Econometrica, Vol. 66, No. 5 (September, 1998), 1223–1230

NOTES AND COMMENTS

HABIT FORMATION AND AGGREGATE CONSUMPTION

BY DAVID A. CHAPMAN[1]

1. INTRODUCTION

DYNAMIC ASSET PRICING MODELS based on the assumptions of complete markets and additively time-separable preferences have had a difficult time explaining both the unconditional and the conditional moments of asset returns. For example, Mehra and Prescott (1985) argue that the unconditional mean of the equity premium poses a "puzzle" for these models because it requires implausibly high levels of risk aversion on the part of the representative agent in order to reconcile these asset returns with the unconditional moments of aggregate consumption.

The "habit formation" preferences described, for example, in Sundaresan (1989), Constantinides (1990), and Detemple and Zapatero (1991) are important examples of attempts to resolve these problems by weakening the time-separability assumption. Equilibrium versions of these pricing models assume that a representative agent has a momentary utility function that depends on both the current rate of consumption *and* an exponentially-weighted index of past consumption rates. High rates of consumption in the recent past imply, ceteris paribus, lower utility from a given current consumption rate. In these models, current consumption decisions have an effect on the entire path of future consumption through the habit index.

This path dependence implies that it is possible for the marginal utility of consumption to be negative in some dates and states of the world if the current consumption level is too high and is unlikely to be sustained in the future. Negative marginal utility (equivalently, negative Arrow-Debreu "state-prices") is an implausible feature of an endowment economy equilibrium, since it implies that the representative agent does not demand all of the aggregate consumption good in those dates and states. A nonlinear joint restriction on the parameters of the utility function and the aggregate endowment process is required to ensure the nonnegativity of marginal utility. The question addressed below is whether or not this restriction is *economically* significant.

A common form of habit formation is used to construct an example of an endowment process that matches the unconditional moments of consumption growth and asset returns (i.e., resolves the equity premium puzzle) but implies negative marginal utility *with probability one*. In essence, this example posits an implausibly high level of habit formation, where "implausibly" means precisely that the agent has negative marginal utility of consumption along the proposed endowment path.

[1] I would like to thank John Heaton, Ehud Ronn, Dan Slesnick, Klaus Toft, two anonymous referees, an editor, participants in the macroeconomics seminar in the Economics Department of the University of Texas at Austin, and (in particular) George Constantinides and Sheridan Titman for helpful comments on earlier versions of this paper. A portion of this work was completed while I was on leave at the William E. Simon Graduate School of Business Administration at the University of Rochester. Of course, I retain the sole responsibility for any remaining errors or inconsistencies.

The implications of this example for model calibration and estimation are clear: An endowment economy with habit formation requires additional distributional assumptions beyond moment conditions, or (at a minimum) it requires checking to see if the proposed parameter combination satisfies the nonlinear restriction that guarantees positive state-prices. While the example below is simple enough to permit an explicit calculation of the nonnegativity condition, this is a problem that is endemic to the endowment economy structure. By contrast, habit formation in a production economy setting, as in Constantinides (1990), avoids the fallacy that generates negative state-prices, but it does so only by placing substantial additional structure on the form of asset returns.

2. THE STATE-PRICE DEFLATOR UNDER HABIT FORMATION

The economic environment is a standard continuous-time, finite-horizon model of security markets. The stochastic properties of equilibrium security prices are computed by solving the optimal consumption-portfolio problem of a representative agent and using a stochastic process for the endowment of the consumption good that mimics the important properties of measured aggregate consumption. Following Detemple and Zapatero (1991), preferences can be summarized by the utility function

$$(1)\qquad U(\{c_t\}) = E\left[\int_0^T \exp(-\beta t) u(c_t, z_t)\, dt\right],$$

where $E[\bullet]$ denotes the unconditional expectations operator and

$$(2)\qquad z_t = z_0 \exp(-\alpha t) + \delta \int_0^t \exp(-\alpha(t-s)) c_s\, ds$$

is the "habit index." The consumption process, c_t for $t \in [0, T]$ is assumed to be square-integrable. β, α, δ are positive constants that define time preference and habit formation, respectively. The persistence of past consumption on current utility is measured by α, and δ measures the "intensity" of habits. An exogenous nonnegative initial condition for the habit index is denoted by z_0.[2]

A common choice for the momentary utility function is

$$(3)\qquad u(c, z) = \frac{(c-z)^\eta}{\eta}$$

for $c \geq z$, $1 - \eta > 0$, and $\eta \neq 0$. If $c < z$, then $u(c, z) = -\infty$. Equation (3) is sometimes referred to as "addictive" habit formation, since current consumption can never fall below the contemporaneous level of the habit index.[3] Detemple and Zapatero (1991) prove that the general form of the state-price deflator, denoted ξ_t, in a model with habit formation defined by (1) and (2) is

$$(4)\qquad u_1(c_t, z_t) + \delta E_t\left[\int_t^T \exp(-(\beta + \alpha)(s-t)) u_2(c_s, z_s)\, ds\right] = \varphi \xi_t$$

[2]See Detemple and Zapatero (1991) for a complete description of the probabilistic structure of the model and for the technical conditions imposed on the momentary utility function, $u(\bullet, \bullet)$.

[3]This is the most common form of "intrinsic" habit formation, in which the temporal dependence is with respect to the consumer's own past consumption choices. "Extrinsic" habit formation, as in Campbell and Cochrane (1997), assumes that the current utility is related to "... the history of aggregate consumption rather than the history of individual consumption." (Campbell and Cochrane (1997, p. 4).)

for all $t \in [0, T]$, where φ is the Lagrange multiplier for the static budget constraint under the equivalent martingale measure.[4]

When evaluated at the aggregate endowment, (4) imposes a joint restriction on the utility function and the stochastic process for the endowment.

STATE-PRICE RESTRICTION: *Given a stochastic process for the aggregate endowment* e_t *and denoting* (2) *evaluated at the exogenous endowment by* z_t^e, *it must be true that*

$$u_1(e_t, z_t^e) + \delta E_t\left[\int_t^T \exp(-(\beta+\alpha)(s-t))u_2(e_s, z_s^e)\,ds\right] > 0 \tag{5}$$

and $e_t > z_t^e$ *for all* $t \in [0, T]$.

This condition implies that the representative agent demands all of the aggregate endowment at all dates. Conversely, if (5) is violated, then the agent's marginal utility of consumption (including the impact of current consumption on future utility) is negative for some dates along the proposed endowment path. This means that the representative agent will not demand all of the good at those dates.[5]

3. AN EXAMPLE WITH NEGATIVE STATE-PRICES

The example examined below will use parameter combinations that satisfy the following conditions: (i) the unconditional mean and standard deviation of the (exogenous) endowment growth process are equal to the unconditional mean and standard deviation of measured consumption growth; and (ii) the unconditional mean of the short-rate and the unconditional mean of the ratio of the (real) risk premium to the real volatility of the return on a broad stock portfolio are equal to their sample counterparts in the U.S. data. These conditions constitute a "solution" to the equity premium puzzle in the sense of reconciling measured consumption growth with the moments of asset returns.

An endowment process that is consistent with finite marginal utility in the addictive habit formation case (and growth over time in the consumption rate) is

$$y_t = y_0 \cdot \exp\left(\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma B_t\right) \tag{6}$$

for $t \in [0, T]$, where $y_t \equiv e_t - z_t^e$; i.e., the *difference* between the endowment and the habit index is defined as a lognormal diffusion. A process of this general form is derived in Constantinides (1990) as the (unique) optimal consumption choice of a representative agent with utility given by an infinite-horizon version of (1) and (3) and faced with an asset allocation decision between a risk-free asset with a constant instantaneous interest rate and a single risky production process that follows a lognormal diffusion.

The form of the state-price deflator in this economy is

$$\xi(y_t, \tau) = y_t^{\eta-1}\left[1 - \frac{\delta}{\kappa}(1 - \exp(-\kappa\tau))\right] \tag{7}$$

[4] The form of (4) is an extension of the standard form of the state-price deflator under time-separable preferences: $u_c(c_t) = \varphi\xi_t$.

[5] See Detemple and Zapatero (1991), p. 1645 for a similar argument about the importance of the "State-Price Restriction."

where

$$\kappa \equiv (1-\eta)\left[\mu + \frac{1}{2}\sigma^2(\eta - 2)\right] + (\alpha + \beta),$$

$\tau \equiv T - t$, $1 - \eta > 0$.[6] In this case, a necessary and sufficient condition for strict positivity of the state-price deflator with probability one is

$$\vartheta \equiv \frac{\delta}{\kappa}(1 - \exp(-\kappa\tau)) < 1. \tag{8}$$

Alternatively, if (8) is violated, then state-prices are negative, with probability one.

The instantaneous real interest rate in a complete markets equilibrium model is equal to $-\mu_\xi(t)/\xi_t$, where $\mu_\xi(t)$ is the instantaneous drift of the state-price deflator (see Duffie (1996)). Using (7), Itô's Lemma implies that

$$r(t) = r = \kappa - (\alpha + \beta) - \frac{\kappa\delta \cdot \exp(-\kappa\tau)}{\kappa - \delta(1 - \exp(-\kappa\tau))}. \tag{9}$$

The Sharpe-ratio process for a given risky asset is defined as

$$SR(t) \equiv \frac{\mu_R(t) - r_t}{\sigma_R(t)} \tag{10}$$

where $\mu_R(t)$ denotes the (instantaneous) expected return on the risky asset, $\sigma_R(t)$ is the (instantaneous) volatility of the risky asset, and r_t is the short-rate. For an individual security, it is equal to $-\sigma_\xi(t)/\xi_t$, where $\sigma_\xi(t)$ is the diffusion coefficient of the state-price deflator. Again, using (7) and Itô's Lemma, it follows that

$$SR(t) = SR = (1 - \eta)\sigma. \tag{11}$$

The endowment process consistent with (6) is defined by the following stochastic differential equation:

$$\frac{de_t}{e_t} = [(\mu + \delta) - (\mu + \alpha)x_t]\,dt + \sigma(1 - x_t)\,dB_t, \tag{12}$$

where $x_t \equiv z_t^e/e_t$ and it is assumed that $\mu + \alpha - \delta - \sigma^2 > 0$.[7] This parameter restriction ensures that the mean of the state variable x_t under its stationary distribution is strictly positive. The unconditional moments of endowment growth can be calculated numerically as[8]

$$\frac{E(de/e)}{dt} \cong (\mu + \delta) - (\mu + \alpha)\int_0^1 x P_\infty^x(x;\psi)\,dx \tag{13}$$

[6] The proof of the form of (7) follows directly from the general form of the state-price deflator, (4), the utility function, (3), the definition of y_t, (6), and Itô's Lemma.

[7] The proof of (12) follows Appendix A in Constantinides (1990). In particular, see Equations (A17) through (A19).

[8] This is similar to the approach used in Constantinides (1990).

and

$$\text{(14)} \qquad \frac{\operatorname{var}(de/e)}{dt} \cong \sigma^2 \int_0^1 (1-x)^2 P_\infty^x(x;\psi)\,dx$$

where $\psi \equiv (\beta, \eta, \alpha, \delta, \mu, \sigma)'$ and $P_\infty^x(x;\psi)$ is the stationary distribution of x_t, which is defined as

$$\text{(15)} \qquad P_\infty^x(x;\psi) = \left(\frac{2\delta}{\sigma^2}\right)^{\nu} \frac{1}{\Gamma(\nu)} \exp\left(\frac{2\delta}{\sigma^2}\right) x^{-(1+\nu)}(1-x)^{\nu-1} \exp\left(-\frac{2\delta}{\sigma^2}\frac{1}{x}\right)$$

where $\nu \equiv (2(\mu + \alpha - \delta)/\sigma^2) - 1$.

The sample moments used to calibrate the model parameters are determined as follows:

1. *Consumption Data*: Annual data on real per capita personal consumption expenditures on nondurables and services from 1948 to 1993 have a sample mean growth rate of 1.85% per year, with a standard error of the mean of 0.19% per year, and a standard deviation of consumption growth of 1.26% per year.

2. *Interest Rate Data*: The real one-month interest rate is defined as the (end-of-month) yield-to-maturity on a one-month Treasury bill minus the ex post inflation rate for that month, defined as the logarithmic first difference in the "Consumer Price Index—All Urban Consumers" (CPI). The yield data comes from the "Riskfree Rate" file on the Center for Research in Security Price's (CRSP) Government Bond Tape. The sample mean for the 551 months of usable data from January 1948 to December 1993 is 0.65% per year, with a standard error of the estimate of 0.17% per year.

3. *Stock Return Data*: Monthly returns on the CRSP value-weighted portfolio were collected for the period from January 1947 to December 1993. Ex post real returns were computed by subtracting the realized inflation rate. Monthly estimates of the volatility of real stock returns were constructed as in Schwert (1989). This volatility series was used to construct a proxy to the Sharpe-ratio process. The sample mean of this series from January 1948 to December 1993 was 0.64, with a standard error of the estimate of 0.11.

Assume that $\tau = 10$ years and let $\beta = 0.037$.[9] If the consumption moments, the mean of the Sharpe-ratio process, and the mean of the short-rate are all set equal to their sample counterparts, there are *no* parameter combinations that simultaneously satisfy the moment conditions and the nonnegativity condition (8)! For example, for choices of $\eta \in [-2.6, -4.6]$, there are choices of μ, σ, δ, and α that satisfy the four moment conditions, but $\vartheta \in [1.3131, 1.0410]$, which means that the state-price deflator is strictly negative (equivalently, marginal utility is negative) with probability one.

A complete characterization of the region of the parameter space consistent with all of the moment conditions (evaluated at their sample means) is beyond the scope of this paper, but the parameter values considered above are not in any sense pathological. In fact, values of $\eta > -2.6$ were inconsistent with the condition $\mu + \alpha - \delta - \sigma^2 > 0$ and the average Sharpe-ratio process, and $\eta < -4.6$ implied no solution to the set of nonlinear restrictions on the parameters implied by the moment conditions alone. It is possible that if the moments of consumption growth and asset returns are allowed to vary from their sample means, there may exist some set of model parameters that is consistent with

[9] The assumption of a ten-year horizon for the representative agent's problem was chosen to be conservative, since the probability of nonnegative state-prices is strictly increasing in τ. The value of β is chosen to be consistent with Constantinides (1990).

strictly positive state-prices.[10] However, the point here is that the nonnegativity constraint is binding over an important region of the parameter space.

The sample moments calculated using the data from 1948 to 1993 are different in two important respects from the moments that commonly define the equity premium puzzle, constructed in Mehra and Prescott (1985) using data from 1889 to 1978. First, post-1948 consumption growth variability is dramatically lower, 1.26% per year versus 3.57% per year. Second, the ratio of the (real) risk premium to the volatility of a broad index of stock returns is dramatically higher in the more recent period, at 0.64 versus 0.44.[11] This is consistent with the evidence in Schwert (1989), which shows that stock return volatility during the 1930's was substantially higher than over any other period in U.S. history.

Using the Mehra and Prescott (1985) sample moments does not change the conclusions formed from the 1948 to 1993 data. If anything, the problem is worse. There are no parameter values that reconcile the point estimates of the mean and standard deviation of consumption growth with the average level of the real risk-free rate or the average Sharpe ratio. There is a narrow range of parameter values that can match the consumption growth moments to the level of the real rate and a Sharpe ratio equal to (roughly) one standard deviation above the mean (i.e., $E(SR(t)) = 0.57$).[12] However, none of these parameter combinations satisfy the nonnegativity condition. Again, the point is not that there are not some combinations of plausible consumption growth and asset return moments that might possibly satisfy the nonnegativity condition, but rather that the condition itself is economically significant and not a pathological case.

4. A PRODUCTION ECONOMY

The results in the previous section raise an obvious question: Why are the calibration exercises in Constantinides (1990) successful when an endowment version of a "similar" model fails so completely? The answer is that the models are not quite as similar as they might appear given the general form of (12). The moments of asset returns and consumption growth are not identified in the same manner, and the production economy generates fundamentally different restrictions on the model parameters.

First, the asset market moments are defined exogenously to be a constant risk-free rate, $\tilde{r}$, and a lognormal diffusion, defined by the drift parameter $\tilde{\mu}$ and the diffusion parameter $\tilde{\sigma}$.[13] Matching the two asset market moments, the risk premium, and the volatility of real risky asset returns, can be done without reference to any of the utility parameters. An explicit assumption used in identifying the asset market moment condi-

[10] If the mean of the Sharpe-ratio process is lowered by one standard deviation, the range of η consistent with a solution to the moment matching exercise is enlarged. However, extensive numerical evaluations of the model for different values of the moment conditions did not find parameter values that both matched the moments and satisfied the nonnegativity condition.

[11] The point estimate of the average Sharpe ratio over the 1889 to 1978 period is constructed using the nine decades of data for the real risk premium and the real volatility of S&P 500 returns reported in Table 1 of Mehra and Prescott (1985). The range of the Sharpe ratios was from 0.01 in the 1929 to 1938 subperiod to 1.40 in 1949 to 1958 subperiod. The standard error of the mean Sharpe ratio is 0.15.

[12] Specifically, $\eta \in [-1.1, -1.4]$, $\delta \in [0.35, 0.48]$, $\alpha \in [0.38, 0.54]$, $\mu \in [0.046, 0.055]$, and $\sigma \in [0.238, 0.271]$.

[13] The '$\tilde{\ }$' is used to distinguish parameters in the Constantinides (1990) production economy from the parameters in the previous section.

tions is that the (instantaneous) return premium on the risky asset and the (instantaneous) volatility of the real return in the risky asset are uncorrelated.

Consumption growth in the Constantinides (1990) production economy is determined by two conditions that are analogous to—but different from—(13) and (14). Endowment growth and the variance of endowment growth in the production economy (defined at the optimum) are

$$\frac{E(d\tilde{e}/\tilde{e})}{dt} \cong (\tilde{\omega} + \tilde{\delta}) - (\tilde{\omega} + \tilde{\alpha}) \int_0^1 \tilde{x} P_\infty^{\tilde{x}}(\tilde{x}; \tilde{\psi})\, d\tilde{x} \tag{16}$$

and

$$\frac{\operatorname{var}(d\tilde{e}/\tilde{e})}{dt} \cong \frac{(\tilde{\mu} - \tilde{r})^2}{(1-\tilde{\eta})^2 \tilde{\sigma}^2} \int_0^1 (1-\tilde{x})^2 P_\infty^{\tilde{x}}(\tilde{x}; \tilde{\psi})\, d\tilde{x} \tag{17}$$

where

$$\tilde{\omega} \equiv \frac{\tilde{r} - \tilde{\beta}}{1 - \tilde{\eta}} + \frac{(\tilde{\mu} - \tilde{r})^2 (2 - \tilde{\eta})}{2(1-\tilde{\eta})^2 \tilde{\sigma}^2}.$$

The precise form of $P_\infty^{\tilde{x}}(\tilde{x}; \tilde{\psi})$ and the state-price deflator in the production economy follow immediately using arguments similar to the ones in the previous section. In particular, the nonnegativity condition (analogous to (8)) is

$$\tilde{\vartheta} \equiv \frac{\tilde{\delta}}{\tilde{\kappa}}(1 - \exp(-\tilde{\kappa}\tau)) < 1 \tag{18}$$

where

$$\tilde{\kappa} \equiv (1 - \tilde{\eta})\left[\tilde{\omega} + \frac{1}{2}\tilde{\sigma}^2(\tilde{\eta} - 2)\right] + (\tilde{\alpha} + \tilde{\beta}).$$

It can be demonstrated—by straightforward calculation—that the parameter values used to resolve the equity premium puzzle in Constantinides (1990) do satisfy the nonnegativity condition.[14]

5. CONCLUSIONS

This note has used a simple example to demonstrate a general problem that can arise in the use of an endowment economy to examine an equilibrium asset pricing model with habit formation. A "reasonable" endowment process *cannot* be specified by relying only on a few moments of measured consumption growth. It is possible to choose an exogenous endowment that matches the mean and variance of measured consumption growth, reasonable moments for asset returns, but implies that the single good in the economy is not demanded by the representative agent at some dates (with probability one).

It should be stressed that this is not an artifact of the simplicity of the endowment process used in this example, although it is this simplicity that permits a (comparatively) easy evaluation of the nonnegativity condition. One solution to this problem is to use a

[14] The values for $\tilde{\vartheta}$ range from 0.3689 to 0.6775.

production economy with sufficient structure on the asset return distribution to enable an endogenous derivation of the aggregate consumption process. The cost of this solution is clearly that it requires a potentially strong (and possibly counterfactual) assumption about asset returns.

Finance Department and Economics Department, Graduate School of Business, The University of Texas at Austin, Austin, TX 78712-1179, U.S.A.; chapman@eco.utexas.edu; http: // www.bus.utexas.edu/ ~ chapmand

Manuscript received July, 1994; final revision received September, 1997.

REFERENCES

CAMPBELL, J. Y., AND J. H. COCHRANE (1997): "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," NBER Working Paper No. 4995 (Revised).

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

DETEMPLE, J. B., AND F. ZAPATERO (1991): "Asset Prices in an Exchange Economy with Habit Formation," *Econometrica*, 59, 1633–1657.

DUFFIE, D. (1996): *Dynamic Asset Pricing Theory*, 2nd Edition. Princeton: Princeton University Press.

MEHRA, R., AND E. C. PRESCOTT (1985): "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15, 145–161.

SCHWERT, G. WILLIAM (1989): "Why Does Stock Market Volatility Change Over Time," *Journal of Finance*, 44, 1115–1153.

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