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

Updated versions of this paper will be posted at http://www.econ.nyu.edu/user/ludvigsons/. Ludvigson acknowledges financial support from the Alfred P. Sloan Foundation. Both authors acknowledge financial support form 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, Chris Flinn, Per Krusell, Lars Hansen, Anthony Lynch, Monika Piazzesi, Masao Ogaki, B. Ravikumar, Martin Schneider, Jim Stock, George Tauchen, Zhenyu Wang and seminar participants at Carnegie Mellon GSIA, Notre Dame, NYU, the University of Chicago, the University of Pennsylvania, the 2004 CIREQ-CIRANO Financial Econometrics Conference at Montreal, the 2004 NBER Spring Asset Pricing meeting, the 2003 NBER Summer Institute, the 2004 American Finance Association Meetings, and the 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. Our approach is to treat the functional form of the habit as unknown, and to estimate it along with the rest of the model's finite dimensional parameters. This semiparametric approach allows us to empirically evaluate a number of interesting hypotheses about the specification of habit-based asset pricing models. Using stationary quarterly data on consumption growth, assets returns and instruments, our empirical results indicate that the estimated habit function is nonlinear, the habit formation is internal, and the estimated time-preference parameter and the power utility parameter are sensible. In addition, our 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

Over the last fifteen years, academics interested in asset pricing have witnessed an explosion of theoretical research aimed at explaining the behavior of expected stock market returns, both in the times series and the cross-section. There are several competing classes of theories, ranging from explanations based on idiosyncratic income shocks, incomplete markets and borrowing constraints,[1] to those based on limited stock market participation,[2] heterogeneity in preferences,[3] nonseparable utility between durable and nondurable consumption,[4] and irrational expectations.[5] Yet a comprehensive survey of this literature reveals that a leading and increasingly pervasive explanation of aggregate stock market behavior is one based on investor preferences. This strand of the literature argues 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, where the habit is a function of lagged and (possibly) contemporaneous consumption.[6]

Given the plethora of competing theories, it would seem important to find some way of empirically evaluating the representative-agent, habit-based asset pricing framework as an explanation for aggregate stock market behavior. For the most part, these models have been "tested" by undertaking a calibration exercise, and then asking whether the calibrated model is capable of matching a select set of asset pricing moments computed from data. Although such exercises are undoubtedly useful as an initial step in the evaluation of asset pricing theories, it's clear that a complete evaluation of these models requires moving beyond calibration, to formal estimation and testing.[7]

It is little wonder that such an empirical investigation has yet to emerge. Consider the range of habit-based asset pricing models cited in footnote 6. All of these models place testable restrictions on the joint behavior of aggregate consumption and asset returns, and each implies 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 relies on a linear specification for the habit stock as a function of past consumption (e.g.,

[1] For example, Constantinides and Duffie (1996), Heaton and Lucas (1996), Krusell and Smith (1997), Constantinides, Donaldson, and Mehra (2002), Kogan and Uppal (2002).

[2] For example, Attanasio, Banks, and Tanner (2002), Brav, Constantinides, and Geczy (2002), Vissing-Jrgensen (2002). Constantinides (2002) provides a survey of this literature.

[3] For example, Abel (1989), Dumas (1989), Grossman and Zhou (1996), Sandroni (1999), Chan and Kogan (2002).

[4] For example, Lustig and Nieuwerburgh (2002); Piazzesi, Schneider, and Tuzel (2002); Yogo (2003).

[5] For example, Barsky and De Long (1993), Barberis, Shleifer, and Vishny (1998), Hansen, Sargent, and Tallarini (1999), Cecchetti, Lam, and Mark (2000).

[6] 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); Wachter (2002), Dai (2003), and Menzly, Santos, and Veronesi (2004). We discuss these papers further below.

[7] Special cases of habit-based asset pricing models have been empirically evaluated: Ferson and Constantinides (1991); Heaton (1995). We discuss these papers further below.

Sundaresan (1989); Constantinides (1990); Heaton (1995); Jermann (1998); Boldrin, Christiano, and Fisher (2001)). By contrast, more recent theoretical work often takes as a starting point the highly nonlinear habit specification that includes current consumption developed in Campbell and Cochrane (1999) (e.g., Campbell and Cochrane (2000); Li (2001); Wachter (2002); and Menzly, Santos, and Veronesi (2004)). These authors parameterize the functional form of the habit so that a calibrated version of their model closely matches a selected set of asset pricing moments calculated from post-war data. Because the habit specifications have not been estimated, however, it is unclear whether they provide a valid description of the data. For example, emphasis on matching different sets of asset pricing moments is likely to lead to different functional forms for the habit; it is unclear how one should choose among these.[8]

These observations raise an important econometric issue for researchers interested in estimation and testing: there are good reasons to think that the true habit specification is unknown, implying that its functional form should be treated not as a given, but as part and parcel of the empirical investigation.

This study evaluates the ability of a general class of habit-based asset pricing models to match the conditional moment restrictions implied by asset pricing theory. Our approach is to treat the functional form of the habit as unknown, and estimate it along with the rest of the model's parameters. The empirical model we explore presumes that investor utility is a power function of the difference between aggregate consumption and a habit level, but allows 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. In essence, our empirical investigation does for estimation and testing what Campbell and Cochrane (1999) did for calibration: we allow the data to "reverse engineer" the functional form of the habit that most closely matches the joint distribution of aggregate consumption growth and asset returns implied by asset pricing theory. Moreover, to avoid potential misspecification, the law of motion of consumption growth and asset returns are left unspecified.

Estimation and testing are conducted by applying a minimum distance procedure to the essential asset pricing condition (a set of Euler equations) corresponding to the habit-based framework we study. These Euler equations deliver a set of restrictions on the joint distribution of aggregate consumption growth and asset returns by dictating that the product of the intertemporal marginal rate of substitution in consumption and each asset return must have a conditional expectation equal to unity. We use the sieve minimum distance (SMD hereafter) estimator for semiparametric condi-

[8] Most studies focus on the mean and standard deviation of excess stock returns and the risk-free rate. An exception is Otrok, Ravikumar, and Whiteman (2002), which asks whether habit models can explain these moments taking into account changes in the temporal distribution of consumption volatility.

tional moment models developed in Newey and Powell (2003) and Ai and Chen (2003) to directly estimate the Euler equations underlying the optimal consumption choice of an investor with access to N asset payoffs. The SMD estimator is an especially appealing estimator for this application because it can be implemented as Generalized Method of Moments (GMM, Hansen (1982)), an approach that will be familiar from prior work in estimating fully parametric, consumption-based asset pricing models (e.g., Hansen and Singleton (1982)).[9]

The "sieve" part of the minimum distance estimator is a procedure for approximating an unknown function by a sequence of parametric functions, with the number of parameters expanding as the sample size grows (Grenander (1981)). The obvious advantage of this approach relative to parametric modeling is that it imposes few restrictions on the form of the joint distribution of the observed data, so there is little room for model misspecification. The cost of the nonparametric approach is that the convergence rate of the resulting estimator is slower than the parametric rate. Nevertheless, we show in the Appendix that all of the parameters in our model (the habit function, the curvature parameter of the power utility function and the time-preference parameter) are identified, and may be consistently estimated using the SMD methodology. In addition, the SMD estimates of the finite dimensional parameters that are part of the power function and time-preference are $\sqrt{T}$ consistent (where T is the sample size), and asymptotically normally distributed.

This approach allows us to empirically investigate a number of interesting hypotheses about the specification of habit-based asset pricing models that have not been previously investigated. One interesting hypothesis concerns whether the habit is better described as a linear function, as in the work of Sundaresan (1989), Constantinides (1990), Heaton (1995), Jermann (1998) and Boldrin, Christiano, and Fisher (2001), or as a nonlinear function, as in the more recent work of Campbell and Cochrane (1999) and the many other researchers who have extended their model to accommodate a variety of settings (e.g., Campbell and Cochrane (2000); Li (2001); Wachter (2002); and Menzly, Santos, and Veronesi (2004)). Campbell and Cochrane (1999) argue that nonlinearities in the habit are crucial for allowing such models to fit key features of asset pricing data, such as time-series predictability of excess stock returns and counter-cyclical variation in the conditional Sharpe ratio for the aggregate stock market. Our empirical results suggest that the functional form of the habit is better described as nonlinear rather than linear, consistent with these more recent modeling strategies.

A second interesting hypothesis concerns the distinction between "internal" and "external" habit formation. The models investigated by Sundaresan (1989), Constantinides (1990), Heaton (1995) and Boldrin, Christiano, and Fisher (2001) are models of internal habit formation, in which the habit is a function of the agent's own past consumption. By contrast, Campbell and Cochrane

[9] Jagannathan, Skoulakis, and Wang (2002) provide several examples illustrating the use of GMM in asset pricing applications.

(1999), Campbell and Cochrane (2000), Li (2001), Shore and White (2002), Wachter (2002), and Menzly, Santos, and Veronesi (2004) investigate 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 have dramatically different implications for optimal tax policy and welfare analysis, as well as for whether such models are capable of resolving long-standing asset-allocation puzzles in the international finance literature (e.g., see Ljungqvist and Uhlig (2000) and Shore and White (2002)). Our empirical results indicate that the data are better described by internal habit formation than external habit formation.

Finally, our approach allows us to assess the quantitative importance of the habit in the power utility specification. Using stationary quarterly data on consumption growth, assets returns and instruments, 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. In addition, our estimated habit function is concave, generates positive intertemporal marginal rate of substitution in consumption, our estimated time-preference parameter is around 0.99 and the estimated power utility parameter is about 0.80 for three different combinations of instruments and asset returns.

How well does the habit-based framework fit the asset pricing data? We evaluate the SMD-estimated habit model and several competing asset pricing models by employing the model comparison distance metric recommended in Hansen and Jagannathan (1997), where all the models are treated as stochastic discount factor (SDF) proxies to the truth. 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 asset pricing model of Fama and French (1993), and the approximately linear, conditional, or "scaled" consumption capital asset pricing model (CCAPM) explored in Lettau and Ludvigson (2001b). We also compare its performance with the classic CAPM of Sharpe (1964) and Lintner (1965) and classic consumption CAPM of Breeden (1979) and Breeden and Litzenberger (1978). Doing so, we find that a SMD-estimated internal habit model can explain a cross-section of size and book-market sorted equity returns better than the Fama-French three-factor model, better than the Lettau-Ludvigson scaled consumption CAPM, better than a SMD-estimated external habit model, and better than the classic CAPM and consumption CAPM models.[10]

To our knowledge, there has been only a small amount of prior work applying sieve nonparamet-

[10] For the internal and external habit models, we treat the SMD-estimated habit functions as a known part of the SDF proxy for those models. As we explain below, this procedure places the habit models at a disadvantage relative to other models because we don't minimize the Hansen-Jagannathan distance with respect to the habit sieve parameters but instead fix them at the values that minimize the SMD criterion. See Section 6 for further discussion.

ric estimation techniques to asset pricing questions. Gallant and Tauchen (1989) employed a "semi-nonparametric" modeling approach based on series expansions to estimate a consumption-based asset pricing model. Gallant, Hansen, and Tauchen (1990) employ the same semi-nonparametric methodology, but apply it to the conditional distribution of a vector of monthly asset payoffs. This procedure allowed the efficient use of conditioning information in the computation of volatility bounds for the intertemporal marginal rate of substitution of consumers (Hansen and Jagannathan (1991)). Following this work, Bansal and Viswanathan (1993) and Bansal, Hsieh, and Viswanathan (1993) use the semi-nonparametric methodology 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 study differs from these in several ways. First, we place more structure on the empirical asset pricing model by embedding the unknown habit function in the more familiar power-utility framework.[11] By contrast, Gallant and Tauchen (1989) treat the entire period-by-period utility function as unknown and approximate it using polynomial series, while Bansal and Viswanathan (1993), Bansal, Hsieh, and Viswanathan (1993) and Chapman (1997) approximate the whole SDF as a function of a few macroeconomic factors. The more structural approach taken in this paper makes it straightforward to 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. Second, Gallant and Tauchen (1989) focused on time series asset pricing properties of their estimated model using a consumption growth series and a small number ($N = 2$) of asset return series. We focus on the cross-sectional asset pricing properties with a larger number (for example, $N = 7, 17$ or 26) of return series, and we also use different conditioning variables. These differences are of relevance because, as shown in the Appendix, larger N and the appropriately chosen conditioning variables make it more likely that the conditions for identification of all the model parameters (the unknown habit function, the power utility parameter and the time-preference parameter) will be satisfied. Indeed, these conditions appear to be well satisfied in our data, since the SMD estimation results are all very similar when different sets of asset returns and instruments are used. Third, Gallant and Tauchen (1989) and Gallant, Hansen, and Tauchen (1990) approximate the transition density underlying the conditional moment restrictions using a Hermite polynomial, whereas we approximate the conditional moments directly using known basis functions of conditioning variables. Their approach is potentially more efficient, while ours is computationally simpler, especially when the number of conditional moment restrictions N is large.[12] Fourth, treating all models as potentially

[11] There are good reasons in asset pricing theory to allow the habit to be an unknown function of contemporaneous consumption. However, this specification makes the estimation and testing very challenging, since it becomes a nasty nonparametric nonlinear ill-posed inverse problem, see e.g., Blundell and Powell (2001), Darolles, Florens, and Renault (2002), Newey and Powell (2003), Ai and Chen (2003).

[12] For the set of Euler equations with N returns, the Gallant and Tauchen (1989) procedure needs to estimate the

misspecified, we conduct an extensive model comparison of the performance of our SMD-estimated habit models by evaluating them against other leading asset pricing models designed to explain the cross-section of asset returns.

The rest of this paper is organized as follows. In the next section we lay out the empirical asset pricing model to be estimated and tested. Section 3 explains the estimation technique and how it is implemented. Section 4 describes the data; Section 5 presents the results of estimation and hypothesis testing about the linearity of the habit function, and Section 6 provides specification tests for internal versus external habit formation, and conducts model comparison when all the competing models can be misspecified. Section 7 concludes. In the Appendix we present identification of all the model parameters, and provide some technical details on the large sample properties of the SMD estimators and the related test statistics.

2 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 relative risk-aversion is constant, hence they have difficulty accounting for the predictability of excess stock returns documented in the empirical asset pricing literature.[13] 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 6 of this paper.

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

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

$(1+N)-$dimensional conditional density of the consumption growth and N asset returns, given the past consumption growth and past returns, subject to the N Euler equations. By contrast, our procedure only needs to compute N least square projections of the product of the SDF and an asset return onto the instruments.

[13] A large literature finds that excess stock returns are forecastable. Shiller (1981), Fama and French (1988), Campbell and Shiller (1988), Campbell (1991), and Hodrick (1992) find that the ratios of price to dividends or earnings have predictive power for excess returns. Harvey (1991) finds that similar financial ratios predict stock returns in many different countries. Lamont (1998) forecasts excess stock returns with the dividend-payout ratio. Campbell (1991) and Hodrick (1992) find that the relative T-bill rate (the 30-day T-bill rate minus its 12-month moving average) predicts returns, while Fama and French (1988) study the forecasting power of the term spread (the 10-year Treasury bond yield minus the one-year Treasury bond yield) and the default spread (the difference between the BAA and AAA corporate bond rates). Lettau and Ludvigson (2001a) forecast returns with a proxy for the log consumption-wealth ratio.

Here X_t is the level of the habit, and δ is the 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 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)).[14]

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 consumption is given by

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

where

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

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, however, 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 as C_t from now on.

The asset pricing model comes from the first-order conditions for optimal consumption choice. These first-order conditions place restrictions on the joint distribution of the intertemporal marginal rate of substitution in consumption and asset returns. They 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 $X_t = f\left(C_t, C_{t-1}, ...C_{t-L}\right)$. It is clear that identification of the unknown habit function using (4) requires ruling out the case $X_t = KC_t$ with a constant $K \in [0, 1)$; this is not a problem since any model of habit formation naturally depends on past consumption. The model (4) is semiparametric in the sense that it contains both finite

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

dimensional and infinite dimensional unknown parameters. In addition, it leaves the law of motion unspecified.

3 Empirical Implementation

Our empirical approach is based on estimation of the conditional moment restrictions (4). Estimation in this setting may be undertaken using the sieve minimum distance (SMD) estimator developed in Newey and Powell (2003) and Ai and Chen (2003).

The idea behind the SMD estimator is that sample analog of the conditional moment (4) can be consistently estimated via minimum distance estimation in a procedure that has two essential parts. First, although the functional form of the conditional distribution implied by (4) is unknown, we may replace the conditional expectation itself with a consistent nonparametric estimator (to be specified later). Second, although the habit function f is an infinite-dimensional unknown parameter, we can approximate it by a sequence of finite-dimensional unknown parameters (sieves) f_{K_T}, where the approximation error decreases as the dimension K_T increases with the sample size T, and where f_{K_T} is estimated jointly with the finite-dimensional parameters $(\delta, \gamma)'$ by minimizing a (weighted) quadratic norm of estimated conditional expectation functions.

Under the assumption of i.i.d. observations, Ai and Chen (2003) show that the SMD estimators of the unknown functions such as f are consistent with rate, that the SMD estimators of the finite-dimensional parameters such as $(\delta, \gamma)'$ are $\sqrt{T}$ consistent and asymptotically normally distributed and the optimally weighted versions are semiparametric efficient. In the Appendix we show that the results on nonparametric consistency and parametric $\sqrt{T}-$asymptotic normality can be easily extended to allow for stationary beta-mixing time series observations.[15] Beta-mixing is one popular measure of temporal dependence for nonlinear time series; see Appendix for the formal definition. It is satisfied by many widely used financial time series models including nonlinear ARCH, GARCH, stochastic volatility and diffusion models; see e.g., Doukhan (1994), Chen, Hansen, and Carrasco (2001) and Carrasco and Chen (2002). Thus, our procedure requires stationary ergodic observations but does not restrict to linear time series specifications, nor do we specify parametric laws of motions of the data.

Before we can estimate the model, we must address two specification issues that arise both from the nature of the data on aggregate consumption and the nature of the moment conditions specific to our application. First, it is clear that consumption is trending over time, so it is necessary to transform the model to use stationary observations on consumption growth. We address this problem by assuming that the unknown function $X_t = f(C_t, C_{t-1}, ...C_{t-L})$ is homogeneous of degree one. The homogeneous of degree one assumption is consistent with the habit models studied in

[15] However, there is currently no general semiparametric efficient result for SMD estimators of time series models of conditional moment restrictions containing unknown functions.

the asset pricing literature cited above, including the complex habit specification investigated in Campbell and Cochrane (1999). 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. In this case, the unknown function X_t may be written

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

which can be redefined as

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

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 identified and estimated along with (δ, γ) using the Euler equation (4) and the SMD procedure. In the Appendix we show that $g, \delta,$ and γ are identified as long as $0 \leq g < 1$, $g(0, ..., 0) = 0$, $g \neq const.$, $\delta > 0$ and $\gamma > 0$.

A second implementation issue concerns the form of the nonparametric specification in (6). A specification such as that in (6) will clearly be infeasible if L is too large, a "curse of dimensionality."[16] One approach to this problem is to estimate a fully nonparametric sieve model (e.g., tensor product linear sieves such as tensor product splines) and to simply limit the number of lags, L, to some small number, such as one.[17] Alternatively, we may employ more lags in our estimation by using a nonlinear sieve (e.g., a neural network) to approximate $g\left(\frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t}\right)$. Such an approach has several important advantages for our application. First, it allows the use of more lags by delivering a relatively fast convergence rate (compared to linear tensor product sieves) when approximating the unknown function g (Chen and White (1999)). Second, the use of a nonlinear sieve is often in practice better able to allow for nonlinearities in the unknown function, something that is particularly important for the habit-based asset pricing literature which has increasingly emphasized nonlinear specifications for the habit. Third, a nonlinear sieve allows for possible non-

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

[17] This approach takes linear combinations of the tensor product of basis functions over each lag of consumption:

$$g\left(\widetilde{C}_t, \widetilde{C}_{t-1}, ...\widetilde{C}_{t-L}\right) \approx \sum_{i=0}^{K_T} \pi_i \prod_{j=0}^{L} B_{i_j}\left(\widetilde{C}_{t-j}\right). \tag{7}$$

Approximations of this form are routinely employed in economic problems which require a numerical solution to a functional equation (for example, in numerical solutions of stochastic growth models), and are known to deliver reasonably accurate results (e.g., Judd (1998), McGrattan (1998)). An important shortcoming of this approach in empirical settings, however, is that approximation based on linear tensor product sieves may have slower convergence rates than approximation based on nonlinear sieves when the domain of g is of high dimension (Chen and White (1999)).

separability between elements of g. Such nonseparability is a feature of well-known habit-based asset pricing models (e.g., Campbell and Cochrane (1999)).

Of course, even using a nonlinear sieve, the number of lags of C_t upon which X_t is estimated to depend must be restricted to some reasonable number relative to the sample size. Nevertheless, 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.

In this paper, we estimate the function $g\left(\frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t}\right)$ using a single-layer smooth Artificial Neural Network (ANN) sieve approximation, defined as

$$g\left(\frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t}\right) \approx \alpha_0 + \sum_{j=1}^{K_T} \alpha_j \psi\left(\gamma_{j,1}\frac{C_{t-1}}{C_t} + \cdots + \gamma_{j,L}\frac{C_{t-L}}{C_t} + \beta_j\right), \tag{8}$$

where $\psi(\cdot)$ is called 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.[18] We denote the unknown parameters to be estimated as $\boldsymbol{\alpha} = (\delta, \gamma, g)' =$ $(\delta, \gamma, \alpha_0, \alpha_1, ..., \alpha_{K_T}, \gamma_{1,1}, ..., \gamma_{1,L}, ..., \gamma_{K_T,1}, ..., \gamma_{K_T,L}, \beta_1, ..., \beta_{K_T})'$. We are not interested in the sieve parameters *per se*, but in the dynamic behavior of the habit stock and marginal utility, which depend on those parameters.

An important advantage of using the particular ANN sieve above is that, as long as consumption is strictly positive, we may easily restrict coefficients 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 in our setting because the sigmoid 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. Imposing this restriction is sometimes difficult even when the habit is specified parametrically, for example, as a linear or polynomial function of past consumption.

We are now in a position to estimate the unknown δ, γ, and g using the conditional moment conditions (4). When the habit stock is a homogeneous of degree one function of current and past

[18] Hornik, Stinchcombe, and White (1989) provide a universal approximation result which states that, asymptotically, we can approximate any continuous function 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 stochastic discount factor of an nonlinear arbitrage pricing model. See Campbell, Lo, and MacKinlay (1997) for additional neural network sieve applications in finance.

consumption, marginal utility, MU_t, takes the form

$$\begin{aligned} MU_t &= C_t^{-\gamma}\left(1-g\left(\frac{C_{t-1}}{C_t},...,\frac{C_{t-L}}{C_t}\right)\right)^{-\gamma} \\ &\quad -C_t^{-\gamma}E_t\left[\sum_{j=0}^{L}\delta^j\left(\frac{C_{t+j}}{C_t}\right)^{-\gamma}\left(1-g\left(\frac{C_{t+j-1}}{C_{t+j}},...,\frac{C_{t+j-L}}{C_{t+j}}\right)\right)^{-\gamma}\frac{\partial X_{t+j}}{\partial C_t}\right], \end{aligned} \tag{9}$$

where,

$$\frac{\partial X_{t+j}}{\partial C_t}=\begin{cases} g_j\left(\frac{C_{t+j-1}}{C_{t+j}},...,\frac{C_{t+j-L}}{C_{t+j}}\right) & \forall j\neq 0 \\ g\left(\frac{C_{t-1}}{C_t},...,\frac{C_{t-L}}{C_t}\right)-\sum_{i=1}^{L}g_i\left(\frac{C_{t-1}}{C_t},...,\frac{C_{t-L}}{C_t}\right)\frac{C_{t-i}}{C_t} & j=0 \end{cases} \tag{10}$$

In the expression directly above, g_i denotes the derivative of g with respect to its ith argument. Together, equations (9) and (10) imply that the stochastic discount factor can be expressed as a function of the gross growth rates of consumption:

$$M_{t+1}=\delta\frac{MU_{t+1}}{MU_t}=\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}\Psi_{t+1}, \tag{11}$$

where,

$$\Psi_{t+1}\equiv\frac{\left(\begin{array}{l}\left(1-g\left(\frac{C_t}{C_{t+1}},...,\frac{C_{t+1-L}}{C_{t+1}}\right)\right)^{-\gamma} \\ \quad -E_{t+1}\left[\sum_{j=0}^{L}\delta^j\left(\frac{C_{t+1+j}}{C_{t+1}}\right)^{-\gamma}\left(1-g\left(\frac{C_{t+j}}{C_{t+1+j}},...,\frac{C_{t+j+1-L}}{C_{t+1+j}}\right)\right)^{-\gamma}\frac{\partial X_{t+1+j}}{\partial C_{t+1}}\right]\end{array}\right)}{\left(\begin{array}{l}\left(1-g\left(\frac{C_{t-1}}{C_t},...,\frac{C_{t-L}}{C_t}\right)\right)^{-\gamma} \\ \quad -E_t\left[\sum_{j=0}^{L}\delta^j\left(\frac{C_{t+j}}{C_t}\right)^{-\gamma}\left(1-g\left(\frac{C_{t+j-1}}{C_{t+j}},...,\frac{C_{t+j-L}}{C_{t+j}}\right)\right)^{-\gamma}\frac{\partial X_{t+j}}{\partial C_t}\right]\end{array}\right)}.$$

The stochastic discount factor, M_{t+1}, is the product of two terms, $\delta\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma}$ and Ψ_{t+1}. The first term is the familiar expression for the intertemporal marginal rate of substitution when preferences are characterized by constant relative risk aversion utility and no habit formation. The second term is a complicated function of expected future, current, and past consumption growth, and is attributable to the presence of X_t in (1).

To obtain an estimable expression, the stochastic discount factor, M_{t+1}, must be rearranged so that the conditional expectation E_t appears only on the outside of (4). The Appendix presents several equivalent expressions of this form; here we present one. Denote the true values of the parameters with an "o" subscript. Combining (11) and (4) and rearranging terms generates a set of N conditional moment conditions:

$$E_t\left\{\left(\delta_o\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma_o}\mathit{\Gamma}_{t+1}R_{i,t+1}-1\right)\Phi_{t+1}\right\}=0,\quad i=1,...,N, \tag{12}$$

where

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

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

Let

$$\mathbf{z}_{t+1} \equiv \begin{array}{l} \Bigg(R_{1,t+1}, ..., R_{N,t+1}, \left\{\frac{C_{t+1+j}}{C_{t+1}}\right\}_{j=1}^{L}, \left\{\frac{C_{t+j}}{C_{t+1+j}}, ..., \frac{C_{t+j+1-L}}{C_{t+1+j}}\right\}_{j=1}^{L}, \frac{C_t}{C_{t+1}}, ... \\ \quad ..., \frac{C_{t+1-L}}{C_{t+1}}, \left\{\frac{C_{t+j}}{C_t}\right\}_{j=1}^{L}, \left\{\frac{C_{t+j-1}}{C_{t+j}}, ..., \frac{C_{t+j-L}}{C_{t+j}}\right\}_{j=1}^{L}, \frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t}\Bigg)' \end{array}.$$

Defining

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

we may write (12) more compactly as

$$E\left\{\rho_i(\mathbf{z}_{t+1}, \delta_o, \gamma_o, g_o)|\mathbf{w}_t^*\right\} = 0, \qquad i = 1, ..., N, \tag{13}$$

where the conditional expectation in (13) is taken with respect to agents' information set at time t, $\mathbf{w}_t^*$. Let $\mathbf{w}_t$ be a $d_w \times 1$ observable subset of $\mathbf{w}_t^*$ that does not contain a constant. Equation (13) implies

$$E\left\{\rho_i(\mathbf{z}_{t+1}, \delta_o, \gamma_o, g_o)|\mathbf{w}_t\right\} = 0, \qquad i = 1, ..., N. \tag{14}$$

In the Appendix we provide sufficient conditions so that the conditional moment restrictions (14) identifies the parameters of interest $\boldsymbol{\alpha}_o = (\delta_o, \gamma_o, g_o)'$.

For any candidate value $\boldsymbol{\alpha} = (\delta, \gamma, g)'$ we let $\rho(\mathbf{z}_{t+1}, \boldsymbol{\alpha}) = (\rho_1(\mathbf{z}_{t+1}, \boldsymbol{\alpha}), ..., \rho_N(\mathbf{z}_{t+1}, \boldsymbol{\alpha}))'$ and $m(\mathbf{w}_t, \boldsymbol{\alpha}) = E\{\rho(\mathbf{z}_{t+1}, \boldsymbol{\alpha})|\mathbf{w}_t\}$. It is obviously true that

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

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 $\widehat{m}(\mathbf{w}, \boldsymbol{\alpha}) = \left(\sum_{t=1}^{T} \rho(\mathbf{z}_{t+1}, \boldsymbol{\alpha}) p^{J_T}(\mathbf{w}_t)'(\mathbf{P}'\mathbf{P})^{-1}\right) p^{J_T}(\mathbf{w})$ 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 its consistency). In this paper we consider a simple SMD estimator

$$\widehat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \frac{1}{T} \sum_{t=1}^{T} \widehat{m}(\mathbf{w}_t, \boldsymbol{\alpha})' \widehat{m}(\mathbf{w}_t, \boldsymbol{\alpha}), \tag{15}$$

which is in fact a two-stage nonlinear Least Squares estimator of $\boldsymbol{\alpha}_o$. By extending the results in Newey and Powell (2003) and Ai and Chen (2003) for i.i.d. data to stationary beta-mixing time series observations, we obtain the consistency of $\widehat{g}$ and root-T consistency and asymptotic normality of $(\widehat{\delta}, \widehat{\gamma})$.

Plugging the sieve least squares estimator $\widehat{m}(\mathbf{w}, \boldsymbol{\alpha})$ into (15) reveals that $\widehat{\boldsymbol{\alpha}}$ is also a GMM (Hansen (1982)) estimator:

$$\widehat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \left[\mathbf{g}(\boldsymbol{\alpha}; \mathbf{y}^T)\right]' \{\mathbf{I}_N \otimes (\mathbf{P}'\mathbf{P})^{-1}\} \left[\mathbf{g}(\boldsymbol{\alpha}; \mathbf{y}^T)\right], \tag{16}$$

where[19]

$$\mathbf{g}(\boldsymbol{\alpha}; \mathbf{y}^T) \equiv \frac{1}{T} \sum_{t=1}^{T} \rho(\mathbf{z}_{t+1}, \boldsymbol{\alpha}) p^{J_T}(\mathbf{w}_t) \tag{17}$$

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

$$E\{\rho_i(\mathbf{z}_{t+1}, \delta_o, \gamma_o, g_o) p_{0j}(\mathbf{w}_t)\} = 0, \qquad i = 1, ..., N, \quad j = 1, ..., J_T. \tag{18}$$

It is well known that as long as the sequence of basis functions $\{p_{0j}(\mathbf{w}_t),\ j = 1, 2, ..., J_T\}$ is dense in the space of square integrable functions of $\mathbf{w}_t$, the conditional moment restrictions (14) hold if and only if the increasing number of unconditional moment restrictions (18) hold. Therefore, one could also estimate $\boldsymbol{\alpha}_o = (\delta_o, \gamma_o, g_o)'$ by minimizing the following GMM criterion with a general weighting matrix $\mathbf{W}$:

$$Q_T(\boldsymbol{\alpha}) = \left[\mathbf{g}(\boldsymbol{\alpha}; \mathbf{y}^T)\right]' \mathbf{W} \left[\mathbf{g}(\boldsymbol{\alpha}; \mathbf{y}^T)\right]. \tag{19}$$

We conjecture that minimizing the GMM criterion (19) using an arbitrary weighting matrix $\mathbf{W}$ would also lead to consistent estimation of $\boldsymbol{\alpha}_o = (\delta_o, \gamma_o, g_o)'$. Unfortunately, even for i.i.d. data, such a consistency result for an arbitrary weighting matrix has not been established when an unknown function g_o of an endogenous variable is involved. Fortunately, minimization of (19) using the particular weighting matrix $\mathbf{W} = \mathbf{I}_N \otimes (\mathbf{P}'\mathbf{P})^{-1}$ is shown to yield consistent estimation of $\boldsymbol{\alpha}_o = (\delta_o, \gamma_o, g_o)'$. 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, hence gives greater weight to moments that are more highly correlated with the instruments $p^{J_T}(\cdot)$.

It is important to note that the general SMD procedure described in Ai and Chen (2003) and stated in the Appendix 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, a specification that is crucial to the nonparametric estimation of the conditional mean (14) using the basis functions $\{p_{0j}(\mathbf{w}_t),\ j = 1, 2, ..., J_T\}$. Estimation of the *conditional* expectation $E\{\rho(\mathbf{z}_{t+1}, \delta, \gamma, g)|\mathbf{w}_t\}$ (as opposed to an unconditional expectation) is in turn necessary to obtain consistent estimates of the nonparametric habit function.

[19] In this paper $\mathbf{y}^T \equiv (\mathbf{z}'_{T+1}, ...\mathbf{z}'_2, \mathbf{w}'_T, ...\mathbf{w}'_1)'$ denotes the vector containing all observations in the sample of size T.

To summarize, the empirical procedure for estimating $\boldsymbol{\alpha}$ is based on the following steps. First, we transform the model so that the observations we employ are stationary beta-mixing, by assuming that the habit is a homogeneous of degree one function of current and past consumption. This allows us to derive an expression for the stochastic discount factor that is a function only of the gross growth rates of consumption. Second, a flexible and robust functional form for the habit is obtained by approximating it using a neural network sieve, whose dimensionality (complexity) grows with the sample size T. Third, we estimate the set of N conditional expectations in (13) by transforming (14) into a set of NJ_T unconditional expectations, multiplying each $\rho_i(\mathbf{z}_{t+1}, \delta, \gamma, g)$ for $i = 1, ...N$ by J_T "instruments," $p^{J_T}(\mathbf{w}_t)$, which are known basis functions of observable variables, $\mathbf{w}_t$. Fourth, we compute the sample average of the NJ_T orthogonality conditions, $\mathbf{g}(\boldsymbol{\alpha}; \mathbf{y}^T)$. Finally, we find estimates of $\boldsymbol{\alpha}$ by setting a weighted sum of the NJ_T sample average moments $\mathbf{g}(\boldsymbol{\alpha}; \mathbf{y}^T)$ as close as possible to the population moment of zero, by minimizing the GMM criterion function (16).[20]

We considered a number of additional implementation issues in our estimation. First, as with any nonlinear estimation procedure, it is necessary to require the parameter space 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 $\gamma \in [.1, 100]$.

Second, we may compute standard errors for δ and γ, but have no way of formulating standard errors for the sieve parameters or the habit function itself. Although we provide consistency of the sieve estimator for time series observations (see the Appendix), a general asymptotic distribution theory for SMD estimators of unknown functions has not been developed, even for i.i.d. observations.[21] But the parametric part of our specification (δ and γ) are root-T asymptotically normally distributed. Moreover, as shown in Ai and Chen (2003) for the i.i.d. case, the asymptotic standard errors for these parameters (δ and γ) may be computed using GMM theory (Hansen (1982)). We present these estimates in Table 1 below.

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

[20] In this paper we only assume that the time series $\{(\mathbf{z}'_t, \mathbf{w}'_t)\}$ is stationary beta-mixing, and hence only consider the inefficient but consistent SMD estimation of $\boldsymbol{\alpha}_o = (\delta_o, \gamma_o, g_o)'$. In the future we plan to study the semiparametric efficient SMD estimation of $\boldsymbol{\alpha}_o$; we conjecture that Ai and Chen's (2003) optimally weighted SMD criterion will still lead to efficient estimation for time series data.

[21] 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.

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. This suggests that time-aggregation may not have a large influence on the estimates from minimum distance procedures. Second, Ferson and Constantinides also note that, at least for the case of linear habit specifications, linear approximations of the Euler equation imply that the effect of time-aggregation is to increase the order of the moving average process followed by the GMM error, in our case $\rho_i(\mathbf{z}_{t+1}, \delta_o, \gamma_o, g_o)$. Of course, the influence of time-aggregation may be more complex for nonlinear specifications; but 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 $\rho_i(\mathbf{z}_{t+1}, \delta_o, \gamma_o, g_o)$.

4 The Data

A detailed description of the data and our sources is provided in the Appendix. Our data are quarterly,[22] and span the period from the fourth quarter of 1952 to the fourth quarter of 2001.

We study three groups of asset returns. All stock return data are taken from Kenneth French's Dartmouth web page (URL provided in the appendix). 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 we consider (Group 2), uses a smaller number of asset returns, 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. Because of the smaller number of asset returns, this estimation has a smaller number of endogenous variables. Finally, we consider a larger cross-section

[22] It is known that the quarterly consumption data contains less measurement errors than the monthly consumption data.

of returns: 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). Again, the portfolios are created from stocks traded on the NYSE, AMEX, NASDAQ, and are constructed as described on Kenneth French's web page.

The focus of this paper is on testing and modeling cross-sections of asset returns, rather than one or two aggregate asset returns. Exploiting the cross-section aids the empirical identification of the unknown habit function. The number of moment conditions must be greater than the number of parameters to be estimated, which is typically fairly large for estimation of an unknown function. Exploiting a cross-section of asset returns is useful because there is often more independent information in a well-chosen cross-section of asset returns than in ever larger numbers of basis functions of a few conditioning variables.

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.

The procedure requires computation of instruments, $p^{J_T}(\mathbf{w}_t)$, which are known basis functions (including a constant function) of conditioning variables, $\mathbf{w}_t$. 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.[23] Although no formal test of instrument relevance has been developed for estimation involving an unknown function, we focus on variables for $\mathbf{w}_t$ that are known to be strong predictors of asset returns and consumption growth in quarterly data. 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 on aggregate stock market indexes and/or consumption growth in quarterly data, as well as for several of the portfolio returns studied here. The first variable is a proxy for the log consumption-wealth ratio studied in (Lettau and Ludvigson (2001a)), where wealth here includes human capital as well as nonhuman capital. This proxy is measured as the cointegrating residual between log consumption, log asset wealth, and log labor income and is denoted $\widehat{cay}_t$.[24] 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 portfolio returns. Two other 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

[23] See Stock, Wright, and Yogo (2002) for a survey of this issue.

[24] 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.

(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, $SPEX$.[25] Time-series regressions using these variables to predict future excess stock returns on aggregate stock market indexes can be found in Lettau and Ludvigson (2001a). 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.[26]

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. The third set of instruments utilizes just a constant and the linear terms of each variable, or five instruments in total. Recall that Group 1 assets contain 17 returns, Group 2 contains 7 returns and Group 3 contains 26 returns. We use the third instrument set when analyzing the larger asset return group, Group 3, and the second instrument set when analyzing the smaller asset return group, Group 2. This is done because 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 $T \to \infty$. Since Group 3 has 26 asset returns, we reduce the number of instruments by using only a constant and the linear transformations of $\mathbf{w}_t$ in this case, so that the total number of moments is similar to that for the estimations on Group 1 and Group 2 assets. Similar reasoning suggests that we use a greater number of instruments when employing the smaller set of asset returns in Group 2.

[25] We focus on these variables not only because they have been found to have significant forecasting power for future excess stock returns, but also 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)).

[26] 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.

5 Empirical Results

5.1 Empirical Estimates

All the tables and figures are reported at the end of the Appendix. Table 1 and Figures 1-10 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.[27] The results reported below were very similar with $L = 4$ and $L = 3$. Thus, we opt for the more parsimonious specification, and in all cases reported below set $L = 3$.[28] We emphasize that our use of three lags is already 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)).[29]

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.

We consider three estimations: the first one using Group 1 asset returns with a constant, the linear and squared values of the elements of $\mathbf{w}_t$ as instruments; the second one using Group 2 asset returns with a constant, the linear, squared and cross terms of the elements of $\mathbf{w}_t$ as instruments; the third one using Group 3 asset returns with a constant and linear terms of $\mathbf{w}_t$ as instruments. The estimates of the time-preference parameter, δ, and the curvature parameter, γ, are presented in Table 1, with asymptotic standard errors in parentheses.

[27] 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.

[28] Taken literally, the choice of $L = 3$ implies that the error term, $\{M_{t+1}R_{i,t+1} - 1\}$, 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.

[29] 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.

Table 1 shows that 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.[30] The estimates for γ are effectively estimates of unity, since the minimized value of the GMM 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 panel 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 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. Note, however, the habit-consumption ratio is not highly volatile, ranging only from 0.97 to 0.974.

The bottom panel of Figure 1 plots the estimated stochastic discount factor over time. The stochastic discount factor (M_t) is given in equation (11), and depends on the conditional expectation of nonlinear functions of consumption growth and the estimated habit. Shown in Figure 1 is an estimate of M_t, using our estimated parameter values with those parts of M_t that appear in expectation as their projection onto the set of instruments used in that estimation. Note that 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).[31] 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

[30] Standard errors for δ and γ are computed using GMM theory (Hansen (1982)), which requires the inversion of the product of two matrixes, the first a function of the first derivatives of the GMM errors with respect to the parameter values, and the second the GMM weighting matrix. In many applications this matrix can be near singular, and it is in ours. To alleviate the instability of the estimator that is attributable to such near-singularity, we add a tiny positive scalar to the roots of the near-singular matrix to be inverted, delivering a "Ridge" estimator for the variance-covariance matrix. In practice this is implemented by adding $1.0e-08$ to the diagonal elements of the matrix to be inverted, a value on the same order of magnitude as the minimized GMM objective function.

[31] 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.

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 too 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, and it serves as a reminder that the estimation in this study places weight on a much larger set of asset pricing restrictions than those implied by the unconditional volatility bound. Our finding is also similar to findings by 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.

Figure 2 plots $C_t - X_t$ against C_t (top panel), X_t against C_t (middle panel) and X_t against C_{t-1} (bottom panel) for the estimation on Group 1 assets with a constant, the linear and squared values of the elements of $\mathbf{w}_t$ as instruments; again, the corresponding plots for the other estimations are similar and are omitted to conserve space. This figure plots the observed values of consumption against our estimate of the habit stock over time. Its purpose is to get a sense of how the habit stock varies over time with consumption. Note that this is not a plot of partial derivatives of the habit with respect to consumption, but merely a plot of two time series, one against the another. Thus, for example, Panel B shows how X_t tends to vary with C_t over time, but it does not hold fixed the values of lagged consumption, which also influence the habit stock.

The top panel of Figure 2 shows that the difference between consumption and habit tends to rise with contemporaneous consumption; the middle panel shows that the habit also increases with consumption, as would be consistent with common notions of habit formation. Thus, the habit tends to rise with consumption, but does not rise one-for-one with consumption. The estimated habit also increases with lagged consumption, as the bottom panel of each figure shows for the case of one-period lagged consumption, again consistent with common notions of habit persistence. Plots of X_t against the second and third lags of consumption (not shown) are similar.

We can also 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 given our estimated X function, we may plot the derivatives as they vary over time with lagged consumption growth. This is done in Figure 3, 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. This result is again consistent with common intuition about the properties of habit formation: the habit depends positively on lagged consumption, but this positive dependence decreases as the consumption becomes more distant.

Although the plots of X_t against C_t and C_{t-1} in Figure 2 look "linear," one cannot make inferences about whether the habit itself is a linear function of current and past consumption on the basis of these time-series plots.[32] The shape of our estimated habit function can be better 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 4 through 6 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 held fixed at their median, 25th, and 75th percentile values in our sample. The three figures are quite similar; we can draw several conclusions from them. 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. 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.

5.2 Is Habit Function Linear?

Figures 4 to 6 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. Sundaresan (1989), Constantinides (1990) and Heaton (1995) model the habit as a linear function of consumption, but Campbell and Cochrane (1999) and many subsequent authors argue that nonlinearities are a key factor determining whether the habit-based framework can match the time-series properties of asset returns and consumption growth. For example, Campbell and Cochrane (1999) argue that a nonlinear specification is necessary to make stable risk-free rates and a time-varying Sharpe ratio consistent with a random walk process for log consumption.[33] Is the habit we estimate here better described as a linear or nonlinear function of consumption growth?

To answer this question, it is useful to think about what linearity implies in this context. It

[32] See the next subsection for a statistical test of the linear habit specification.

[33] Campbell and Cochrane specify their nonlinear habit as a pure externality, however nonlinearities may also be important in internal habit specifications. In this section we test for nonlinearity in the internal habit specification; in the next section we undertake model specification tests aimed at determining whether internal or external habit formation better describes the data.

implies that $\frac{\partial X_{t+i}}{\partial C_t}$ is constant for all $i \geq 0$, and in particular

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

The second derivative of the habit function with respect to each value of lagged consumption must be identically zero everywhere, that is for all possible values of lagged consumption, not merely those observed in our sample. We can get an idea of whether the second derivatives are zero by plotting the estimated values of (20) based on our estimates for the cases above. This is done in Figures 7 through 9 for $i = 1$. Notice from (10) 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 7 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 8 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 9 plots the same for the model estimated on Group 3 assets with a constant, the linear terms of $\mathbf{w}_t$ as instruments. The second partial derivatives are everywhere negative; taken together with the estimates of the first partial derivatives (e.g., Figure 3), 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. Even though the numbers are less than one in absolute value, they are nonzero. We should expect a small value for g_{11} merely because the second derivative of the sigmoid function itself is always less than one, while the sieve parameters, the squared values of which multiply the second derivative of the sigmoid function, are also typically small in absolute value in order to keep the habit less than consumption. The point is that the units of g_{ii} are naturally small, so we should expect values for the second derivative that are significantly less than one in absolute value. What one should look for in these figures is whether the derivatives are *identically* zero everywhere, as would be required by a linear habit function. This is not evident from Figures 7 through 9.

Ideally we would like to construct a consistent statistical test of whether the second derivatives are identically zero. Unfortunately, this is impractical because the convergence rates for any test statistic based on second derivatives of an unknown function are known to be very slow. Nevertheless, we can provide a formal statistical 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 $g_{11}\left(\frac{C_t}{C_{t+1}}, ..., \frac{C_{t+1-L}}{C_{t+1}}\right)$.

We wish to test whether the mean of g_{11}, denoted as

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

is different from zero. 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 the result in Ai and Chen (2004) and under mild regularity conditions, $\sqrt{T}(\widehat{\mu}_g - \mu_g)$ 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 such as the Newey-West HAC estimator, then one could use $\sqrt{T}\widehat{\mu}_g/\widehat{\sigma}_{11}$ as a test statistic, which has a standard normal limiting distribution under the null hypothesis $\mu_g = 0$. However the asymptotic variance σ_{11}^2 takes a complicated form and the calculation of any consistent estimator $\widehat{\sigma}_{11}^2$ is complicated. Instead, we use a bootstrap procedure to directly compute an empirical, 95% confidence region for μ_g. We estimate the distribution of μ_g by bootstrapping the transformed raw data consisting of consumption growth rates, asset returns and conditioning variables which can be regarded as drawn from stationary beta-mixing processes. 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.[34] We then conduct SMD estimation on 500 bootstrap samples so formed, which delivers 500 estimates of μ_g.

The results are as follows. There is very little variation in the variance of the mean across bootstrap samples. This is not surprising, as the Figures above also suggest there is very little variation in the second derivative estimates over time. This results in very strong rejections of $\mu_g = 0$. For the estimation on Group 1 assets with a constant, 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

[34] 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).

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.

6 Specification and Model Comparison

In the introduction we discussed a number of interesting specification and model comparison issues. Here we investigate two of them. First, we ask whether habit formation is better described as internal or external habit formation. Second, we undertake a model comparison of the empirical performance of our estimated habit models with several leading and classic asset pricing models.

6.1 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 the existence of a constant risk-free rate, R^f, which equals the inverse of the rate of time preference, $1/\delta$. Cochrane (2001) shows that the asset pricing implications of such a model when the habit is external are observationally equivalent to those when the habit is internal. This example can be understood by inspecting (2), (3) and (4). When $R^f = 1/\delta$, the first-order condition for optimal consumption choice becomes

$$MU_t = E_t\{MU_{t+1}\}. \tag{22}$$

When habit formation is external, regardless of its functional form, marginal utility is

$$MU_t = (C_t - X_t)^{-\gamma}. \tag{23}$$

By contrast, when habit formation is internal, (3) gives the expression for MU_t. The internal habit expression (3) shows that if the $\frac{\partial X_{t+j}}{\partial C_t}$ terms on the right-hand-side are all constant, as is the case with habits linear in past consumption, the solution $(C_t - X_t)^{-\gamma} = E_t\left[(C_{t+1} - X_{t+1})^{-\gamma}\right]$ will lead to an expression for MU_t that satisfies (22). By plugging this solution into (3), it is straightforward to show that marginal utility in the internal habit formation case is proportional to $(C_t - X_t)^{-\gamma}$, and therefore to external habit formation marginal utility. It follows that–in this simple example with a constant risk-free rate $R^f = 1/\delta$ and linear habits–the asset pricing implications of each specification, which derive from the intertemporal *ratio* of marginal utilities, MU_{t+1}/MU_t, are equivalent.

This equivalency breaks down if habits are nonlinear in past consumption. In this case, the $\frac{\partial X_{t+j}}{\partial C_t}$ terms in (3) are not constant, but instead vary with lagged consumption. As a consequence, with nonlinear habit formation, marginal utility in the internal habit case will not be proportional to $(C_t - X_t)^{-\gamma}$, and the asset pricing implications of internal and external habit specifications may differ. In the last subsection we tested whether the habit function we estimate is linear; our test results strongly reject linear habit in favor of nonlinear habit. Moreover, even if the habit function is linear in current and past consumption, this equivalency will generally break down when consumption growth and $\mathbf{w}_t$ are serially dependent and when $R^f \neq 1/\delta$. The distinction between internal and external habits is important, not only because it is likely to have asset pricing implications at the aggregate level, but also because the two paradigms have dramatically different social welfare and tax policy consequences (Ljungqvist and Uhlig (2000)).[35]

One way to assess which specification better describes the data is simply to compute the value of the minimized SMD criterion function when habit formation is restricted to be external and compare it with that of the internal habit cases estimated above. This is done by estimating the model on the same set of moments imposing the restriction that $MU_t = (C_t - X_t)^{-\gamma}$. Doing so for each estimation described above, we find that the minimized SMD criterion is several orders of magnitude larger when marginal utility is restricted to external habit formation. For example, using Group 3 asset returns with a constant and the linear terms of $\mathbf{w}_t$ as instruments, the external habit case is found to be 1.2177e-04, compared to 1.3424e-07 for the internal habit case, about 1000 times larger. The estimations using Group 1 and Group 2 assets produced similar results. Ideally, of course, this comparison would be made on statistical grounds, but unfortunately criterion-based statistical tests have not been developed for procedures involving an unknown function.[36] Still, the sheer magnitude of the difference in minimized criteria suggests it unlikely the values would be judged the same statistically by any test.

An alternative way to address the question posed in the title of this subsection is to directly

[35] Campbell and Cochrane (1999) note that there are two possible interpretations of the representative agent habit-based paradigm. One is that the representative-agent preferences could take the same form as the underlying preferences of individual agents, another is that they could result from aggregation of heterogeneous agents with quite different preferences. Unfortunately, it is very difficult to conclusively determine which interpretation is more empirically plausible because of the severely limited nature of household-level consumption data. One careful study of habit formation on micro data is Dynan (2000). However, Dynan's conclusions bear out these data limitations as they pertain only to food consumption and to the simplest linear model of habit formation.

[36] Note that one could use the subsampling procedure to provide a statistical comparison between the minimized SMD criteria of the external habit and the internal habit, however, since the subsampling procedure is computationally very time-consuming, we do not implement it here.

test the Euler equations corresponding to the external habit formation:

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

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

$$E\left(\delta\left(\frac{C_{t+1}-X_{t+1}}{C_t-X_t}\right)^{-\gamma}R_{i,t+1}\widetilde{F}_{i,t+1}-1\mid \mathbf{w}_t\right)=0,\qquad i=1,...,N, \tag{25}$$

where

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

We note that both the external and the internal habit conditional moment restrictions are nested into

$$E\left(\delta\left(\frac{C_{t+1}-X_{t+1}}{C_t-X_t}\right)^{-\gamma}R_{i,t+1}\widetilde{F}_{i,t+1}(\beta;\delta,\gamma,g)-1\mid \mathbf{w}_t\right)=0,\qquad i=1,...,N, \tag{26}$$

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

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

It is clear that the external habit corresponds to (26) with $\beta=0$; the internal habit corresponds to (26) with $\beta=1$. Note that we still assume that X_t is a homogeneous of degree one function of current and lagged consumptions with finite lag length L. Nevertheless, we allow for one set of parameter values $(\delta^{ex},\gamma^{ex},g^{ex}(\cdot))$ satisfying the external habit conditional moment restrictions (24) and another set of parameter values $(\delta^{in},\gamma^{in},g^{in}(\cdot))$ satisfying the internal habit conditional moment restrictions (25). In particular we allow that $g^{ex}\left(\frac{C_{t-1}}{C_t},...,\frac{C_{t-L}}{C_t}\right)\neq g^{in}\left(\frac{C_{t-1}}{C_t},...,\frac{C_{t-L}}{C_t}\right)$.

We estimate all the unknown parameters $(\delta,\gamma,\beta,g(\cdot))$ corresponding to the generalized conditional moment restrictions (26) by the SMD method employed before, using the same ANN sieve (8) to approximate the unknown $g(\cdot)$, the same three asset groups with the associated sets of instruments. Again, the SMD estimates of $(\delta,\gamma,\beta,g(\cdot))$ converge in probability to the pseudo-true values $(\delta^*,\gamma^*,\beta^*,g^*(\cdot))$ identified by the generalized conditional moment restrictions (26); and the SMD estimates of finite-dimensional parameters δ,γ,β are root-T consistent and asymptotically normally distributed. Let $\widehat{\sigma}^2_\beta$ 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;[37] 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 $\widehat{\sigma}_\beta = 0.057$, we can easily reject the null hypothesis of $\beta = 0$ but fail to reject the null hypothesis of $\beta = 1$. It is clear that the stochastic discount factor is much better approximated by internal habit formation than external habit formation.

6.2 Model Comparison

We have estimated a habit-based consumption asset pricing model, allowing the habit to be a flexibly specified function of current and past consumption. Of interest is the question of how well habit-based models explain asset pricing data relative to other models that have been explored in the literature. We seek a methodology which recognizes that competing models are mere approximations of reality and therefore by definition false, and allows us to investigate which model provides the best approximation of the data. Such a methodology is provided by Hansen and Jagannathan (1997), who develop a way to compare asset pricing models when it is understood that the competing stochastic discount factors do not correctly price all portfolios. As Hansen and Jagannathan emphasize, pricing errors (given by the sequence $\{E(M_{t+1}R_{i,t+1}) - 1\}$, for any candidate M, and a set of N asset returns indexed by i) can arise either because the model is viewed formally as an approximation, or because the empirical counterpart to the theoretical specification is measured with error. In their approach, all stochastic discount factor models are treated as misspecified proxies for the true stochastic discount factor, and the relevant question is which model contains the least specification error.

We apply this approach to assess pricing errors for the habit-based framework considered in this paper, and compare its performance along this dimension to a number of alternative asset pricing models. Hansen and Jagannathan suggest that we compare the pricing errors of various candidate stochastic discount factor 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

[37] 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.

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. Hansen and Jagannathan (1997) also show that d_T gives the least-square distance between the candidate stochastic discount factor and the closest point to it in the set of all admissible stochastic discount factors (all stochastic discount factors that price assets correctly). 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.

To apply this procedure to the habit-based framework, we treat our SMD estimate of the habit as a proxy for the true habit, and minimize $\mathbf{g}_T^{HJ}(\boldsymbol{\theta})$ corresponding to the asset pricing model in (1-4) over the parameters $\boldsymbol{\theta} = (\delta, \gamma)'$ using the same quarterly sample that was used in the SMD estimation. Notice that we cannot obtain consistent estimates of the infinite dimensional habit function g using the Hansen-Jagannathan procedure.[38] Thus, we treat the habit estimated from the SMD procedure as part of the stochastic discount factor proxy, and compute the HJ distance for the habit models by only choosing the two scalar parameters δ and γ to minimize $\mathbf{g}_T^{HJ}(\boldsymbol{\theta})$. Recall that the SMD minimization gives greater weights to moments that are more highly correlated with the instruments $p^{J_T}(\mathbf{w}_t)$, while the Hansen-Jagannathan minimization matches unconditional moments. Because we fix the habit sieve parameters at their minimized SMD criterion values, (rather than reestimating them along with δ, γ to minimize the HJ distance), the habit model is placed at a disadvantage relative to the other models we use as a comparison, since for the comparison models all free parameters are chosen to minimize the HJ distance.[39]

[38] This is true both theoretically and as a practical matter. Theoretically, we cannot even identify the unknown habit function from a fixed finite set of unconditional population moments. Practically, even if we treated the neural net sieve function as a fully parametric specification, rather than recognizing that it is an approximation of an unknown function, it is not possible to choose the sieve parameters of the habit function to minimize $\mathbf{g}_T^{HJ}(\boldsymbol{\theta})$, which necessarily contains a large number of parameters. The reason is that doing so would require the use of significantly more asset returns in the estimation, requiring the second moment return weighting matrix to be of a sufficiently large dimension so as to render the weighting matrix nearly singular once inverted.

[39] One way to see that fixing the sieve parameters at their minimized SMD criterion values does not help the habit models in fitting the unconditional moments studied in this section, is to compare the HJ distance for the habit models when *all* of their parameters (including δ and γ) are fixed at their minimized SMD criterion values, with the HJ distance computed by minimizing $\mathbf{g}_T^{HJ}(\boldsymbol{\theta})$ over δ and γ. If this procedure did not place the habit models at a disadvantage, the HJ distances for these models, with parameters fixed at the values which minimize the SMD

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

$$M_{t+1} = \theta_0 + \sum_{i=1}^{k} \theta_i F_{i,t+1}, \tag{27}$$

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.[40] 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. The standard CCAPM of Breeden (1979) uses just the consumption growth rate as the single observable factor, but performs poorly empirically. By contrast, Lettau and Ludvigson (2001b) find that the scaled CCAPM performs about as well as the Fama-French model in explaining average returns on portfolios double-sorted on the basis of size and book equity-to-market equity on the aggregate stock market.

One possible interpretation of the Fama-French and scaled CCAPM models is that they approximate the stochastic discount factor of a consumption CAPM with habit formation of the type

criterion, should be no larger than the HJ distance for the same set of moments computed by minimizing $\mathbf{g}_T^{HJ}(\boldsymbol{\theta})$ over δ and γ. In fact, however, fixing the parameters at their SMD-estimated values makes the HJ distance metric much larger for the two sets of asset returns (six size/book-market returns, and six size/book-market returns plus the T-bill) upon which the models are evaluated below. For example, for the internal habit model and the six return case, the HJ distance with all parameters fixed at their SMD estimates is twice as large (equal to 0.339) as the HJ distance computed by minimizing $\mathbf{g}_T^{HJ}(\boldsymbol{\theta})$ over δ and γ (reported in Table 2 below), whereas it is almost three times as large (equal to 0.4768) for the seven return case.

[40] 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).

that generates time-varying risk aversion (Campbell and Cochrane (2000), Lettau and Ludvigson (2001b)). Because these models are not explicit structural models of the stochastic discount factor under habit formation, however, such a proposition can only be considered a conjecture. By estimating and evaluating a fully structural model of the stochastic discount factor under habit formation, we may provide direct empirical evidence on whether the empirical success of the Fama-French and scaled CCAPM models can be reasonably attributed to a representative-agent, habit-based asset pricing framework. 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 (27), $k = 1$) (Breeden and Litzenberger (1978)), and to those of the CAPM (with the market return R_m the single variable factor in (27), $k = 1$). For all linear models, the unknown coefficients θ_0 and θ_i are estimated by minimizing the corresponding squared HJ distance, $\mathbf{g}_T^{HJ}(\boldsymbol{\theta})$, for that model.

We evaluate the specification errors of the asset pricing models described above using a time-series on two alternative sets of quarterly returns: (i) the six equity returns on portfolios double-sorted on size and book-to-market characteristics provided by Fama and French, and (ii) these six equity portfolio returns plus the three-month Treasury bill rate. 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 seem to absorb the roles of leverage, earnings-to-price ratio and many other factors governing average stock return differentials. We include the Treasury-bill rate to assess how well the models explain average returns on a set of assets that also includes non-equity returns. Although Fama and French (1992) evaluate the CAPM on 25 size and book-market sorted portfolios, we follow Hansen and Jagannathan (1997) and evaluate the specification error of each model on smaller sets of six or seven portfolio returns. We do so because computation of the Hansen-Jagannathan distance requires an estimate of the second moment matrix of returns, and experience tells us that this matrix is poorly estimated in time-series samples of the size currently available when the number of portfolios, N, is too large (e.g., see Hansen, Heaton, and Yaron (1996), Ahn and Gadarowski (2004), Lettau and Ludvigson (2001b)). Our own experience shows that when N is reduced to about six or seven, estimates of the second moment matrix are reliable, but that larger numbers of test assets produce nearly singular weighting matrices. For this reason, we do our analysis on the two sets of assets mentioned above. The six size and book-to-market returns should continue to provide a good summary measure of the cross-section of U.S. equity returns, albeit at a more aggregated level than the 25 portfolios sorted on this basis. Below we present results from using the habit estimate of g generated from SMD estimation on Group 2 assets (six size and book-market returns plus the T-bill rate).[41]

[41] Estimation of the Hansen-Jagannathan distance for the internal habit model requires an estimate of the stochastic discount factor (11), which contains the conditional expectation of several terms in the denominator and numerator

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. The estimates of the curvature parameter γ, when freely estimated to minimize the Hansen-Jagannathan criterion function, are substantially larger than those using the SMD estimation, equal to 26 when the model is evaluated on the equity portfolios alone, and 25 when the Treasury bill is included. The larger values for γ are not surprising because the Hansen-Jagannathan procedure places emphasis on unconditional mean returns whereas the SMD procedure emphasizes conditional moments. Fitting unconditional moments requires a more volatile discount factor (Hansen and Jagannathan (1991)), which can be generated by a higher value for γ.

We also report the HJ distance for external habit formation, based on SMD estimation that restricts the last term in (3) to be zero. As for the internal habit case, we treat the habit estimated from the SMD procedure on the six size and book-market returns plus the T-bill as part of the stochastic discount factor proxy, and compute the HJ distance for the model by choosing the finite dimensional parameters δ and γ to minimize $\mathbf{g}_T^{HJ}(\boldsymbol{\theta})$. The resulting estimates of the curvature parameter γ, are equal to 37 when the Treasury bill is included, and 62 when it is excluded. Estimates of δ are more similar to the internal habit case, equal to 0.9 when the Treasury bill is included and 0.6 when it is omitted.

Table 2 reports the measure of specification error given by the Hansen-Jagannathan distance ("HJ Dist"), $d_T \equiv \sqrt{\mathbf{g}_T^{HJ}(\widehat{\boldsymbol{\theta}})}$, for all the models discussed above. There are interesting differences in the estimated specification error across the models. Regardless of whether the Treasury bill rate is included in the set of test assets, by a wide margin 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 models are assessed on equity returns alone, the scaled CCAPM delivers the second lowest specification error, equal to 0.21, but performs relatively worse when the T-bill is added (specification error = 0.35). For equity returns, the Fama-French model and the external habit model have almost identical values of the HJ distance, with the former 0.262 and the latter 0.261. The classic CCAPM and CAPM have errors substantially larger, equal to 0.31 and

of Ψ_{t+1}. An estimate of the stochastic discount factor is formed by replacing the terms in (11) over which a conditional mean is taken with their fitted values from a regression of those terms on the set of instruments used in the SMD estimation of g using Group 2 assets (six size and book-market equity returns, and the Treasury bill rate). Just as the terms themselves are functions of δ and γ, so are the expressions for their fitted values, and $\mathbf{g}_T^{HJ}(\boldsymbol{\theta}) \equiv \mathbf{w}_T'(\boldsymbol{\theta})\,\mathbf{G}_T^{-1}\mathbf{w}_T(\boldsymbol{\theta})$ can easily be minimized over $\boldsymbol{\theta} = (\delta, \gamma)'$.

0.34, respectively. But when the Treasury bill is included in the set of test assets, the external habit model fairs the worst, with a specification error of 0.43, the largest for any model. These results are particularly encouraging for the internal habit framework, especially since the model is placed at a disadvantage because the habit function is not reestimated to minimize the HJ distance.

We note that Table 2 only reports the point estimates of the HJ distances for the six competing asset pricing models without taking into account the statistical uncertainty. 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 permits the data to be governed by stationary beta-mixing processes, allowing a wide variety of nonlinear time-series processes such as diffusion models, stochastic volatility, nonlinear ARCH, GARCH, Markov switching, and many more. Luckily, we can apply the "reality check" method developed in White (2000) and perform statistical model comparison tests of the six competing models as follows. 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 $d_j^2 \equiv \mathbf{g}_j^{HJ}(\boldsymbol{\theta}_j^*)$ as the population squared HJ distance associated with model 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 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 $j = 1, 2, ..., 6$. By slightly modifying the result of Hansen, Heaton, and Luttmer (1995), we have, under mild regularity conditions, that

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

where $(Z_2, ..., Z_6)'$ is distributed as $\mathcal{N}(0, \Omega)$ for a positive semi-definite variance Ω. This justifies our applications of White's reality check test (White (2000)), $\mathcal{T}^W$, and Hansen's modified reality check test (Hansen (2003)), $\mathcal{T}^H$:

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

Both tests have complicated null limiting distributions. To implement the reality check test $\mathcal{T}^W$, White (2000) suggested to approximate its null limiting distribution using the stationary bootstrap of Politis and Romano (1994). Lately several papers including Hansen (2003) and Corradi and Swanson (2003) point out that the block bootstrap can also approximate its null limiting distribution. In the following we present two bootstrap test statistics: White's original bootstrap test $\mathcal{T}^{W,b}$ and a modified bootstrap test $\mathcal{T}^{H,b}$ suggested in Hansen (2003); both are implemented using the same block bootstrap procedure that we have employed in the test of linear habit.

Step 1: Draw a $b-$th block bootstrap sample of size T from the transformed raw data,[42] then compute the bootstrap estimates $d_{T,j}^{2,b} = \min_{\boldsymbol{\theta}_j} \mathbf{w}_{T,b}^j(\boldsymbol{\theta}_j)' \mathbf{G}_{T,b}^{-1} \mathbf{w}_{T,b}^j(\boldsymbol{\theta}_j)$ for $j = 1, 2, ..., 6$;

Step 2: Compute the $b-$th bootstrap estimates of the tests:

$$\mathcal{T}^{W,b} \equiv \max_{j=2,...,6} \sqrt{T} \left\{ (d_{T,1}^{2,b} - d_{T,j}^{2,b}) - (d_{T,1}^2 - d_{T,j}^2) \right\}.$$

$$\mathcal{T}^{H,b} \equiv \max \left(\max_{j=2,...,6} \sqrt{T} \left\{ (d_{T,1}^{2,b} - d_{T,j}^{2,b}) - (d_{T,1}^2 - d_{T,j}^2) 1\{ d_{T,1}^2 - d_{T,j}^2 \geq -\frac{\log(\log T)}{4\sqrt{T}} \} \right\}, 0 \right).$$

Step 3: Repeat Steps 1 - 2 for $b = 1, ..., B$ with $B = 500$ (say). Compute the bootstrap estimates of the $p - value$

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

The test $\mathcal{T}^W$ (or $\mathcal{T}^H$) using the squared HJ-distance is to reject null hypothesis if $\widehat{p}_W$ (or $\widehat{p}_H$) is close to zero and not to reject null if $\widehat{p}_W$ (or $\widehat{p}_H$) is close to one. We implemented both bootstrap tests. For the six return case $\widehat{p}_W = 0.9780$ and $\widehat{p}_H = 0.9120$, while for the six return plus T-bill case $\widehat{p}_W = 0.9926$ and $\widehat{p}_H = 0.8868$. Both suggest that the model 1 (the internal habit model) is the best according to the squared HJ distance measure.

Hansen and Jagannathan (1997), and more recently Wang and Zhang (2003), have noted that it is sometimes of interest to identify a stochastic discount factor from some initial set of asset returns and use it in pricing other assets, such as derivative claims on the payoffs of the initial assets. When this is the objective, Hansen and Jagannathan show that an alternate distance metric, one that restricts the set of admissible stochastic discount factors to be positive, should be used for model comparison. Table 3 reports this alternate distance metric, denoted "HJ$^+$ Dist," for all the models discussed above. The SMD-estimated internal habit model continues to beat all the other models by a wide margin, according to this metric.[43] The only models for which HJ$^+$ is substantially different

[42] Recall that the transformed raw data consists of consumption growth rates, asset returns, conditioning variables and various factors that can be regarded as drawn from stationary beta-mixing processes.

[43] We also perform the reality check model comparison tests of White and Hansen using the squared HJ$^+$-distance. This is simply done by replacing $d_{T,j}$ and $d_{T,j}^b$ in $\mathcal{T}^W$, $\mathcal{T}^H$ and Steps 1-3 by the estimates of HJ$^+$-distances $d_{T,j}^+$ and $d_{T,j}^{+b}$. Both bootstrap tests again suggest that the internal habit is the best according to the squared HJ$^+$ distance measure.

from HJ, are the scaled CCAPM and the classic CCAPM. For example, the scaled CCAPM metric 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.

Why does the internal habit model perform better than the external habit model? The stochastic discount factors 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).[44] Such time-series properties are likely to be important determinants of whether the estimated models can rationalize any time-variation in conditional moments of returns. This point is emphasized by Campbell and Cochrane (1999) who note that the slow-moving and persistent behavior of their calibrated habit model allows it to match evidence for long-horizon predictability in stock returns. Here, however, models are tested not on their ability to explain time-variation in selected conditional moments of returns, but on their ability to explain moment restrictions about the cross-section of asset returns implied by the theory itself, in this case a cross-section of unconditional mean returns on size and book-market sorted equity portfolios. Such theoretical moment restrictions dictate that any cross-sectional variation in unconditional mean excess returns must be proportional to the covariance of the stochastic discount factor with each return.[45]

What seems to drive the superior empirical performance of the internal habit SDF over the external SDF is the presence of the forward looking terms in marginal utility, present in internal habit formation but not in external (see (3)). Those forward-looking terms are captured empirically as projections on to instruments which are correlated with the nonlinear functions of future endogenous variables that appear in our moment conditions. As a diagnostic, we re-computed the HJ distance for the internal model using instead the ex-post values of the future terms in (3) and found that, in that case, the performance of the internal habit model was much closer to that of the external habit model. Thus, a key feature of the empirical success of the internal habit model lies with the forward-looking nature of marginal utility for that model.

[44] These numbers are for the stochastic discount factors estimated when computing the HJ distances–where only δ and γ are reestimated to minimize $g_T^{HJ}(\boldsymbol{\theta})$–not for the original SMD estimate of the stochastic discount factor.

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

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

The relative performance of the two models can also be intuitively understood by examining how the stochastic discount factors of each model (internal and external habit) covaries with the test asset returns. Table 4 shows sample estimates of the unconditional covariances, for each of the six size and book-market sorted portfolio returns. For reference, the top panel of Table 4 shows the average quarterly excess return on each portfolio over the three-month Treasury bill rate.

The first aspect of note is 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). Accordingly, the bottom panel of Table 4 shows that a big part of the explanation for why the internal habit model does better than the external 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.[46] Returns simply covary more with the estimated internal habit SDF than they do with the external habit SDF. The internal model also gets the cross-sectional patterns right. For example, the biggest 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 the latter. This is not the case for the external habit model, where these two covariances are about the same. Such patterns in the unconditional moments of returns and stochastic discount factors help explain intuitively why the internal habit model performs better than the external habit model according to the tests above.

To close this section, we note that when there is only one asset, a natural gauge of model misspecification is the pricing error associated with that asset. Hansen and Jagannathan (1997) stress that when there is more than one asset, comparing pricing errors is not possible without taking a stand on the relative importance of the various assets. A weighted average of the pricing errors is the natural measure of model misspecification. Nevertheless, it is sometimes useful to examine the individual pricing errors for each test asset, if only because it gives a sense of which assets are better priced than others. For reference, Table 5 gives the raw (unweighted) pricing errors for the Fama-French model and for the SMD-estimated internal habit model, when these models are evaluated on the size and book-market sorted returns.

7 Conclusion

Theories of asset pricing have developed in a number of new and interesting directions in recent years. Nevertheless, it could be argued that the theoretical possibilities multiply more rapidly than

[46] 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.

their empirical evaluation: formal estimation and testing of these new models is less common and often lags well behind their development. Data limitations, identification problems, and general econometric pitfalls are among the likely culprits responsible for the relative paucity of empirical work. In the case of the burgeoning literature on asset pricing theories of habit formation, empirical study is immediately confronted by the lack of agreement over the functional form of the habit specification. When such lack of agreement is present, econometric theory dictates that we treat the functional form of the habit, not as a given, but as an unknown parameter to be estimated along with the rest of the model's parameters.

In this article, we empirically evaluate a general class of representative-agent asset pricing models that derive their most salient implications from the presence of habit formation in investor preferences. Rather than choosing, from this literature, a particular functional form for the habit, we treat the habit specification as unknown and estimate it along with parameters governing curvature of the subutility function and the rate of time-preference. The resulting empirical model of investor utility is semiparametric, and consequently imposes few restrictions on the functional form of the habit in matching the joint distribution of aggregate consumption and asset returns implied by theory.

This semiparametric approach allows us to empirically evaluate a number of interesting hypotheses about habit-based asset pricing models that have previously not been evaluated. First, our results suggest that—conditional on the power utility framework—preferences are far from time separable: a flexibly specified habit constitutes 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, we strongly reject the hypothesis that habits are a pure externality governed by the consumption of everyone else in the economy; models of habits 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. We use the methodology of Hansen and Jagannathan (1997) to compare a stochastic discount factor proxy using a SMD-estimated habit formation, with proxies from a variety of alternative linear (or approximately linear) models that have been explored in the asset pricing literature. According to the Hansen and Jagannathan (1997) distance metric, a SMD-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.

There is at least one possible extension of our analysis that could be undertaken in future work. The specification of the habit may be treated as a recursive function of past habits, e.g., $X_t = r(C_t, C_{t-1}, X_{t-1})$, thereby allowing the habit stock to implicitly depend on an infinite number of past consumption lags. Such a specification would permit a change in focus, to an analysis of habit models and long-horizon aggregate stock-market returns, in which an extremely slow-moving habit is likely to be more important. The difficulty with this type of recursive estimation is that it might be slow to converge. Not only must the unknown habit, X_t, be estimated, but the recursive functional $r(\cdot)$ must also be estimated nonparametrically. The econometric theoretical results required to execute such an estimation have yet to be developed. Nevertheless, the empirical work in this paper is a natural starting place for such an investigation, because the estimation of X_t nonparametrically comprises one step in the recursive estimation procedure. We pursue this interesting work in future research.

8 Appendices

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. Appendix 3 describes the general sieve minimum distance (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

$$M_{t+1} = \delta_o \frac{MU_{t+1}}{MU_t}, \tag{28}$$

where

$$MU_t = \frac{\partial U}{\partial C_t} = (C_t - X_t)^{-\gamma_o} - E_t \left[\sum_{j=0}^{L} \delta_o^j (C_{t+j} - X_{t+j})^{-\gamma_o} \frac{\partial X_{t+j}}{\partial C_t} \right] \tag{29}$$

$$= (C_t - X_t)^{-\gamma_o} E_t \left\{ 1 - \sum_{j=0}^{L} \delta_o^j \left(\frac{C_{t+j} - X_{t+j}}{C_t - X_t} \right)^{-\gamma_o} \frac{\partial X_{t+j}}{\partial C_t} \right\}, \tag{30}$$

where

$$X_t = C_t f_o \left(1, \frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t} \right) = C_t g_o \left(\frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t} \right), \tag{31}$$

$$C_t - X_t = C_t \left\{ 1 - g_o \left(\frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t} \right) \right\} \tag{32}$$

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:

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

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

$$E_t \left(\delta_o \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma_o} R_{i,t+1} F_{i,t+1} - 1 \right) = 0, \qquad i = 1, ..., N, \tag{33}$$

with

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

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

Identification:

It is obvious that the presence of internal habit formation (i.e. $\widetilde{F}_{i,t+1}$ a nonlinear function of $(\delta_o, \gamma_o, g_o)$) will make the identification of $(\delta_o, \gamma_o, g_o)$ easier than the external habit (i.e. $\widetilde{F}_{i,t+1} = 1$), also that more lags (i.e. $L > 1$) will make the identification of $(\delta_o, \gamma_o, g_o)$ easier than the model with $L = 1$. Therefore it suffices that we study identification of the unknown true parameters $(\delta_o, \gamma_o, g_o)$

satisfying the conditional moment (33) under the special case of external habit with $L = 1$. The conditional moment restrictions for this special case becomes:

$$E\left(h_o\left(\frac{C_t}{C_{t+1}};\frac{C_{t-1}}{C_t}\right)R_{i,t+1}-1\mid \mathbf{w}_t\right)=0,\quad i=1,...,N, \tag{34}$$

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

$$\begin{aligned} h_o\left(\frac{C_t}{C_{t+1}};\frac{C_{t-1}}{C_t}\right) &= \delta_o\left(\frac{C_{t+1}}{C_t}\right)^{-\gamma_o}\frac{\left(1-g_o\left(\frac{C_t}{C_{t+1}}\right)\right)^{-\gamma_o}}{\left(1-g_o\left(\frac{C_{t-1}}{C_t}\right)\right)^{-\gamma_o}} \\ \text{with}\quad 0 &\leq g_o<1,\ g_o\neq const.,\ \gamma_o>0,\ \delta_o>0. \end{aligned} \tag{35}$$

We note that the conditional moment restriction (34) 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 (34), and we then use the semiparametric specification (35) and the identified $h_o()$ function to identify $(\delta_o, \gamma_o, g_o)$.

The conditional moment restriction (34) is very similar to the equation (2.3) in Newey and Powell (1988). In the following we denote $f(\frac{C_t}{C_{t+1}}, R_{j,t+1}|\mathbf{w}_t)$ as the conditional density of $(\frac{C_t}{C_{t+1}}, R_{j,t+1})$ given $\mathbf{w}_t$. 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 (34) is equivalent to:

there is a j from $\{1, ..., N\}$ such that

$$\begin{aligned} 0 &= \int\int \theta(y_{t+1};y_t)R_{j,t+1}f(y_{t+1},R_{j,t+1}|\mathbf{w}_t)dy_{t+1}dR_{j,t+1} \\ &= \int \theta(y_{t+1};y_t)\left\{\int R_{j,t+1}f(y_{t+1},R_{j,t+1}|\mathbf{w}_t)dR_{j,t+1}\right\}dy_{t+1} \end{aligned} \tag{36}$$

implies $\theta(y_{t+1}; y_t) = 0$ almost surely. Obviously (36) is equivalent to

$$0=\int \theta(y_{t+1};y_t)\left\{\int\frac{R_{j,t+1}}{E[R_{j,t+1}|\mathbf{w}_t]}f(y_{t+1},R_{j,t+1}|\mathbf{w}_t)dR_{j,t+1}\right\}dy_{t+1},$$

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

the j-th return adjusted conditional density $\int \frac{R_{j,t+1}}{E[R_{j,t+1}|\mathbf{w}_t]}f(\frac{C_t}{C_{t+1}}, R_{j,t+1}|w_t)dR_{j,t+1}$ of $\frac{C_t}{C_{t+1}}$ given $\mathbf{w}_t$ is complete.

See Newey and Powell (1988) for additional sufficient condition for the "completeness" in terms of exponential families. We note that the larger N, the easier to find such a j-th return satisfying the identification condition (36).

Next we show that $(\delta_o, \gamma_o, g_o)$ is identified given the identified $h_o()$ function and the specification (35). 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 (35) implies

$$\frac{h'_2\left(\frac{C_t}{C_{t+1}};\frac{C_{t-1}}{C_t}\right)}{\gamma\times h\left(\frac{C_t}{C_{t+1}};\frac{C_{t-1}}{C_t}\right)}=\frac{-g'\left(\frac{C_{t-1}}{C_t}\right)}{1-g\left(\frac{C_{t-1}}{C_t}\right)}\quad \textit{almost surely}, \tag{37}$$

$$\frac{h'_1\left(\frac{C_t}{C_{t+1}};\frac{C_{t-1}}{C_t}\right)}{\gamma\times h\left(\frac{C_t}{C_{t+1}};\frac{C_{t-1}}{C_t}\right)}=\left(\frac{C_t}{C_{t+1}}\right)^{-1}+\frac{g'\left(\frac{C_t}{C_{t+1}}\right)}{1-g\left(\frac{C_t}{C_{t+1}}\right)}\quad \textit{almost surely}. \tag{38}$$

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 (37) and (38), 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 (35):

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

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 (37) 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 (14). 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:

$$\mathcal{G}_T \equiv \left\{ g\left(x_1, ..., x_L\right) = \alpha_0 + \sum_{j=1}^{K_T} \alpha_j \psi\left(\Sigma_{l=1}^{L} \gamma_{j,l} x_l + \beta_j\right),\ 0 \leq g < 1 \right\}, \tag{39}$$

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:

$$\min_{\boldsymbol{\alpha} \in [\underline{\delta}, \overline{\delta}] \times [\underline{\gamma}, \overline{\gamma}] \times \mathcal{G}_T} \frac{1}{T} \sum_{t=1}^{T} \widehat{m}(\mathbf{w}_t, \boldsymbol{\alpha})' \widehat{\Sigma}(\mathbf{w}_t) \widehat{m}(\mathbf{w}_t, \boldsymbol{\alpha}), \tag{40}$$

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:

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

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 (40) 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 (41), the SMD criterion (40) becomes the GMM criterion (16).

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: $\beta(j) = \int \sup_{0 \leq \phi \leq 1} |E[\phi(Y_j)|Y_0 =$

$y] - E[\phi(Y_j)]|dF(y)$. Many financial time series econometric models satisfy *beta*-mixing; see, e.g., Doukhan (1994) for nonlinear ARX(p,q) and nonlinear ARCH models, Chen, Hansen and Carrasco (2001) for diffusion models, and Carrasco and Chen (2002) for GARCH and stochastic volatility models.

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 $\widehat{Q}(\theta) = \frac{1}{T}\sum_{t=1}^{T} \widehat{m}(\mathbf{w}_t, \boldsymbol{\alpha})'\widehat{m}(\mathbf{w}_t, \boldsymbol{\alpha})$, their $Q(\theta) = E\{m(\mathbf{w}_t, \boldsymbol{\alpha})'m(\mathbf{w}_t, \boldsymbol{\alpha})\}$, their parameter θ is our $\boldsymbol{\alpha}$, 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 $\widehat{Q}(\theta)$ and $Q(\theta)$ 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

$$\sup_{[\underline{\delta},\overline{\delta}]\times[\underline{\gamma},\overline{\gamma}]\times\mathcal{G}} \left| \frac{1}{T} \sum_{t=1}^{T} \widehat{m}(\mathbf{w}_t, \boldsymbol{\alpha})'\widehat{m}(\mathbf{w}_t, \boldsymbol{\alpha}) - E\{m(\mathbf{w}_t, \boldsymbol{\alpha})'m(\mathbf{w}_t, \boldsymbol{\alpha})\} \right| = o_p(1). \tag{42}$$

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 (**??**).

These unknown pseudo-true values $(\delta^*,\gamma^*,\beta^*,g^*(\cdot))$ can again be estimated by the SMD method using the same ANN sieve (8) 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 (41) 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 $g^o_{11}()$ denote the second partial derivative of the true $g_o()$ with respect to its first argument. Then

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

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. Then, under some mild additional conditions, we have:

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

The asymptotic normality result of $\sqrt{T}\left(\widehat{\mu}_g - \mu_g\right) \longrightarrow^D \mathcal{N}(0, \sigma^2_{11})$ can be obtained by directly applying Ai and Chen (2004). Nevertheless, the above asymptotic expansion indicates that its limiting variance σ^2_{11} will involve terms like g^o_{11} and f_{11}, hence is complicated.

References

Abel, A. B. (1989): "Asset Prices Under Heterogeneous Beliefs: Implications for the Equity Premium," Unpublished paper, Wharton School, University of Pennsylvania.

——— (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.

Ahn, S. C., and C. Gadarowski (2004): "Small Sample Properties of the GMM Specification Test Based on the Hansen-Jagannathan Distance," *Journal of Empirical Finance,* forthcoming, 11, 109–132.

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

——— (2004): "Estimation of Smooth Functionals of Semiparametric Conditional Moment Restrictions with Different Conditioning Information Sets," Unpublished paper, New York University.

Attanasio, O. P., J. Banks, and S. Tanner (2002): "Asset Holding and Consumption Volatility," *Journal of Political Economy,* forthcoming.

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.

Barberis, N., A. Shleifer, and R. W. Vishny (1998): "A Model of Investor Sentiment," *Journal of Financial Economics*, 49(3), 307–43.

Barsky, R. B., and B. De Long (1993): "Why Does the Stock Market Fluctuate?," *Quarterly Journal of Economics*, 108(2), 291–311.

Blundell, R., and J. L. Powell (2001): "Endogeneity in Nonparametric and Semiparametric Regression Models," Unpublished paper, University of California, Berkeley.

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.

Brav, A., G. M. Constantinides, and C. C. Geczy (2002): "Asset Pricing with Heterogeneous Consumers and Limited Participation: Empirical Evidence," *Journal of Political Economy,* forthcoming, 110.

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 (1988): "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies*, 1, 195–227.

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

Cecchetti, S. G., P. Lam, and N. C. Mark (2000): "Mean Reversion in Equilibrium Asset Prices," *American Economic Review*, 80(3), 398–418.

Chan, Y. L., and L. Kogan (2002): "Catching Up With the Joneses: Heterogeneous Preferences and the Dynamics of Asset Prices," *Journal of Political Economy,* forthcoming.

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.

——— (2002): "Rational Asset Pricing," *The Journal of Finance*, 57(4), 1567–1591.

Constantinides, G. M., J. B. Donaldson, and R. Mehra (2002): "Junior Can't Borrow: A New Perspective on the Equity Premium Puzzle," *Quarterly Journal of Economics*, 117, 269–296.

Constantinides, G. M., and D. Duffie (1996): "Asset Pricing With Heterogeneous Consumers," *Journal of Political Economy*, 104, 219–40.

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.

Darolles, S., J. Florens, and E. Renault (2002): "Nonparametric Instrumental Regression," Unpublished paper, University of Montreal.

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

Dumas, B. (1989): "Two-Person Dynamic Equilibrium in the Capital Market," *Review of Financial Studies*, 2(2), 157–188.

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.

Grossman, S. J., and Z. Zhou (1996): "Equilibrium Analysis of Portfolio Insurance," *Journal of Finance*, 51(4), 1379–1403.

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

Hansen, L. P., J. Heaton, and A. Yaron (1996): "Finite Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics*, 14, 262–280.

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., T. J. Sargent, and T. D. Tallarini (1999): "Robust Permanent Income and Pricing," *Review of Economic Studies*, 66(4), 873–907.

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.

Heaton, J., and D. Lucas (1996): "Evaluating the Effects of Incomplete Markets on Risk Sharing and Asset Pricing," *Journal of Political Economy*, 104(3), 443–87.

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.

Jagannathan, R., G. Skoulakis, and Z. Wang (2002): "Generalized Method of Moments: Applications in Finance," *Journal of Business and Economic Statistics*, 20(4), 470–481.

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

Judd, K. L. (1998): *Numerical Methods in Economics*. The MIT Press, Cambridge, Massachusetts.

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

Kogan, L., and R. Uppal (2002): "Asset Prices in a Heterogenous-Agent Economy with Portfolio Constraints," Unpublished Paper, Sloan School of Management, MIT.

Krusell, P., and A. A. Smith (1997): "Income and Wealth Heterogeneity, Portfolio Choice, and Equilibrium Asset Returns," *Macroeconomic Dynamics*, 1(2), 387–422.

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,* forthcoming.

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.

Lustig, H., and S. V. Nieuwerburgh (2002): "Housing Collateral, Consumption Insurance and Risk Premia," Unpublished manuscript, University of Chicago.

McGrattan, E. R. (1998): "Application of Weighted Residual Methods to Dynamic Economic Models," Federal Reserve Bank of Minneapolis Research Department Staff Report 232.

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.

Otrok, C., B. Ravikumar, and C. H. Whiteman (2002): "Habit Formation: A Resolution of the Equity Premium Puzzle?," *Journal of Monetary Economics*, 49, 1261–288.

Piazzesi, M., M. Schneider, and S. Tuzel (2002): "Housing, Consumption, and Asset Pricing," Unpublished paper, UCLA.

Politis, D., and J. Romano (1994): "The Stationary Bootstrap," *American Statistical Association*, 89, 1303–1313.

Sandroni, A. (1999): "Asset Prices and the Distribution of Wealth," *Economic Letters*, 64(2), 203–207.

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.

Vissing-Jrgensen, A. (2002): "Limited Asset Market Participation and Intertemporal Substitution," *Journal of Political Economy,* forthcoming.

Wachter, J. (2002): "Habit Formation and Returns on Bonds and Stocks," Unpublished paper, Stern School of Business, New York University.

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.

Yogo, M. (2003): "A Consumption-Based Explanation of the Cross Section of Expected Stock Returns," Unpublished paper, Harvard University.

9 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.

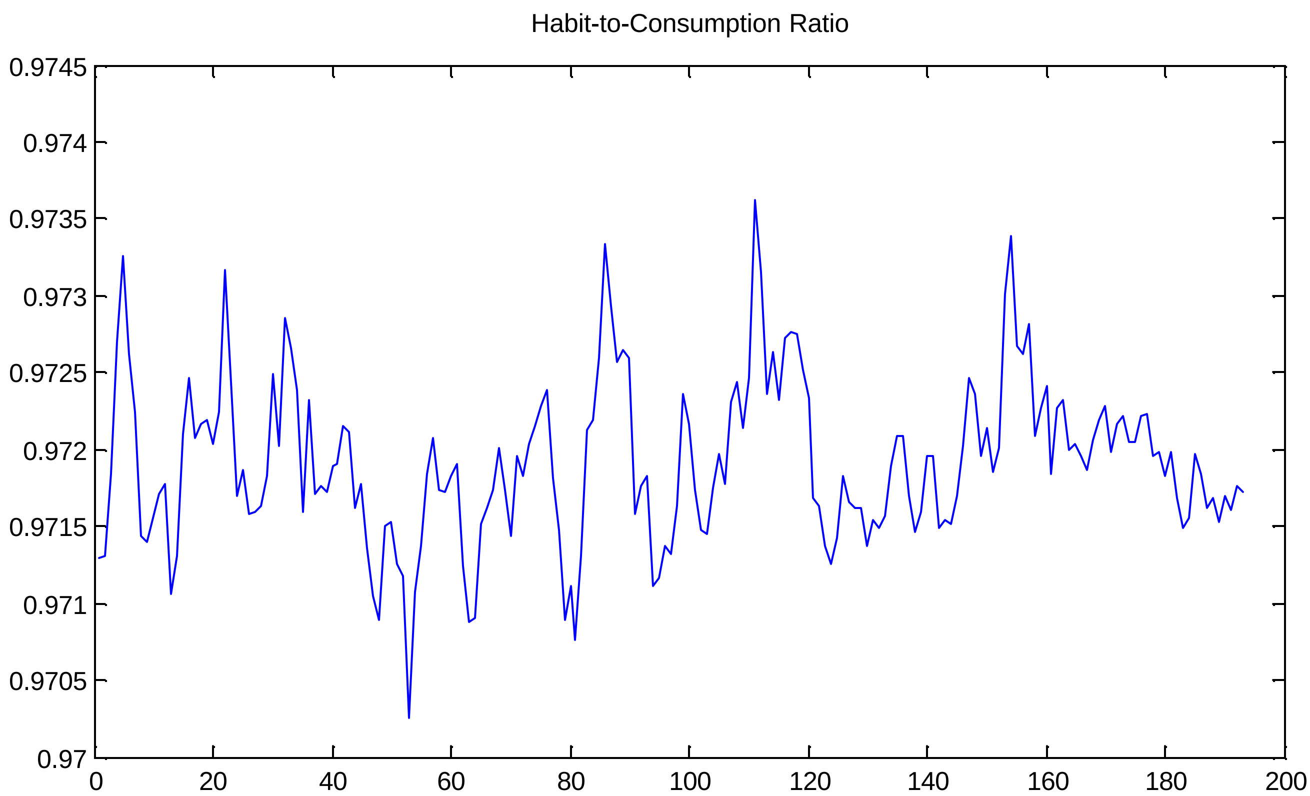
Table 5

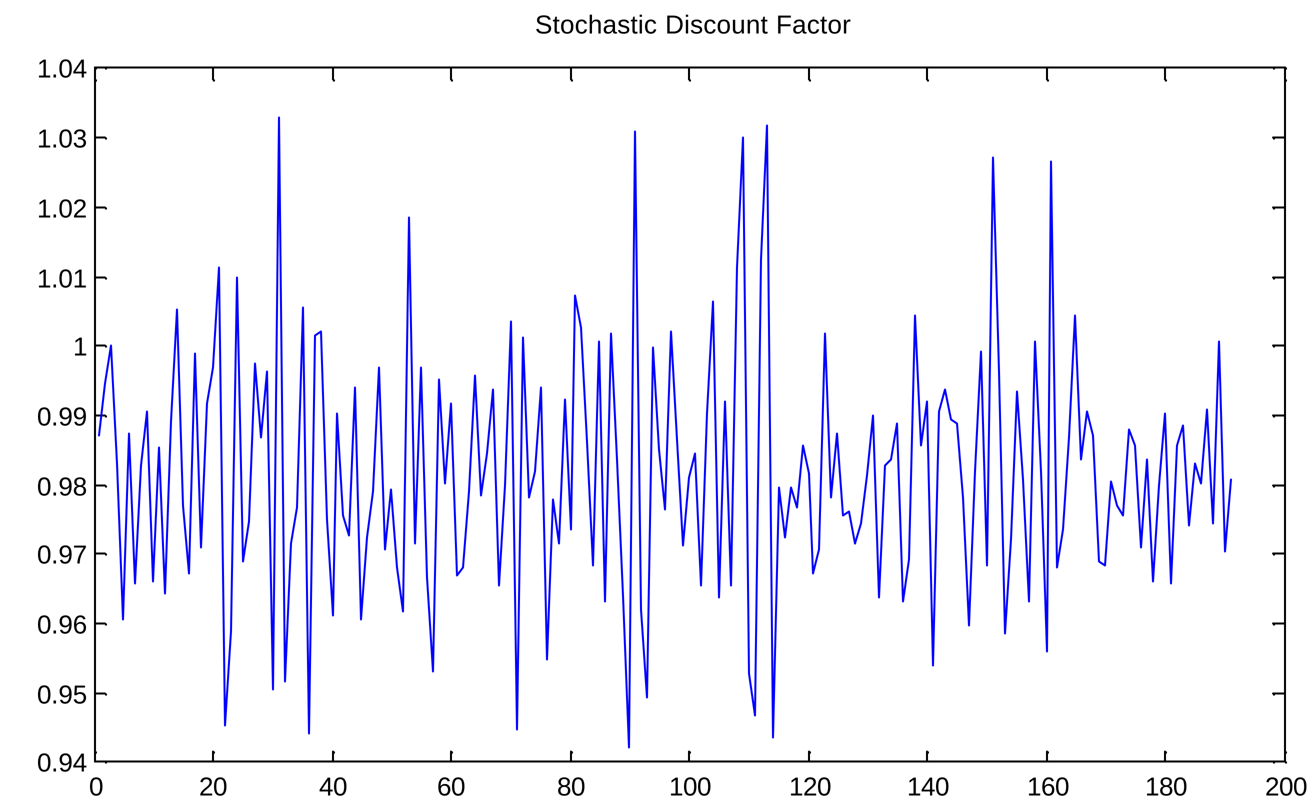
Pricing Errors for 6 Size/BM Returns

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.

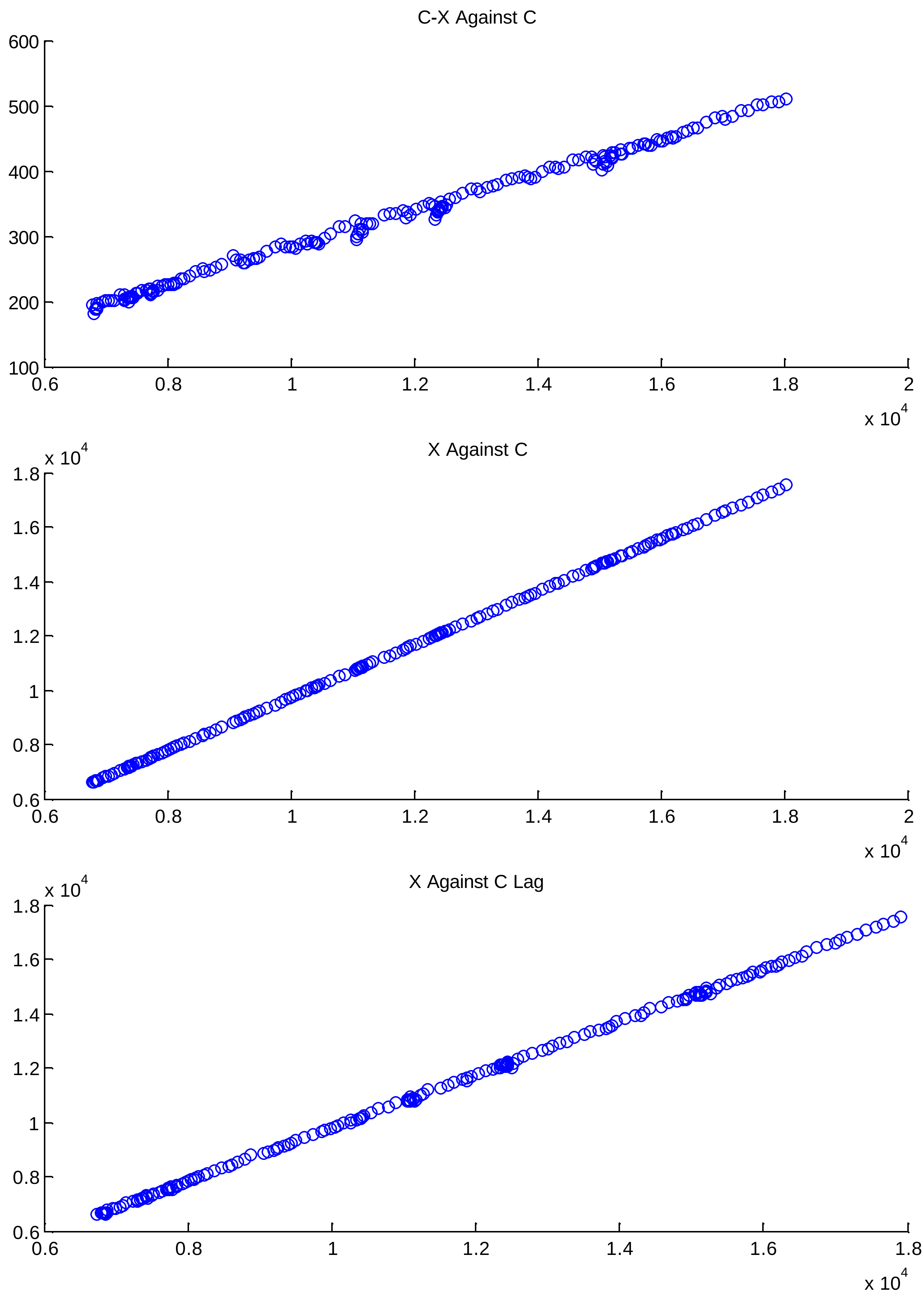
FIGURE 1





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.

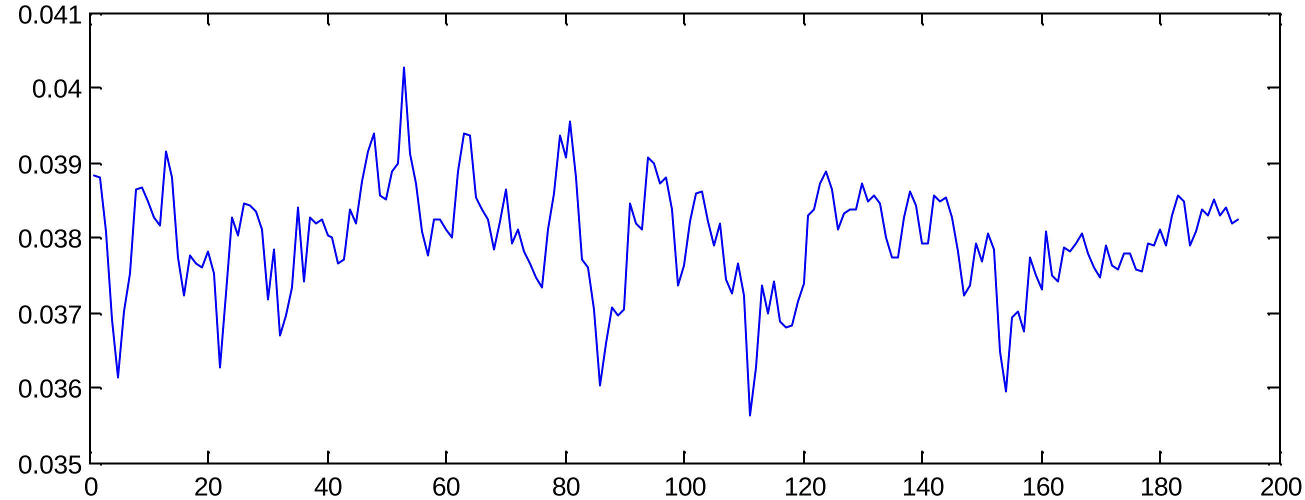
FIGURE 2



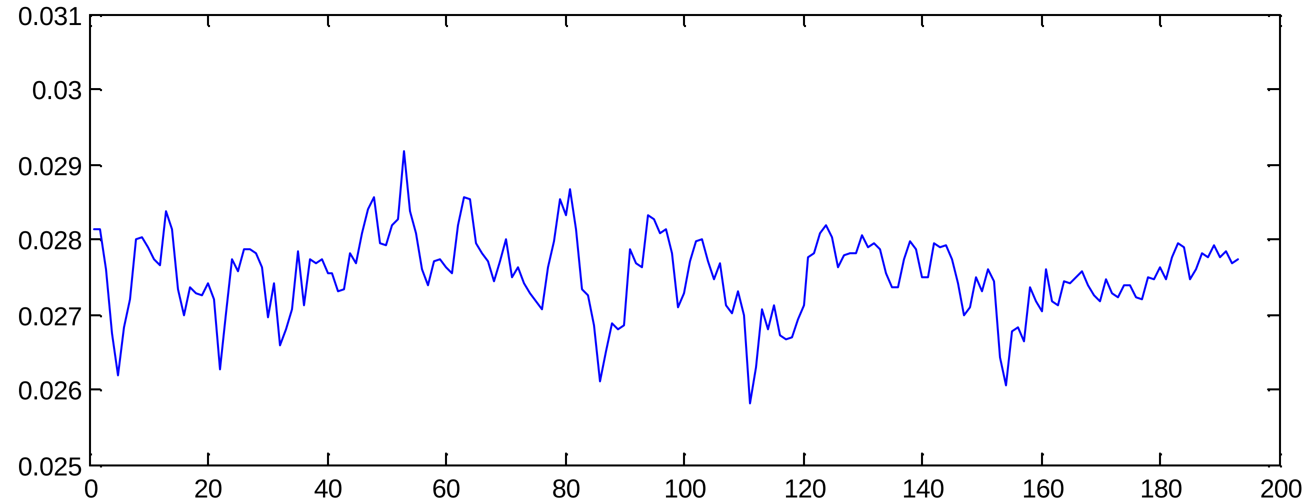
Notes : X is the estimated habit, C is consumption. Estimates use Group 1 assets, linear and squared instruments.

FIGURE 3

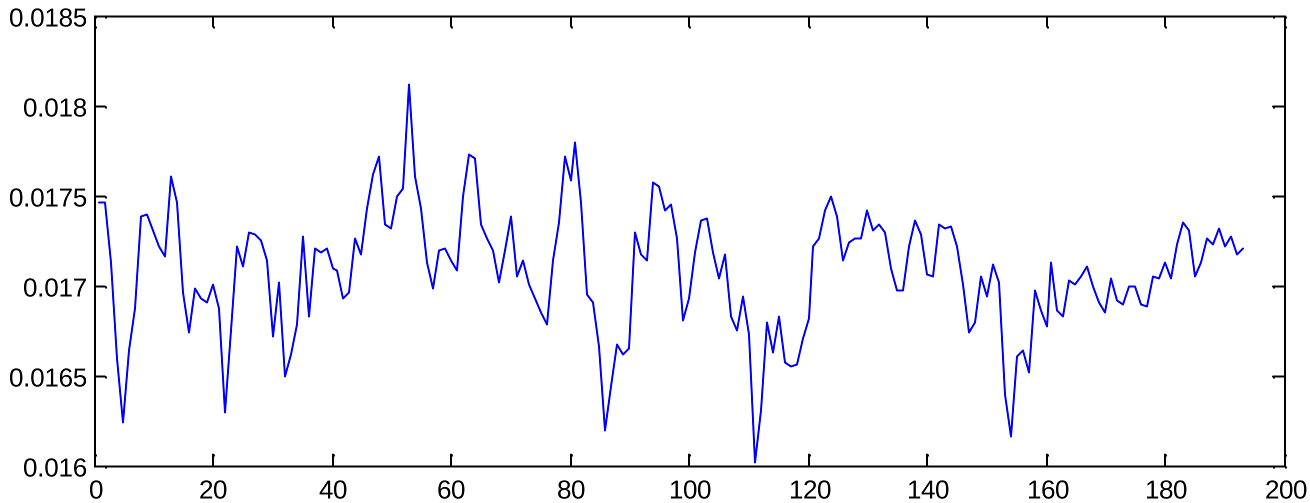
First Derivative Of Habit One Period Ahead With Respect to Consumption



First Derivative Of Habit Two Periods Ahead With Respect to Consumption

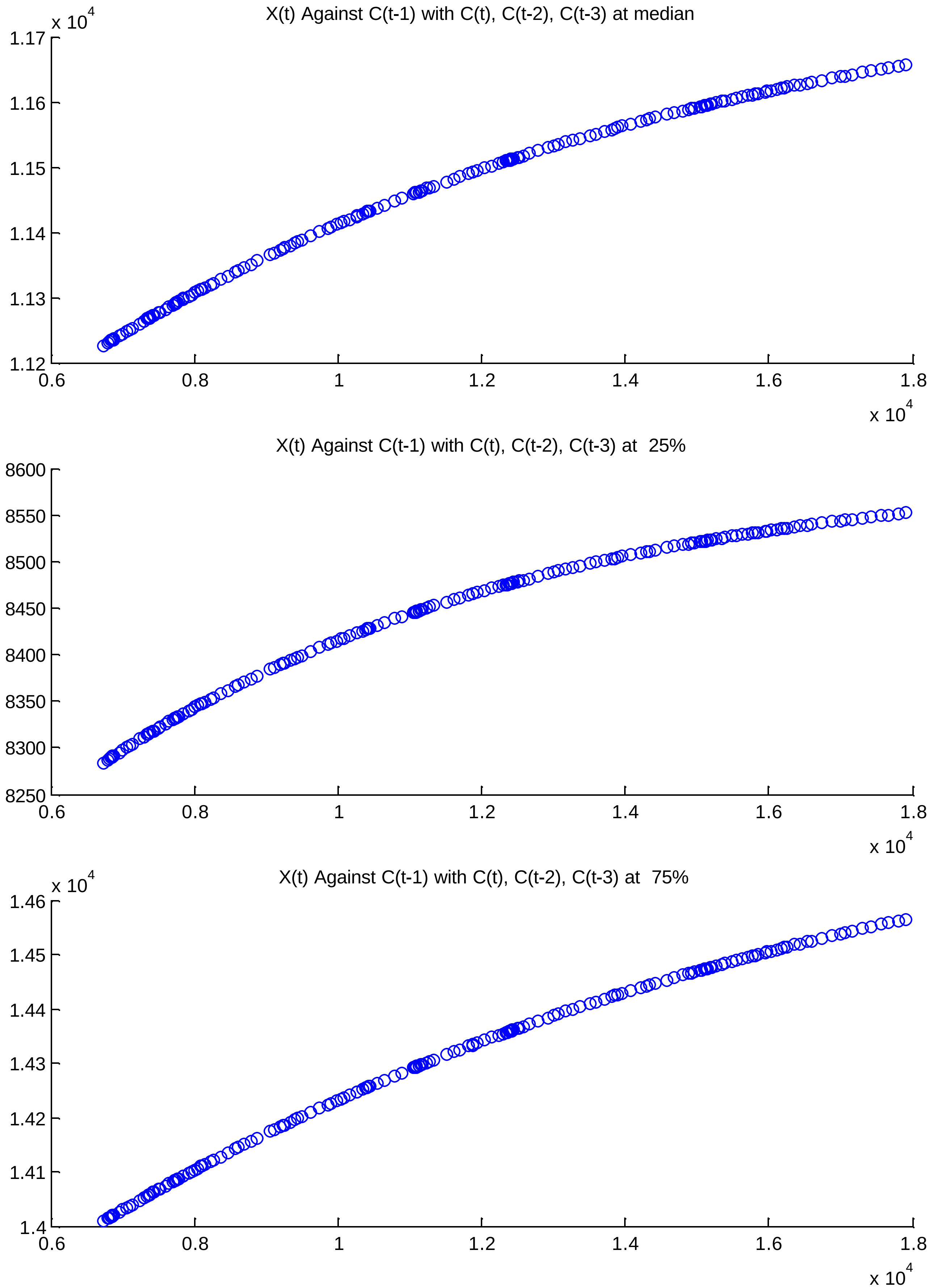


First Derivative Of Habit Three Periods Ahead With Respect to Consumption



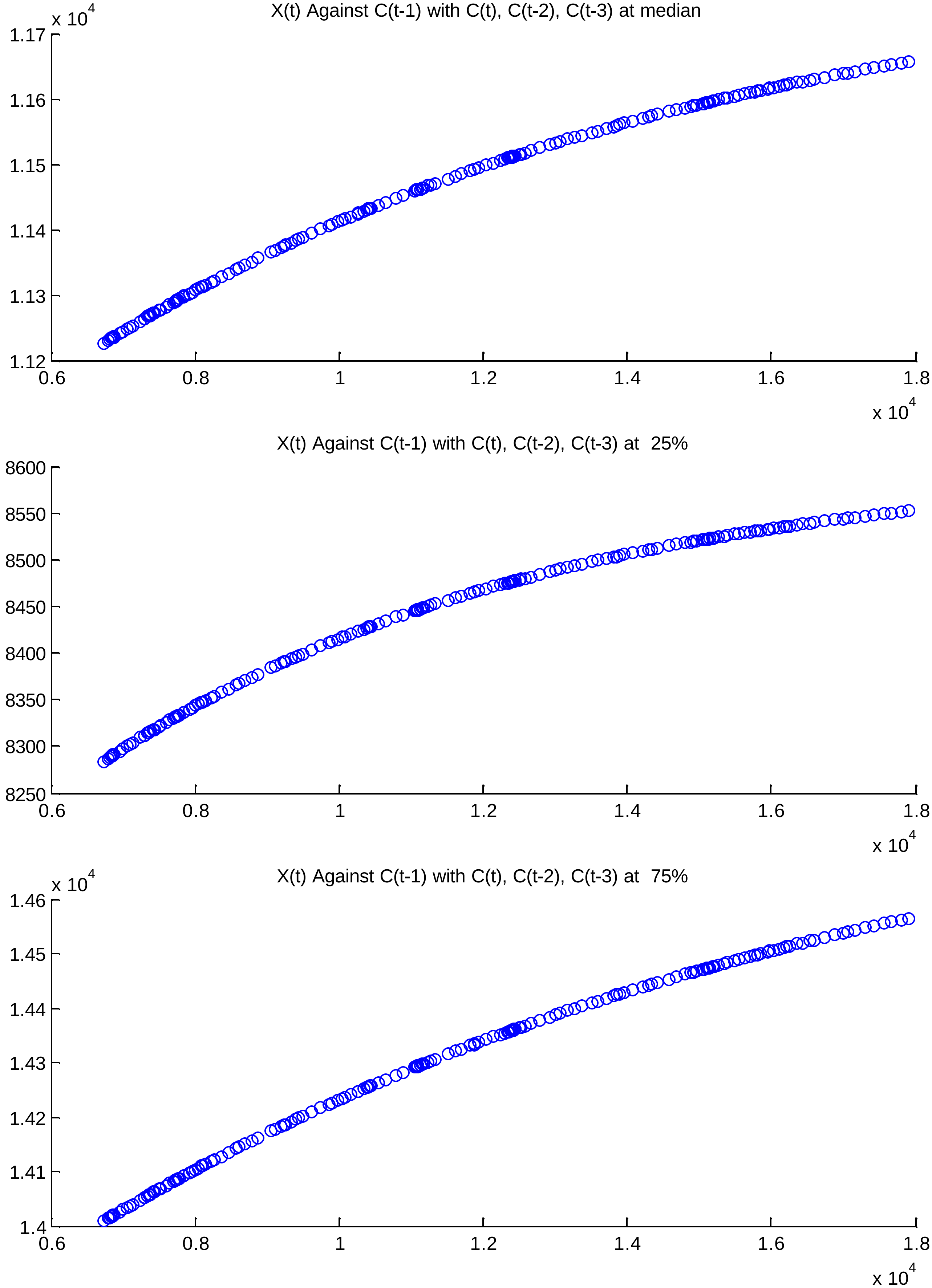
Notes : Estimates use Group 1 assets, linear and squared instruments.

FIGURE 4



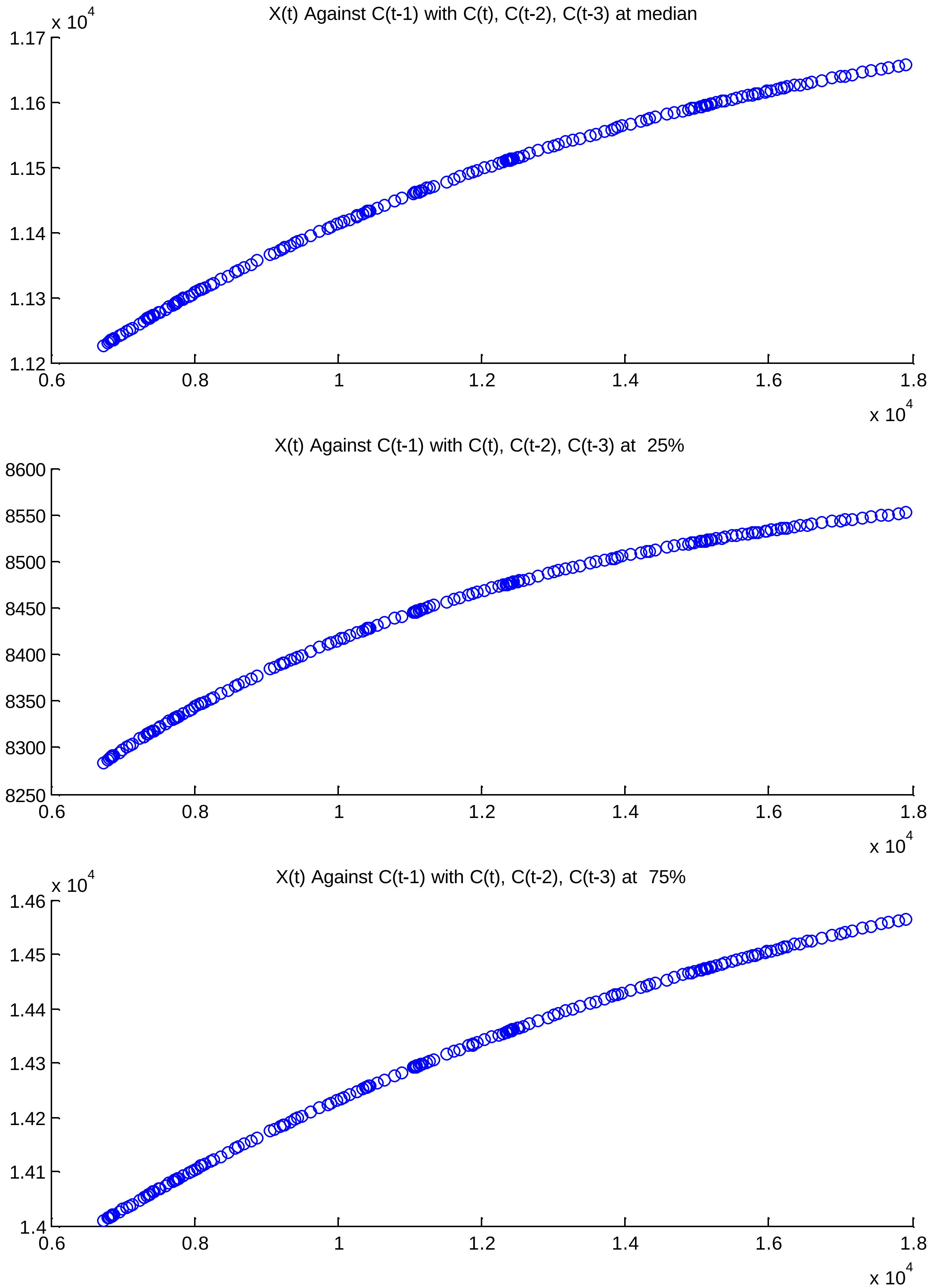
Notes : X is the estimated habit, C is consumption. Estimates use Group 1 assets, linear and squared instruments.

FIGURE 5



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

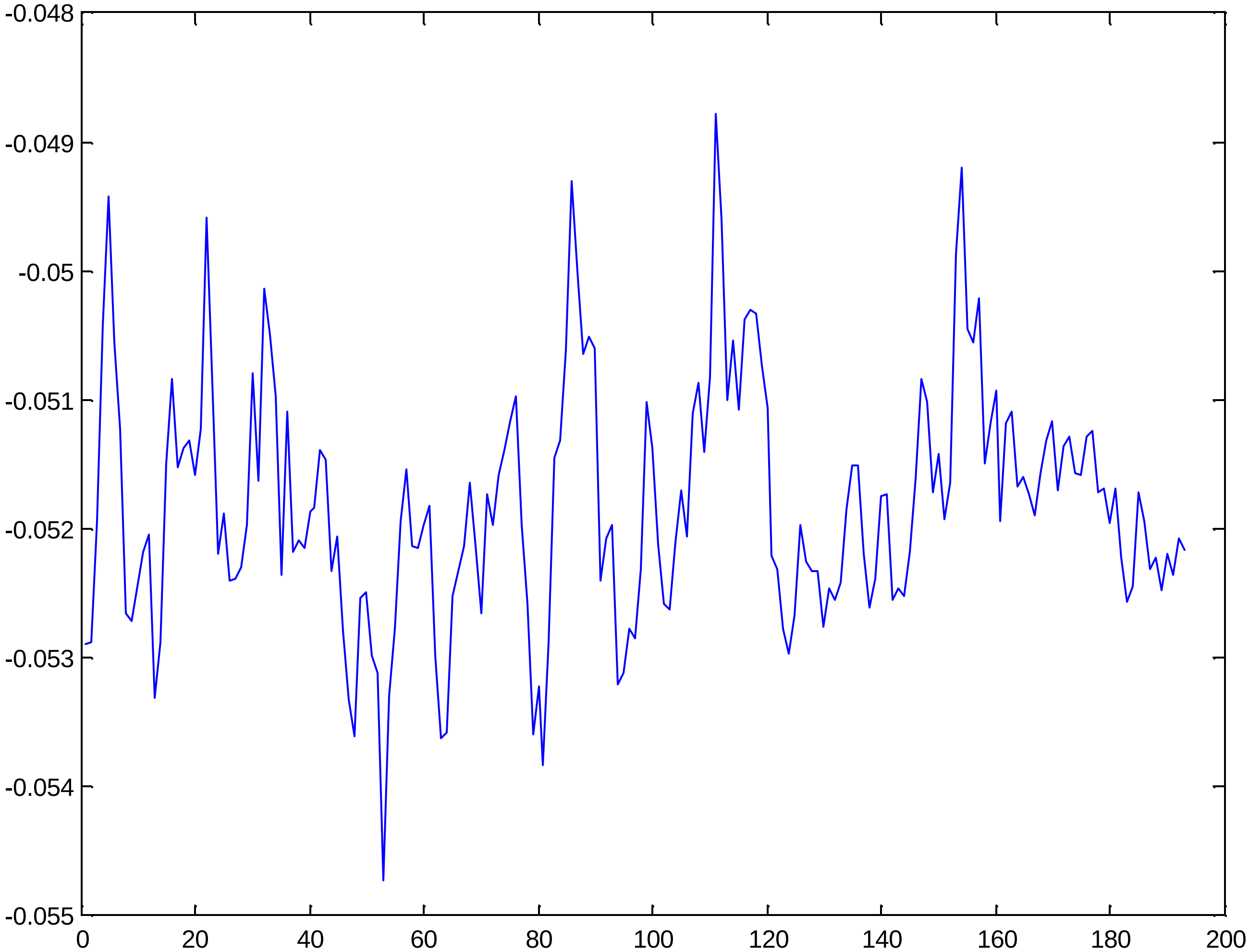
FIGURE 6



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

FIGURE 7

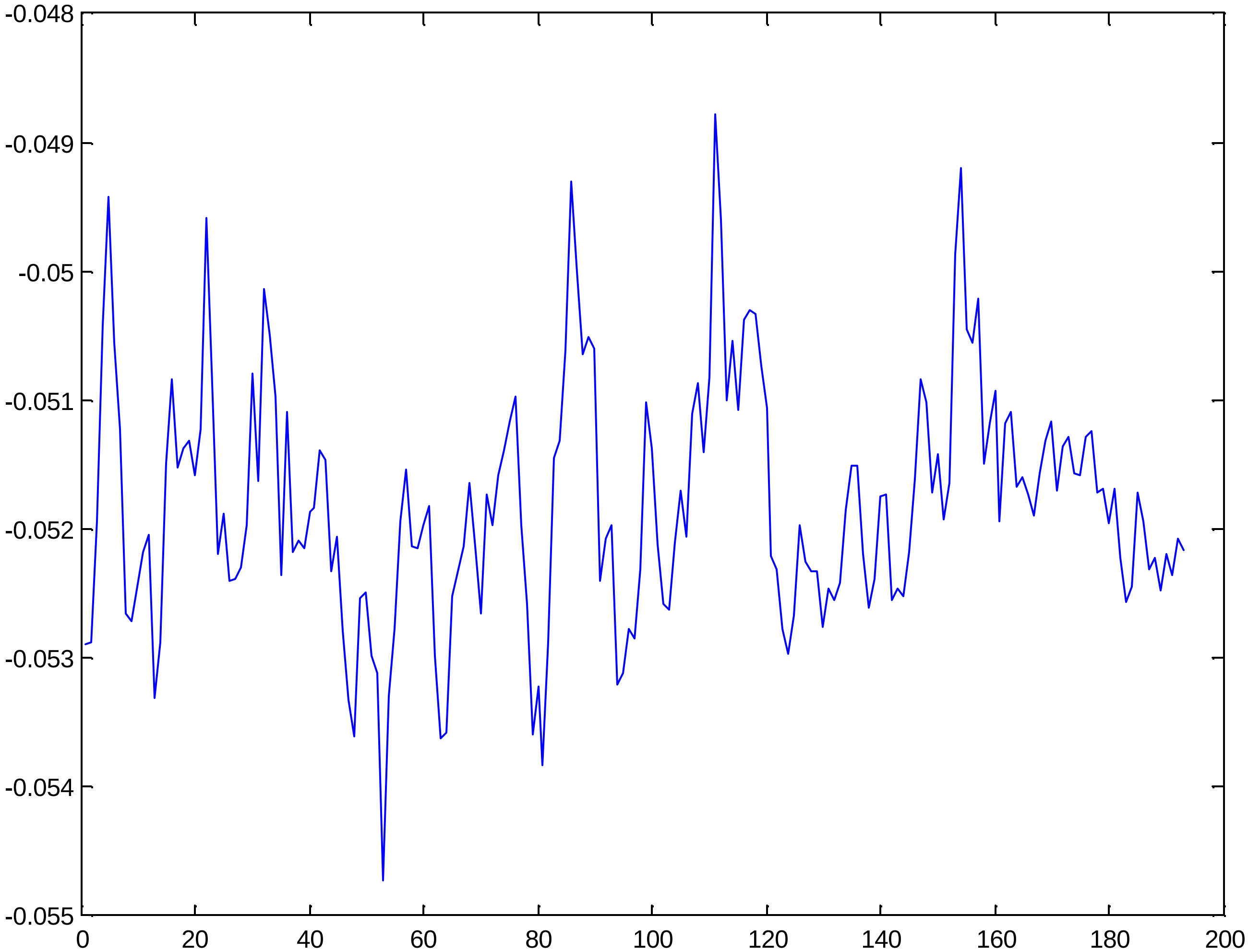
Second Derivative Of Habit One Period Ahead With Respect To Consumption



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

FIGURE 8

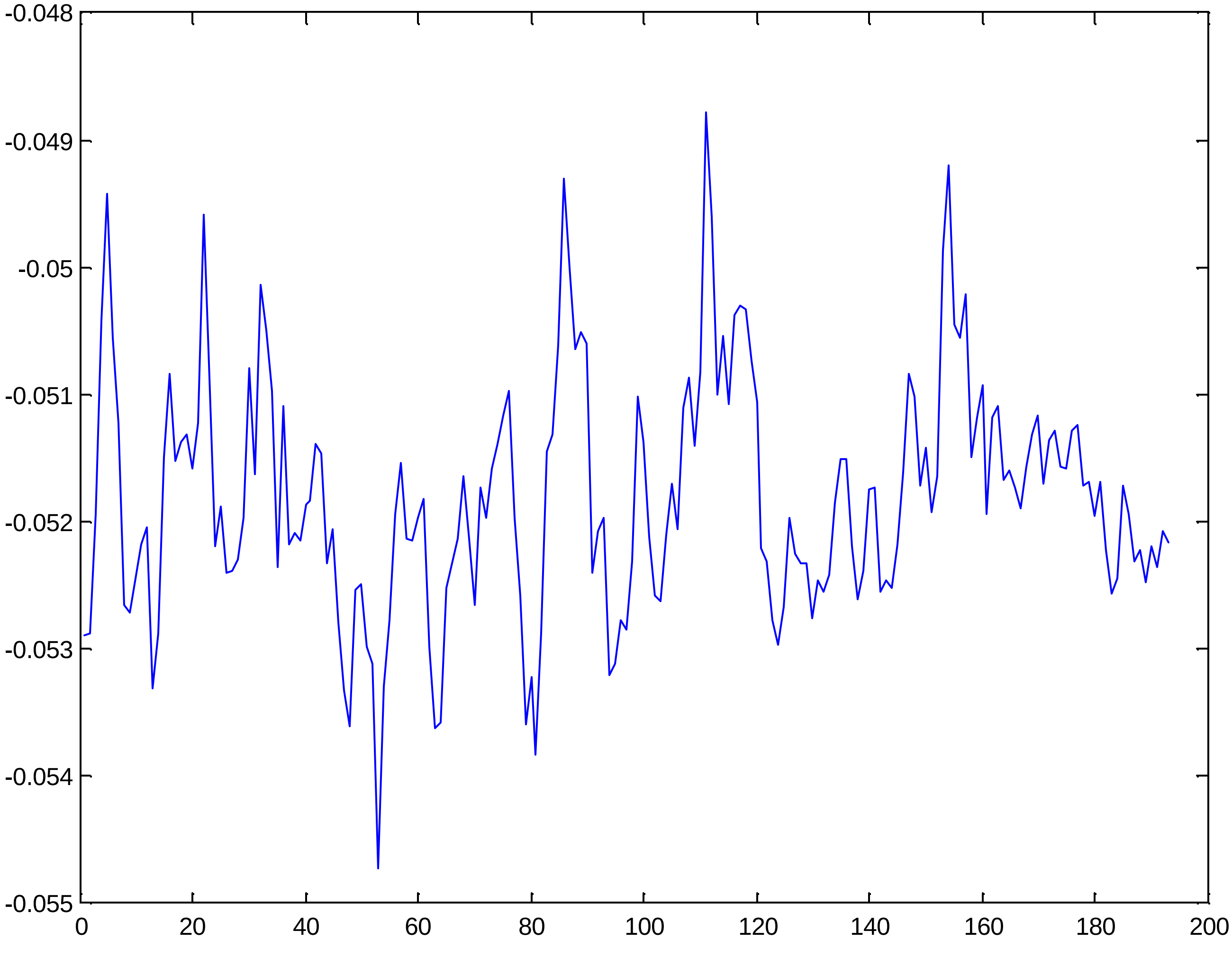
Second Derivative Of Habit One Period Ahead With Respect To Consumption



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

FIGURE 9

Second Derivative Of Habit One Period Ahead With Respect To Consumption



Notes : Estimates use Group 3 assets, linear instruments.