

Non-Time-Separable Utility: Habit Formation

I. Introduction

Thus far, we have considered time-separable lifetime utility specifications such as

$$E_t \left[\int_t^T U[C(s), s] ds \right] \quad (1)$$

where $U[C(s), s]$ is often taken to be of the form

$$U[C(s), s] = e^{-\rho(s-t)} u[c(s)] \quad (2)$$

so that utility at date s depends only on consumption at date s and not consumption at previous or future dates. As discussed previously, there is evidence that this type of utility specification has difficulties matching the empirical characteristics of U.S. consumption and the returns on the stock market relative to a risk-free asset (the equity risk premium). The lack of empirical support has led researchers to explore different, non-time-separable specifications for lifetime utility.

In these notes we consider utility functions in which *past* consumption plays a role in determining current utility. Such non-time-separate utility functions are said to display “habit persistence.” We summarize two models of this type that are based the articles of George Constantinides (1990) “Habit Formation: A Resolution of the Equity Premium Puzzle,” *Journal of Political Economy* 98, p.519-543 and of John Campbell and John Cochrane (1999) “By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior,” *Journal of Political Economy* 107, p.205-251. These models provide an interesting contrast in terms of their assumptions regarding the economy’s aggregate supplies of assets and the techniques we use to solve them. The model in the Constantinides paper is a simple example of a Cox,

Ingersoll, and Ross *Econometrica* (1985) production economy where asset supplies are perfectly elastic. It is solved using a Bellman equation approach. In contrast, the Campbell - Cochrane paper assumes a Lucas *Journal of Economic Theory* (1978) endowment economy where asset supplies are perfectly inelastic. Its solution is based on the economy's stochastic discount factor.

II. Assumptions of the Constantinides Model

A.1. Technology:

A single capital-consumption good can be invested in up to two different technologies. The first is a risk-free technology whose output, $\eta_1(t)$, follows the process

$$d\eta_1/\eta_1 = r dt. \tag{3}$$

The second is a risky technology whose output, $\eta_2(t)$, follows the process

$$d\eta_2/\eta_2 = \mu dt + \sigma dw. \tag{4}$$

Note that the specification of technologies fixes the expected rates of return and variances of the safe and risky investments. In this setting, individuals' asset demands determine equilibrium quantities of the assets supplied rather than asset prices. Since r , μ , and σ are assumed to be constants, there is a constant investment opportunity set.

A.2 Preferences:

Representative agents maximize expected utility of consumption, $c(t)$, of the form

$$E_0 \left[\int_0^\infty e^{-\rho t} \gamma^{-1} [c(t) - x(t)]^\gamma dt \right] \tag{5}$$

where

$$x(t) \equiv e^{-at} x_0 + b \int_0^t e^{-a(t-s)} c(s) ds. \tag{6}$$

Note that if $x_0 = b = 0$, utility is time-separable with constant relative risk aversion parameter $1 - \gamma$. For $b \neq 0$, the variable $x(t)$, which is an exponentially weighted average of past consumption, can be thought of as a “subsistence level” of consumption. Because current utility depends not only on current consumption but on past consumption, through $x(t)$, it is not time-separable, but exhibits habit persistence. An increase in consumption at date t decreases current marginal utility but increases the marginal utility of consumption at future dates. Of course, there are more general ways of modeling habit persistence, for example, $u[c(t), z(t)]$ where $z(t)$ is any function of past consumption levels. However, (5) and (6) is an analytically convenient specification.

A.3 Additional Parametric Assumptions:

$$1 - \gamma > 0 \tag{7}$$

$$W_0 > \frac{x_0}{r + a - b} > 0 \tag{8}$$

$$r + a > b > 0 \tag{9}$$

$$\rho - \gamma r - \frac{\gamma(\mu - r)^2}{2(1 - \gamma)\sigma^2} > 0 \tag{10}$$

$$0 \leq m \equiv \frac{\mu - r}{(1 - \gamma)\sigma^2} \leq 1 \tag{11}$$

where W_0 is the initial wealth of the representative individual. The reasons for making these parametric assumptions are the following. Condition (7) is required for utility to be concave. Note that $c(t)$ needs to be greater than $x(t)$ for the individual to avoid infinite marginal utility.¹ Conditions (8) and (9) ensure that an admissible (feasible) consumption and portfolio choice

¹Note that $\lim_{c(t) \rightarrow x(t)} [c(t) - x(t)]^{-(1-\gamma)} = \infty$.

strategy exists that enables $c(t) > x(t)$. To see this, note that the dynamics for the individual's wealth is given by

$$dW = \{[(\mu - r)\alpha(t) + r]W - c(t)\} dt + \sigma\alpha(t)W dw \quad (12)$$

where $\alpha(t)$, $0 \leq \alpha(t) \leq 1$ is the proportion of wealth that the individual invests in the risky technology. Now if $\alpha(t) = 0$ for all t , that is, one invests only in the riskless technology, and consumption equals a fixed proportion of wealth, $c(t) = (r + a - b)W(t)$, then

$$dW = \{rW - (r + a - b)W\} dt = (b - a)W dt \quad (13)$$

which is a first order differential equation in W having initial condition $W(0) = W_0$. Its solution is

$$W(t) = W_0 e^{(b-a)t} > 0 \quad (14)$$

so that wealth always stays positive. This implies $c(t) = (r + a - b)W_0 e^{(b-a)t} > 0$ and

$$\begin{aligned} c(t) - x(t) &= (r + a - b)W_0 e^{(b-a)t} - \left[e^{-at}x_0 + b \int_0^t e^{-a(t-s)}(r + a - b)W_0 e^{(b-a)s} ds \right] \\ &= (r + a - b)W_0 e^{(b-a)t} - \left[e^{-at}x_0 + b(r + a - b)W_0 e^{-at} \int_0^t e^{bs} ds \right] \\ &= (r + a - b)W_0 e^{(b-a)t} - \left[e^{-at}x_0 + (r + a - b)W_0 e^{-at}(e^{bt} - 1) \right] \\ &= e^{-at} [(r + a - b)W_0 - x_0] \end{aligned} \quad (15)$$

which is greater than zero by assumption (8).

Condition (10) is a transversality condition. It ensures that if the individual follows an optimal policy (which will be derived below), the expected utility of consumption over an infinite horizon is finite. As will be seen, condition (11) ensures that the individual wishes to

put positive levels of wealth in both the safe and risky technologies, that is, the individual's optimal portfolio choice has an interior solution. m turns out to be the optimal choice of the risky asset portfolio weight for the time-separable constant relative risk aversion case.²

III. Consumption and Portfolio Choice in the Constantinides Model

The solution technique presented here is different from that found in Constantinides's Appendix A, which, he mentions, is based on a martingale approach presented in the working paper M. Davis and A. Norman (1987) "Portfolio Selection with Transactions Costs," Imperial College, London. The current notes use a dynamic programming approach, similar to S. Sundaresan (1989) "Intertemporally Dependent Preferences and the Volatility of Consumption and Wealth," *Review of Financial Studies* 2, p.73-89, which is based on R. Merton (1971) "Optimum Consumption and Portfolio Rules in a Continuous-Time Model," *Journal of Economic Theory* 3, p.373-413.

The individual's maximization problem is

$$\max_{\{c, \alpha\}} E_t \left[\int_t^\infty e^{-\rho s} \gamma^{-1} [c(s) - x(s)]^\gamma ds \right] \equiv e^{-\rho t} J(W(t), x(t)) \quad (16)$$

subject to the intertemporal budget constraint given by equation (12). Given the assumption of an infinite horizon, we can simplify the analysis by separating out the factor of the indirect utility function that depends on calendar time, t . The "discounted" indirect utility function depends on two state variables, wealth, $W(t)$, and the state variable $x(t)$, the current subsistence level of consumption. Since there are no changes in investment opportunities (μ , σ , and r are all constant), there are no other relevant state variables. Similar to wealth, $x(t)$ is not completely exogenous but depends on past consumption. We can work out its dynamics using equation (6):

$$dx/dt = -ae^{-at}x_0 + bc(t) - ab \int_0^t e^{-a(t-s)}c(s) ds, \quad \text{or} \quad (17)$$

²Also, see R. Merton (1971) "Optimum Consumption and Portfolio Rules in a Continuous-Time Model," *Journal of Economic Theory* 3, p.373-413 or J. Ingersoll (1987) *Theory of Financial Decision Making* Rowman and Littlefield, Totowa, NJ, p.275.

$$dx = [bc(t) - ax(t)] dt. \quad (18)$$

Thus, changes in $x(t)$ are instantaneously deterministic. The Bellman equation is then

$$\begin{aligned} 0 &= \max_{\{c, \alpha\}} \{u(c(t), x(t), t) + L[e^{-\rho t} J]\} \\ &= \max_{\{c, \alpha\}} \{e^{-\rho t} \gamma^{-1} (c - x)^\gamma + e^{-\rho t} J_W [((\mu - r)\alpha + r)W - c] \\ &\quad + \frac{1}{2} e^{-\rho t} J_{WW} \sigma^2 \alpha^2 W^2 + e^{-\rho t} J_x (bc - ax) - \rho e^{-\rho t} J\}. \end{aligned} \quad (19)$$

The first order conditions with respect to c and α are:

$$\begin{aligned} (c - x)^{\gamma-1} &= J_W - bJ_x, \quad \text{or} \\ c &= x + [J_W - bJ_x]^{\frac{1}{\gamma-1}}, \end{aligned} \quad (20)$$

and

$$\begin{aligned} (\mu - r)WJ_W + \alpha\sigma^2 W^2 J_{WW} &= 0, \quad \text{or} \\ \alpha &= \frac{(\mu - r)}{\sigma^2} \frac{J_W}{(-WJ_{WW})}. \end{aligned} \quad (21)$$

Note that the additional term $-bJ_x$ in (20) reflects the fact that an increase in current consumption has the negative effect of raising the level of subsistence consumption, which decreases future utility. The form of (21), which determines the portfolio weight of the risky asset, is more traditional.

Substituting (20) and (21) back into (19), we obtain the equilibrium partial differential equation:

$$\frac{1 - \gamma}{\gamma} [J_W - bJ_x]^{\frac{\gamma}{1-\gamma}} - \frac{J_W^2}{J_{WW}} \frac{(\mu - r)^2}{2\sigma^2} + (rW - x)J_W + (b - a)xJ_x - \rho J = 0. \quad (22)$$

For the time-separable, constant relative risk aversion case ($a = b = x = 0$), we showed in

earlier notes that a solution for J is of the form $J(W) = kW^\gamma$, and since $u = e^{-\rho t}c^\gamma/\gamma$, and $u_c = e^{-\rho t}J_W$, optimal consumption is proportional to wealth:

$$c^* = (\gamma k)^{\frac{1}{(\gamma-1)}} W = W \left[\rho - r\gamma - \frac{1}{2} \left(\frac{\gamma}{1-\gamma} \right) \frac{(\mu-r)^2}{\sigma^2} \right] / (1-\gamma) \quad (23)$$

and

$$\alpha^* = m \quad (24)$$

where m is defined above in condition (11).

These results for the time-separable case might suggest a functional form for the non-time-separable case that looks like

$$J(W, x) = k_0[W + k_1x]^\gamma. \quad (25)$$

Making this guess, substituting it into (22), and setting the coefficients on x and W equal to zero, we find

$$k_0 = \frac{(r+a-b)h^{\gamma-1}}{(r+a)\gamma} \quad (26)$$

where

$$h \equiv \frac{r+a-b}{(r+a)(1-\gamma)} \left[\rho - \gamma r - \frac{\gamma(\mu-r)^2}{2(1-\gamma)\sigma^2} \right] > 0 \quad (27)$$

and

$$k_1 = -\frac{1}{r+a-b}. \quad (28)$$

Using equations (20) and (21), this implies

$$c^* = x(t) + h \left[W(t) - \frac{x(t)}{r+a-b} \right] \quad (29)$$

and

$$\alpha^* = m \left[1 - \frac{x(t)/W(t)}{r+a-b} \right]. \quad (30)$$

Interestingly, since $r+a > b$, by assumption, the individual always demands less of the risky asset compared to the case of no habit persistence. Thus we would expect lower volatility of wealth over time.

In order to find the dynamics of c^* , consider the change in the term $\left[W(t) - \frac{x(t)}{r+a-b} \right]$. Recall that the dynamics of $W(t)$ and $x(t)$ are given in equations (12) and (18), respectively. Using these, one finds

$$d \left[W(t) - \frac{x(t)}{r+a-b} \right] = \left\{ [(\mu-r)\alpha^* + r]W - c^* - \frac{bc^* - ax}{r+a-b} \right\} dt + \sigma \alpha^* W dw. \quad (31)$$

Substituting in for α^* and c^* from (29) and (30), one obtains

$$d \left[W(t) - \frac{x(t)}{r+a-b} \right] = \left[W(t) - \frac{x(t)}{r+a-b} \right] [n dt + m\sigma dw] \quad (32)$$

where

$$n \equiv \frac{r-\rho}{1-\gamma} + \frac{(\mu-r)^2(2-\gamma)}{2(1-\gamma)^2\sigma^2}. \quad (33)$$

Using this and (29), one can show (see Appendix A in Constantinides)

$$\frac{dc}{c} = \left[n + b - \frac{(n+a)x}{c} \right] dt + \left(1 - \frac{x}{c} \right) m\sigma dw. \quad (34)$$

Constantinides' Theorem 2 specifies parametric conditions for which the ratio $\frac{x}{c-x}$ has a stationary distribution. However, one sees from the stochastic term in (34), $\left(1 - \frac{x}{c} \right) m\sigma dw$, that consumption growth is smoother than in the case of no habit persistence. This is the intuition for why habit persistence can imply very smooth consumption paths, even though risk aversion, γ , may not be of a very high magnitude. The lower demand for the risky asset, relative to the time-separable case, can result in a higher equilibrium excess return on the risky asset and, hence, help explain the ‘‘puzzle’’ of a large equity premium.

IV. Assumptions of the Campbell - Cochrane Model

A.1 Technology:

Campbell and Cochrane consider a discrete-time endowment economy. Date t aggregate consumption, which also equals aggregate output, is denoted C_t , and it is assumed to follow an independent and identically distributed lognormal process

$$\ln(C_{t+1}) - \ln(C_t) = g + \nu_{t+1} \quad (35)$$

where $\nu_{t+1} \sim N(0, \sigma^2)$.

A.2 Preferences:

It is assumed that there is a representative agent who maximizes expected utility of the form

$$E_0 \left[\sum_{t=0}^{\infty} \delta^t \frac{(C_t - X_t)^{1-\gamma} - 1}{1-\gamma} \right] \quad (36)$$

where $\gamma > 0$ and X_t denotes the “habit level.” X_t is related to past consumption in the following manner. Define the “surplus consumption” ratio, S_t , as

$$S_t \equiv \frac{C_t - X_t}{C_t} \quad (37)$$

Then the log of surplus consumption is assumed to follow the auto-regressive process³

$$\ln(S_{t+1}) = (1 - \phi) \ln(\bar{S}) + \phi \ln(S_t) + \lambda(S_t) \nu_{t+1} \quad (38)$$

³This process is locally equivalent to $\ln(X_t) = \phi \ln(X_{t-1}) + \lambda \ln(C_t)$ or $\ln(X_t) = \lambda \sum_{i=0}^{\infty} \phi^i \ln(C_{t-i})$. The reason for the more complicated form in (38) is that it ensures that consumption is always above habit since $S > 0$. This precludes infinite marginal utility.

where

$$\lambda(S_t) = \frac{1}{\bar{S}} \sqrt{1 - 2 [\ln(S_t) - \ln(\bar{S})]} - 1 \quad (39)$$

and

$$\bar{S} = \sigma \sqrt{\frac{\gamma}{1 - \phi}} \quad (40)$$

The lifetime utility function in (36) looks somewhat similar to (5) of the Constantinides model. However, while Constantinides assumes that an individual's habit level depends on his or her own level of past consumption, Campbell and Cochrane assume the an individual's habit level depends on everyone else's past consumption. Thus, in the Constantinides model, the individual's choice of consumption, c_t , affects his future habit level, x_s , for all $s > t$, and he takes this into account in terms of how it affects his expected utility when he chooses c_t . This type of habit formation is referred to as *internal* habit. In contrast, in the Campbell and Cochrane model, the individual's choice of consumption, C_t , does not affect her future habit level, X_s , for all $s > t$, so that she views X_t as exogenous when choosing C_t . This type of habit formation is referred to as *external* habit or "keeping up with the Joneses."⁴ The external habit assumption simplifies the representative agent's decision making because habit becomes an exogenous state variable that depends on aggregate, not the individual's, consumption.

V. Consumption, Portfolio Choice, and Asset Pricing in the Campbell - Cochrane Model

Because habit is exogenous to the individual, the individual's marginal utility of consumption is

⁴A similar modeling was developed by A. Abel (1990) "Asset Prices under Habit Formation and Catching Up with the Joneses," *American Economic Review* 80, p.38-42.

$$u_c(C_t, X_t) = (C_t - X_t)^{-\gamma} = S_t^{-\gamma} C_t^{-\gamma} \quad (41)$$

and the representative agent's stochastic discount factor is

$$m_{t,t+1} = \delta \frac{u_c(C_{t+1}, X_{t+1})}{u_c(C_t, X_t)} = \delta \left(\frac{S_{t+1}}{S_t} \right)^{-\gamma} \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma} \quad (42)$$

If we define r_{ft} as the continuously-compounded risk-free real interest rate between dates t and $t + 1$, then it equals

$$r_{ft} = -\ln(E_t[m_{t,t+1}]) = -\ln(\delta) + \gamma g - \frac{1}{2}\gamma(1 - \phi) \quad (43)$$

which, by construction, turns out to be constant over time. One can also derive a relationship for the date t price of the market portfolio of all assets, denoted P_t . Recall that since we have an endowment economy, aggregate consumption equals the economy's aggregate output, which equals the aggregate dividends paid by the market portfolio. Therefore,

$$P_t = E_t[m_{t,t+1}(C_{t+1} + P_{t+1})] \quad (44)$$

or, equivalently, one can solve for the price - dividend ratio for the market portfolio.

$$\frac{P_t}{C_t} = E_t \left[m_{t,t+1} \frac{C_{t+1}}{C_t} \left(1 + \frac{P_{t+1}}{C_{t+1}} \right) \right] = \delta E_t \left[\left(\frac{S_{t+1}}{S_t} \right)^{-\gamma} \left(\frac{C_{t+1}}{C_t} \right)^{1-\gamma} \left(1 + \frac{P_{t+1}}{C_{t+1}} \right) \right] \quad (45)$$

As in the Lucas model, this stochastic difference equation can be solved forward to obtain

$$\frac{P_t}{C_t} = \delta E_t \left[\left(\frac{S_{t+1}}{S_t} \right)^{-\gamma} \left(\frac{C_{t+1}}{C_t} \right)^{1-\gamma} \left(1 + \delta \left(\frac{S_{t+2}}{S_{t+1}} \right)^{-\gamma} \left(\frac{C_{t+2}}{C_{t+1}} \right)^{1-\gamma} \left(1 + \frac{P_{t+2}}{C_{t+2}} \right) \right) \right] \quad (46)$$

$$\begin{aligned}
&= E_t \left[\delta \left(\frac{S_{t+1}}{S_t} \right)^{-\gamma} \left(\frac{C_{t+1}}{C_t} \right)^{1-\gamma} + \delta^2 \left(\frac{S_{t+2}}{S_t} \right)^{-\gamma} \left(\frac{C_{t+2}}{C_t} \right)^{1-\gamma} + \dots \right] \\
&= E_t \left[\sum_{i=1}^{\infty} \delta^i \left(\frac{S_{t+i}}{S_t} \right)^{-\gamma} \left(\frac{C_{t+i}}{C_t} \right)^{1-\gamma} \right]
\end{aligned}$$

The solutions can then be computed numerically by simulating the lognormal processes for C_t and S_t .

In this model, note that the coefficient of relative risk aversion is given by

$$\eta_t = -\frac{C_t u_{cc}}{u_c} = \frac{\gamma}{S_t} \quad (47)$$

which is time-varying and is relatively high when S_t is relatively low, that is, when consumption is low (a recession). This allows the model to explain a high risk-premium on risky assets (the market portfolio). To see this, recall the relationship between the “Sharpe ratio” and the coefficient of relative risk aversion when consumption is lognormally distributed:

$$\left| \frac{E[r_i] - r_f}{\sigma_{r_i}} \right| \leq \eta_t \sigma_c \quad (48)$$

Campbell and Cochrane show that the model can match the equity risk-premium because the average level of η_t can be set fairly high. Moreover, the model predicts that the equity risk-premium increases during a recession (when η_t is high), a phenomenon that seems to be present in the post-war U.S. stock market.