

10 Tables

Table 1

SMD Estimates of δ and γ

Assets	Instruments	δ	γ
Group 1	1, $\mathbf{w}_t$, squared terms	0.9850	0.757
		(0.005)	(0.107)
Group 2	1, $\mathbf{w}_t$, squared, cross terms	0.9875	0.789
		(0.005)	(0.077)
Group 3	1, $\mathbf{w}_t$	0.9847	0.811
		(0.006)	(0.149)

Notes: The table reports SMD parameter estimates and asymptotic standard errors in parentheses.

Table 2

Specification Errors for Alternative Models: HJ

	6 size/BM	6 + T-bill
Model	HJ Dist	HJ Dist
(1)	(2)	(3)
Internal Habit	0.172	0.179
External Habit	0.261	0.425
Fama-French	0.262	0.282
Scaled CCAPM	0.208	0.352
CCAPM	0.307	0.403
CAPM	0.339	0.416

Notes: For each model labeled in column 1, the table reports the Hansen-Jagannathan distance ("HJ Dist") evaluated on equity returns alone (column 2) or equity returns plus Treasury bill rate (column 3).

Table 3

Specification Errors for Alternative Models: HJ^+

	6 size/BM	6 + T-bill
Model	HJ^+ Dist	HJ^+ Dist
(1)	(2)	(3)
Internal Habit	0.177	0.180
External Habit	0.289	0.455
Fama-French	0.262	0.287
Scaled CCAPM	0.810	0.601
CCAPM	0.372	0.618
CAPM	0.340	0.418

Notes: For each model in column 1, "HJ^+ Dist" is the distance between the model proxy and the family of admissible nonnegative stochastic discount factors. In column 2, test assets are equity returns; in 3, test assets are equity returns plus T-bill rate.

Table 4

Covariances of SDFs with Returns

	Average Excess Return	Average Return
Tbill		1.0035
S1B1	0.0177	1.0213
S1B2	0.0274	1.0309
S1B3	0.0322	1.0357
S2B1	0.0185	1.0221
S2B2	0.0195	1.0230
S2B3	0.0260	1.0295
	SDF internal habit	SDF external habit
	Covariance	
S1B1	-0.0189	-0.0103
S1B2	-0.0248	-0.0096
S1B3	-0.0292	-0.0100
S2B1	-0.0200	-0.0074
S2B2	-0.0182	-0.0039
S2B3	-0.0272	-0.0083

Notes : This table reports average returns for the portfolios in the left column, and covariance of the stochastic discount factors of internal and external habit models with each excess return. The sample spans the period 1952:Q4 -2001:Q2.

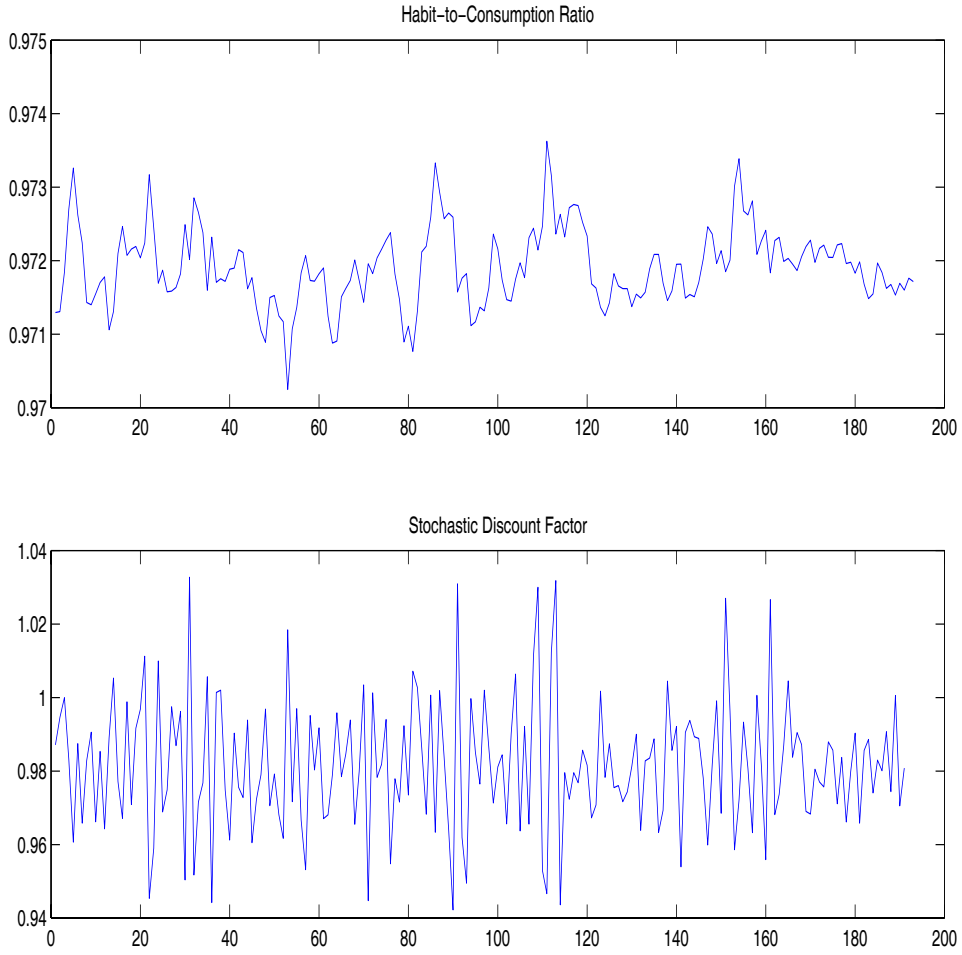
Table 5

Pricing Errors for 6 Size/BM Returns

Return	Fama-French	Internal Habit
	HJ Dist = 0.26	HJ Dist = 0.172
S1B1	-0.0031	-0.0052
S1B2	0.0005	-0.0001
S1B3	0.0009	-0.0005
S2B1	0.0026	-0.0041
S2B2	-0.0029	-0.0008
S2B3	-0.0015	-0.0027

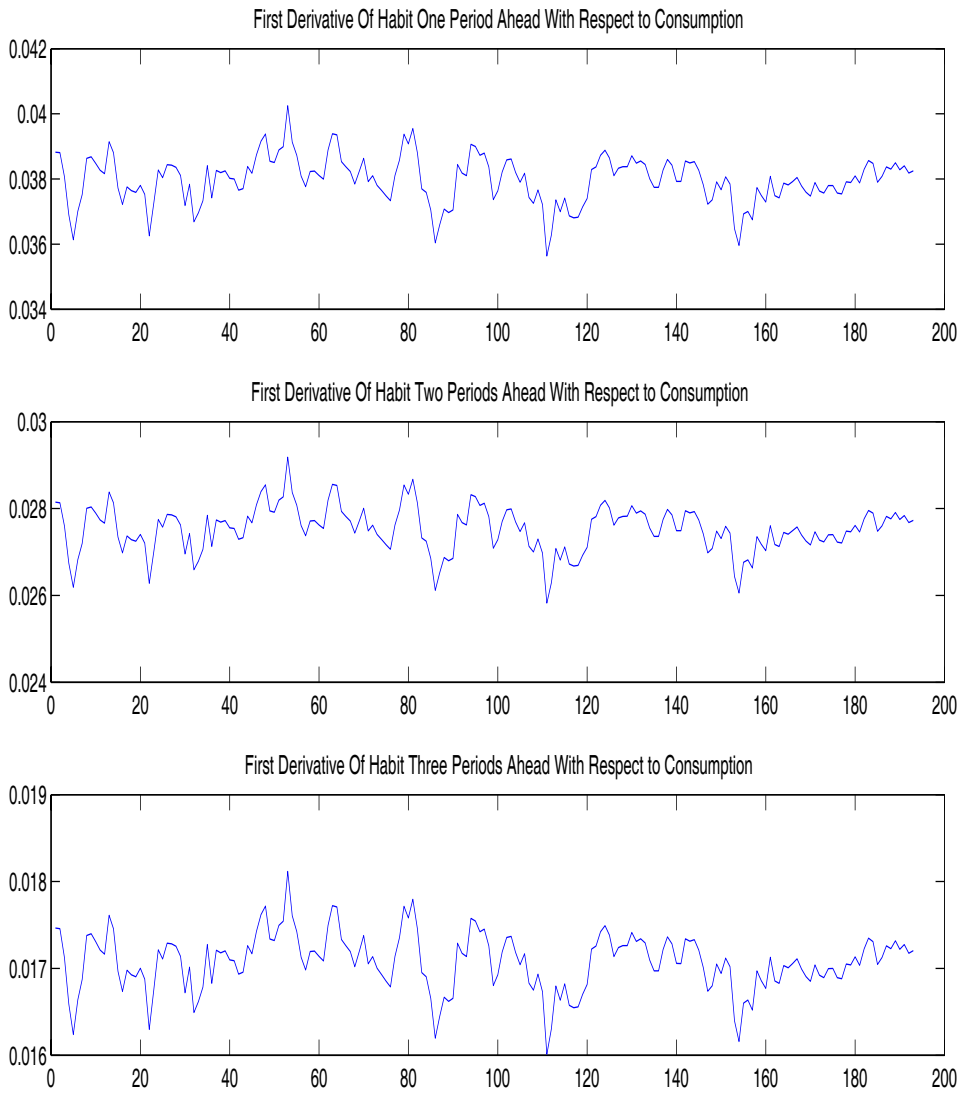
Notes: This Table reports pricing errors from the Hansen-Jagannathan minimization using six size/book-market portfolio equity returns. Errors for the Fama-French model and internal habit model are reported. The internal habit model uses SMD estimates on Group 2 assets for g.

Figure 1



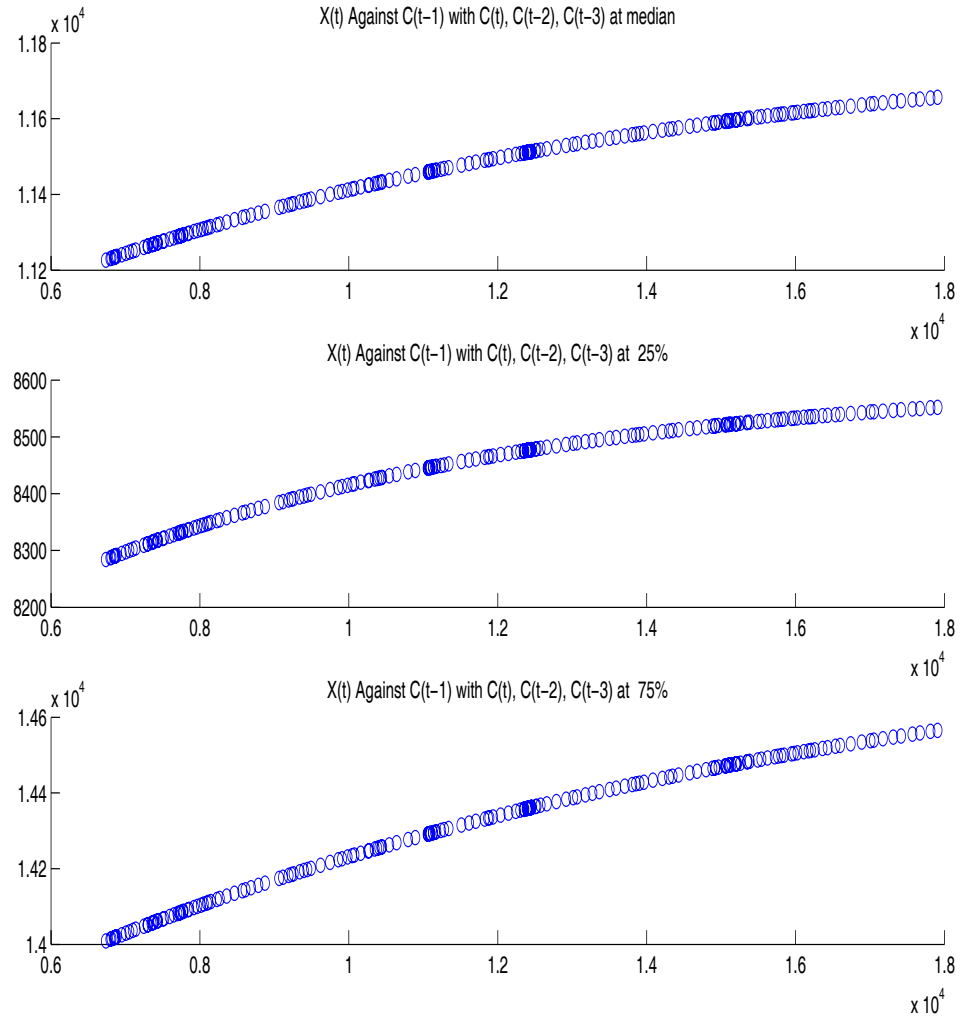
Notes: This figure plots the estimated habit-consumption ratio (top panel) and estimated stochastic discount factor (bottom panel) using Group 1 assets, linear and squared instruments.

Figure 2



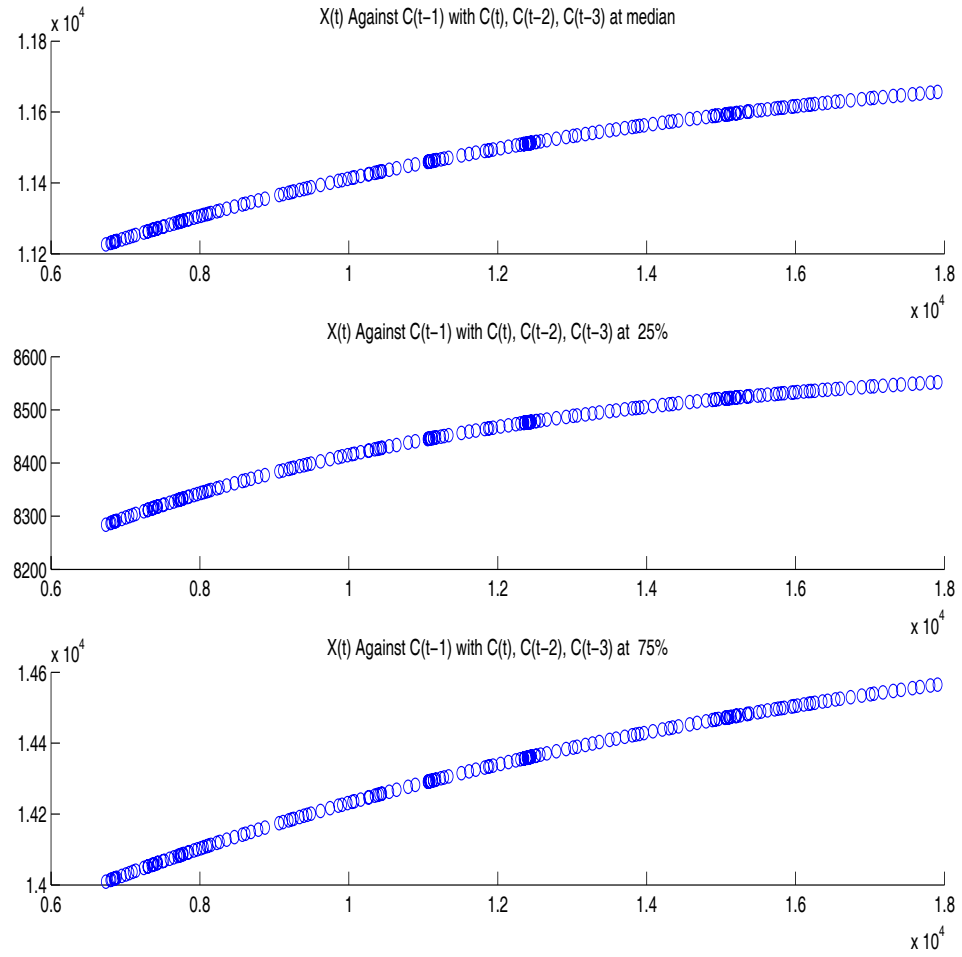
Notes: Estimates use Group 1 assets, linear and squared instruments.

Figure 3



Notes: $X(t)$ is the estimated habit, $C(t)$ is consumption. Estimates use Group 1 assets, linear and squared instruments.

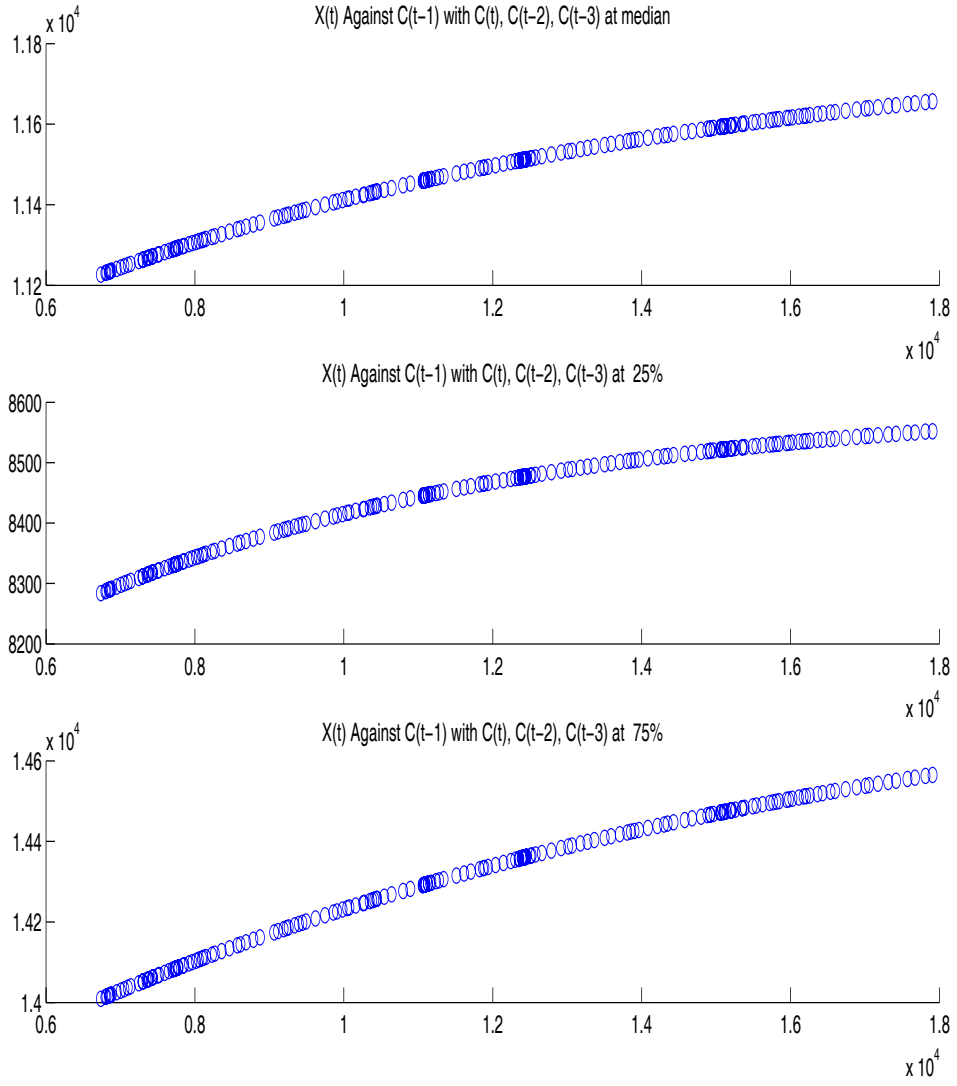
Figure 4



Notes: $X(t)$ is the estimated habit, $C(t)$ is consumption. Estimates use Group 2 assets, linear, squared and cross term instruments.

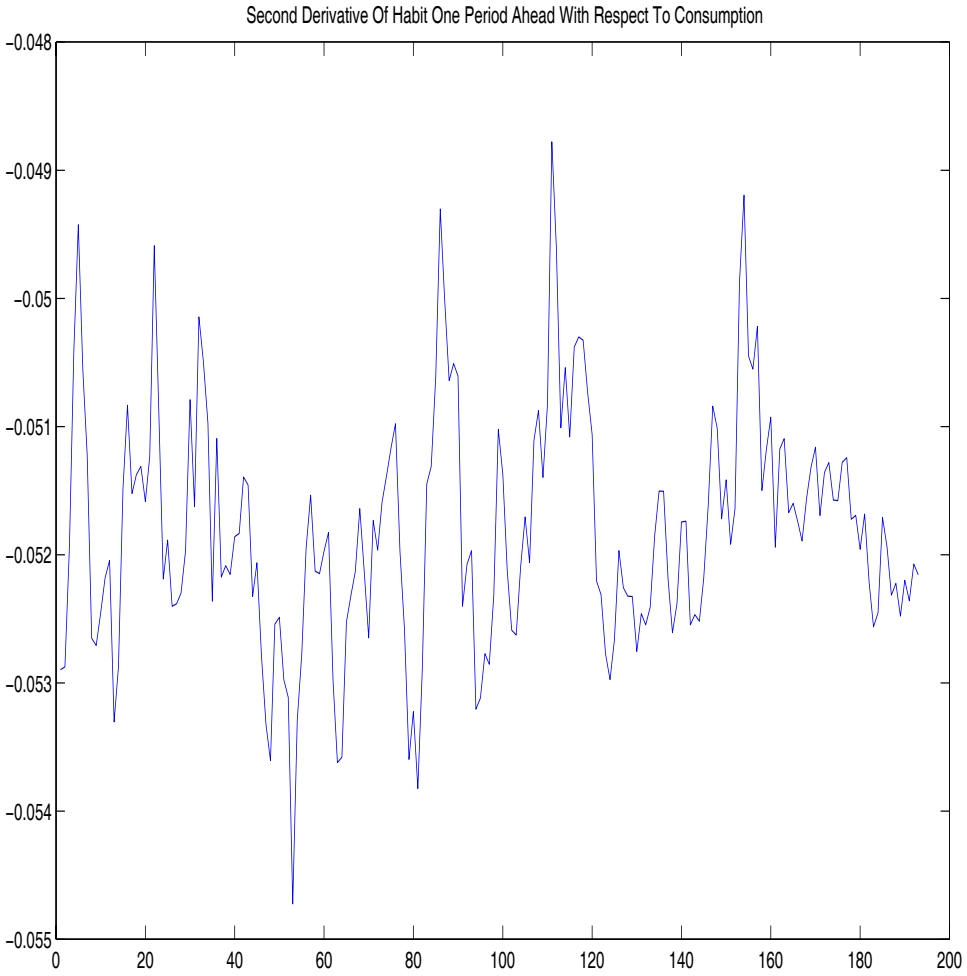
Figure 5

NoteNoat



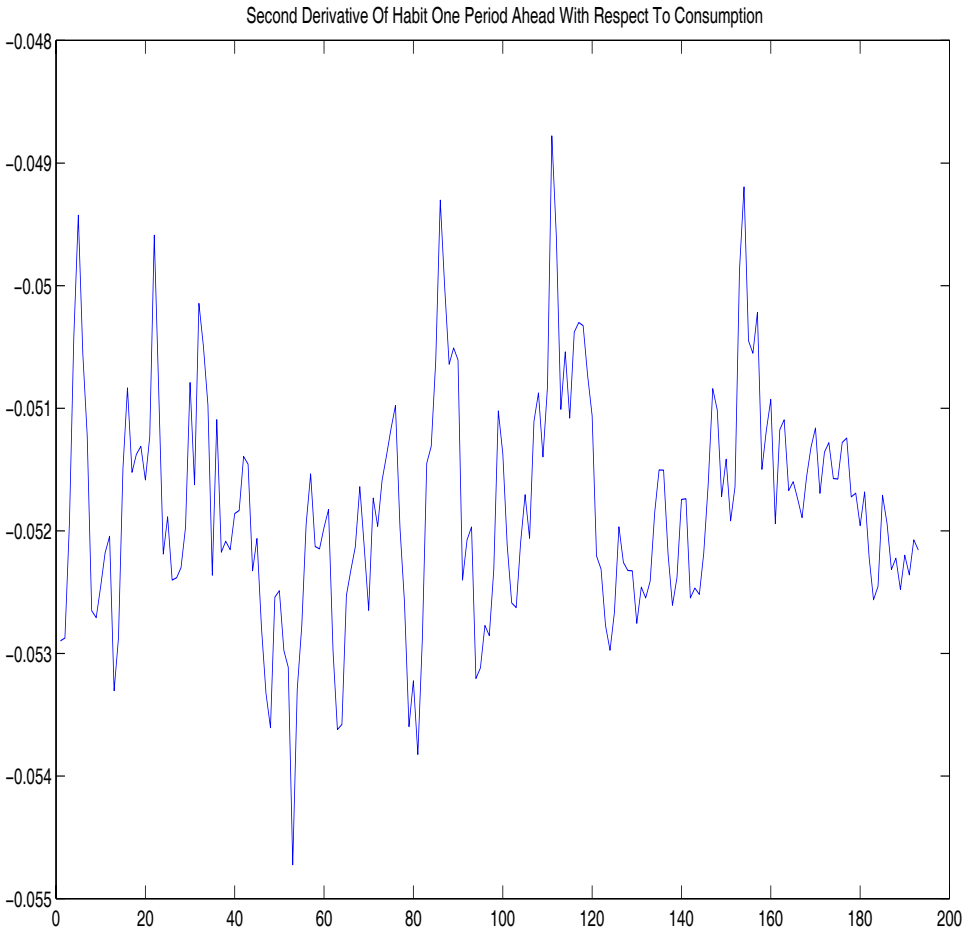
Notes: $X(t)$ is the estimated habit, $C(t)$ is consumption. Estimates use Group 3 assets, linear instruments.

Figure 6



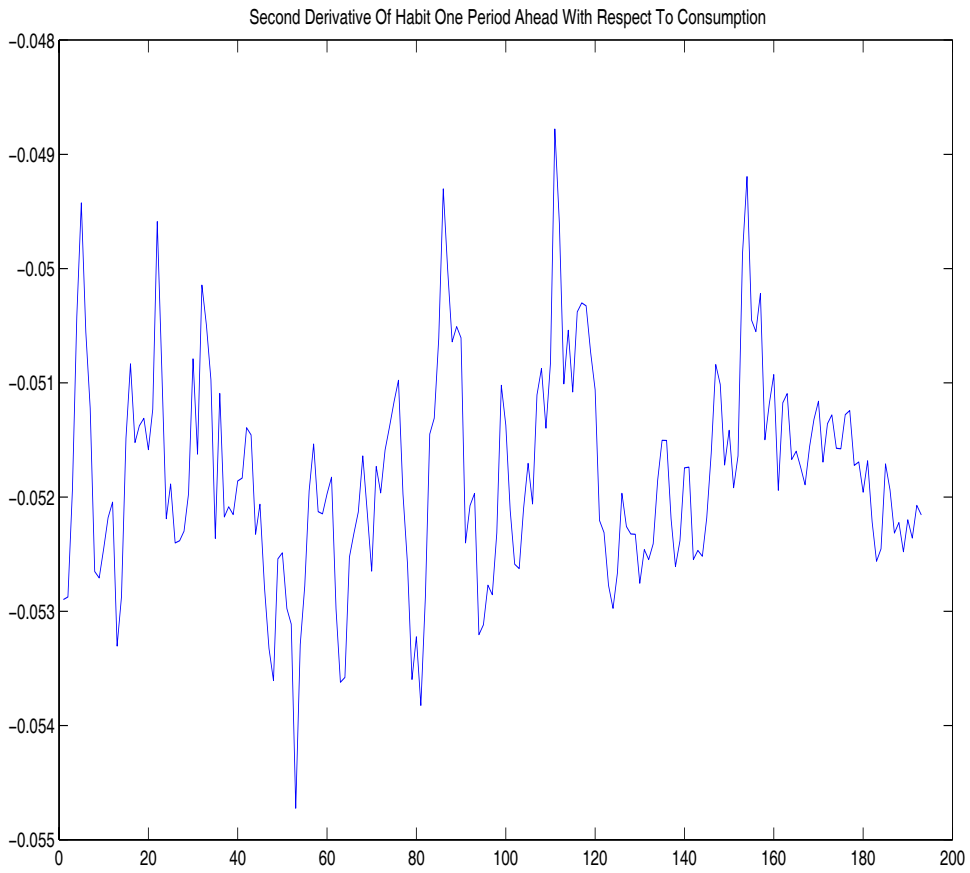
Notes: Estimates using Group 1 assets, linear and squared instruments.

Figure 7



Notes: Estimates using Group 2 assets, linear, squared and cross term instruments.

Figure 8



Notes: Estimates using Group 3 assets, linear instruments.