Land of Addicts? An Empirical Investigation of Habit-Based Asset Pricing Models*

Xiaohong Chen
New York University

Sydney C. Ludvigson
New York University and NBER

Preliminary and Incomplete

Comments Welcome

First draft: September 26, 2002
This draft: June 19, 2003

*Chen: Department of Economics, New York University, 269 Mercer Street, 7th Floor, New York, NY 10003; Email: xiaohong.chen@nyu.edu, Tel: (212) 998-8970; Fax: (212) 995-4186. Ludvigson: Department of Economics, New York University, 269 Mercer Street, 7th Floor, New York, NY 10003; Email: sydney.ludvigson@nyu.edu, Tel: (212) 998-8927; Fax: (212) 995-4186. 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 and from the National Science Foundation. Both authors acknowledge financial support form the C.V. Starr Center at NYU. We are especially grateful to John Y. Campbell, John Cochrane, and Lars P. Hansen for helpful suggestions on this project. We also thank Kenneth French for providing the portfolio return data, Chris Flinn, Per Krusell, Anthony Lynch, Mark Gertler and seminar participants at NYU and Carnegie Mellon GSIA for useful comments. Jinyong Kim and Artem Voronov provided excellent research assistance. Any errors or omissions are the responsibility of the authors.

Land of Addicts? An Empirical Investigation of Habit-Based Asset Pricing Models

Abstract

A leading explanation of aggregate stock market behavior suggests 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 some function of lagged and (possibly) contemporaneous consumption. But theory does not provide precise guidelines about the parametric functional relationship between the habit and aggregate consumption. This makes formal estimation and testing challenging; at the same time, it raises an empirical question about the functional form of the habit that best explains asset pricing data.

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 parameters. 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. This semiparametric approach allows us to empirically evaluate a number of interesting hypotheses about the specification of habit-based asset pricing models, and to formally test the framework's ability to explain stock return data relative to other models that have proven empirically successful.

JEL: G12, C14, C52

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] and irrational expectations.[4] 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.[5]

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 excercise, 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.[6]

[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-Jørgensen (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, Barsky and De Long (1993), Barberis, Shleifer, and Vishny (1998), Hansen, Sargent, and Tallarini (1999), Cecchetti, Lam, and Mark (2000).

[5] 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 (2003). We discuss these papers further below.

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

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 5. 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., 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 (2003)). 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.

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 and asset returns implied by asset pricing theory.

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 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 conditional moment models developed in 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)).[7]

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 (Grenader (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 will provide assumptions under which the sieve estimator for our application is consistent with a reasonably fast convergence rate (under certain metric), while 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 hypothesis concerns whether the habit is better described as a linear function, as in the work of Sundaresan (1989), Constantinides (1990), Heaton (1995), Jermann

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

(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 accomodate a variety of settings (e.g., Campbell and Cochrane (2000); Li (2001); Wachter (2002); and Menzly, Santos, and Veronesi (2003)). 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 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 (1999), Campbell and Cochrane (2000), Li (2001), Shore and White (2002), Wachter (2002), and Menzly, Santos, and Veronesi (2003) 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 Jonses." 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 some long-standing asset-allocation puzzles in the international finance literature (e.g., see Ljungqvist and Uhlig (2000) and Shore and White (2002)). We develop a consistent statistical test of the null hypothesis of external habit formation against the alternative of internal habit formation. Our results indicate that we may overwhelmingly reject the external habit specification in favor of internal habit formation.

Finally, our approach allows us to assess the quantitative importance of the habit in the power utility specification. Our 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.

How well does the habit-based framework fit the asset pricing data? We formally evaluate the habit-based framework by employing non-nested model comparison tests based on computation of the Hansen-Jagannathan distance (Hansen and Jagannathan (1997)) for several asset pricing models. This procedure is designed to ask how well the estimated habit model satisfies the restrictions implied by asset pricing theory compared to that of other asset pricing models that have demonstrated empirical success in explaining expected stock market returns. The intention of these tests is not to determine whether the habit-based model is literally true (since all models are approximations of reality and therefore by definition false), but rather to test the hypothesis that the habit-based paradigm provides a better approximation of the data than do other asset pricing models proposed in the literature. By allowing the habit to depend flexibly on current and past consumption, we minimize the chance that the model is rejected merely because the functional form of the habit is misspecified.

The nonnested tests we pursue compare the 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 (CCAPM) explored in Lettau and Ludvigson (2001b). This procedure allows us to explicitly test whether the empirical success of these models can be reasonably attributed to the representative-agent, habit-based asset pricing framework now prevalent in the literature.

To our knowledge, there has been only a small amount of prior work applying sieve (or series) nonparametric estimation techniques to asset pricing questions. Gallant and Tauchen (1989), building off the work of Elbadawi, Gallant, and Souza (1983) and Gallant and Nychka (1987), employed a "seminonparametric" modeling approach based on series expansions to estimate a consumption-based asset pricing model. Gallant, Hansen, and Tauchen (1990) employ the same seminonparametric methodology used in Gallant and Tauchen (1989), 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) use the seminonparametric methodology to estimate a nonlinear arbitrage-pricing model. Our study differs from these in three ways. First, in contrast to Gallant and Tauchen (1989) and Bansal and Viswanathan (1993), we place more structure on the empirical model by embedding the unknown habit function in the more familiar power-utility framework. 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) approximate the whole pricing kernel as a function of a few macroeconomic factors. The more structural approach taken in this paper allows us to investigate a number of interesting hypotheses about the habit-based framework studied in the theoretical literature that have not been investigated elsewhere. Second, 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 moment directly using known basis functions. Third, we provide asymptotic justification for our results by extending the work of Newey and Powell (2003) on the consistency of the purely nonparametric SMD estimator and the work of Ai and Chen (2003) on the $\sqrt{T}$ asymptotic normality of the semiparametric SMD estimator.

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. Section 6 concludes.

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 the *ratio* of consumption to the habit stock, as in Abel (1990) and Abel (1999). Ratio models of external habit formation have difficulty accounting for the predictability of excess stock returns documented in the empirical asset pricing literature, since they imply relative risk-aversion is constant.[8] By contrast, difference

[8]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

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 5 of this paper.

Identical agents maximize the utility function

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

Here X_t is the level of the habit, and δ is the 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 $X_t < C_t$. 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)).[9]

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

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

and where E_t is the expectation operator conditional on information available at time t. When the habit is external, agents maximize (1) but ignore the impact of today's consumption on tomorrow's habits, since the habit in this specification merely plays the role of an externality. In this case, only the first term on the right-hand-side of (3), $(C_t - X_t)^{-\gamma}$, is part

yield) and the default spread (the difference between the BAA and AAA corporate bond rates). Lewellen (1999) and Vuolteenaho (2000) forecast returns with an aggregate book-market ratio. Lettau and Ludvigson (2001a) forecast returns with a proxy for the log consumption-wealth ratio.

[9]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., $X_t = r\left(C_t, C_{t-1}, X_{t-1}\right)$.

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, $\theta = (\delta, \gamma)'$, and a single unknown function $X_t = f\left(C_t, C_{t-1}, ...C_{t-L}\right)$. We denote the true unknown parameters of interest as $\alpha_o = (\theta_o, f_o)$. The model (4) is semiparametric in the sense that it contains both finite dimensional and infinite dimensional unknown parameters.

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 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 parameter θ 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 estimator of θ is $\sqrt{T}$ consistent and asymptotically normally distributed. In addition, the sieve estimator of f is shown to be consistent, with a convergence rate faster than $T^{-1/4}$ under certain metric. These results are extended to models with stationary $\beta-$mixing observations in the Appendix of this paper.[10] Thus, as with any estimation of finite moment conditions, our procedure requires stationary observations, but allows for flexible linear serial correlation of the stationary time-series used in (4).

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, such as 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 the asset pricing and 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 $g : R^L \to R$ is an unknown function of the gross growth rates of consumption, with domain space reduced by one dimension relative to f. Note that g now replaces f as the unknown function to be estimated using the SMD procedure.

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

[10] $\beta-$mixing is one measure of temporal dependence. Its formal definition for stationary Markov processes is as follows. Let $\{Y_t\}$ be a stationary Markov process with invariant measure Q. The $\beta-$mixing coefficients are given by: $\beta_t = \int \sup_{0\leq\phi\leq1} |E[\phi(Y_t)|Y_0 = y] - E[\phi(Y_t)]| dQ(y)$. The process $\{Y_t\}$ is said to be beta-mixing if $\lim_{t\to\infty} \beta_t = 0$. For a survey of strong mixing conditions, see Bradley (1986).

dimensionality."[11] 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.[12] This strategy has been employed in nonparametric setting by Gallant and Tauchen (1989), who estimate the nonseparability term of the utility function using a linear polynomial sieve over one lag of consumption growth, and in a parametric setting by Ferson and Constantinides (1991) who estimate a model in which the habit is a linear function of one lag of consumption. 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) for the parameters of the nonparametrically estimated 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 nonseparabilities between elements of g. Such nonseparabilities are a feature of many 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

[11]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.

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

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

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 accurate results (e.g., McGrattan (1998)). An important shortcoming of this approach in empirical settings, however, is that approximation based on linear tensor product sieves is known to have slower convergence rates than approximation based on nonlinear sieves when the domain of g is of high dimension (Chen and White (1999)).

consumption have the greatest influence on the habit stock. Thus, the estimation procedure we propose can 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 using a single-layer smooth Artificial Neural Network (ANN) sieve taking the functional form

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

where $\psi(\cdot)$ is a scalar function from the class of functions that includes all probability distribution functions with compact support. A common choice for ψ is the sigmoid function, $\psi(x) = (1 + e^{-x})^{-1}$, a specification we use here. This provides 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.[13] We denote the unknown parameters to be estimated as $\boldsymbol{\alpha} = (\theta, g)' =$ $(\delta, \gamma, \alpha_0, \alpha_1, ..., \alpha_{K_T}, \gamma_{1,0}, ...\gamma_{1,L}, ..., \gamma_{K_T,0}, ...\gamma_{K_T,L}, \beta_1, ..., \beta_{K_T})'$. 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 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. 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 difficult when the habit is specified parametrically, for example, as a linear function of past consumption. By contrast, 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. By constraining the α coefficients in (8) to sum to a number less than one in absolute value, we insure that the flexibly estimated habit-consumption ratio, $g\left(\frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t}\right)$, is always less than unity, and not just in a particular sample.

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

[13] Hornik, Stinchcombe, and White (1989) provide a universal approximation result that justifies the use of neural network approximation; 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.

marginal utility, MU_t, takes the form

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

where,

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

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:

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

where,

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

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 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 on the outside of (4). The Appendix presents several equivalent expressions of this form; here we present one. Combining (11) and (4) and rearranging terms generates a set of N conditional moment conditions that must hold for each asset $i = 1, ...N$:

$$E_t \left\{ \left(\delta_o \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma_o} \digamma_{t+1} R_{i,t+1} - 1 \right) \Phi_{t+1} \right\} = 0, \tag{12}$$

where,

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

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

Let

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

Defining

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

we may write (12) more compactly as

$$E_t \left\{ \rho_i \left(\mathbf{z}_{t+1}, g_o, \delta_o, \gamma_o \right) \right\} = 0 \qquad i = 1, ...N. \tag{13}$$

The SMD estimator can be implemented as GMM (Hansen (1982)). The conditional expectation in (13) is taken with respect to agents' information set at time t, $\mathbf{w}_t^*$,

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

Let $\mathbf{w}_t$ be a $D \times 1$ observable subset of $\mathbf{w}_t^*$. Equation (14) implies

$$E \left\{ \rho_i \left(\mathbf{z}_{t+1}, g_o, \delta_o, \gamma_o \right) | \mathbf{w}_t \right\} = 0 \qquad i = 1, ...N. \tag{15}$$

As in any GMM estimation, we assume the conditional expectation (15) uniquely identifies the parameters of interest. Equation (15) holds if and only if

$$E \left\{ \rho_i \left(\mathbf{z}_{t+1}, g_o, \delta_o, \gamma_o \right) q(\mathbf{w}_t) \right\} = 0 \tag{16}$$

for all $q(\cdot)$ bounded, continuous functions of $\mathbf{w}_t$. The unknown parameters $\boldsymbol{\alpha}_o = (g_o, \delta_o, \gamma_o)'$ may be estimated semiparametrically by minimizing a criterion function based on the following increasing number of unconditional moment conditions

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

where $\{p_{oj}(\mathbf{w}_t),\ j = 1, 2, ...J_T\}$ is a sequence of known basis functions mapping $R^D \rightarrow R$. Note that J_T must grow with the sample size to insure consistency of the method.

Let $p^{J_T}(\cdot) \equiv (p_{o1}(\cdot), ..., p_{oJ_T}(\cdot))'$ and the $T \times J_T$ matrix $\mathbf{P} \equiv \left(p^{J_T}(w_1), ..., p^{J_T}(w_T)\right)'$. Stacking the equations in (17) for $i = 1, ..., N$ produces a $(NJ_T) \times 1$ vector of orthogonality conditions that can be used to estimate $\boldsymbol{\alpha}_o = (g_o, \delta_o, \gamma_o)'$:

$$E\left\{\mathbf{h}\left(\boldsymbol{\alpha}_o, \mathbf{z}_{t+1}, \mathbf{w}_t\right)\right\} = 0,$$

where

$$\mathbf{h}\left(\boldsymbol{\alpha}_o, \mathbf{z}_{t+1}, \mathbf{w}_t\right) = \begin{bmatrix} \rho_1(\mathbf{z}_{t+1}, g_o, \delta_o, \gamma_o) p^{J_T}(\mathbf{w}_t) \\ \rho_2(\mathbf{z}_{t+1}, g_o, \delta_o, \gamma_o) p^{J_T}(\mathbf{w}_t) \\ . \\ . \\ . \\ \rho_N(\mathbf{z}_{t+1}, g_o, \delta_o, \gamma_o) p^{J_T}(\mathbf{w}_t) \end{bmatrix}. \tag{18}$$

The sample average value of $\mathbf{h}\left(\boldsymbol{\alpha}, \mathbf{z}_{t+1}, \mathbf{w}_t\right)$ is

$$\mathbf{g}(\boldsymbol{\alpha}; \mathbf{y}_T) \equiv (1/T) \sum_{t=1}^{T} \mathbf{h}\left(\boldsymbol{\alpha}, \mathbf{z}_{t+1}, \mathbf{w}_t\right), \tag{19}$$

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

$$Q(\boldsymbol{\alpha}) = [\mathbf{g}(\boldsymbol{\alpha}; \mathbf{y}_T)]' \mathbf{W} [\mathbf{g}(\boldsymbol{\alpha}; \mathbf{y}_T)], \tag{20}$$

which is minimized numerically with respect to $\boldsymbol{\alpha}$. Ai and Chen (2003) show that consistent estimates of $\boldsymbol{\alpha}$ and of the conditional moments in (13) may be obtained with this minimization by setting

$$\mathbf{W} = \mathbf{I} \otimes (\mathbf{P}'\mathbf{P})^{-1}, \tag{21}$$

where $\mathbf{I}$ is the $N \times N$ identity matrix. In this paper we extend these results so that semiparametric estimation of conditional moment conditions such as (4) may be applied to problems with serially correlated data. Minimization of the (20) using the weighting matrix (21) 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. The minimization gives greater weight to moments that are more highly correlated with the instruments $p^{J_T}(\cdot)$. The appendix gives a general expression for the sieve minimum distance estimator.

It is important to note that arbitrary weighting matrixes cannot be used and implemented as GMM in this setting. The SMD estimation procedure described here collapses to a case of GMM only when the specific weighting matrix (21) is employed, a specification that is crucial to the nonparametric estimation of the conditional mean (13) using the basis functions $\{p_{oj}(\mathbf{w}_t),\ j = 1, 2, ...J_T\}$. Estimation of the *conditional* expectation (as opposed to an unconditional expectation) is in turn necessary to obtain consistent estimates of the nonparametric habit function (see Ai and Chen (2003)). This implies that it is not possible to evaluate the semiparametric model we study (either initially or in subsequent analysis) using estimation techniques that may be more familiar in the asset pricing literature, such as (for example) the cross-sectional methodology proposed by Fama and MacBeth (1973), which is essentially GMM estimation with the identity weighting matrix.

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 in the sense of 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 (15) into a set of NJ_T unconditional expectations, multiplying each $\rho_i(\mathbf{z}_{t+1+L}, g_x, \delta, \gamma)$ 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 (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 $\psi(\cdot)$ 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, $\delta \in (0, 1]$, and the curvature parameter $\gamma \in [.1, 20]$.

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 a theory for the consistency and rates of convergence of the sieve estimator (see the Appendix), a general asymptotic distribution theory for sieve estimators has not been developed, even for i.i.d. observations. But the coefficients of the parametric part of our specification (δ and γ) are asymptotically normally distributed, implying that asymptotic standard errors for these parameters may be computed using GMM theory (Hansen (1982)). We present these estimates 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 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. 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 $\rho_i(\mathbf{z}_{t+1}, g_o, \delta_o, \gamma_o)$. 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 $\rho_i(\mathbf{z}_{t+1}, g_o, \delta_o, \gamma_o)$.

4 The Data

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

We study two 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), contains 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. 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 just a few aggregate asset returns. The reason: exploiting the cross-section aids the empirical identification of the unknown habit function. More moments yield more stable estimates of the unknown function. Increasing the number of asset returns increases the number of moments and helps in this regard.[14]

[14] Alternatively, we could raise the number of moments by reducing the number of asset returns but increase the number of basis functions of conditioning variables. Reducing to just the six equity returns listed above and using a greater number of polynomial basis functions of our conditioning variables produced results

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 of observable variables, $\mathbf{w}_t$. We focus on variables for $\mathbf{w}_t$ that predict asset returns in quarterly data, which should provide powerful tests of the conditional moment restrictions implied by the theory. We choose three variables that have been shown elsewhere to have significant forecasting power for excess stock returns on aggregate stock market indexes in quarterly data. The first is the proxy for the log consumption wealth ratio studied in (Lettau and Ludvigson (2001a)), which is measured as the cointegrating residual between log consumption, log asset wealth, and log labor income and denoted $\widehat{cay}_t$.[15] This variable has strong forecasting power for excess stock returns over horizons ranging from one quarter to several years. 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 (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, $SPEX$.[16] Time-series regressions using these variables to predict future excess stock returns can be found in Lettau and Ludvigson (2001a). Thus, $\mathbf{w}_t = [1, \widehat{cay}_t, RREL_t, SPEX_t]'$.[17]

Since the error term $\rho_i(\mathbf{z}_{t+1}, g_o, \delta_o, \gamma_o)$ is orthogonal to the information set $\mathbf{w}_t$, any non-

similar to those reported below. A worry with doing this, however, is that there is more independent information in a well-chosen cross-section of asset returns than in ever larger numbers of basis functions of the same few conditioning variables.

[15]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.

[16]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 they drive out many of the other popular forecasting variables studied in the papers cited above, such as many financial ratios, term spreads and default spreads in quarterly forecasts (Lettau and Ludvigson (2001a)).

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

linear transformation $p^{J_T}(\mathbf{w}_t)$ can be used as valid instruments. We use power series as instruments, focusing on three different specifications. Each specification includes a constant (vector of ones). The first specification also includes the linear terms plus the squared terms of each variable; the second includes the linear terms plus the pair-wise cross product of each variable. These two specifications create seven instruments which we use when studying the asset returns in Group 1. The third set of instruments includes just the linear terms of each variable, or four instruments in total. We use this instrument set when analyzing the the larger asset return group, Group 2. Note that 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, but at a slower rate than the sample size T, so that $NJ_T/T \to 0$ and $NJ_T \to \infty$ as $T \to \infty$. Since Group 2 has 26 asset returns, we reduce the number of instruments by using only the linear transformations of $\mathbf{w}_t$ in this case, so that the total number of moments is similar to that for the estimation on Group 1 assets.

5 Empirical Results

5.1 Empirical Estimates

Tables 1-2 (in the text) and Figures 1-9 (at the end of the paper) present the results of estimating and testing the habit framework presented above, using the instruments and test assets described in the previous section. 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 $L = 3$. 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), Gallant and Tauchen (1989)).

For the dimensionality of the ANN sieve, (8), we set $K_T = 3$. Because asymptotic theory only provides guidance about the *rate* at which 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 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, $\boldsymbol{\alpha}$, 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 two estimations using Group 1 asset returns based on different polynomial basis functions of $\mathbf{w}_t$: those using the linear and squared values of the elements of $\mathbf{w}_t$ and those using the linear and cross-terms. Using Group 2 assets we use just the linear terms (that is just $\mathbf{w}_t$ itself) as instruments. The estimates of the rate of time-preference, δ, and the curvature parameter, γ, for these estimations are presented in Table 1, with asymptotic standard errors in parentheses.

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 close to one, and the curvature parameter is between $\gamma = 0.7$ and $\gamma = 0.8$. The standard errors indicate that these variables are estimated precisely.[18] 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.

[18]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.

Table 1

SMD Estimates of δ and γ

Assets	Instruments	δ	γ
Group 1	$\mathbf{w}_t$ & squared terms	0.9850	0.707
		(0.005)	(0.089)
Group 1	$\mathbf{w}_t$ & cross terms	0.9850	0.787
		(0.005)	(0.079)
Group 2	$\mathbf{w}_t$	0.9847	0.851
		(0.006)	(0.048)

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

To get a sense of how important the habit is in the power utility specification, Figures 1, 3 and 5, top panels, plot the habit-consumption ratio over time. Figure 1 plots the case where the instruments are $\mathbf{w}_t$ and squared terms; Figure 3 plots the case where the instruments are $\mathbf{w}_t$ and cross terms; in these cases Group 1 assets were used to estimate the model. Figure 5 plots the case where the instruments are $\mathbf{w}_t$, in which Group 2 assets were used to estimate the model. Just as the point estimates of δ and γ are similar across specifications, the estimates of the quantitative importance of the habit are similar; they demonstrate significant evidence in favor of habit formation, conditional on the power utility framework.

Consider first the two estimations on Group 1 assets. For these cases, 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 for these cases is not highly volatile, ranging only from 0.97 to 0.974.

The bottom panels of Figures 1 and 3 plot the estimated stochastic discount factor over time for estimations on Group 1 assets. The stochastic discount factor is given in equation (11), and depends on the conditional expectation of nonlinear functions of consumption growth and the estimated habit. Panel B of each figure plots an estimate of M_t, using our estimated parameter values and estimating those parts of M_t that appear in expectation as their projection onto the set of instruments used in that estimation.[19] The relative stability of the habit-consumption ratio for these cases translates into a relatively stable stochastic discount factor: in both cases the mean is slightly less than one (0.98), while the standard deviation is 0.016 in Figure 1, and 0.012 in Figure 3. Although these values are relatively small, they are nonetheless significantly larger than the standard deviation of quarterly consumption growth, equal to 0.0045 in this sample. Nevertheless, it's clear 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.

Now consider the estimation on Group 2 assets. Figure 5, top panel, shows that habit is also quantitatively important for the estimation on Group 2 assets. The habit is again estimated to be about 97% of consumption on average, reinforcing the evidence in favor of habit formation in the power utility framework. Preferences do not appear to be time-separable. An estimate of the stochastic discount factor, shown in the bottom panel of Figure 5, is also very similar, having a mean of 0.981 and a standard deviation of 0.012.

Figures 2, 4 and 6 plot $C_t - X_t$ against C_t (top panel), X_t against C_t (middle panel) and X_t against C_{t-1} (bottom panel). Figure 2 plots these values for the estimation on Group 1 assets in which the instruments are $\mathbf{w}_t$ and squared terms; Figure 4 plots the same for the estimation on Group 1 assets in which the instruments are $\mathbf{w}_t$ and cross terms; Figure 6 plots the same for the estimation on Group 2 assets in which the instruments are $\mathbf{w}_t$. These figures

[19] The estimates of the sieve coefficients in each case below imply that marginal utility, MU_t, is strictly positive at all dates in our sample.

plot the observed values of consumption against our estimate of the habit stock over time. The purpose of these plots is to get a sense of how the habit stock varies with consumption. Note that these are not plots of partial derivatives of the habit with respect to consumption, but merely plots of two time series against one another. Thus, for example, Panel B shows how X_t tends to vary with C_t over time, but it does not hold constant the values of lagged consumption, which also influence the habit stock.

The top panels of Figures 2 and 4 and 6 show 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 will not be constants, and they will vary with the lags of consumption growth. Given our estimated X function, we may plot the derivatives as they vary over time with lagged consumption growth. This is done in Figures 7 through 9. Figure 7 plots the partial first derivatives of X_{t+1} for the model estimated on Group 1 assets and linear and squared terms of the instruments; Figure 8 plots the same for the model estimated on Group 1 assets and linear and cross terms of the instruments; Figure 9 plots the same for the model estimated on Group 2 assets and the linear instruments. The three plots are similar. 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 with as consumption becomes more distant.

Although the plots of X_t against C_t and C_{t-1} look “linear,” it should be noted that we cannot make inferences about whether the habit itself is a linear function of current and past consumption on the basis of these plots. The habit, X_t, is, in general, a nonlinear function that allows for nonseparabilities between its arguments. Thus plots of X_t against C_t are plots of the estimated time series $X_t = C_t g\left(\frac{C_{t-1}}{C_t}, ..., \frac{C_{t-L}}{C_t}\right)$ against C_t, which cannot be used to infer whether X_t is linear in current and past consumption.

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. 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 implies that $\frac{\partial X_{t+i}}{\partial C_t}$ is constant for all $i \geq 0$, or equivalently that

$$\frac{\partial^2 X_{t+i}}{\partial C_t^2} \equiv 0 \qquad \forall i \geq 1. \tag{22}$$

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 (22) based on our estimates for the cases above. This is done in Figures 10 through 12 for $i = 1$. Notice from (10) that $\partial^2 X_{t+i}/\partial C_t^2$ will take the form $(1/C_{t+i})\, g_{ii}$, where g_{ii} denotes the derivative of g_i with respect to the ith element. Obviously $(1/C_{t+i})\, g_{ii}$ cannot be identically zero unless g_{ii} 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 corresponding to $\partial^2 X_{t+1}/\partial C_t^2$.

Figure 10 plots the normalized second derivative of X_{t+1} for the model estimated on Group 1 assets and linear and squared terms of the instruments; Figure 11 plots the same for the model estimated on Group 1 assets and linear and cross terms of the instruments; Figure

12 plots the same for the model estimated on Group 2 assets and the linear instruments. The second partial derivatives are everywhere negative; taken together with the estimates of the first partial derivatives presented in Figures 7 through 9, this implies that the habit is increasing in lagged consumption, but at a decreasing rate. 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 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 10 through 12.

Ideally we would like to construct a formal 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

$$\mu \equiv E\left[\frac{\partial^2 X_{t+i}}{\partial C_t^2}\right]. \tag{23}$$

Clearly if (22) is everywhere identically equal to zero, its mean (23) must be zero, although the reverse need not hold. Nevertheless, if we find that the mean (23) is statistically different from zero, we may conclude that (22) is not true, and the habit function is nonlinear. We present the results of such a test now. A complete description of the asymptotic justification for this test is given in the Appendix (to be included).

The test uses the result that the sample mean of the second derivative of an unknown function is normally distributed, and converges to its true value at rate proportional to $\sqrt{T}$. With this result, it follows that the square of the ratio of the estimated mean of $\partial^2 X_{t+i}/\partial C_t^2$ to its estimated standard deviation has a Chi-square (1 degree of freedom) distribution. To compute this statistic, we need an estimate of the standard deviation of $\partial^2 X_{t+i}/\partial C_t^2$, which appears in the denominator of the statistic. The asymptotic approximation of the standard deviation is of complicated form, however, because it involves the second partial derivative of the joint density with respect to the growth rates of consumption (see the Appendix), which must be estimated nonparametrically. Moreover, in finite samples, we are likely to obtain more accurate estimates of both the mean and standard deviation of $\partial^2 X_{t+i}/\partial C_t^2$ by bootstrapping the raw data to form empirical estimates of these moments. The resulting test statistic based on bootstrapped moments may be compared with critical values from the Chi-square distribution. To do so, we again focus on the normalized second derivative corresponding to $\partial^2 X_{t+1}/\partial C_t^2$, which is g_{11}. 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.[20] We then conduct SMD estimation on 500 bootstrap samples so formed, which delivers 500 estimates of the mean of g_{11}, denoted μ_g. We then compute the mean and standard deviation of μ_g over bootstrap samples to generate a test statistic with which we can compare to the Chi-square critical values. This test statistic is the ratio of $\left(E_b\left(\mu_g\right)/\sqrt{\text{Var}_b\left(\mu_g\right)}\right)^2$, where the subscript "$b$" denotes the estimate formed across bootstrap simulations.

The results are as follows. There is very little variation in the variance of the mean across bootstrap samples. This is not surprising, as Figures 10 through 12 also suggest there is very little variation in the second derivative estimates over time. This results in very large test statistics and strong rejections that the mean is zero. For the estimation on Group 1 assets, linear and squared instruments, the value of the statistic is 6360.43; on Group 1 assets and linear and cross-term instruments, the statistic is 6075.88; for Group 2 assets and linear instruments, the statistic is 5630.63. The p-values for these tests are each less than 0.00001. Thus, we strongly reject the hypothesis that the mean of g_{11} is zero, implying that (22) cannot be true and the function estimated is nonlinear.

The shape of our estimated habit function can also be illustrated by plotting X_t as a function of lagged consumption, C_{t-1}, holding fixed current consumption, C_t, and the other lags of consumption, $C_{t-2}, ...C_{t-L}$. Figure 13 plots precisely this relation for C_t, $C_{t-2}, ...C_{t-L}$ alternately held fixed at their median, 25th, and 75th percentile values in our sample. The figure reinforces several conclusions drawn above. First, the estimated habit is nonlinear;

[20] 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 variance and bias is $T^{1/3}$.

this is evident from the curved shape of the function and from the finding that the shape depends on where in the domain space the arguments C_t, $C_{t-2}, ...C_{t-L}$ are located. Second, the estimated habit is always increasing in past consumption. Third, the estimated habit is increasing at a decreasing rate in past consumption.

5.2 Hypothesis Tests

The introduction discussed a number of interesting hypotheses pertinent to models with habit formation. Here we provide formal tests of two of them: whether the habit is external or internal, and model comparison tests for how well the habit paradigm fits the asset pricing data.

5.2.1 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, considers an example in which the habit is a distributed lag of past consumption, and 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. If the habit is nonlinear, however, the asset pricing implications of external and internal habit models need not be identical.[21] Here we investigate whether the asset pricing implications are better described by external or internal habit formation. The distinction is important not only because it affects the stochastic discount factor and

[21] This can be understood by inspecting (3). Suppose that there is a constant risk-free rate equal to the rate of time preference, δ. Then, when habit formation is of the external variety, the first order conditions for optimal consumption choice imply

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

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 in (3) are all constant, as with habits linear in lagged consumption, guessing the solution $(C_t - X_t)^{-\gamma} = \omega E_t \left[(C_{t+1} - X_{t+1})^{-\gamma}\right]$, for arbitrary constant ω, implies that marginal utility in (3) is proportional to $(C_t - X_t)^{-\gamma}$, and therefore to external habit formation marginal utility. It follows that the asset pricing implications of each specification, which derive from the intertemporal *ratio* of marginal utilities, are equivalent. If the $\frac{\partial X_{t+j}}{\partial C_t}$ terms are not constant, but instead vary with lagged consumption, MU_t in (3) will not in general be proportional to $(C_t - X_t)^{-\gamma}$, and the asset pricing implications of the two specifications may differ.

therefore has asset pricing implications, but also because the externality inherent in external habit formation implies the two paradigms have dramatically different social welfare and tax policy consequences (Ljungqvist and Uhlig (2000)).

One way to assess which specification better describes the data is simply to compute the value of the minimized 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 the last term on the right-hand-side of (3) be excluded from marginal utility. Doing so for each estimation described above, we find that the minimized GMM criterion is several orders of magnitude larger when marginal utility is restricted to external habit formation. For example, using Group 2 asset returns and linear instruments, the external habit case is found to be 1.1944e-04, compared to 1.2631e-07 for the internal habit case, about 1000 times larger. The estimations using Group 1 assets produced similar results. Ideally, of course, this comparison would be made on statistical grounds. Unfortunately, criterion-based statistical tests have not been developed for procedures involving an unknown function. 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 section is to estimate the general model of internal habit formation and test the restriction on preferences implied by external habit formation. This approach exploits the nested nature of external habit preferences and is based on the conditional moments implied by the asset pricing model. Unlike the criterion based comparison discussed above, it affords the advantage of providing a formal statistical assessment. To do so, we use the consistent hypothesis testing methodology for testing conditional moment conditions in semiparametric models developed in Chen and Fan (1999). A detailed explanation of this procedure as it is used in our application is given in the appendix (to be included) at the end of this paper. In the appendix we show that the conditional moment restrictions for the asset pricing model we estimate can be written as:

$$E_t\left(\delta\left(\frac{C_{t+1}-X_{t+1}}{C_t-X_t}\right)^{-\gamma} R_{i,t+1}\widetilde{F}_{i,t+1}-1\right)=0, \qquad i=1,...,N, \tag{24}$$

where

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

Equation (24) forms the basis of a statistical test for external habit formation. Notice that, for external habit formation, $\widetilde{F}_{i,t+1} = 1$, so the internal habit model we estimate nests the external habit model as a special case. The external habit formation implies

$$E_t\left(\delta\left(\frac{C_{t+1} - X_{t+1}}{C_t - X_t}\right)^{-\gamma} R_{i,t+1} - 1\right) = 0, \qquad i = 1, ..., N. \tag{25}$$

It follows that we may evaluate whether external habit formation is a plausible description of the data by testing whether the conditional moment (25) is statistically different from zero. Below we provide test statistics for each $R_{i,t+1}$ and each of the SMD estimations discussed above.

We employ a testing procedure for conditional moment conditions involving functions of unknown form, and use it to assess whether (25) is statistically different from zero using our estimated habit function, $\widehat{X}_{t+1}$. The null hypothesis is that the conditional mean in (25) is satisfied when $\widetilde{F}_{i,t+1} = 1$, which coincides with external habit formation. The alternative is that (25) is not satisfied, which coincides with the general internal habit formation specification we estimate. The test statistic itself is a nonparametric estimate of the conditional moment in (25), and measures the difference between this function and the zero function by a weighted integrated squared distance. The limiting distribution of this statistic is of complicated form, so we apply a conditional Monte-Carlo approach to approximate its critical values. We refer the reader to the appendix for the details. Tables 2 and 3, below, present the 95% simulated confidence interval and the associated empirical test statistic, for each conditional moment and each estimation described above. In the tables below, portfolio returns for size and book-to-market sorted portfolios are denoted SiBj, where i indicates the size quantile (1 indicating the smallest category, 2 the second smallest, and so on), and j indicates the book-to-market quantile (1 indicating the lowest book-market ratio, 2 the second lowest, and so on). Industry portfolio returns are abbreviated according to categories provided on Kenneth French's web site.

Tables 2 and 3 show that, no matter what estimation case we analyze, and no matter which moment, this test uniformly rejects the null of external habit formation. All the conditional moments corresponding to every asset return fails to satisfy (25). The test statistics are many orders of magnitude smaller than the lower tail of the 95% confidence interval under the null of external habit formation. This provides striking evidence against the external habit specification, and indicates that we may accept the alternative of internal habit formation with very high confidence. Below we also use the Hansen and Jagannathan (1997) procedure to assess how well the external specification fares relative to the internal specification in explaining equity returns.

Table 2

Tests For External Habit Formation

Est. 1			Est. 2		
Return	95% CI	Statistic	Return	95% CI	Statistic
Tbill	[1.8e-05, 7.0e-04]	1.9e-33*	Tbill	[1.6e-05, 6.0e-04]	2.5e-33*
S1B1	[1.4e-04, 5.0e-03]	1.6e-33*	S1B1	[1.3e-04, 5.1e-04]	1.4e-33*
S1B2	[7.3e-05, 2.7e-03]	3.1e-33*	S1B2	[7.1e-05, 2.4e-04]	2.9e-33*
S1B3	[4.9e-005, 1.6e-03]	5.0e-33*	S1B3	[6.4e-05, 2.4e-04]	5.8e-33*
S2B1	[2.1e-04, 8.1e-03]	8.2e-34*	S2B1	[1.9e-04, 7.2e-04]	7.9e-34*
S2B2	[1.0e-04, 3.8e-03]	2.1e-33*	S2B2	[1.2e-05, 4.6e-04]	1.1e-33*
S2B3	[6.3e-05, 2.4e-03]	2.9e-33*	S2B3	[8.2e-05, 3.2e-04]	2.4e-33*
NoDur	[1.3e-04, 4.9e-03]	2.3e-33*	NoDur	[1.3e-04, 5.0e-03]	1.5e-33*
Durl	[2.3e-04, 8.9e-03]	1.7e-33*	Durl	[4.3e-04, 1.7e-02]	2.1e-33*
Manuf	[1.3e-04, 5.1e-03]	1.6e-33*	Manuf	[1.3e-04, 4.8e-03]	7.4e-34*
Enrgy	[5.9e-05, 2.2e-03]	4.5e-34*	Enrgy	[5.4e-05, 2.0e-03]	9.3e-34*
HiTec	[2.4e-04, 9.4e-03]	2.3e-33*	HiTec	[3.2e-04, 1.2e-02]	3.4e-33*
Telcm	[1.8e-04, 7.1e-03]	1.4e-33*	Telcm	[1.6e-04, 6.30e-03]	3.5e-34*
Shops	[1.4e-04, 4.9e-03]	1.2e-33*	Shops	[1.2e-04, 4.7e-03]	1.3e-33*
Hlth	[5.1e-04, 2.7e-03]	3.5e-33*	Hlth	[2.5e-04, 9.6e-03]	6.0e-33*
Utils	[1.1e-04, 4.3e-03]	1.0e-33*	Utils	[1.1e-04, 4.1e-03]	4.8e-34*
Other	[1.5e-04, 6.0e-03]	4.1e-34*	Other	[8.9e-05, 3.3e-03]	1.4e-33*

Notes: Tests for external habit, as described in text. The null hypothesis is external habit formation. '*' indicates null is rejected at better than 5% level. "Est. 1" is estimation on Group 1 assets, linear and squared instruments. "Est. 2" is estimation on Group 1 assets, linear and cross-term instruments.

Table 3

Tests For External Habit Formation

Est. 3					
Return	95% CI	Statistic	Return	95% CI	Statistic
Tbill	[1.6e-05, 6.4e-04]	1.4e-33*	S3B3	[1.0e-04, 3.9e-02]	2.0e-33*
S1B1	[4.1e-04, 1.5e-02]	1.3e-33*	S3B4	[8.1e-05, 3.0e-04]	3.5e-33*
S1B2	[1.1e-04, 3.8e-03]	3.0e-33*	S3B5	[1.1e-04, 4.5e-03]	4.3e-33*
S1B3	[7.4e-005, 2.8e-03]	4.6e-33*	S4B1	[1.6e-04, 6.1e-03]	8.6e-34*
S1B4	[1.3e-04, 4.9e-03]	8.9e-34*	S4B2	[1.4e-04, 5.6e-03]	1.2e-33*
S1B5	[1.2e-04, 4.6e-03]	6.3e-33*	S4B3	[1.2e-04, 4.8e-03]	3.9e-33*
S2B1	[8.2e-05, 2.6e-03]	4.9e-34*	S4B4	[3.1e-05, 1.1e-03]	2.2e-33*
S2B2	[7.8e-05, 2.9e-03]	4.7e-33*	S4B5	[3.3e-05, 1.3e-03]	4.7e-33*
S2B3	[5.6e-05, 2.1e-03]	4.6e-33*	S5B1	[2.7e-04, 1.0e-02]	1.4e-33*
S2B4	[5.3e-05, 1.9e-03]	6.2e-33*	S5B2	[2.1e-04, 8.2e-03]	7.4e-34*
S2B5	[9.4e-05, 3.6e-03]	1.2e-32*	S5B3	[1.3e-04, 5.0e-03]	1.7e-33*
S3B1	[1.1e-04, 4.4e-03]	1.5e-33*	S5B4	[1.4e-04, 5.0e-03]	1.3e-33*
S3B2	[5.6e-05, 2.0e-03]	3.3e-33*	S5B5	[1.6e-04, 6.4e-03]	1.88e-34*

Notes: Tests for external habit, as described in text. The null hypothesis is external habit formation. '*' indicates null is rejected at better than 5% level. "Est. 3" is estimation on Group 2 assets, linear instruments.

5.2.2 Model Comparison Tests

We have estimated a habit-based asset pricing framework, 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 in some quantifiable sense relative to other models that have been explored in the literature. We seek a methodology that recognizes that competing models are mere approximations of reality and therefore may be misspecified, and allows us to investigate which model is least misspecified. 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, $\boldsymbol{\alpha}$, to minimize the quadratic form $\mathbf{g}_T^{HJ} \equiv \mathbf{w}_T'(\boldsymbol{\alpha})\,\mathbf{G}_T^{-1}\mathbf{w}_T(\boldsymbol{\alpha})$, where $\mathbf{w}_T(\boldsymbol{\alpha})$ is the sample average of pricing errors, and $\mathbf{G}_T^{-1}$ is the sample second moment matrix of the N asset returns upon which the models are evaluated. Hansen and Jagannathan show that the square root of this minimized quadratic form gives the maximum pricing error per unit norm on any portfolio of the N assets studied, and delivers a measure of misspecification suitable for model comparison.[22] We refer to the square root of this minimized quadratic form as the *Hansen-Jagannathan distance*, or HJ distance for short.

An 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. Thus, using the second moment matrix helps to avert this problem.

To apply this procedure to the habit-based framework, we treat our estimate of the habit as a *proxy* for the true habit, and minimize $\mathbf{g}_T^{HJ}$ corresponding to the asset pricing model in (1-4) over the parameters δ and γ using the same quarterly sample that was used in the SMD estimation. 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 model by choosing the finite dimensional parameters δ and γ to minimize $\mathbf{g}_T^{HJ}$. Below we present results from using

[22] Hansen and Jagannathan (1997) also show that $\sqrt{\mathbf{g}_T^{HJ}}$ gives the least-square distance between the candidate stochastic discount factor and the closest point to it in the set of all stochastic discount factors that price assets correctly.

the habit estimate generated by SMD estimation on Group 2 assets, as discussed above. Results using the habit estimates from SMD estimation on Group 1 assets are similar.

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 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

$$M_{t+1} = a + \sum_i b_i F_{i,t+1}, \tag{26}$$

where $F_{i,t+1}$ are variable factors, and the coefficients a and b_i are treated as free parameters to be estimated. Fama and French develop an empirical three-factor model, 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.[23] 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, $\widehat{cay}_t$, $\widehat{cay}_t \cdot \Delta \log C_{t+1}$, 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.

[23] 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).

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 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 an educated 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.

Finally, we also compare specification errors of these models to those of the standard CCAPM (with consumption growth the single variable factor in (26)), and to those of the CAPM (with the market return, R_m the single variable factor). For all linear models, the unknown coefficients a and b_i are estimated by minimizing the corresponding quadratic form, $\mathbf{g}_T^{HJ}$, 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. 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 encountered here when the number of portfolios, N, is too large (e.g., see Hansen, Heaton, and Yaron (1996), Ahn and Gadarowski (1999), Lettau and Ludvigson (2001b)). Our own experience suggests that when N is reduced to about 6, given the time-series samples currently encountered in quarterly data, estimates of the second moment matrix are reliable. Thus, we do our analysis on the smaller number of size/book-to-market portfolios provided by Fama and French, a set that should continue to provide a good summary measure of the cross-sectional variety of U.S. equity returns, albeit at a more aggregated level.

Before discussing how each model fares according to specification error, we note that the estimates of δ generated from minimizing $\mathbf{g}_T^{HJ}$ for the general internal habit model are similar to those estimated using the SMD procedure but generally smaller, equal to 0.74 when the model is evaluated on the equity portfolios alone, and 0.90 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 18.5 when the model is evaluated on the equity portfolios alone, and 9.2 when the Treasury bill is included. One possible reason for this latter finding is that the Hansen-Jagannathan procedure places greater emphasis on unconditional mean returns than does the SMD procedure, which 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 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 $\mathbf{g}_T^{HJ}$. This produces much larger estimates of the curvature parameter γ, equal to 40 when the Treasury bill is included, and 67 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 0.7 when it is omitted.

Table 4 reports the measure of specification error given by the Hansen-Jagannathan distance ("HJ Dist") for all the models discussed above. There are interesting differences in the estimated specification error across the models. For the analysis carried out on equity returns alone (column 2 of Table 4), the smallest specification error, 0.208, is generated by the scaled CCAPM, but the internal habit model delivers a value for this measure of specification error that is almost as small, equal to 0.213. All of the other models (the external habit model, the Fama-French model, the CCAPM and the CAPM) have considerably larger specification errors. 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 0.263. The CCAPM and CAPM substantially larger than that. Thus, for equity returns, the specification error is smaller with internal habit formation than external habit formation, consistent with the tests reported above. These results are particularly encouraging for the habit-based framework, and suggest a possible structural interpretation of the empirical success of the scaled CCAPM and Fama-French models in explaining the behavior of equity returns documented elsewhere.

When the Treasury bill rate is added to the set of six equity returns, the results are different (column 3 of Table 4). For these portfolios inclusive of the Treasury bill rate, the Fama-French model has the smallest estimated specification error, 0.282. The scaled CCAPM has relatively more difficulty explaining this set of portfolio returns: its specification error measure is found to be 0.352. Both the habit-based models (internal and external) deliver the largest specification error. At first glance, this evidence would appear to suggest that the habit-based model has greater difficulty pricing the Treasury-bill rate than do the other models we compare it to.

Upon closer inspection, however, it is not clear that such a conclusion is warranted, for the following reason. The habit is estimated using SMD procedure, which, as discussed above, by necessity places the most weight on moments that deliver the best estimates of the conditional mean, i.e. are most highly correlated with the instruments. In our application, those moments are precisely the ones associated with the equity returns because our instruments are forecasting variables for stock returns; they have little forecasting power for short-term interest rates.[24] Moreover, none of the instruments have significant forecasting power for consumption growth. It follows that it is the moments associated with equity returns that get the most weight in the SMD estimation; the moment associated with the Treasury bill rate gets relatively little weight. As a consequence, the habit proxy we bring to the Hansen-Jagannathan procedure is essentially estimated to fit the moments associated with equity returns, not the moment associated with the Treasury bill rate.

By contrast, the Hansen-Jagannathan procedure, used to *evaluate* each model, necessarily gives far more weight to the Treasury bill rate than it does to the equity returns when the

[24]It is well known that the forecastable component of short-term interest rates is much smaller than that for equity returns; for example, see the evidence in Campbell, Lo, and MacKinlay (1997), Chapter 8. By contrast, Lettau and Ludvigson (2001b) find that $\widehat{cay}_t$ is a strong forecasting variable for stock returns sorted into categories based on size and book-to-market characteristics.

former is included in the set of test assets. This occurs by mere virtue of the former's smaller sampling variability. The up-shot of this is that the habit-model is placed at a disadvantage relative to all of the other models when it is evaluated on a set of returns that includes the Treasury bill rate. Contrary to what is done for the habit model, all of the other models' free parameters are estimated as part of the Hansen-Jagannathan procedure, and are therefore always set to match the very same moments used to evaluate the model. When the Hansen-Jagannathan procedure is applied to the set of equity returns alone, the habit model is placed on a more level playing field with the other models because, in that case, both the SMD and Hansen-Jagannathan procedures emphasize moments associated with equity returns, so that estimation and evaluation of the habit framework is done on a roughly comparable set of moments. But when the Treasury bill rate is included in the set of returns, the habit model is effectively estimated to match one set of moments and evaluated on another, a standard that the other models are never held to. For this reason, we are reluctant to place great emphasis on the results reported in column 3 of Table 4.

In summary, if we want to assess how well the habit-based model explains asset returns relative to the other models discussed above, it is more appropriate to focus on the results using only equity returns, those reported in column 2 of Table 3. The specification errors obtained when including the Treasury bill rate are not comparable across models as a result of the different weighting schemes inherent in the SMD and Hansen-Jagannathan procedures for our application.

Table 4

Specification Errors for Alternative Models

Model	6 size/BM	6 + T-bill
	HJ Dist	HJ Dist
(1)	(2)	(3)
Internal Habit	0.213	0.459
External Habit	0.263	0.426
Fama-French	0.262	0.282
Scaled CCAPM	0.208	0.352
CCAPM	0.307	0.403
CAPM	0.340	0.420

Notes: For each model labeled in the left-hand-column, 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).

We conclude this section by noting an alternative approach to evaluating relative performance: compare pricing errors on the primitive test assets by computing a distance measure using the identity matrix instead of the second moment matrix to weight pricing errors. When evaluated in this way, results (not reported) show that the habit models perform considerably worse than the Fama French model and the scaled CCAPM, and about the same as the CAPM and CCAPM. As mentioned, this approach has the disadvantage relative to the Hansen-Jagannathan methodology of making the objective function dependent on the initial choice of primitive test assets. For this reason, many researchers use only the HJ distance as a measure of specification error, as we have done here. More generally, however, judgement and taste play a role in determining which of the several possible measures of model misspecification to emphasize. The HJ distance gives the maximum pricing error per unit norm of any portfolio (including those that take big long and short positions) of the primitive assets. It is entirely possible for the pricing error of the worst performing portfolio to deliver one ranking among competing models, while an equally weighted average of pricing errors on the original test assets delivers another. As an aside, we note it is somewhat remarkable that the estimated habit models do as well as they do in any of these tests: the analysis of data that contain endogenous variables,[25] when extended to nonlinear, semi-nonparametric settings, has only recently been contemplated, and it is well known that obtaining consistent estimators under these circumstances is extremely challenging both as a matter of theory and implementation (for example, Blundell and Powell (2001); Newey and Powell (2003)). These challenges arguably place endogenous, nonlinear, semiparametric models at a significant disadvantage relative to other models. It is therefore notable that the habit-based models so estimated perform well pricing equity returns according to one robust and widely used metric of model misspecification.

6 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 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

[25] By endogenous variables we mean observable explanatory variables that may be correlated with the unobserved error term.

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 resoundingly reject the hypothesis that habits are a pure externality governed by the consumption of everyone else in the economy, in favor of the hypothesis that habits are based on own-consumption.

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 under flexibly estimated habit formation, with proxies from a variety of alternative linear (or approximately linear) models that have been explored in the asset pricing literature. It is encouraging for the habit-based framework that it does well in explaining portfolios of equity returns double-sorted on size and book-to-market characteristics. This finding, by itself, provides an interesting insight into the type of structural discount factor models that may underly some of the empirical success of leading empirical asset pricing models.

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., $X_t = r\left(C_t, C_{t-1}, X_{t-1}\right)$, 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 is hard to implement. Not only must the unknown habit, X_t, but be estimated, but in addition the recursive functional, $r(\cdot)$, must be estimated nonparametrically. The econometric theoretical results required to execute such an estimation have yet to be developed. The empirical work in this paper is a natural starting place for such an investigation, however, because the estimation of X_t nonparametrically would comprise one step in the recursive estimation procedure. We leave this interesting extension to future research.

Appendix

1. Derivation of Empirical Moment Condition

Recall that the conditional moment restrictions:

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

where

$$M_{t+1}=\delta\frac{MU_{t+1}}{MU_t}, \tag{27}$$

where

$$\begin{aligned} MU_t &= \frac{\partial U}{\partial C_t}=(C_t-X_t)^{-\gamma}-E_t\left[\sum_{j=0}^{L}\delta^j\left(C_{t+j}-X_{t+j}\right)^{-\gamma}\frac{\partial X_{t+j}}{\partial C_t}\right] && (28)\\ &= (C_t-X_t)^{-\gamma}E_t\left\{1-\sum_{j=0}^{L}\delta^j\left(\frac{C_{t+j}-X_{t+j}}{C_t-X_t}\right)^{-\gamma}\frac{\partial X_{t+j}}{\partial C_t}\right\}, && (29)\end{aligned}$$

where

$$\begin{aligned} X_t &= C_t f\left(1,\frac{C_{t-1}}{C_t},...,\frac{C_{t-L}}{C_t}\right)=C_t g\left(\frac{C_{t-1}}{C_t},...,\frac{C_{t-L}}{C_t}\right), && (30)\\ C_t-X_t &= C_t\left\{1-g\left(\frac{C_{t-1}}{C_t},...,\frac{C_{t-L}}{C_t}\right)\right\} && (31)\end{aligned}$$

hence

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

and

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

This can be expressed three different ways:

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

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

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

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

$$E_t\left(\delta\left(\frac{C_{t+1}-X_{t+1}}{C_t-X_t}\right)^{-\gamma}[R_{i,t+1}+h]-\delta^2\left(\frac{C_{t+2}-X_{t+2}}{C_t-X_t}\right)^{-\gamma}hR_{i,t+1}-1\right)=0$$

which coincides with their expression.

Alternatively we can write the conditional moment restrictions as:

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

with

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

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

Alternatively

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

with

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

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

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

2. Asymptotic Justification: Consistency and Convergence Rates of SMD Estimator (to be included)

3. Hypothesis Tests for External versus Internal Habit Formation (to be included)

4. 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.

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 (1999): "Small Sample Properties of the Model Specification Test Based on the Hansen-Jagannathan Distance," Unpublished paper, Arizona State University.

Ai, C., and X. Chen (2003): "Efficient Sieve Minimum Distance Estimation of Semiparametric Conditional Moment Models," *Econometrica,* forthcoming.

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

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.

Bradley, R. C. (1986): "Basic Properties of Strong Mixing Conditions," in *Dependence in Probability and Statistics*, ed. by E. Eberlein, and M. Taqqu, pp. 165–192. Birkhauser, Boston, Basel and Stuttgart.

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.

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.

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.

Chen, X., and Y. Fan (1999): "Consistent Hypothesis Testing in Semiparametric and Nonparametric Models for Econometric Time Series," *Journal of Econometrics*, 91, 373–401.

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.

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

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

Elbadawi, I., A. R. Gallant, and G. Souza (1983): "An Elasticity Can Be Estimated Consistently Without Prior Knowledge of Functional Form," *Econometrica*, 51(6), 1731–1751.

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.

Fama, E. F., and J. MacBeth (1973): "Risk, Return and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81, 607–636.

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 D. W. Nychka (1987): "Semi-nonparametric Maximum Likelihood Estimation," *Econometrica*, 55(2), 363–390.

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

Grenader, 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 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.

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.

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.

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.

Lewellen, J. W. (1999): "The Time Series Relations Among Expected Return, Risk, and book-to-market," *Journal of Financial Economics*, 54, 5–53.

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

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.

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 (2003): "The Time Series of the Cross Section of Asset Prices," Unpublished paper, Graduate School of Business, University of Chicago.

Newey, W. K., and J. Powell (2003): "Nonparametric Instrumental Regressions," *Econometrica* forthcoming.

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

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.

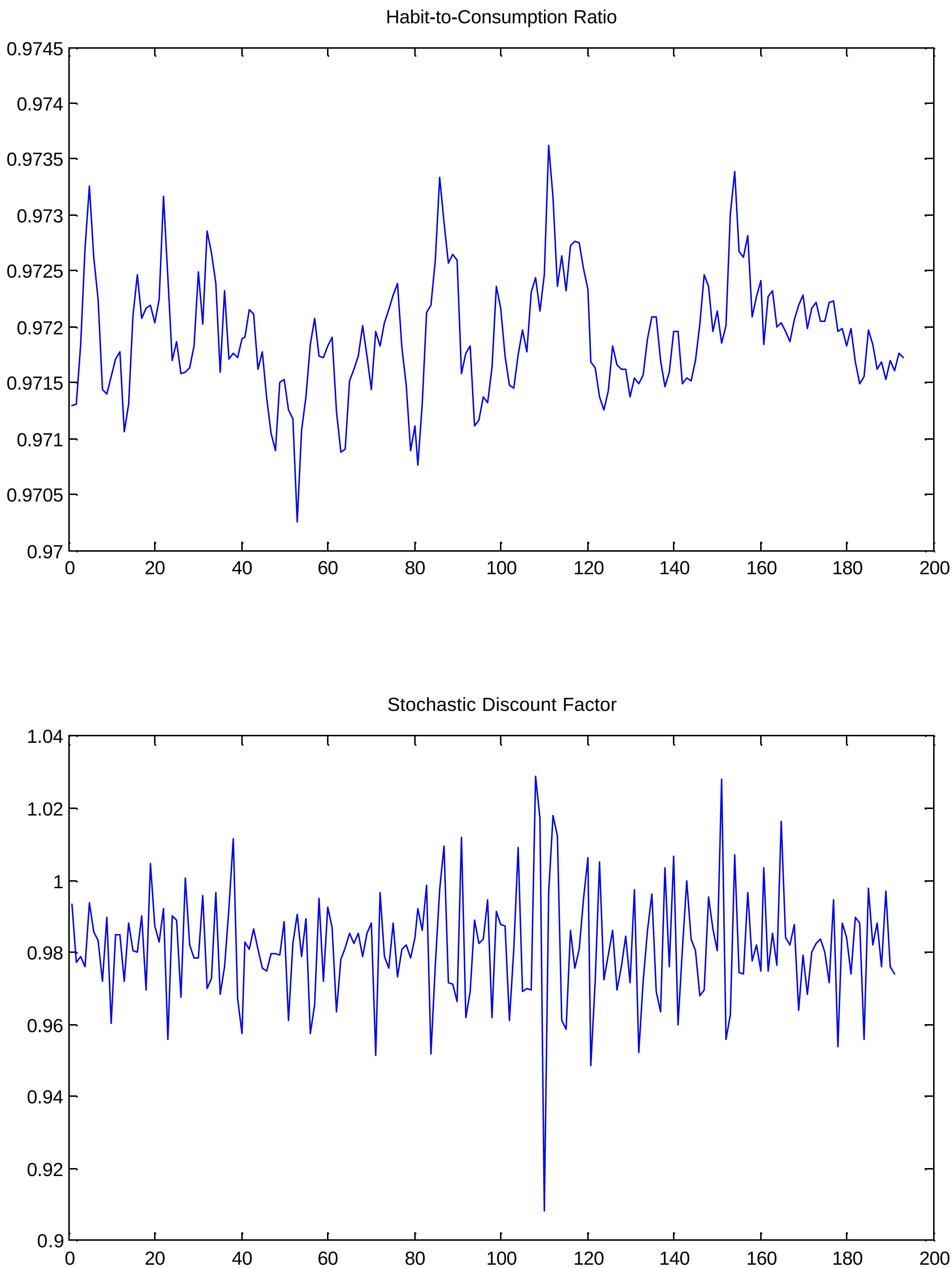
Sundaresan, S. (1989): "Intertemporally Dependent Preferences and the Volatility of Consumption and Wealth," *The Review of Financial Studies*, 2, 73–89.

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

Vuolteenaho, T. (2000): "Understanding the Aggregate Book-Market Ratio and its Implications to Current Equity-Premium Expectations," Unpublished paper, Harvard University.

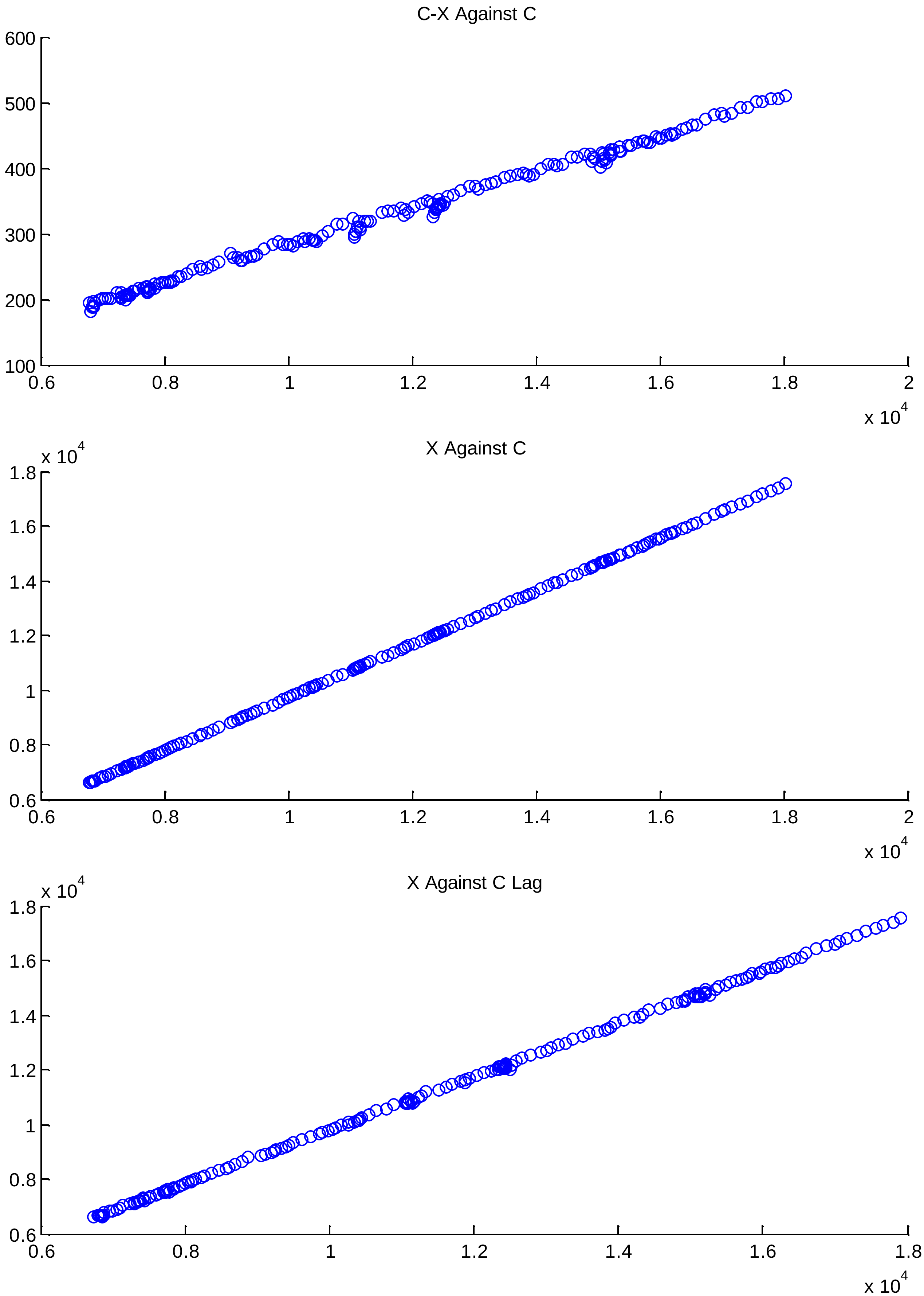
Wachter, J. (2002): "Habit Formation and Returns on Bonds and Stocks," Unpublished paper, Stern School of Business, New York University.

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.

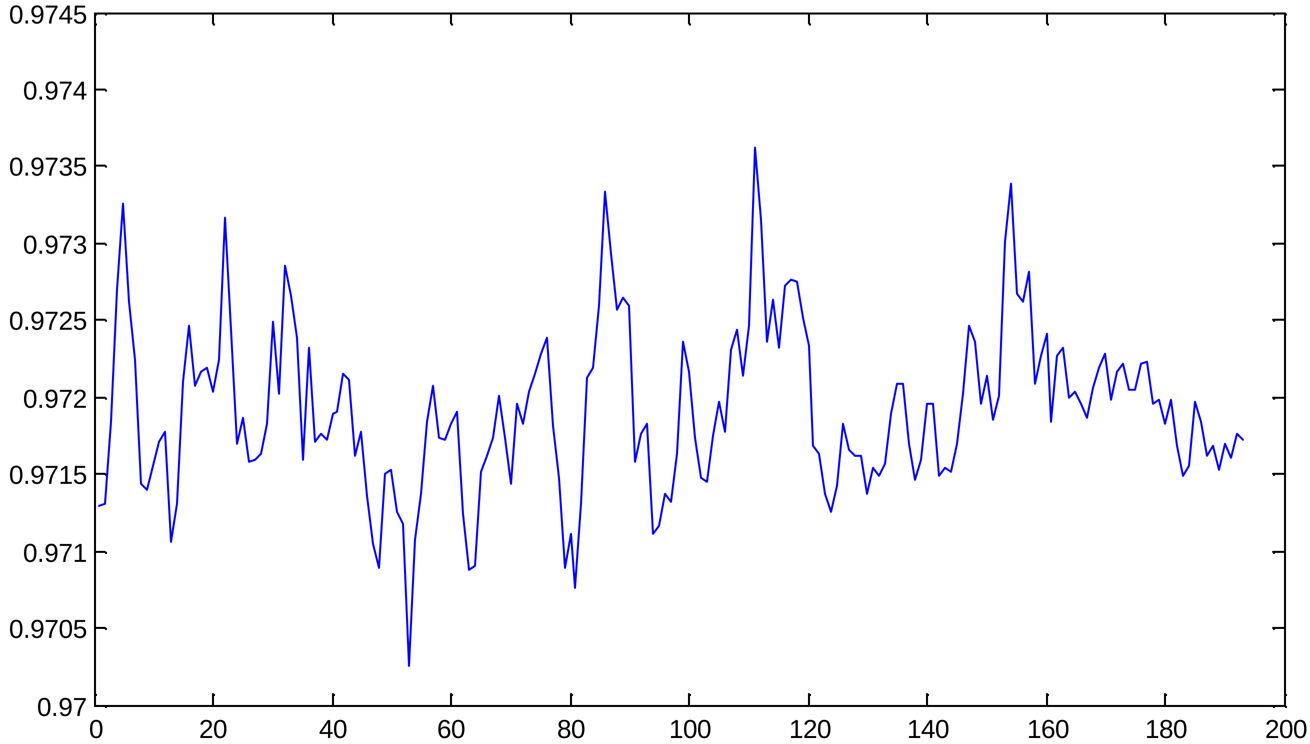
FIGURE 2



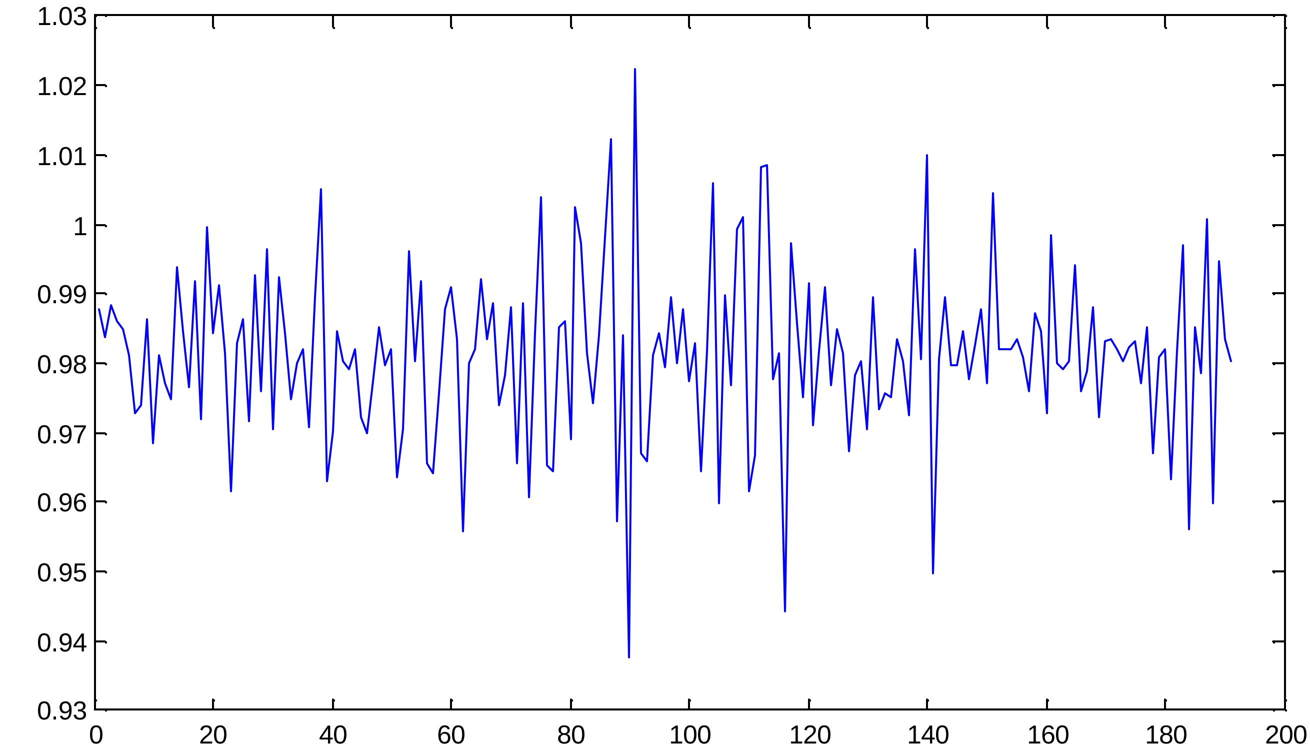
Notes : X is the estimated habit, C is consumption. Estimates use Group 1 assets, linear and squared instruments.

FIGURE 3

Habit-to-Consumption Ratio

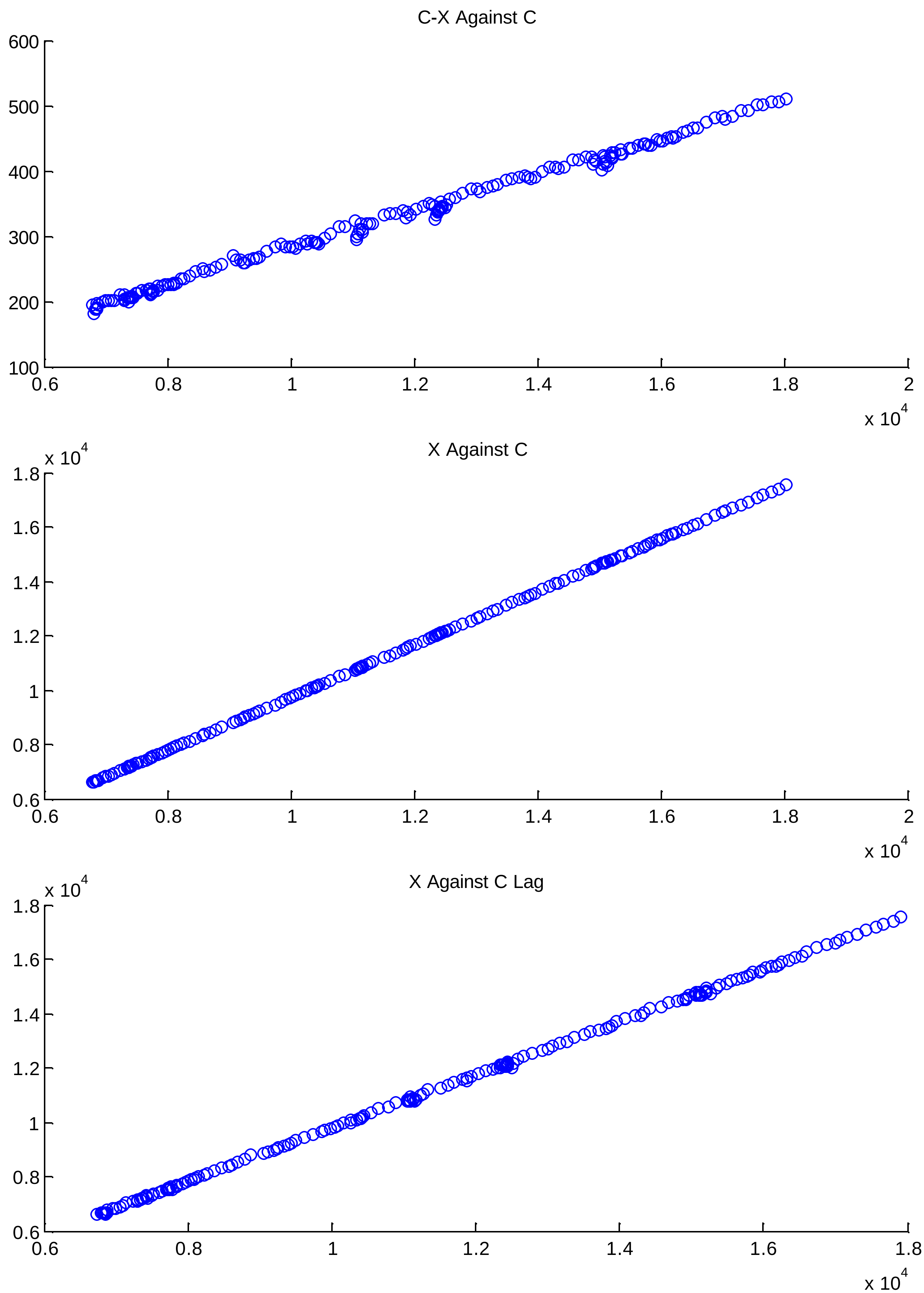


Stochastic Discount Factor



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

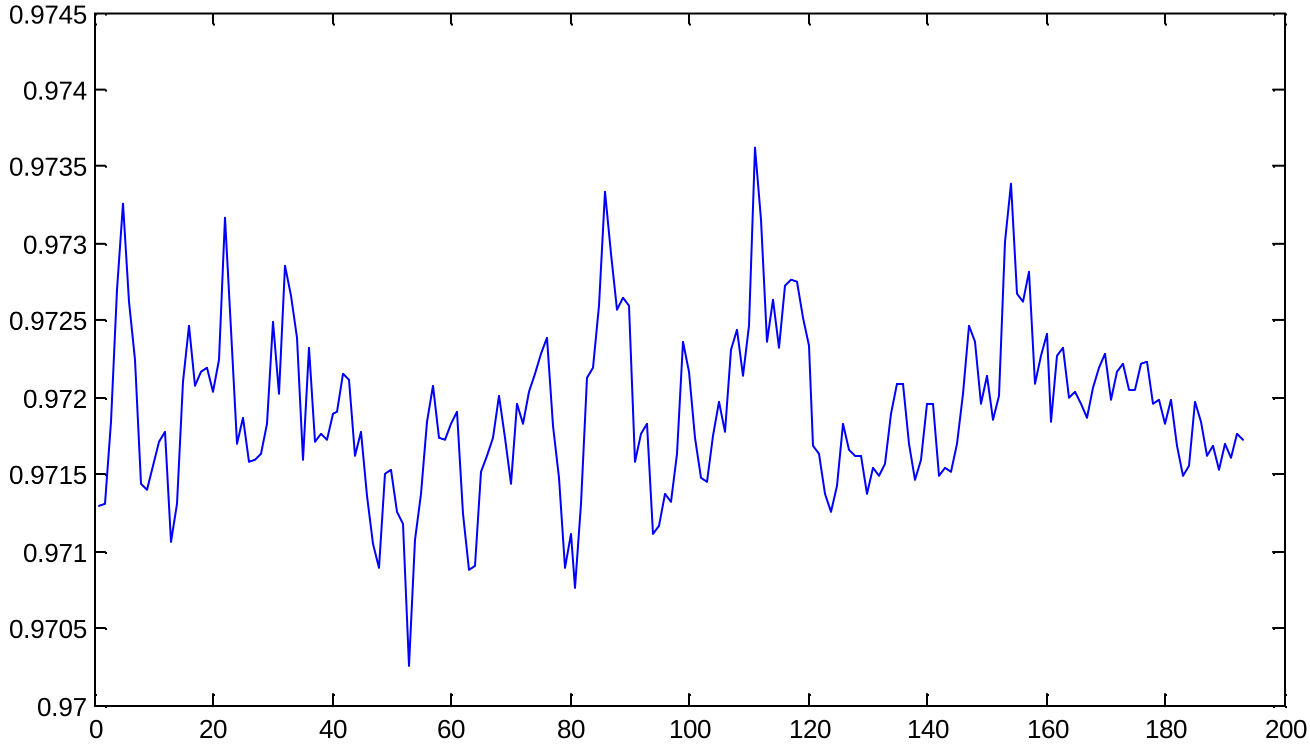
FIGURE 4



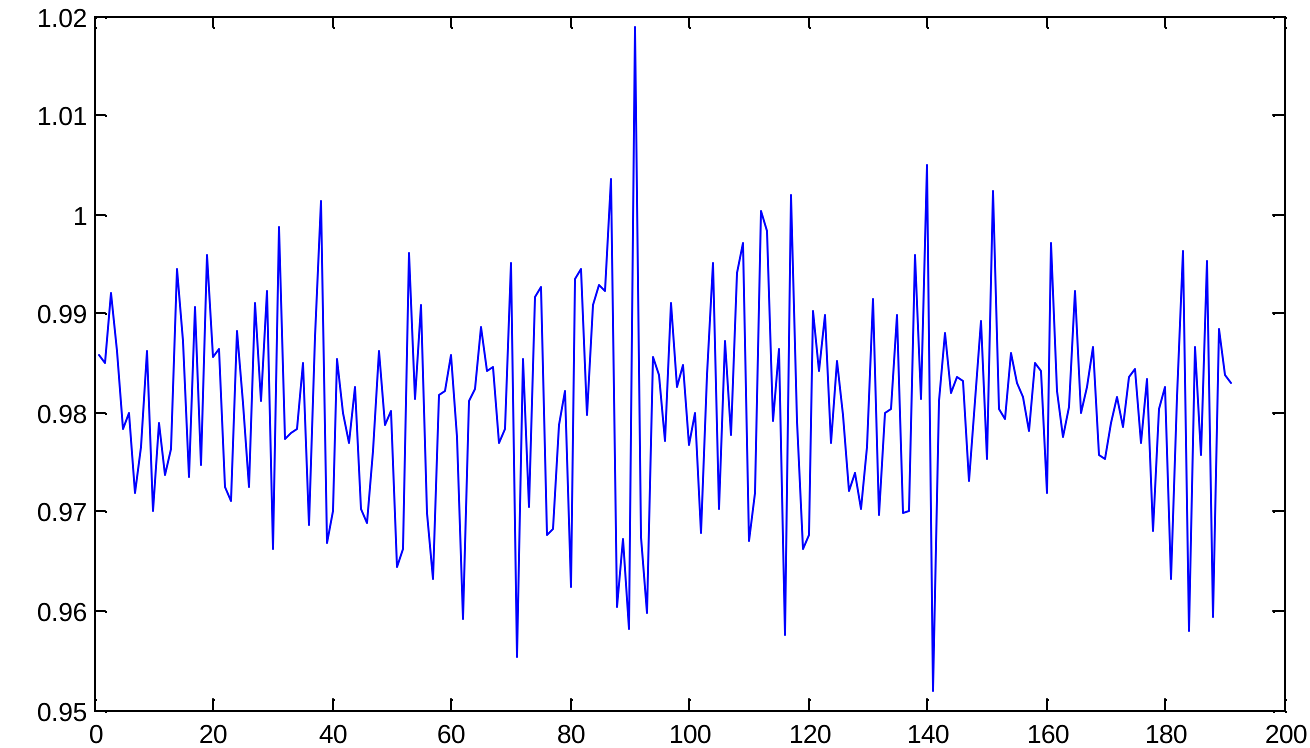
Notes : X is the estimated habit, C is consumption. Estimates use Group 1 assets, linear and cross term instruments.

FIGURE 5

Habit-to-Consumption Ratio

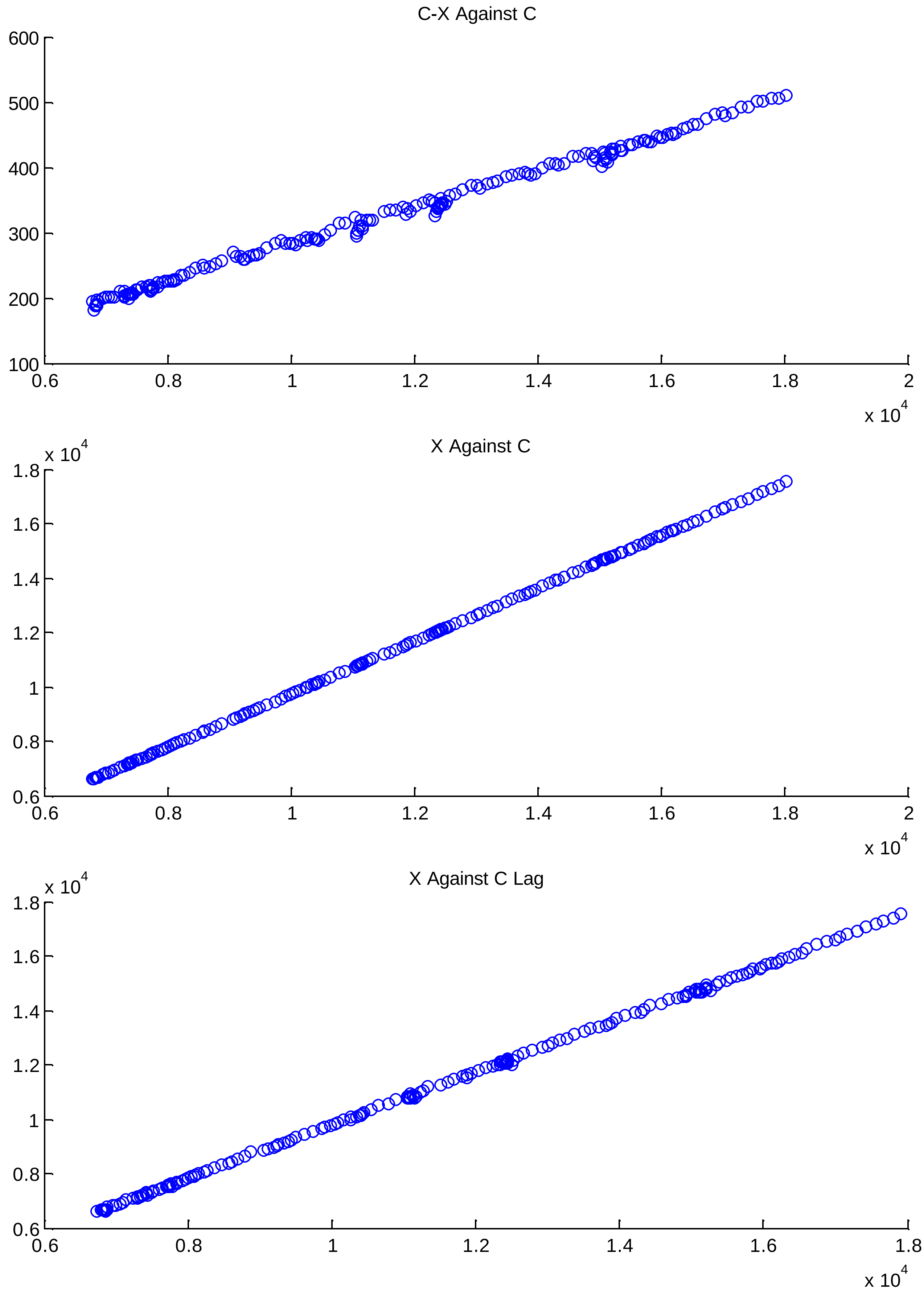


Stochastic Discount Factor



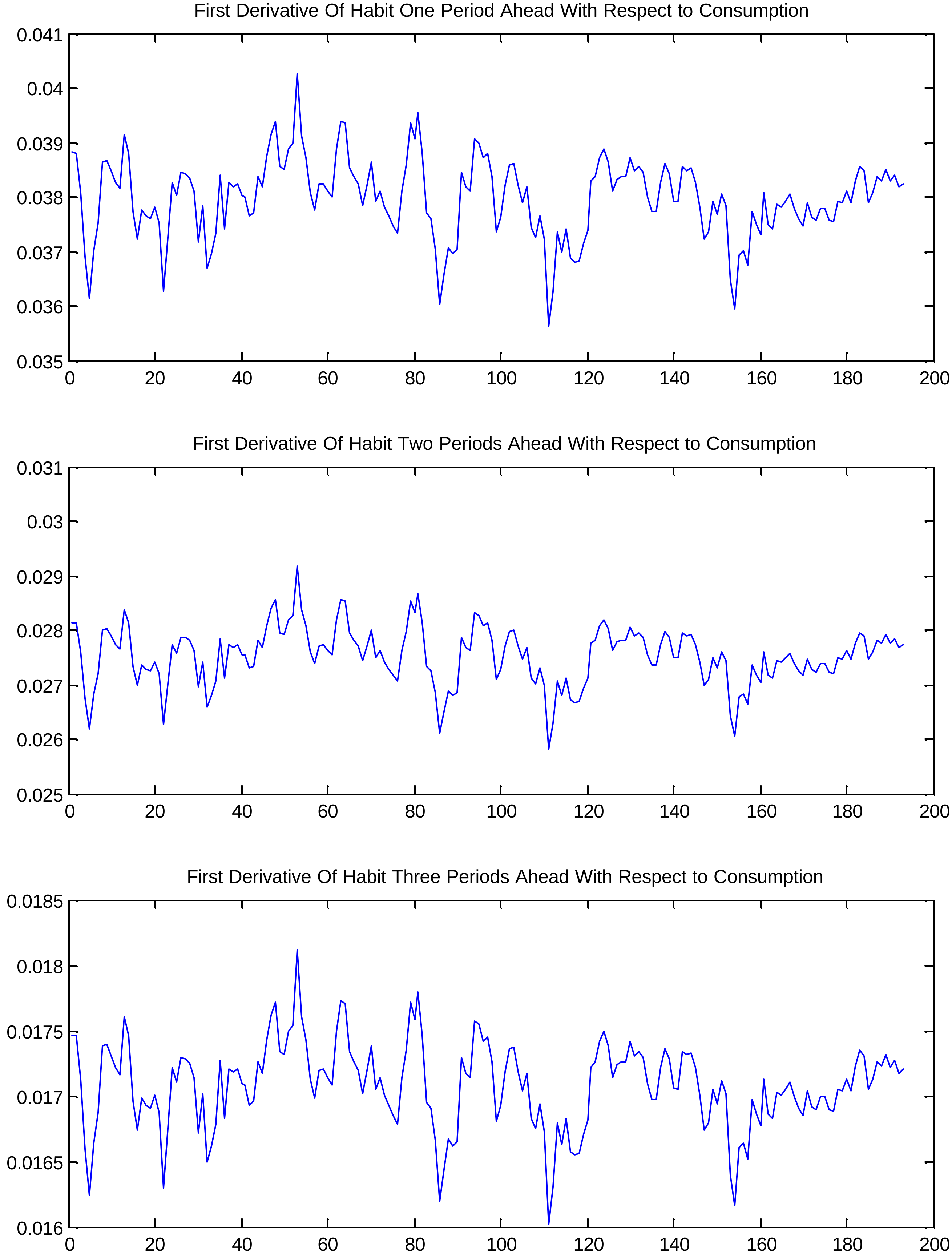
Notes : This figure plots the estimated habit-consumption ratio (top panel) and estimated stochastic discount factor (bottom panel) using Group 2 assets, linear instruments.

FIGURE 6



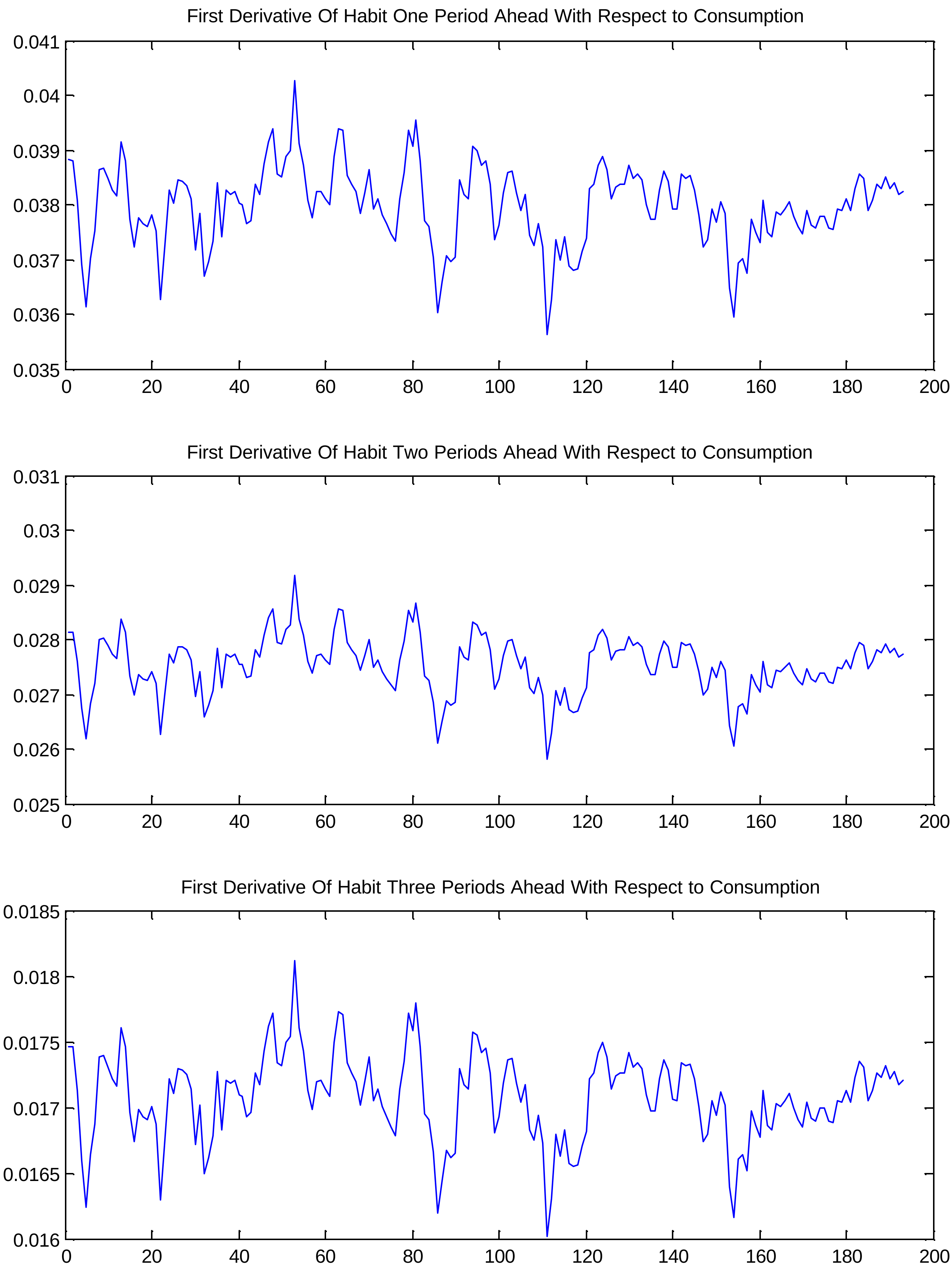
Notes : X is the estimated habit, C is consumption. Estimates use Group 2 assets, linear instruments.

FIGURE 7



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

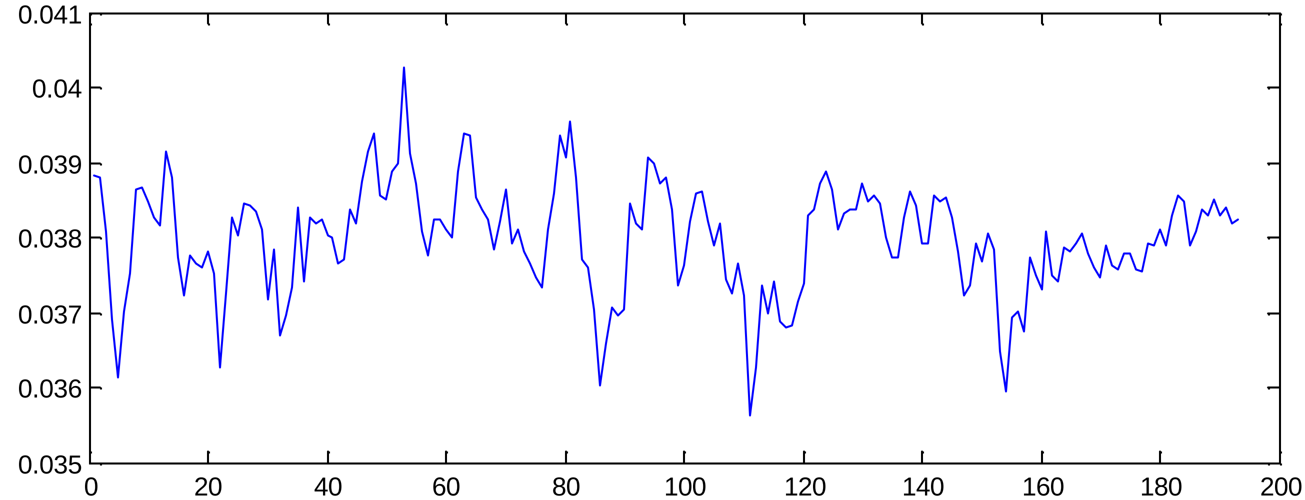
FIGURE 8



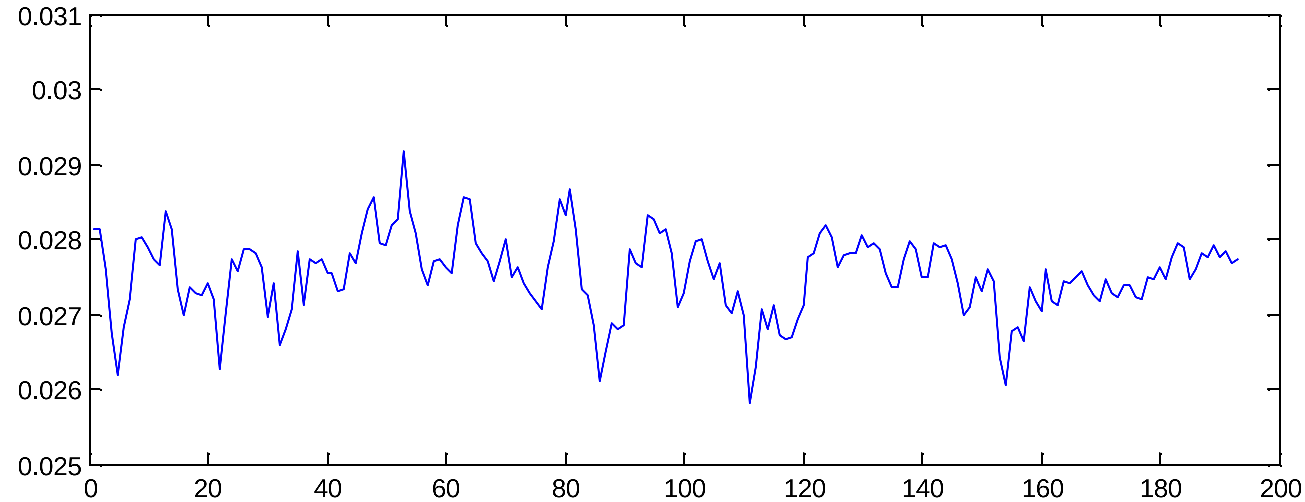
Notes : Estimates use Group 1 assets, linear and cross term instruments.

FIGURE 9

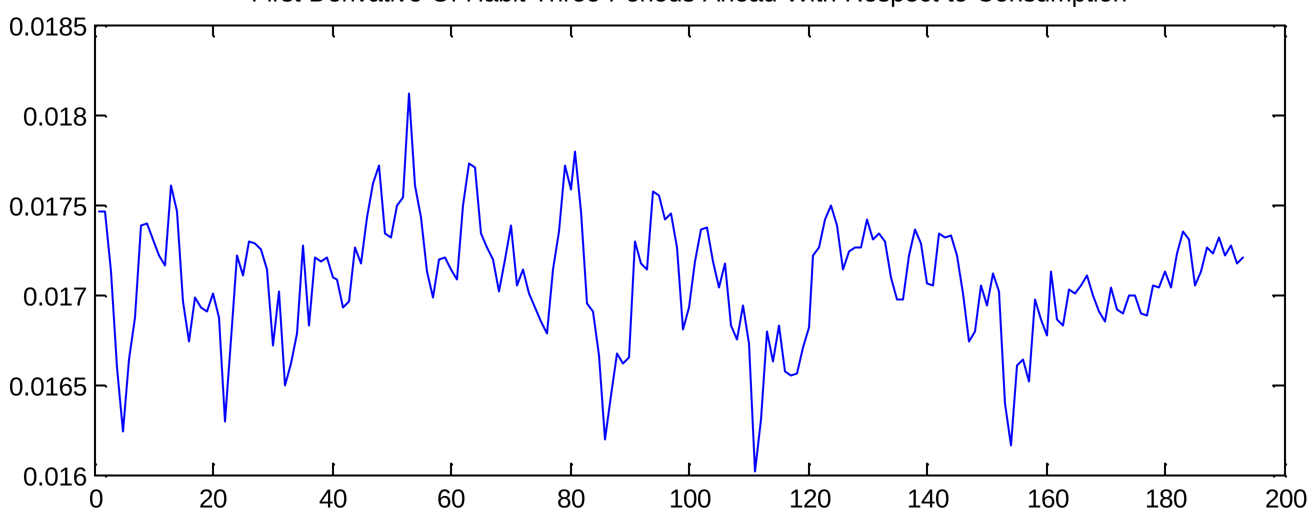
First Derivative Of Habit One Period Ahead With Respect to Consumption



First Derivative Of Habit Two Periods Ahead With Respect to Consumption



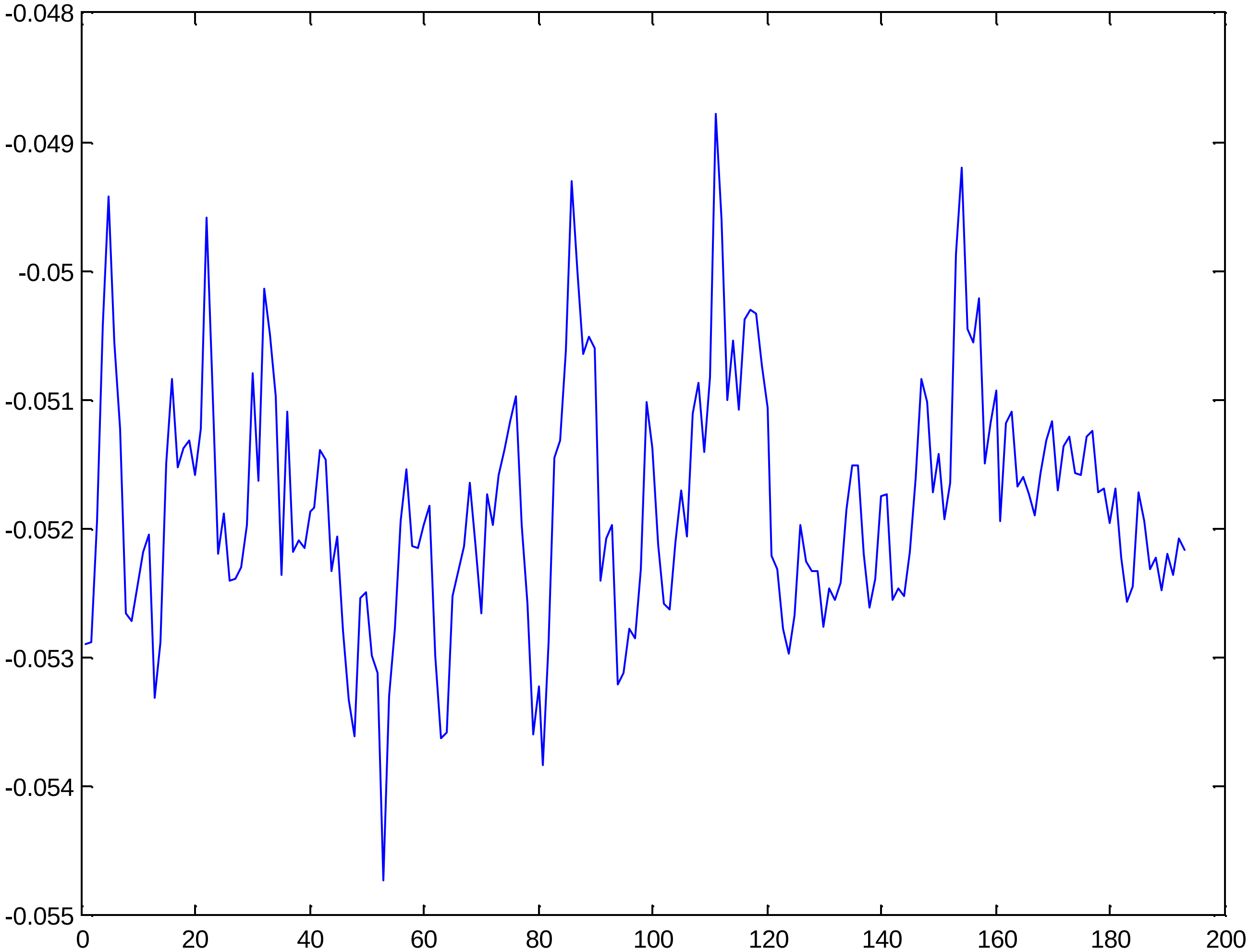
First Derivative Of Habit Three Periods Ahead With Respect to Consumption



Notes : Estimates use Group 2 assets, linear instruments.

FIGURE 10

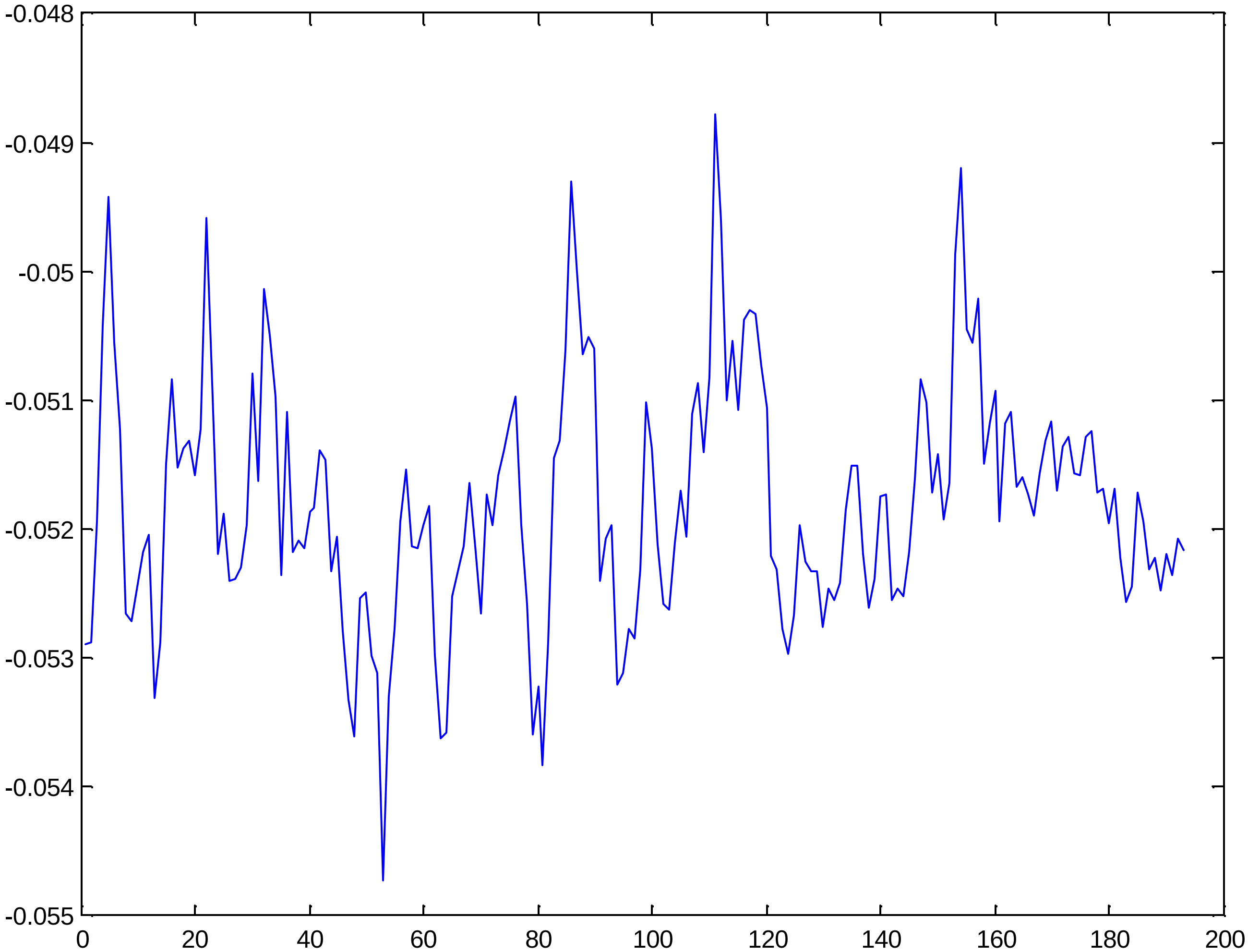
Second Derivative Of Habit One Period Ahead With Respect To Consumption



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

FIGURE 11

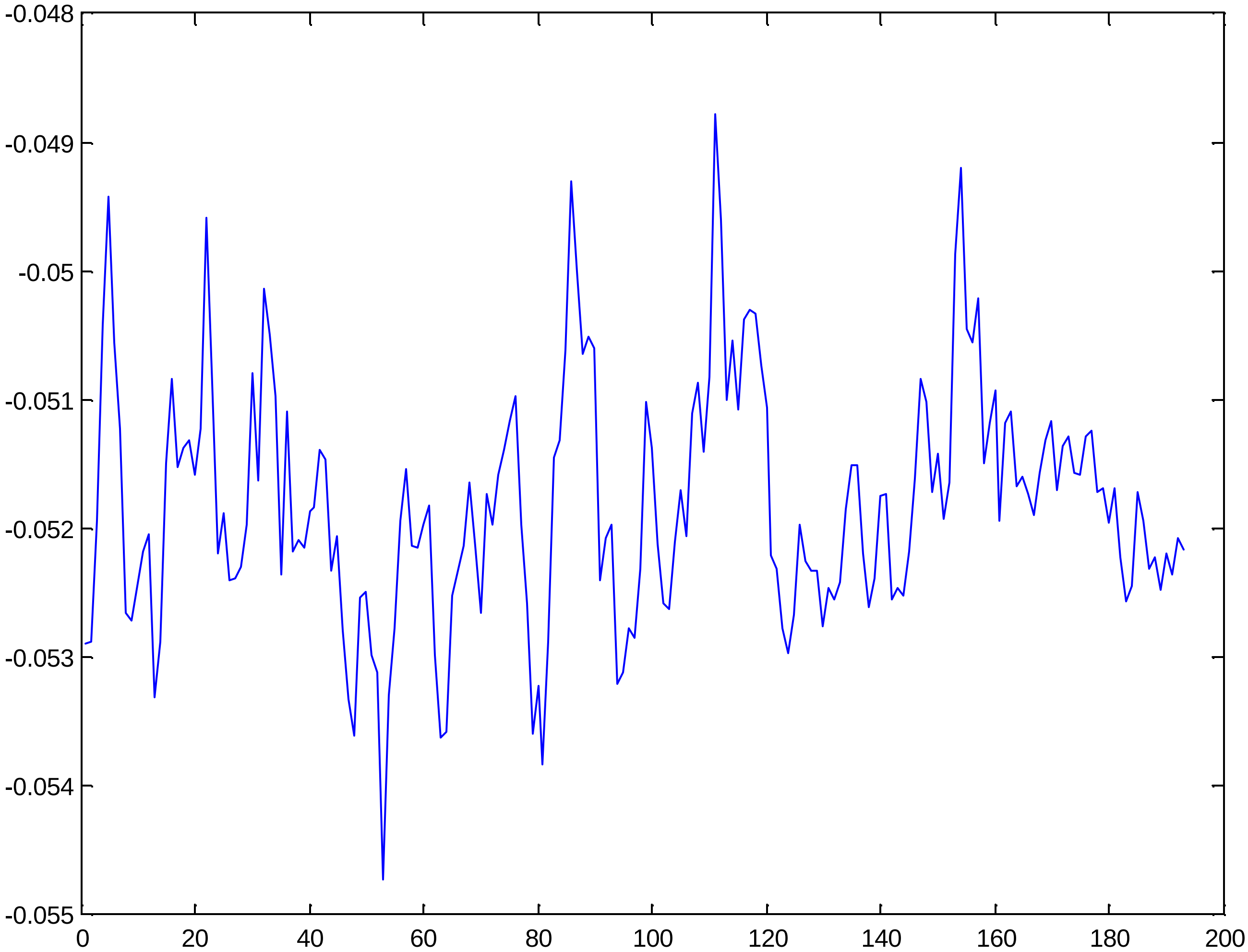
Second Derivative Of Habit One Period Ahead With Respect To Consumption



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

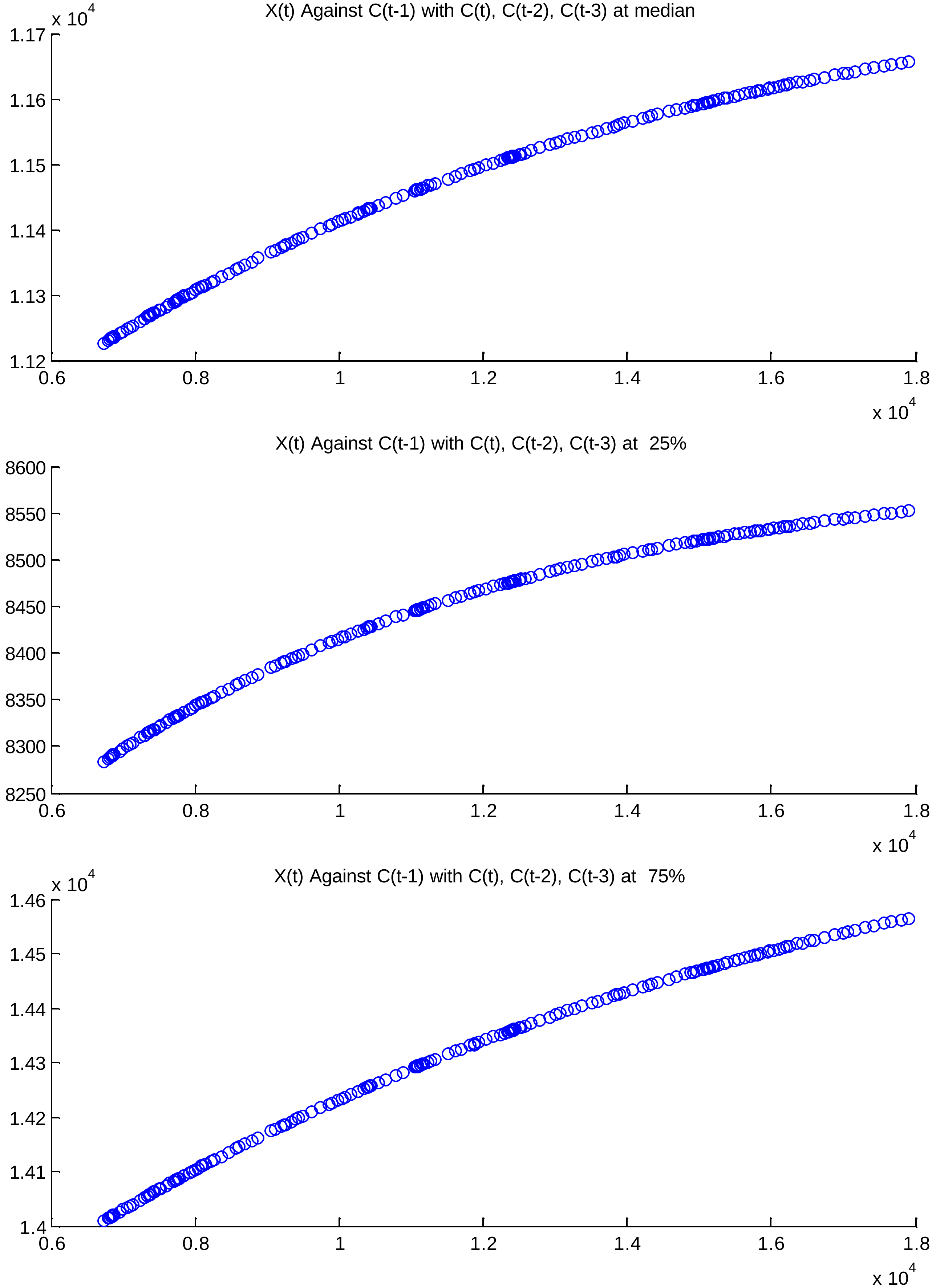
FIGURE 12

Second Derivative Of Habit One Period Ahead With Respect To Consumption



Notes : Estimates use Group 2 assets, linear instruments.

FIGURE 13



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