ASSET PRICING IN A NEOCLASSICAL MODEL WITH LIMITED PARTICIPATION

By QIANG DAI[1]

September 19, 2001

[1]Visiting assistant professor of finance, Graduate School of Business, Stanford University, and assistant professor of finance, Stern School of Business, New York University. Email: dq@stanford.edu. Phone: (650) 723-1463. I would like to thank David Backus, Jacob Boudoukh, Stephen Brown, John Campbell, Michael Chernov, John Cochrane, Tom Cooley, Jerome Detemple, Ned Elton, Marti Gruber, Vincenzo Quadrini, Alex Shapiro, and Suresh Sundaresan for valuable discussions, and seminar participants at a Columbia/NYU joint seminar, an NBER asset pricing meeting, and a Stern Economics Macro seminar, for valuable comments. I would like to thank especially Ken Singleton, Darrell Duffie, Yakov Amihud, Amnon Levy, Jun Liu, and Anthony Lynch for research discussions and valuable inputs.

Abstract

In this paper, I show that habit formation is perhaps not what it is commonly perceived to be: an extension of preference specification for the representative agent. Rather, it captures a dynamic interaction between aggregate financial income and aggregate labor income. I also show that existing specifications of consumption habit can be extended to incorporate a stochastic shock, which is interpreted as the labor income shock. As a result of these two innovations, I show that a habit formation model can explain the equity premium, equity volatility, and riskfree rate puzzles simultaneously, and provide an equilibrium justification for the predictability of equity and bond returns by dividend/price ratio and term spreads – all in terms of observable sample moments of aggregate dividend income and labor income growth rates and reasonable values of the risk aversion coefficient and the subjective discount rate.

To substantiate these claims, I present an extension of the Breeden-Lucas CCAPM by incorporating a particular form of heterogeneity assumption and a particular form of limited participation assumption. The resulting model features a richer technological specification (from the perspective of a production economy) or a richer endowment specification (from the perspective of an exchange economy), but retains standard assumptions of constant relative risk aversion, complete markets, and frictionless trading from the perspective of the marginal investor.

Contents

1 Introduction

Consumption-based capital asset pricing models (CCAPM) developed by Lucas [1978] and Breeden [1979] seek to explain asset returns in terms of aggregate consumption behavior. Such models have not withstood vigorous empirical tests (see, e.g., Hansen and Singleton [1982]). Empirical challenges to the standard CCAPM formulation often take the form of an asset pricing puzzle or anomaly. These include the *equity premium puzzle* of Mehra and Prescott [1985], the *risk-free rate puzzle* of Weil [1989], the *equity market volatility puzzle* (see, e.g., Hansen and Jagannathan [1991] and Campbell [1999]), as well as the ability to predict expected returns by a number of economic variables, such as term spread and dividend/price ratio (see, e.g., Fama [1984] and Fama and Bliss [1987], and Fama and French [1989]).

In this paper, I propose an extension of CCAPM based on two basic ideas. First, at the aggregate level, the economy consists of two fictitious agents: an investor and a worker. The investor receives all of the financial income and no labor income, and the worker receives all of the labor income income and no financial income. Since the nature of financial income and labor income are distinctly different, this assumption provides a particular way of capturing *income heterogeneity*. Second, labor income received by the worker in exchange for participating in the risky production is implicitly determined by an ex ante risk-sharing arrangement between the investor and the worker in such a manner that there is no further gain for the worker from self-insurance. Since the net demand by the worker for financial assets (for the purpose of self-insurance) is zero, the worker behaves as if he does not have access to the financial markets: he consumes all of his labor income, no more and no less. I refer to this behavior as a *limited participation* assumption.

The heterogeneity assumption is made operational by constructing a production-based, continuous-time dynamic asset pricing model based on the optimizing behavior of the investor, taking as given a dynamic interaction between the investor's consumption choice (i.e., dividend policy) and the worker's wage demand. The limited participation assumption is made operational by assuming that the investor invests all of her wealth in the risky production process. In another word, the net borrowing or lending by the investor is zero. This provides the market clearing condition required for a general equilibrium characterization of the asset markets. Analytical tractability is accomplished by adopting the standard power utility specification for the investor, complemented by a "linear-quadratic" specification of the labor income process that helps preserve a homothetic structure. The behavior of asset returns are completely determined by the joint processes of aggregate dividend income and aggregate labor income through optimality and equilibrium restrictions. Empirical implications of the model are examined in Section 4.

The model resolves the equity premium, equity volatility, and the riskfree rate puzzle simultaneously based on the following intuitions. First, since equities are effectively priced by the aggregate dividend growth rate rather than the aggregate consumption growth rate, the risk aversion parameter required to explain the observed equity premium is low. Second, since the volatility of the labor income growth rate is an order of magnitude lower than that of the dividend growth rate, and the labor's share of output is high, the volatility of the dividend growth rate is an order of magnitude higher than that of the aggregate consumption growth rate. On average, the ratio of the two is given approximately by the *operating leverage*, namely, the ratio of the aggregate consumption and the aggregate dividend income. Finally, since the risk aversion parameter required to explain the equity premium puzzle is sufficiently low, the elasticity of intertemporal substitution is sufficiently high. Consequently, the riskfree rate is sufficiently low under a reasonable value of the subjective discount rate.

The model also contains a non-trivial term structure of real interest rates, with the riskless rate being a monotonically increasing function of the (per capita) aggregate labor income, and the market price of risk being a monotonically decreasing function of the (per capita) aggregate labor income. Three important empirical implications are highlighted in the paper. First, since the riskless rate is a deterministic function of the labor income, its volatility is low. Second, since the riskless rate and the market price of risk are negatively correlated, expected bond returns and real term spreads are positively correlated, possibly paving the way for a resolution of the expectations puzzle. Finally, since the expected equity returns and dividend/price ratio are also functions of the (per capita) aggregate labor income, equity returns are predictable not only by the dividend price ratio, but also by the real term spreads.

The model is motivated (in Section 2) and solved (in Section 3) as a production economy. It is also possible to reformulate it as an endowment economy – with an interesting twist. Section 5 examines some theoretical issues and provides some useful comments on the existing asset pricing literature from this alternative perspective.

In the course of explaining asset pricing puzzle, the paper gives an economic interpretation of consumption habit. Specifically, after a change of variables, the model can be interpreted as a model of habit formation, with the aggregate labor income being interpreted as an (additive) consumption habit. The similarities and differences between this model and existing (additive) habit formation models,[1] including Sundaresan [1989], Constantinides [1990], Detemple and Zapatero [1991], and Campbell and Cochrane [1999]), will be discussed in depth in Sections 4 and 5 in order to highlight the marginal contributions of this paper.

2 A Neoclassical Model with Limited Participation

In this section, I will outline the economic structure captured by this paper and discuss in depth various key assumptions, and their economic motivations. This discussion sets the stage for a parametric formulation of a general equilibrium asset pricing model and the analytical characterization of its solution in Section 3.

Consider a closed economy with a risky production technology and a riskless storage technology. The risky technology takes two inputs: capital and labor. Accordingly, there are two infinitely lived agents: an investor and a worker. The investor supplies financial capital $K(t)$ and controls the production process, embodied in a representative production firm. The worker supplies labor in exchange for a labor income $X(t)$. Obviously, $X(t)$ is the labor cost to the production firm. Throughout the paper, the terms "labor income" and "labor cost" will be used interchangeably.

The state of the economy is described by the pair $(K(t), X(t))$, which is assumed to be

[1]See Footnote 5 for a comment on models with multiplicative consumption habit.

jointly Markovian, conditional on a stochastic process $Z(t)$:

$$dK = [\mu_e(K,X)K - Z]\,dt + \sigma_e(K,X)K\,dB_K, \tag{1}$$
$$dX = \nu(X,Z)\,dt + \psi(X,Z)\,dB_X, \tag{2}$$

where $\mu_e(K,X)$ is the conditional expected return of the risky technology, $\sigma_e(K,X)$ is the conditional volatility of the risky return, and $Z(t)$ is the "dividend" drawn from the financial stock $K(t)$ to finance current consumption by the investor.

Equation (1) is a standard budget constraint, and equation (2) describes the evolution of the labor cost (implicitly borne by the investor) or equivalently the labor income (accruing to the worker). From the perspective of the investor, equations (1)–(2) jointly determine the investment environment. We refer to these equations collectively as the *budget set*. The financial press is replete with reports and commentaries on how cost of labor affects a firm's financial performance and equity return. This budget set is one way of capturing a direct link between labor cost and asset returns.

In principle, there are two individual optimization problems: both the investor and the worker are rational agents maximizing expected utility. However, I will abstract away from the worker's problem by appealing to a *limited participation* assumption. Specifically, I will assume that, as a result of an *ex ante* risk-sharing agreement between the worker/employee and the investor/employer, the labor income that the worker receives is such that there is no further gain from self-insurance through asset markets. In another word, the worker consumes exactly X, no more and no less. A direct implication of this assumption is that the net supply of the riskless asset and other contingent claims such as long-term bonds from the worker are identically zero. Consequently, the net demands for these assets by the investor must be zero.[2] These are precisely the market clearing conditions we need for a general equilibrium characterization of the asset markets.

To facilitate further theoretical development, it proves useful to introduce an alternative formulation of the budget set. To this end, let us suppose that the investor's problem has already been solved under the assumption of complete markets and frictionless trading, and denote the unique state price density or pricing kernel by $M(t)$. Using $M(t)$, we can price any risky security. In particular, a contingent claim paying a "dividend" stream $X(t)$ has an "ex-dividend" price $N(t)$, given by

$$N(t) = M(t)^{-1} E_t \left[\int_t^\infty M(s)X(s)ds \right] = \mathbb{N}(K(t), X(t)),$$

where the second equality obtains because of the Markovian structure of the budget set. For lack of better terms, we will refer to $N(t)$ as the *value of human capital*. Define $W \equiv K + N = K + \mathbb{N}(K,X) = \mathbb{W}(K,X)$ and suppose that, given X, the function $\mathbb{W}(\cdot, X)$ is

[2]In another word, the investor has no financial leverage in equilibrium, in stark contrast to the limited participation model of Basak and Cuoco [1998]. Basak and Cuoco [1998] assume that a subset of individual investors do not participate in the stock market, but lend money to those who do, and the resulting financial leverage assumed by the stockholders explain the equity premium puzzle. In contrast, financial leverage as a mechanism for resolving the equity premium puzzle is explicitly excluded in my model.

invertible, with the inverse given by $\mathbb{K}(\cdot, X)$. Then the state vector (K, X) can be replaced by the state vector (W, X) – by substituting out K everywhere by $\mathbb{K}(W, X)$.[3]

Throughout the paper, I will also assume that the population grows at a constant rate of G: $H(t) = H(0)e^{Gt}$, where $H(t)$ is the size of the population at time t. Furthermore, I assume that all per capita level variables are stationary. In particular, if we define $w(t) \equiv \frac{W(t)}{H(t)}$ and $x(t) \equiv \frac{W(t)}{H(t)}$, then the state vector $(w(t), x(t))$ is stationary. All subsequent theoretical development is based on the per capita variables (denoted by lower case letters). In particular, the per capita financial income is denoted by $z(t) \equiv \frac{Z(t)}{H(t)}$, and the per capita value of financial stock is denoted by $k(t) \equiv \frac{K(t)}{H(t)}$. Similarly, the per capita value of human capital is denoted by $n(t) \equiv \frac{N(t)}{H(t)}$. All rates of returns and yields in the per capita economy are obtained from those in the original economy by a reduction of G and all rates of mean reversion are obtained from the original economy by an increment of G. All contemporaneous differences in rates of return and yields and all contemporaneous ratios such as proportional conditional volatilities in the per capita economy are the same as those in the original economy.

In Appendix A, I show that, in terms of the state vector (w, x), the investor's problem can be stated as follows:

$$\max_{\alpha(t), z(t): t \geq 0} E_0 \left[\int_0^\infty u(z(t), t)\, dt \right], \tag{3}$$

subject to

$$\begin{aligned} dw &= [\alpha(\mu(w,x) - r(w,x))w + r(w,x)w - z - x]\, dt + \alpha\sigma(w,x)w\, dB_w, & (4)\\ dx &= \nu(x,z)\, dt + \psi(x,z)\, dB_x, & (5) \end{aligned}$$

where B_w and B_x are standard Brownian motions with constant correlation δ. Furthermore, the market clearing condition is

$$\alpha(t) = 1, \tag{6}$$

which is equivalent to the condition that the net supply of riskless asset is zero. Equations (3)–(6) constitute a generic description of the general equilibrium asset pricing model studied in this paper. This model differs from standard permanent income models in several important respects.

First, the labor income process in permanent income models is typically assumed to be autonomous Markovian processes. This is not the case here. The explicit dependence of $\nu(x, z)$ and $\psi(x, z)$ on the choice variable z captures a strategic interaction between the investor and the worker. One of the maintained assumptions, $\nu_z(x, z) > 0$, captures the idea that the worker may be willing to offer a lower wage demand in "bad" states of the economy (low values of z) in exchange for the investor's willingness to accept a higher wage demand in "good" states of the economy (high values of z). This is consistent, in spirit, with Klein

[3] Obviously, this procedure can be done in reverse: start with primitive assumptions about the dynamics for (W, X) and solve the model. Let the pricing kernel be $M(t)$ and optimal control be $Z^* = Z(W, X)$. The value of financial capital is given by $K(t) = M(t)^{-1} E_t \left[\int_t^\infty M(s) Z^*(s)\, ds \right] = \mathbb{K}(W, X)$. If $\mathbb{K}(\cdot, X)$ is invertible with inverse $\mathbb{W}(\cdot, X)$, we can restate the problem in terms of the new state vector (K, X) by substituting out $W = \mathbb{W}(K, X)$ everywhere.

[1950]'s wage demand function: that the wage demand is a weighted sum, or distributed lag of past output implies that the expected growth rate of wage income is increasing with current output.[4]

Second, the utility function in permanent income models is defined in terms of aggregate consumption $c(t) = z(t) + x(t)$. This is not the case here. Since the model describes the optimizing behavior of the investor, rather than the representative agent, it is natural to assume that the utility function is defined in terms of the (per capita) financial income $z(t)$. There is no reason to believe that the worker's consumption $x(t)$ should enter the investor's utility function, unless these two agents are one and the same.

Even though the utility function is defined in terms of $z(t)$, nothing prevents us from using $c(t)$ as a dummy variable in lieu of $z(t)$. Through a simple change of variable: $z = c - x$, the model can be rewritten as

$$\max_{\alpha(t),c(t):t\geq 0} E_0 \left[\int_0^\infty u(c(t) - x(t), t)\, dt \right], \tag{7}$$

subject to

$$\begin{aligned} dw &= [\alpha(\mu(w,x) - r(w,x))w + r(w,x)w - c]\; dt + \alpha\sigma(w,x)w\, dB_w, &(8)\\ dx &= \nu(x, c-x)\, dt + \psi(x, c-x)\, dB_x. &(9)\end{aligned}$$

This formulation suggests a different interpretation of the model: if we mistakingly think of the marginal agent as the representative agent consuming $c(t)$, then the agent acts as if she develops a consumption habit $x(t)$. In another word, the concept of consumption habit may simply be an outdated label for the aggregate labor income![5] To hit the point home, suppose that the following simple parameterization of the labor income process is adopted: $\nu(x, c-x) = bc - ax$, and $\psi(x, c-x) = 0$, and that the investor has constant relative risk aversion. Then the model is isomorphic to the internal habit formation model of Constantinides [1990].

Key maintained assumptions are summarized as follows.

Assumption 1 (Wealth Independence of Riskless Rate) *The riskless rate $r(w,x)$ is independent of w, i.e., $r(w,x) = r(x)$.*

This assumption has two motivations. First, a volatile riskless rate is a common malaise of exchange-based asset pricing models. To get rid of this undesirable feature, Campbell and Cochrane [1999] impose a parametric restriction so that the riskless rate is strictly constant, which renders the term structure of interest rates trivial. This paper offers a model in which such an extreme assumption is avoided and a realistic term structure of interest rates can

[4]Klein's wage demand function also implies that the expected growth rate of wage income is decreasing in the current level of wage income, which is consistent with another maintained assumption in my model that $\nu_x(x,z) < 0$.

[5] The multiplicative habit model, proposed by Abel [1990], can also be interpreted in the same spirit. Let $s \equiv \frac{z}{c} = 1 - \frac{x}{c}$ be the investor's share of output, and $y \equiv \frac{1}{s} = \frac{c}{c-x}$, then the investor's utility function can be written as $u(z,t) = u(c/y, t)$. Thus, there is a formal equivalence between a model with additive habit x and a multiplicative habit y.

also be accommodated. Second, since labor income growth is known to be smooth over time, the riskless rate also has a low volatility if it depends only the labor income.

To characterize the nature of the state-dependence of the riskless rate, I will take the view that the riskless rate also represents the worker's productivity when he does not participate in the risky production. Thus, it is intimately related to the income or consumption process when the worker pursues the outside opportunity, which is characterized below.[6]

Assumption 2 (Worker's Outside Opportunity) *When the worker does not participate in the risky production (which means $k = z = 0$), the labor income process is deterministic and declining over time. That is, for $\forall x \geq 0$,*

$$\nu(x,0) = l_0(x) \leq 0, \quad \psi(x,0) = 0.$$

Furthermore, $l_0(0) = 0$ and $l_0'(x) < 0$ for $x > 0$.

Henceforth, the value of human capital associated with the outside opportunity, $f(x) \equiv n(0,x)$, is referred to as the *reservation value of human capital.* When $k = 0$, $\mu(w,x) = r(x)$, and equation (4) implies that $df(x) = [r(x)f(x) - x]\,dt$. This leads to the following lemma:

Lemma 1 *If Assumptions 1–2 hold and the budget constraint (4) holds at $k = 0$ or $w = f(x)$, then for $\forall x \geq 0$,*

$$l_0(x) = \frac{r(x)f(x) - x}{f'(x)}, \tag{10}$$

or equivalently,

$$r(x) = \frac{x + l_0(x)f'(x)}{f(x)}. \tag{11}$$

Given the reservation value $f(x)$, equation (10) imposes a restriction on the labor income growth rate given the riskless rate, and equation (11) imposes a restriction on the riskless rate given the labor income process. If both $l_0(x)$ and $r(x)$ are taken as given, then either of the two equations effectively imposes a restriction on the reservation value of human capital $f(x)$. For an example, if $l_0(x) = -\kappa x$, $\kappa > 0$, and $r(x) = \bar{r}$, where $\bar{r}$ is a constant, then either equation (10) or equation (11) implies that $f(x) = \eta x$, where $\eta = \frac{1}{\bar{r}+\kappa}$.

In Appendix B, I show how the reservation value $f(x)$ emerges from a worker's optimization problem when he pursues the outside opportunity. Without prior restrictions on the worker's preference function, $f(x)$ can take on arbitrary functional forms. Rather than specifying the worker's preference function as a primitive assumption, it is more convenient to impose directly the following restrictions on $f(x)$.

Assumption 3 (Reservation Value of Human Capital) *For $\forall x \geq 0$,*

$$f(0) = 0, \quad f'(x) > 0, \quad f''(x) \leq 0, \quad A'(x) \geq 0,$$

where $A(x) \equiv -\frac{xf''(x)}{f'(x)} \leq 0$.

[6] See Appendix B for an elaboration on the worker's problem when he pursues the outside opportunity. This problem is distinct from the worker's "self-insurance" problem when the worker participates in the risky production and receives a risky labor income. The latter problem has been rendered trivial by the limited participation assumption.

The following lemma gives a strong result pertaining to the state-dependent nature of the riskless rate.

Lemma 2 (State-dependence of $r(x)$) *Let $l_0(x) = -\kappa x$, where $\kappa > 0$. Suppose that Assumption 3 holds. Then $r'(x) > 0$ for $\forall x > 0$.*

The proof is given in Appendix B.1. In order for $r(x)$ to be monotonically increasing in x, it suffices that $f''(x) < 0$ and $A'(x) = 0$ (e.g., $f(x) = \eta \frac{1-e^{-\epsilon x}}{\epsilon}$, where $\epsilon > 0$),

Turning to the behavior of the labor income process when the worker participates in the risky production, I start with the following assumption.

Assumption 4 (Labor Income with Risky Production) *For $\forall x \geq 0$ and $\forall z \geq 0$,*

$$\nu_z(x,z) \geq 0, \quad \psi_z(x,z) \geq 0.$$

The monotonicity of $\nu(x,z)$ in z captures two intuitions. First, in order to induce the worker to participate in the risky production, the investor may have to offer or the worker may be able to demand an *incentive for participation*, in the form of a higher rate of labor income growth when z is higher. This can be captured by assuming that $\nu(x,z)$ contains a piece proportional to z. Second, since there is a residual labor income risk, $\psi(x,z) \neq 0$, the worker needs to be compensated for bearing the risk. For reasons elaborated below, this means that $\nu(x,z)$ should also contain a piece of the form: $-\frac{1}{2}\frac{f''(x)}{f'(x)}\psi(x,z)^2$. The following assumption puts more structure on the labor income process.

Assumption 5 (Expected Rate of Labor Income Growth) *For $\forall x \geq 0$ and $\forall z \geq 0$,*

$$\nu(x,z) = l_0(x) + l_1(x)z - \frac{1}{2}\frac{f''(x)}{f'(x)}\psi(x,z)^2, \tag{12}$$

where $l_1(x) \geq 0$.

The linkage between the conditional mean $\nu(x,z)$ and the conditional volatility $\psi(x,z)$ of the labor income process is induced entirely by the curvature of $f(x)$. If the investor is not required to compensate the worker for bearing the labor income risk, she may be tempted to shift risk to the worker (by judicious choice of z). An appropriate amount of risk compensation removes this temptation.

To illustrate ideas, suppose that the investor's indirect utility function can be written as $V[w - f(x)]$, and the worker's (reservation) indirect utility function can be written as $U(f(x))$ (ignoring the time dependence for simplicity), where both $V(\cdot)$ and $U(\cdot)$ are at least increasing.[7] Ito's lemma implies

$$\begin{aligned} dV &= \left[\ldots - \nu(x,z)V'f'(x) + \frac{1}{2}\psi(x,z)^2\left[V''f'(x)^2 - V'f''(x)\right]\right]dt + \ldots, \\ dU &= \left[\nu(x,z)U'f'(x) + \frac{1}{2}\psi(x,z)^2\left[U''f'(x)^2 + U'f''(x)\right]\right]dt. \end{aligned}$$

[7] It will be shown later that the investor's indirect utility function takes exactly the form $V(w - f(x), t)$. The "splitting" of the total wealth w into $w - f(x)$ and $f(x)$ mirrors the "splitting" of the total income or consumption c into $z = c - x$ and x. It is important to note however, when the stock value of z is positive, i.e., $k > 0$, the value of human capital $n(k,x)$ is in general different from $f(x) = n(0,x)$, the reservation value. An obvious conjecture is that $n(k,x) \geq f(x)$ for $\forall k > 0$, which can be explicitly verified.

A concave $f(x)$ implies that labor income volatility can be beneficial to the investor, as the term $-\frac{1}{2}\psi(x,z)^2 V' f''(x)$ is positive and can dominate the utility cost $\frac{1}{2}\psi(x,z)^2 V'' f'(x)^2$ in some states of the world. In contrast, a concave $f(x)$ implies that labor income volatility unambiguously leads to a utility cost for the worker, as the term $\frac{1}{2}\psi(x,z)^2 V' f''(x)$ is negative and increases with $\psi(x,z)$. The asymmetric effect of the concavity of $f(x)$ is neutralized under Assumption 5, because in both cases the third term in $\nu(x,z)$ cancels exactly the term associated with $f''(x)$ (note that the third term in $\nu(x,z)$ reduces the expected growth rate of V but increases the expected growth rate of U, because V decreases with $f(x)$ and U increases with $f(x)$).

Starting from Section 3, I address asset pricing issues using a parametric specification based on a *linear-quadratic structure*, which obtains under the following assumption.

Assumption 6 (Volatility of Labor Income Shock) *For $\forall x \geq 0$ and $\forall z \geq 0$,*

$$\psi(x,z) = l_3(x)z,$$

where $l_3(0) = 0$.

Assumptions 5 and 6 imply that $\nu(x,z) = l_0(x) + l_1(x)z + l_2(x)z^2$, where

$$l_2(x) = -\frac{1}{2}\frac{f''(x)}{f'(x)} l_3(x)^2. \tag{13}$$

Thus, $\psi(x,z)$ is linear in z and $\nu(x,z)$ is quadratic in z.

This linear-quadratic structure can be relaxed at the expense of losing analytical tractability. A potentially interesting example outside the linear-quadratic structure is obtained by assuming that $\psi(x,z) = l_3(x)\sqrt{z}$. According to Assumption (5), the third term in $\nu(x,z)$ is also linear in z and can thus be absorbed into the second term in $\nu(x,z)$. Consequently, the labor income process is "conditionally affine":

$$dx = [l_0(x) + l_1(x)z]\,dt + l_3(x)\sqrt{z}dB_x. \tag{14}$$

Under mild regularity conditions, this process is well-defined, as long as z is non-negative. In a production-based setting, z is endogenously determined. Unless the model can be solved analytically, it is difficult to characterize and impose conditions under which the optimal control z is non-negative. This difficulty is easily circumvented in an exchange setting, where the positivity of the endowment process z can be exogenously imposed. See Section 5 for a discussion of the endowment economy approach.

The final maintained assumption is the utility specification of the marginal agent in this economy, namely, the investor. For simplicity and also for a direct comparison to popular models in the existing literature, I shall adopt the CRRA specification, which is utterly conventional.

Assumption 7 (Investor's Utility Function) *For $\forall z \geq 0$,*

$$u(z,t) = e^{-\rho t}\frac{z^\gamma}{\gamma},$$

where $\rho > 0$ and $\gamma < 0$.

In Section 3, I will show that, under a parametric specification consistent with Assumptions 1–7, the model admits an analytical solution. Furthermore, in Section 5, I will show how to formulate the model as an endowment economy. I choose to present the model as a production economy, because the primitive assumptions are easier to motivate (as I have tried to accomplish so far in this section) and asset pricing implications are easier to characterize. I conclude this section by elaborating the last point.

To facilitate the discussion, I first present two general results in the form of two lemmas. Assumptions 1–7 are sufficient, but not necessary, for these lemmas to hold.

Lemma 3 (Marginal Utility of Wealth) *Suppose that the model defined by equations (3)–(6) admits a unique equilibrium with value function $V(w, x, t)$ and optimal policies $\alpha^* = \alpha(w, x)$ and $z^* = z(w, x)$.*

Let $\mathcal{G}(w, x) \equiv w\sigma(w, x)'\alpha(w, x)$. Then, the expected rate of growth of the marginal utility of wealth, V_w, is equal to $-r(w, x)$, if and only if

$$wr_w + \mathcal{G}(w, x)\Lambda_w = 0. \tag{15}$$

The proof is given in Appendix C. An immediate implication is that, if both the riskless rate $r(w, x)$ and the Sharpe ratio $\Lambda(w, x)$ are independent of the per capita total wealth w, the marginal utility of wealth, V_w, can potentially be interpreted as the pricing kernel in this economy (whose expected rate of growth is $-r(w, x)$). This is confirmed by the next lemma.

Lemma 4 (Pricing Kernel) *Consider the model defined by equations (3)–(6). Suppose that*

1. *In addition to claims to the production technology and the riskless asset, the investor can trade a contingent claim with dividend $D(t)$ and ex-dividend price $P(t)$, with the total return given by*

$$\frac{dP(t) + D(t)dt}{P(t)} = \mu_P(t)dt + \sigma_P(t)dB_p(t),$$

   where $corr(B_p(t), B_w(t)) = \delta_{pw}$ and $corr(B_p(t), B_x(t)) = \delta_{px}$.

2. *The contingent claim is not redundant: $\delta_{px} \neq \delta\delta_{pw}$.*

3. *The net supply of the contingent claim is zero.*

4. *$r_w = \Lambda_w = 0$, i.e., $r(w, x) = r(x)$ and $\Lambda(w, x) = \Lambda(x)$.*

5. *There exists a unique equilibrium with value function $V(w, x, t)$.*

Then $m(t) \equiv V_w(w, x, t)$ is the unique equilibrium pricing kernel, in the sense that[8]

$$\mu_P(t) - r(t) = -cov\left(\frac{dP(t) + D(t)dt}{P(t)}, \frac{dm(t)}{m(t)}\right). \tag{16}$$

for any security with dividend stream $D(t)$ and ex-dividend price $P(t)$ (including claims to the production firm).

[8] Throughout the paper, the instantaneous covariance cov$(\cdot)$ is always normalized by dt.

The proof is given in Appendix D. By design, the additional contingent claim completes the market, thereby pinning down the unique equilibrium pricing kernel.

If conditions are met so that these two lemmas hold, then the pricing kernel is given by $m(t) = V_w(w, x, t)$, i.e., the marginal utility of wealth, rather than $u_z(z, t)$, the utility gradient (see Duffie and Skiadas [1994]). In most existing asset pricing models, including the venerable Merton's model and the CIR model, the equality $V_w(w, x, t) = u_z(z, t)$ holds and is commonly referred to as the *envelope condition*. In my model, this envelope condition breaks down (because of the explicit dependence of the labor income process on the control variable z) (except for some special cases). Lemma 4 shows that, in my model, pricing is straightforward in the production setting, once the value function is obtained (either analytically as in this paper or numerically in some other cases). In contrast, pricing becomes a formidable challenge in the exchange setting.[9]

3 Stochastic Growth and Asset Pricing

In this section, I will present the main theoretical result of the paper: the explicit characterization of a unique general equilibrium of asset markets, under a particular parameterization of the model defined by equations (3)–(6).

To begin, let us first consider the following simple parameterization of the model:

$$\max_{\alpha(t), z(t): t \geq 0} E_0 \left[\int_0^\infty e^{-\rho t} \frac{z(t)^\gamma}{\gamma} dt \right], \tag{17}$$

subject to the constraints

$$\begin{aligned} dw &= [\alpha [\mu(w, x) - r(x)] w + r(x) w - z - x] dt + \alpha \sigma(w, x) w dB_w, && (18) \\ dx &= \left(bz - \kappa x + \frac{1}{2} \epsilon \beta^2 z^2 \right) dt + \beta z dB_x, && (19) \end{aligned}$$

where $\rho > 0$, $\gamma < 0$, $b > 0$, $\kappa > 0$, $\beta > 0$, $\delta = \mathrm{corr}(dB_w, dB_x) < 0$, and $\epsilon > 0$. It is easy to verify that, if

$$r(x) = \frac{1 - \kappa f'(x)}{f(x)/x}, \quad f(x) = \eta \frac{1 - e^{-\epsilon x}}{\epsilon}, \tag{20}$$

where $\eta = \frac{1}{r_0 + \kappa} > 0$ and $r_0 = r(0)$, the parametric specification satisfies Assumptions 1–7. The model is completely specified by eight parameters: $(\rho, \gamma, b, \kappa, \beta, \delta, \epsilon, r_0)$.

Following the proof for a more general case to be given shortly, we can show that, if $w_0 > f(x_0)$, then the inter-temporal optimality conditions (i.e., the Hamilton-Jacobian-Bellman equation) and the market clearing condition $\alpha = 1$ are satisfied for the above model, with the indirect utility function given by $V(w, x, t) = e^{-\rho t} \phi \frac{(w - f(x))^\gamma}{\gamma}$, where $\phi > 0$.

[9] The consumption habit in Campbell and Cochrane [1999] is assumed to be "external", rather than "internal", precisely because in their exchange setting, pricing is much more involved with an "internal" habit. In my paper, the labor income process must be interpreted as an *internal* habit. Introducing stochastic shocks to the internal habit further escalates the computational difficulty in the exchange setting.

Furthermore, the optimal consumption policy is given by $z^* = h(x)(w - f(x))$, and the equilibrium return is given by

$$\mu(w,x) = r(x) + \Lambda(x)\ \sigma(w,x), \tag{21}$$

$$\sigma(w,x) = \frac{w - f(x)}{w} g(x). \tag{22}$$

where the real functions $h(x)$, $g(x)$, and $\Lambda(x)$ are endogenously determined. Note that $\Lambda(x)$ is the Sharpe ratio, which is independent of w. Thus, Lemma 4 applies, and the pricing kernel in this model is given by

$$m(t) = V_w(w,x,t) = e^{-\rho t}\phi(w - f(x))^{\gamma-1} = e^{-\rho t}\phi\left(\frac{z}{h(x)}\right)^{\gamma-1}.$$

Since both z and x are observable, at least in principle, this model is readily testable using the Euler equation approach. Take a time grid $t = 1, 2, \ldots, T$, and consider a security that pays dividend D_t with ex-dividend price P_t. The one-period return of the security, $\tilde{R}_{t+1} \equiv \frac{P_{t+1}+D_{t+1}}{P_t}$, must satisfy the following Euler equation $E_t\left[\frac{m(t+1)}{m(t)}\tilde{R}_{t+1}\right] = 1$, or more explicitly,

$$E_t\left[\left(\frac{z_{t+1}}{z_t}\frac{h(x_t)}{h(x_{t+1})}\right)^{\gamma-1}\tilde{R}_{t+1}\right] = 1. \tag{23}$$

Since the expression inside the conditional expectation can be computed directly from observable macro and financial variables, the equation can be used to construct sample moment conditions (when applied to different assets) as testable over-identifying restrictions, along the line of Hansen and Singleton [1982].

If x were an autonomous process, the pricing kernel would be equal to the utility gradient $u_z = e^{-\rho t}z^{\gamma-1}$, and the Euler equation would be given by[10]

$$E_t\left[\left(\frac{z_{t+1}}{z_t}\right)^{\gamma-1}\tilde{R}_{t+1}\right] = 1. \tag{24}$$

In this case, the extra state variable x, whatever its economic origin, would enter the Euler equation only through the information set. In contrast, in my model, per capita labor income enters the Euler equation directly through the function $h(x)$, as well as indirectly as a conditioning variable. This is a key distinction between this model and many existing empirical implementations of asset pricing models.

Other asset pricing implications of the model can be readily derived in a straightforward manner. Before we do, however, a minor technical detail must be taken care of. It can be shown that, under the maintained sign restrictions on the model parameters, the endogenously determined functions $h(x)$, $g(x)$, and $\Lambda(x)$ exist only if x does not exceed a critical value $x_m > 0$, beyond which $h(x)$, $g(x)$, and $\Lambda(x)$ become complex-valued functions, which

[10] It is often debated whether z in the Euler equation should be aggregate dividend or aggregate consumption. In my model, the aggregate dividend is clearly the more sensible choice (ignoring other non-human income as most asset pricing models do).

does not make economic sense. The economic reason why this occurs is as follows. As a general equilibrium implication, the Sharpe ratio is monotonically decreasing with labor income. As labor income increases, the Sharpe ratio eventually becomes zero. Further increases in labor income renders risky production unprofitable compared to the riskless investment, which is always available to the investor. Thus, x_m represents the technological limit for the level of labor income. Any labor income beyond this limit can not be sustained (the investor will not invest for a negative expected equity premium).

The problem with the naive specification (17)–(20) is that, as long as $0 < x_m < \infty$, the volatility of the labor income growth does not vanish except when $z = 0$. Thus, as long as $z > 0$, there is a positively probability that the labor income process will diffuse above x_m, which renders the model invalid. Fortunately, this defect can be easily remedied by modifying the model slightly. The following model does the job.

Let $r(x)$ and $f(x)$ be given by equation (20). The investor solves

$$\max_{\alpha(t),z(t):t\geq 0} E_0\left[\int_0^\infty e^{-\rho t}\frac{z(t)^\gamma}{\gamma}dt\right], \tag{25}$$

subject to the constraints

$$dw = [\alpha[\mu(w,x) - r(x)]w + r(x)w - z - x]\,dt + \alpha\sigma(w,x)wdB_w, \tag{26}$$

$$dx = \left(b\hat{l}_1(x)z - \kappa x + \frac{1}{2}\epsilon\beta^2\hat{l}_3(x)^2z^2\right)dt + \beta\hat{l}_3(x)zdB_x, \quad x \in [0, x_m], \tag{27}$$

where $\rho > 0$, $\gamma < 0$, $b > 0$, $\kappa > 0$, $\beta > 0$, $\delta = \text{corr}(dB_w, dB_x) < 0$, $\epsilon > 0$, and $x_m > 0$. In equilibrium, $\alpha = 1$.

In this model, the labor income process x stays within the interval $[0, x_m]$. Its boundary behavior is regulated by two functions $\hat{l}_1(x)$ and $\hat{l}_3(x)$, which are assumed to be 1 in the interval $[0, x_m]$, except when x is sufficiently close to both boundaries. As $x \to x_m$, $\hat{l}_1(x)$ and $\hat{l}_3(x)$ approach zero at a sufficiently fast rate. This precipitous decline in $\hat{l}_1(x)$ and $\hat{l}_3(x)$ toward zero as $x \to x_m$ reduces the risk that the investor is exposed to by just the right amount so that the investor stays invested in the risky technology (when $x = x_m$, the investor is not exposed to any risk, and enjoys an expected return equal to the riskless rate).

Although this specification is more general than equations (17)–(19), parameters involving the boundary behavior of $\hat{l}_1(x)$ and $\hat{l}_3(x)$ are largely irrelevant, except for the location x_m of the upper boundary (we will see why x_m is an important exogenous parameter soon). In essence, this is a nine-parameter model.

In solving the model, I will denote $b\hat{l}_1(x)$ by $l_1(x)$ and $\beta\hat{l}_3(x)$ by $l_3(x)$ for notational simplicity. Most results are valid for a general specification of $l_1(x)$ and $l_3(x)$. Only when we discuss the boundary behavior will the functional forms of $\hat{l}_1(x)$ and $\hat{l}_3(x)$ come into play.

Proposition 1 (Partial Equilibrium) *Consider the model defined by equations (25)–(27) and equation (20). Suppose that the expected return and the conditional volatility are given by equations (21) and (22) respectively and $w_0 - f(x_0) > 0$. Then*

1. *The indirect utility function is given by*

$$V(w,x,t) = e^{-\rho t}\phi\frac{[w - f(x)]^\gamma}{\gamma}, \ \phi > 0. \tag{28}$$

2. *The optimal portfolio allocation to the risky production technology is given by*

$$\alpha^* = \frac{\Lambda + (1-\gamma)\delta l_3(x) h(x) f'(x)}{(1-\gamma)\sigma(w,x)} \frac{w - f(x)}{w}. \tag{29}$$

3. *The optimal consumption policy (for the investor) is given by*[11]

$$z^* = z(w,x) = h(x)[w - f(x)], \tag{30}$$

where

$$\frac{h(x)^{\gamma-1}}{\phi} = 1 + l_1(x) f'(x) + (1-\gamma) l_3(x)^2 h(x) f'(x)^2 - (1-\gamma)\delta l_3(x) g(x) f'(x). \tag{31}$$

4. *The Sharpe ratio of the risky production technology is given by*

$$\Lambda(x) = (1-\gamma)\left[g(x) - \delta l_3(x) h(x) f'(x)\right] - \frac{\Phi(x)}{g(x)}, \tag{32}$$

where

$$\begin{aligned} \Phi(x) &= \frac{1-\gamma}{2} g(x)^2 - (1-\gamma)\delta l_3(x) h(x) f'(x) g(x) + \Psi(x), \\ \Psi(x) &= r(x) - \frac{\rho}{\gamma} + \frac{1-\gamma}{\gamma}\frac{h(x)^\gamma}{\phi} + \frac{1-\gamma}{2} l_3(x)^2 h(x)^2 f'(x)^2. \end{aligned}$$

Proposition 2 (General Equilibrium) *Consider the model defined by equations (26)–(27) and equation (20). Suppose that the net supply of the riskless asset is zero and $w_0 - f(x_0) > 0$. Then, the model admits an equilibrium $(\mu(w,x), \sigma(w,x); z^*)$, in the sense that, given $(\mu(w,x), \sigma(w,x))$, z^* satisfy the optimality conditions and markets clear $\alpha^* = 1$. Furthermore,*

1. *The equilibrium consumption policy is given by equation (30).*

2. *The conditional moments of the equilibrium return, $\mu(w,x)$ and $\sigma(w,x)$, are given by equations (21) and (22) respectively.*

3. *The equilibrium Sharpe ratio of the risky production $\Lambda(x)$ is given by*

$$\Lambda(x) = (1-\gamma)\left[g(x) - \delta l_3(x) h(x) f'(x)\right], \tag{33}$$

where, for $\forall x \in [0, x_m]$, $h(x)$ and $g(x)$ jointly solve equations (31) and (34):

$$g(x)^2 - 2\delta l_3(x) h(x) f'(x) g(x) + \frac{2\Psi(x)}{1-\gamma} = 0. \tag{34}$$

[11] For ease of exposition, I will use $z = z(w,x)$ to denote the optimal policy z^* so long as no confusion arises.

In order for x_m to be a proper boundary, a necessary condition is that $g(x_m) = 0$, which implies that $\Psi(x_m) = 0$. Since $l_1(x_m) = l_3(x_m) = 0$, we also have $h(x_m)^{\gamma-1} = \phi$ and $r(x_m) - \frac{\rho}{\gamma} + \frac{1-\gamma}{\gamma}\frac{h(x_m)^\gamma}{\phi} = 0$. It follows that

$$\phi = \left[\frac{\rho - \gamma r(x_m)}{1-\gamma}\right]^{\gamma-1}, \tag{35}$$

provided that $\rho - \gamma r(x_m) > 0$. Since the function $r(x)$, $r_0 = r(0)$, and x_m are exogenously given, $r(x_m)$ is also exogenously determined. Thus, the above equation determines the endogenous parameter ϕ.

Given ϕ and the exogenous parameters, the functions $h(x)$ and $g(x)$ can be solved for each x, in particular, when $x = \bar{x}$, the steady-state value of x, we can solve for $\bar{h} = h(\bar{x})$ and $\bar{g} = g(\bar{x})$, both are implicit functions of ϕ, and therefore x_m. Thus, the location of the upper boundary is an important determinant of the steady-state mean, and consequently the steady-state value of the riskless rate and equity premium. In empirical calibration exercises, we can use the steady-state value (i.e. unconditional means) of the riskless rate or equity premium to determine the state-steady value of the labor income $\bar{x}$, which in turn identifies the upper bound x_m. Most asset pricing implications can be derived once $\bar{x}$ is known. Thus, we rarely need to solve for x_m explicitly (one exception is when we try to simulate the model).

Although the no-bankruptcy restriction is not imposed in solving the model, Corollary 1 asserts that it is not binding:

Corollary 1 (Equilibrium Wealth Process) *Suppose that Proposition 2 holds. Then, in equilibrium,*

$$\begin{aligned}\frac{d[w - f(x)]}{w - f(x)} = &\left[\frac{r(x)-\rho}{1-\gamma} + \frac{2-\gamma}{2(1-\gamma)^2}\Lambda(x)^2 + \frac{2-\gamma}{2}(1-\delta^2)l_3(x)^2h(x)^2f'(x)^2\right]dt \\ &+ g(x)dB_w - l_3(x)h(x)f'(x)dB_x.\end{aligned} \tag{36}$$

If $w_0 - f(x_0)$ is non-negative, then $w - f(x)$ is non-negative. Since $f(x)$ is non-negative, w must also be non-negative. Equation (36) is a key analytical result, from which the equilibrium dividend and consumption processes and the pricing kernel can all be derived.

Since $r(x)$ and $\Lambda(x)$ are independent of w, Lemma 4 implies that the pricing kernel is given by $m(t) = V_w$. This leads to the following corollary.

Corollary 2 (Pricing Kernel) *The pricing kernel $m(t)$ is given by*

$$\frac{dm(t)}{m(t)} = -r(x)dt - \Lambda^w(x)dB_w(t) - \Lambda^x(x)dB_x(t), \tag{37}$$

where $\Lambda^w(x) = -\frac{wV_{ww}}{V_w}\alpha^*\sigma = (1-\gamma)g(x)$ and $\Lambda^x(x) = -\frac{V_{wx}}{V_w}l_3(x)z^* = -(1-\gamma)l_3(x)h(x)f'(x)$.

Alternatively, writing $B_x = \delta B_w + \sqrt{1-\delta^2}B_\perp$, where $B_\perp \perp B_w$, we have

$$\frac{dm(t)}{m(t)} = -r(x)dt - \Lambda(x)dB_w(t) - \Lambda_\perp(x)dB_\perp(t), \tag{38}$$

where $\Lambda(x)$ is given by equation (33), and

$$\Lambda_\perp(x) = -(1-\gamma)\sqrt{1-\delta^2}l_3(x)h(x)f'(x). \tag{39}$$

The proof is trivial. Equation (37) or (38) can also be obtained from equation(36), noting that $m(t) = e^{-\rho t}\phi(w - f(x))^{\gamma-1}$.

The pricing kernel can be used to price any security. In particular, (normalizing $m(0) = 1$)

$$
\begin{aligned}
k(w, x) &= E\left[\int_0^\infty z(w(s), x(s))m(s)ds\middle| w(0) = w, x(0) = x\right], && (40)\\
n(w, x) &= E\left[\int_0^\infty x(s)m(s)ds\middle| w(0) = w, x(0) = x\right], && (41)
\end{aligned}
$$

give the value of financial capital (stock value of $z(w, x)$) and the value of human capital (stock value of x).

Summing equations (40) and (41) yields an accounting identity:

$$
k(w, x) + n(w, x) = E\left[\int_0^\infty [z(s) + x(s)]\, m(s)ds\middle| w(0) = w, x(0) = x\right] = w, \qquad (42)
$$

which is what we expect, since w is the stock value of the aggregate consumption $c = z + x$. One application of this accounting identity is the determination of the initial wealth level, w_0. Applying equation (42) to the initial values, we have

$$
k(w_0, x_0) + n(w_0, x_0) = w_0. \qquad (43)
$$

Since $k_0 \equiv k(w_0, x_0)$ and x_0 are exogenously given, w_0 is *endogenously* determined by solving $k_0 + n(w_0, x_0) = w_0$.

Applying standard continuous-time asset pricing techniques (see, e.g., Duffie [1996]), we can show that k and n can be determined by the following partial differential equations (PDEs) (it suffices to solve one of them due to the accounting identity (42)):

$$
\begin{aligned}
0 = {} & [r(x)w - z - x]\frac{\partial k}{\partial w} + \frac{w^2\sigma(w, x)^2}{2}\frac{\partial^2 k}{\partial w^2} + \delta l_3(x)zw\sigma(w, x)\frac{\partial^2 k}{\partial x \partial w} \\
& + \left[l_0(x) + \tilde{l}_1(x)z + l_2(x)z^2\right]\frac{\partial k}{\partial x} + \frac{l_3(x)^2z^2}{2}\frac{\partial^2 k}{\partial x^2} - r(x)k(w, x) + z,
\end{aligned} \qquad (44)
$$

or

$$
\begin{aligned}
0 = {} & [r(x)w - z - x]\frac{\partial n}{\partial w} + \frac{w^2\sigma(w, x)^2}{2}\frac{\partial^2 n}{\partial w^2} + \delta l_3(x)zw\sigma(w, x)\frac{\partial^2 n}{\partial x \partial w} \\
& + \left[l_0(x) + \tilde{l}_1(x)z + l_2(x)z^2\right]\frac{\partial n}{\partial x} + \frac{l_3(x)^2z^2}{2}\frac{\partial^2 n}{\partial x^2} - r(x)n(w, x) + x,
\end{aligned} \qquad (45)
$$

where $\tilde{l}_1(x) = l_1(x) - [\delta\Lambda(x) + \sqrt{1-\delta^2}\Lambda_\perp(x)]l_3(x)$, and $z = h(x)[w - f(x)]$.

In preparation for subsequent development, I conclude this section by giving a formal definition of the "steady-state means" of some key variables.

Definition 1 (Steady-state Means) *1. The steady-state means of z and x, denoted by $\bar{z}$ and $\bar{x}$ respectively, are the solution to the following joint equations:*

$$
\begin{aligned}
bz - \kappa x + \frac{1}{2}\epsilon\beta^2z^2 &= 0, && (46)\\
\frac{r(x) - \rho}{1-\gamma} + \frac{2-\gamma}{2(1-\gamma)^2}\Lambda(x)^2 + \frac{2-\gamma}{2}(1-\delta^2)\beta^2h(x)^2f'(x)^2 &= 0. && (47)
\end{aligned}
$$

2. *The steady-state mean of c is $\bar{c} = \bar{z} + \bar{x}$.*

3. *The steady-state mean of w is given by $\bar{w} = \frac{\bar{z}}{h(\bar{x})} + f(\bar{x})$.*

4. *The steady-state mean of any function $F(w, x)$ is given by $\bar{F} = F(\bar{w}, \bar{x})$. Thus, we have $\bar{r} = r(\bar{x})$, $\bar{f} = f(\bar{x})$, $\bar{h} = h(\bar{x})$, $\bar{g} = g(\bar{x})$, $\bar{\Lambda} = \Lambda(\bar{x})$, $\bar{k} = k(\bar{w}, \bar{x})$, $\bar{n} = n(\bar{w}, \bar{x})$, etc.*

Note that $\bar{x}$ and $\bar{z}$ are chosen so that the expected growth rate of x and $w - f(x)$ are zero at $x = \bar{x}$ and $z = \bar{z}$. In general, it is not necessarily true that the expected growth rate of a function $F(w, x)$ is true at $w = \bar{w}$ and $x = \bar{x}$, because of the Jensen terms. Such Jensen terms are typically negligible, so that the expected growth rate for each variable is close or equal to zero at its steady-state mean.

4 Asset Pricing Puzzles Revisited

In this section, I revisit asset pricing puzzles using the model given by equations (17)–(20). Although this model is slightly defective in terms of its boundary behavior, it is easily remedied as shown in the previous section. The refinement dealing with the boundary behavior of the labor income process do not affect the analytical form of the solution in the relevant state space, i.e., where $\hat{l}_1(x) = \hat{l}_3(x) = 1$. When $0 < \bar{x} < x_m$, we can always assume, without loss of generality, that $\hat{l}_1(\bar{x}) = \hat{l}_3(\bar{x}) = 1$.

Using the dummy variable $c = z + x$, we can reformulate the model as a parametric specification of the habit formation model defined by equations (7)–(9):

$$\max_{\alpha(t), c(t): t \geq 0} E_0 \left[\int_0^\infty u(c(t) - x(t), t)\, dt \right], \tag{48}$$

subject to

$$\begin{aligned} dw &= [\alpha(\mu(w,x) - r(w,x))w + r(w,x)w - c]\, dt + \alpha\sigma(w,x)w\, dB_w, & (49) \\ dx &= \left(bc - ax + \frac{1}{2}\epsilon\beta^2(c-x)^2 \right) dt + \beta(c-x)dB_x, & (50) \end{aligned}$$

where $\rho > 0$, $\gamma < 0$, $a \equiv b + \kappa > b > 0$, $\beta > 0$, $\delta = \text{corr}(dB_w, dB_x) < 0$, and $\epsilon > 0$, with $r(x)$ and $f(x)$ given by equation (20). For convenience, I will denote this model by the acronym SHABIT (**S**tochastic **HABIT**).

The habit formation model of Constantinides [1990] is obtained by setting both $\beta = 0$ and $\epsilon = 0$. Henceforth, this model is referred to as CHABIT (**C**onstantinides **HABIT**). An intermediate case between SHABIT and CHABIT is obtained by letting $\beta = 0$ but keeping $\epsilon > 0$. The budget set in this model is exactly the same as CHABIT, except that the riskless rate is state-dependent (a different technological specification). Henceforth, I will refer to this model as DHABIT (Constantinides **HABIT** with a state-**D**ependent riskless rate).[12] Since the correlation δ is irrelevant when $\beta = 0$, DHABIT has seven free parameters:

[12] To summarize, SHABIT stands for the full model with $\beta > 0$, $\epsilon > 0$; DHABIT stands for the intermediate model with $\beta = 0$, $\epsilon > 0$; and CHABIT stands for the most restrictive model with $\beta = \epsilon = 0$.

$(\rho, \gamma, a, b, \epsilon, r_0, x_m)$. When $\epsilon = 0$, $r(x) = r_0 = \bar{r}$, and $x_m = \infty$ (the functions $h(x)$ and $g(x)$ are constant, so that there is no restriction on the state space of the labor income process). Thus, CHABIT has five free parameters: $(\rho, \gamma, a, b, \bar{r})$.

In examining the empirical implications of the model, I begin by using models CHABIT and DHABIT to address the equity premium, equity volatility, and the riskfree rate puzzles. Particular attention is given to the issue of parameter identification from both empirical observations and model restrictions. Sample moments are obtained from the annual U.S. data for the postwar period, including the average growth rate of aggregate consumption $G = 1.89\%$, the volatility of the consumption growth rate $\bar{\sigma}_c = 1.5\%$, the volatility of the dividend growth rate $\bar{\sigma}_z = 11.2\%$, the expected real market return to equity $\bar{\mu}_K = 7.85\%$, the volatility of market return to equity $\bar{\sigma}_k = 15.9\%$, and the average real riskless rate $\bar{R} = 0.94\%$. The average growth rate is used to obtain the rates and yields in the per capita economy, where the expected equity return is $\bar{\mu}_k = \bar{\mu}_K - G = 5.96\%$, and the average real riskless rate is $\bar{r} = \bar{R} - G = -0.95\%$. Thus, for the per capita economy, there are five relevant sample moments are $(\bar{\sigma}_c, \bar{\sigma}_z, \bar{\sigma}_k, \bar{\mu}_k, \bar{r}) = (1.5\%, 11.2\%, 15.9\%, 5.96\%, -0.95\%)$. The implied Sharpe ratio of the equity return is given by $\bar{\Lambda} = 0.4341$. Throughout the discussion, I will assume that the steady-state value of a variable is equal to the sample mean of this variable.

4.1 *Models CHABIT and DHABIT*

Evaluated at the steady-state means $\bar{z}$ and $\bar{x}$, both models DHABIT and CHABIT provide three theoretical restrictions and five moment restrictions: (i) the equilibrium restriction on the market price of risk (33); (ii) the optimality conditions on $h(x)$ and $g(x)$ (31) and (34); (iii) the conditional means of the state variables must be zero; (iv) the conditional volatilities of the state variables must be equal to the sample volatilities; and (v) the AR(1) coefficient of the ratio $\xi \equiv \frac{x}{z^*}$ is equal to one minus the mean reversion coefficient of ξ.[13]

Specifically, using the equilibrium restriction $\Lambda(x) = (1-\gamma)g(x)$ to eliminate the function $g(x)$, we have

$$
\begin{aligned}
\frac{h(x)^{\gamma-1}}{\phi} &= 1 + bf'(x), && (51) \\
\Lambda(x)^2 &= -2(1-\gamma)\left[r(x) - \frac{\rho}{\gamma} + \frac{1-\gamma}{\gamma}\frac{h(x)^\gamma}{\phi}\right], && (52) \\
0 &= r(x) - \rho + \frac{2-\gamma}{2(1-\gamma)}\Lambda(x)^2, && (53) \\
0 &= bz - (a-b)x = bc - ax, && (54) \\
\sigma_z(x) &= \frac{\Lambda(x)}{1-\gamma}, && (55) \\
\sigma_c(x) &= \frac{c-x}{c}\sigma_z(x), && (56) \\
AR(1) &= 1 - (a-b) + \sigma_z(x)^2, && (57)
\end{aligned}
$$

[13] Ito's lemma implies $d\xi = (b - \kappa(x)\xi)\,dt - \sigma_z(x)\xi dB_w$, where $\kappa(x) = (a-b) + \mu_z(x) - \sigma_z(x)^2$. Thus, the mean reversion coefficient is $(a-b) + \mu_z(x) - \sigma_z(x)^2$.

These equations also hold for CHABIT, except that in this model, the functions $r(x)$, $f'(x)$, $\Lambda(x)$ (hence $\sigma_z(x)$), and $h(x)$ are constant.

In both models, the equity premium, the equity volatility, and the riskfree rate puzzles are resolved immediately by choosing three parameters:

- **Equity Premium Puzzle:** Equation (55) implies that

$$1-\gamma = \frac{\bar{\Lambda}}{\bar{\sigma}_z} = \frac{0.4341}{11.2\%} = 2.7274.$$

  Intuitively, due to limited participation assumption, assets are priced by the dividend growth rate, which is an order of magnitude more volatile than the consumption growth rate. Thus, the risk aversion parameter required to fit the equity premium puzzle is an order of magnitude smaller than the CCAPM used by Mehra and Prescott [1985].

- **Equity Volatility Puzzle:** Equations (54), (55) and (56) imply that

$$\frac{a}{a-b} = \frac{\bar{c}}{\bar{c}-\bar{x}} = \frac{\bar{\sigma}_z}{\bar{\sigma}_c} = \frac{11.2\%}{1.5\%} = 7.47.$$

  This ratio can be interpreted as the *operating leverage* of the risky production.[14] Intuitively, because the labor income growth is deterministic, the investor bears all of the production risk. The volatility of the investor's income growth is amplified by the operating leverage.

- **Riskfree Rate Puzzle:** Equation (53) implies that

$$\rho = \bar{r} + \frac{2-\gamma}{2(1-\gamma)}\bar{\Lambda}^2 = -0.95\% + \frac{1+2.7274}{2\times 2.7274} \times 0.4341^2 = 0.1193.$$

  Intuitively, when the risk aversion parameter is low, the elasticity of intertemporal substitution is high. Therefore, there is no longer a conflict between a positive discount rate and a low riskless rate.

In fitting the three puzzles, we gave ourselves a break by not using the actual income and consumption levels to pin down the labor's share of income. This is in part motivated by the fact that determining the relative level of financial and labor income involves serious measurement issues, related to such questions as how to distinguish labor income and financial income (which portion of Bill Gates' income is due to his "labor" as a Microsoft employee, and which portion of his income is due to his ownership of the firm?), how to sample the population, how to account for interest income, how to account for housing and other non-financial, non-human assets, and how to account for taxation and government spending. If we were able to come up with an accurate measure of the labor's share of the output, $\frac{\bar{x}}{\bar{c}} = \frac{b}{a}$, which in all likelihood would be different from the value calibrated above, then the two models can no longer fit exactly the three puzzles simultaneously using only three parameters

[14] The ratio $\frac{\bar{x}}{\bar{c}} = \frac{b}{a} = 0.866$ can be interpreted as the *labor's share* of output, which is roughly consistent with historical estimates (see, e.g., Ibbotson and Brinson [1987]). Heaton and Lucas [1996] assume that the labor's share is 85%.

ρ, γ, and $\frac{b}{a}$. One way to restore the simultaneous and exact fit for these three puzzles is to allow the labor income growth to be stochastic and correlated with the dividend income growth, which is exactly what the model SHABIT does. In summary, because we do not pretend to be able to measure the relative level of financial and labor income, equation (54) does not have any empirical content, and serves only to relate the ratio of $\bar{x}$ and $\bar{c}$ to the ratio of the parameters b and a in other restrictions.

Since the AR(1) coefficient of ξ is a sample moment, the two models also share the same value of $a-b$.

We are left with two unused restrictions, (51) and (52), to identify the endogenous parameter ϕ in CHABIT, or the three exogenous parameter (r_0, ϵ, x_m) in DHABIT. There seems to be an over-identification for CHABIT and an under-identification for DHABIT. Fortunately, this is not the case. To see this, note that equation (51) and (52) imply

$$\frac{\bar{\Lambda}^2}{1-\gamma} = \bar{h}(1+b\eta) - \bar{r}, \tag{58}$$

where[15] $\eta = \frac{1}{\bar{r}+a-b}$. On the other hand, the PDE (44) can be solved exactly, and the equity price is given by $k(w,x) = \frac{w-f(x)}{1+b\eta}$. Thus, the dividend/price ratio is given by $\frac{\bar{z}}{\bar{k}} = \bar{h}(1+b\eta)$. Since all per capita variables are stationary with zero growth (the restrictions (53)–(54) ensure that the state variables have zero growth rate), the capital gain in the per capita economy is zero. Thus, the expected equity return in the per capita economy is equal to the dividend/price ratio: $\bar{\mu}_k = \frac{\bar{z}}{\bar{k}}$. Given these results, Equation (58) can be rewritten as $\frac{\bar{\Lambda}^2}{1-\gamma} = \bar{\mu}_k - \bar{r} = \bar{\Lambda}\bar{\sigma}_k$, or

$$\bar{\sigma}_k = \frac{\bar{\Lambda}}{1-\gamma}. \tag{59}$$

Since k is a constant proportion of z, we have $\sigma_k = \sigma_z$, and equation (59) is identical to equation (55),[16] which means that one of the two restrictions (51) and (52) is redundant and ϕ can be identified exactly.

Turning to the model DHABIT, all of the above discussion applies, except that, in this model, equation (44) can no longer be solved analytically due to the state-dependence of $r(x)$. However, when the economy is close to the steady state, a linear approximation gives

$$k \approx \frac{w-f(x)}{1+bf'(\bar{x})},$$

where $f(x)$ is given by equation (20). Thus, the expected equity return is approximately given by $\bar{\mu}_k = \frac{\bar{z}}{\bar{k}} = \bar{h}(1+bf'(\bar{x}))$, and equation (59) still holds (approximately). With only one remaining restriction, say equation (51), the remaining exogenous parameters ϵ, r_0, and x_m (in this model, the endogenous parameter ϕ is deterministically related to the exogenous parameter x_m) can not be separately identified.

[15]For CHABIT, $\epsilon = 0$. Equation (20) therefore implies that $f(x) = \eta x$, and $r(x) = \frac{1-(a-b)\eta}{\eta} = \bar{r}$.

[16]In the data, $\bar{\sigma}_k = 15.9\% \neq \bar{\sigma}_z = 11.2\%$. Letting labor income growth to be stochastic and not perfectly correlated with the dividend growth introduces a wedge between the two. Thus we can potentially use the difference between $\bar{\sigma}_k$ and $\bar{\sigma}_z$ to help identify the stochastic portion of the labor income process. See Section 4.2.

Given r_0, ϵx_m can be identified by equation (51), but ϵ and x_m still can not be separately identified. This is because in this model (due to its homothetic structure) rates of return, yields, and proportional volatilities are independent of the overall scale of income, consumption, and wealth: if we scale up the level of all income, consumption, and wealth variables by a factor of, say, ϵ_0, the structure of the model remains the same, except that ϵ is scaled down by ϵ_0. So long as we don't care about the absolute level of income, consumption, and wealth (including price) variables, scaling $\epsilon_0 = \epsilon$ effectively normalizes ϵ to 1. Henceforth, we impose this normalization by setting $\epsilon = 1$.

If we observe the real interest rate, then r_0 can be identified by estimating the sensitivity of the real riskless rate with labor income. Alternatively, if the real riskless rate is set by a monetary authority, in the manner say $r - \bar{r} = \tau(x - \bar{x})$ (which is reminiscent of a Taylor rule), then r_0 can be inferred from the policy parameter τ. Under both of these scenarios, DHABIT is completely identified (under the normalization $\epsilon = 1$).

The above discussion suggests that both CHABIT and DHABIT do quite a reasonable job of explaining the equity premium, the equity volatility, and the riskfree rate puzzle, under a judicious choice of the parameters $(\rho, 1-\gamma, \frac{b}{a})$. The restrictions that identify these parameters are the same for both models, and therefore the calibrated values of these parameters are identical for both models. However, the exact nature of the fit should be interpreted with caution. It masks the fact that some relevant empirical restrictions have been ignored. First, the labor's share of output in principle determines $\frac{b}{a}$, which would take away this degree of freedom from both models to fit the equity volatility puzzle. Second, both models imply that the volatility of equity return and that of the dividend growth rate are the same, which is clearly counter-factual.

4.2 *Model SHABIT*

The problems associated with CHABIT and DHABIT can be resolved by allowing the labor income process to be stochastic. This leads us to the model SHABIT. The first problem is resolved because when the labor income growth is stochastic, the amplification of the dividend growth depends not only on the operating leverage, but also on the volatility of the labor income shock and the correlation between the dividend income shock and labor income shock.[17] Second, a stochastic labor income shock also introduces a wedge between the volatility of the equity return and the volatility of the dividend growth rate, because the expected equity return becomes state-dependent and stochastic and furthermore contains an extra state-dependent premium arising from the hedging demand associated with the labor income risk.

SHABIT also offers a much richer term structure of real interest rates. In DHABIT, the

[17] To see this, note that $c = z + x$ implies $\frac{dc}{c} = \frac{z}{c}\frac{dz}{z} + \frac{x}{c}\frac{dx}{x}$, which in turn implies:

$$\sigma_c^2 = \left(1 - \frac{x}{c}\right)^2 \sigma_z^2 + \left(\frac{x}{c}\right)^2 \sigma_x^2 + 2\left(\frac{x}{c}\right)\left(1 - \frac{x}{c}\right)\delta_{xz}\sigma_z\sigma_x,$$

where for $y = c, z, x$, σ_y is the conditional volatility of the growth rate $\frac{dy}{y}$, and δ_{xz} is the correlation between the dividend growth rate and the labor income growth rate. Given $\sigma_c = 1.5\%$, $\sigma_z = 11.2\%$, if $\frac{x}{z} \neq 0.866$, then it must be that $\sigma_x \neq 0$.

riskless rate is deterministic, which implies that the term structure of real interest rates is close to being trivial (returns on long term bonds can still be stochastic because they are affected by the production shocks). In contrast, the riskless rate is stochastic in SHABIT, and the endogenously determined bond returns and term spreads exhibit realistic properties. For an example, the correlation between the expected bond returns and the term spreads is consistent with the expectations puzzle (in terms of the sign, but not necessarily the magnitude). This correlation is intimately related to the the correlation between the aggregate dividend growth rate and the aggregate labor income growth rate.

In principle, implications for equity returns and bond returns are inter-related in this model. However, the parameters of the model corresponding to the observed economy are such that some rough form of separation is discernible. In a nutshell, the production shock and the associated risk premium appear to have a first order effect on equity returns (the equity premium and volatility puzzles) but only a second order effect on bond returns (predictability of bond returns by dividend/price ratio), while the labor income shock and the associated risk premium appear to have a second order effect on equity returns (predictability of equity returns by term spreads) but a first order effect on bond returns (stochastic nature of real interest rates and the expectations puzzle).

4.2.1 Parameter Identification

Compared to DHABIT, SHABIT has two additional parameters: β and δ. The first parameter can be identified from the volatility of the labor income growth rate, and the second can be identified from the correlation between the labor income growth rate and the dividend growth rate. To derive the correlation from the model, we need to characterize the equilibrium dividend growth rate.

From equations (30) and (36), we have

$$\frac{dz^*}{z^*} = \mu_z(z^*, x)dt + \sigma_{zw}(z^*, x)dB_w + \sigma_{z\perp}(z^*, x)dB_\perp, \tag{60}$$

where

$$\begin{aligned}
\mu_z(z^*, x) &= \left[\frac{r(x)-\rho}{1-\gamma} + \frac{2-\gamma}{2(1-\gamma)^2}\Lambda(x)^2 + \frac{2-\gamma}{2}(1-\delta^2)l_3(x)^2h(x)^2f'(x)^2\right] \\
&+ \left[\frac{h''(x)}{h(x)}\frac{\psi(x,z^*)^2}{2} + \frac{h'(x)}{h(x)}\psi(x,z^*)\frac{\delta\Lambda(x)+\sqrt{1-\delta^2}\Lambda_\perp(x)}{1-\gamma}\right] + \frac{h'(x)}{h(x)}\nu(x,z^*), \\
\sigma_{zw}(z^*, x) &= \frac{\Lambda(x)}{1-\gamma} + \delta\frac{h'(x)}{h(x)}\psi(x,z^*) = \frac{\Lambda(x)}{1-\gamma}, \\
\sigma_{z\perp}(z^*, x) &= -\sqrt{1-\delta^2}l_3(x)h(x)f'(x) + \sqrt{1-\delta^2}\frac{h'(x)}{h(x)}\psi(x,z^*),
\end{aligned}$$

where $\nu(x,z) = bz - \kappa x + \frac{1}{2}\epsilon\beta^2z^2$, $\psi(x,z) = l_3(x)z$, $l_3(x) = \beta$, $\Lambda(x)$ is given by equation (33), and $\Lambda_\perp$ is given by equation (39).

Thus, the two additional moment restrictions required to identify β and δ are given by:

$$\begin{aligned}\bar{\sigma}_x &= \beta\frac{\bar{z}}{\bar{x}}, \\ \bar{\delta}_{xz} &= \left[\sigma_{zw}(\bar{z},\bar{x})^2 + \sigma_{z\perp}(\bar{z},\bar{x})^2\right]^{-1/2}\left[\delta\sigma_{zw}(\bar{z},\bar{x}) + \sqrt{1-\delta^2}\sigma_{z\perp}(\bar{z},\bar{x})\right],\end{aligned}$$

where $\bar{\sigma}_x$ is the sample volatility of the labor income growth rate, and $\bar{\delta}_{xz}$ is the sample correlation between the labor and dividend income growth rates.

Other parameters are identified as in DHABIT, except that the four moment conditions (53)–(56) generalize to

$$\begin{aligned}0 &= \mu_z(\bar{z},\bar{x}), \\ 0 &= \nu(\bar{x},\bar{z}), \\ \bar{\sigma}_z &= \sqrt{\sigma_{zw}(\bar{z},\bar{x})^2 + \sigma_{z\perp}(\bar{z},\bar{x})^2}, \\ \bar{\sigma}_c &= \sqrt{[\sigma_{zw}(\bar{z},\bar{x}) + \delta\psi(\bar{x},\bar{z})]^2 + \left[\sigma_{z\perp}(\bar{z},\bar{x}) + \sqrt{1-\delta^2}\psi(\bar{x},\bar{z})\right]^2},\end{aligned}$$

for $\beta \neq 0$, and the two optimality conditions (31) and (34) contain terms related to δ and β. In practice, δ and β are identified jointly with the rest of the parameters.

4.2.2 Predictability

Lemma 2 implies that if $\epsilon \neq 0$ (for both DHABIT and SHABIT), the riskless rate must be monotonically increasing in x. We will now show that the functions $h(x)$, $g(x)$, and $\Lambda(x)$ are all state-dependent and, for sufficiently small β, monotone.

Corollary 3 (Predictability) *Suppose that Proposition 2 holds. Then, for sufficiently small β, and $\forall x$, such that $\hat{l}_1(x) = \hat{l}_3(x) = 1$, $h'(x) > 0$, $g'(x) < 0$, and $\Lambda'(x) < 0$.*

This result means that the dividend/price ratio (related to $h(x)$), the conditional volatility of dividend growth rate (related to $g(x)$), and the equity premium (related to both $\Lambda(x)$ and $g(x)$) are all predictable by x, and *by each other*.

The fact that $r'(x) > 0$ and $\Lambda'(x) < 0$ implies that the riskless rate and the market price of risk are negatively correlated. In a one-factor model, if r is mean reverting under the risk-neutral measure, then it is negatively correlated with term spreads. Consequently, we expect that the expected returns on real bonds and real term spreads are positively correlated. This implies that the expectations hypothesis should *not* hold for the real term structure, and the slope coefficients in the expectations regressions should be less than 1, and possibly negative. To make this point formally, we need to examine how the model price real bonds.

4.2.3 Term Structure of Real Interest Rates

Since the pricing kernel is explicitly known, the term structure of real interest rates can be readily characterized. Let $P_T(w, x; t)$ be the price of a real zero-coupon bond with maturity

T, and $y_T(w, x; t) = -\frac{1}{T} \log P_T(w, x)$ be the continuously compounded zero yield. Without loss of generality, we can write

$$\frac{dP_T}{P_T} = \mu_T dt + \sigma_{Tw} dB_w + \sigma_{Tx} dB_x \tag{61}$$

where μ_T is the instantaneous expected return on the bond, $\sigma_{Tw} = \sigma \frac{\partial \log P_T}{\partial \log w}$, and $\sigma_{Tx} = l_3(x) z \frac{\partial \log P_T}{\partial x}$. Since P_T must be priced by $m(t)$, it follows that

$$\mu_T - r = (\sigma_{Tw} + \delta \sigma_{Tx}) \Lambda(x) + \sqrt{1-\delta^2} \sigma_{Tx} \Lambda_\perp(x). \tag{62}$$

In general, both production risk and labor income risk affect the expected bond returns. I examine them separately.

1. *Labor Income Risk and Expectations Puzzle*

To isolate the effect of labor income risk, let $\beta \neq 0$, and fix z at its steady-state mean $\bar{z}$ (effectively eliminating the effect of production risk). This produces a one-factor model of the real term structure: $r = r(x)$, and

$$dx = \left[b\bar{z} - \kappa x + \frac{\epsilon \beta^2}{2} \bar{z}^2 \right] dt + \beta \bar{z} dB_x, \ 0 < x < x_m, \tag{63}$$

with market price of risk for B_x given by $\lambda = \delta \Lambda(x) + \sqrt{1-\delta^2} \Lambda_\perp(x)$.

This is a nonlinear model, and the riskless rate is bounded from both above and below. Consequently, bond prices cannot be computed analytically. To facilitate the discussion, it is convenient to take a linear approximation of this model (linearizing around the steady-state mean of x and keeping only the lowest non-zero terms controlled by β):

$$\begin{aligned} r &\approx \bar{r} + \bar{r}_x (x - \bar{x}), \quad dx \approx \kappa(\bar{x} - x) dt + \beta \bar{z} dB_x, \\ \lambda &\approx \left[\delta \bar{\Lambda} + \sqrt{1-\delta^2} \bar{\Lambda}_\perp \right] + \left[\delta \bar{\Lambda}_x + \sqrt{1-\delta^2} \bar{\Lambda}_{\perp x} \right] (x - \bar{x}). \end{aligned} \tag{64}$$

where $\bar{x} = \frac{b\bar{z}}{\kappa}$. Expressing directly in terms of r, we have

$$dr = \kappa(\bar{r} - r) dt + \bar{\sigma} dB_x, \quad \lambda = \frac{\lambda_0 + \lambda_1 r}{\bar{\sigma}}, \tag{65}$$

where $\bar{\sigma} = \bar{r}_x \beta \bar{z}$, and

$$\begin{aligned} \lambda_0 &= \bar{\sigma} \left[\delta \bar{\Lambda} + \sqrt{1-\delta^2} \bar{\Lambda}_\perp \right] - \bar{r} \lambda_1 \approx \delta \beta \bar{z} \bar{\Lambda} \bar{r}_x - \bar{r} \lambda_1 < 0, \\ \lambda_1 &= \beta \bar{z} \left[\delta \bar{\Lambda}_x + \sqrt{1-\delta^2} \bar{\Lambda}_{\perp x} \right] \approx \delta \beta \bar{z} \bar{\Lambda}_x > 0. \end{aligned} \tag{66}$$

The risk-neutral dynamics of the riskless rate follow immediately:

$$dr = \tilde{\kappa}(\tilde{r} - r) dt + \bar{\sigma} d\tilde{B}_x, \tag{67}$$

where $\tilde{B}_x$ is a standard Brownian motion under the risk-neutral measure, $\tilde{\kappa} = \kappa + \lambda_1 > \kappa$ is the mean reversion coefficient under the risk-neutral measure, and $\tilde{r} = (\kappa \bar{r} - \lambda_0)/\tilde{\kappa}$ is the risk-neutral long-run mean of the riskless rate.

The linearization makes the ubiquitous influence of labor income risk transparent. The volatility of the riskless rate ($\bar{\sigma}$), the constant part of the market price of risk (λ_0), and the state-dependent part of the market price of risk (λ_1) are all proportional to β. All else equal, a larger β leads to a more volatile riskless rate, a steeper mean yield curve, and greater violation of the expectations hypothesis.

When r_t is interpreted as the nominal riskless rate, equation (65) is a model Dai and Singleton [2001] propose to explain the violation of the expectations puzzle in the nominal term structure. They show that when $\kappa > 0$, $\lambda_0 < 0$, and $\lambda_1 > 0$, the model implies (i) an upward-sloping mean yield curve (because $\lambda_0 < 0$); and (ii) a positive correlation between the expected excess return on zero-coupon bonds and the term spread (hence violation of the expectations puzzle in the correct direction). Dai and Singleton [2001] do not provide an economic explanation for why λ_0 should be negative and λ_1 should be positive.

In my model, the sign restrictions $\lambda_0 < 0$ and $\lambda_1 > 0$ arise naturally. We have already shown that, under very general conditions, $r_x > 0$ (pro-cyclical riskless rate), $\Lambda_x < 0$ (counter-cyclical Sharpe ratio), $\bar{z} > 0$, and $\bar{\Lambda} > 0$. Thus, the sign restrictions on λ_0 and λ_1 follow immediately from the negative correlation between the production shock and the labor income shock, i.e., $\delta\beta < 0$.

Since the expectations puzzle relates to nominal yields, and real yields are not directly observable in the U.S. (until very recently), a case cannot be made that labor income risk is solely responsible for violation of the expectations hypothesis, although it is certainly part of the reason. As McCallum [1994] shows theoretically and Kugler [1997] shows empirically, monetary policy can have a direct impact on the behavior of nominal yields. In particular, interest rate smoothing rules can induce or modify both the mean reversion coefficient and the market prices of risk, thereby changing at least the extent of the expectations puzzle.

2. *Production Risk and the Real Term Structure*

Interestingly, in the stochastic habit formation model, the real term structure can also be influenced by production risk. The effect is subtle, however, because production risk does not directly enter the real riskless rate – it affects only the *conditional distribution* of the riskless rate.

To isolate the effect of production shock, we eliminate labor income risk by letting $\beta = 0$. In this case, equation (62) reduces to

$$\mu_T - r = \sigma_{Tw}\Lambda. \tag{68}$$

The expected return on the bond does not vanish because it remains exposed to production risk through the dependence of the conditional moments of r (or x) on the financial income z. It can be shown, however, that the conditional volatility of the bond return and therefore the expected excess bond return are nearly zero at both short and long ends of the maturity range.

To see this, let $f_T = -\frac{\partial \log P_T}{\partial T}$ be the instantaneous forward rate. Clearly,

$$\frac{\partial \sigma_{Tw}}{\partial T} = -\sigma \frac{\partial f_T}{\partial \log w}. \tag{69}$$

As $T \to 0$, we must have $\mu_T \to r$, and $f_T \to r$. Since r is independent of wealth w, we must have

$$\lim_{T\to 0}\sigma_{Tw}=0,\ \lim_{T\to 0}\frac{\partial\sigma_{Tw}}{\partial T}=0. \tag{70}$$

Under reasonable assumptions (the state variables are mean reverting under the risk-neutral measure), we expect that, as $T\to\infty$, $f_T\to\bar{f}$, where $\bar{f}$ is a constant, and $\sigma_{Tw}\to 0$. It follows that

$$\lim_{T\to\infty}\sigma_{Tw}=0,\ \lim_{T\to\infty}\frac{\partial\sigma_{Tw}}{\partial T}=0. \tag{71}$$

Equations (70) and (71) indicate that the volatility of the bond return and therefore the risk premium induced by production risk become appreciable only in the intermediate maturities. This is perhaps a structural reason why we observe a hump in the term structure of volatility around two years of maturity for nominal yields.[18] The size and location of the hump depends on the parameter b, which allows z to influence the real term structure through the drift of x, the mean reversion coefficient κ, which suppresses the influence of z at long maturities, and the risk aversion parameter $1-\gamma$, which determines the size of the risk premium through Λ (see equation (68)).

5 Exchange Formulation of the Model

Consider the following self-insurance problem by the investor (who enjoys complete markets and frictionless trading):

$$\max_{z(t):t\geq 0} E_0\left[\int_0^\infty u(z(t),t)\right],$$

subject to the constraint that

$$E_0\int_0^\infty z(t)m(t)dt\leq E_0\int_0^\infty z^*(t)m(t)dt,$$

where $z^*(t)$ is the endowment process, and $m(t)$ is the pricing kernel to be determined in equilibrium (with market clearing condition $z=z^*$).

A simple way of introducing labor income x is to assume that the endowment process $z^*(t)$ is given by

$$\frac{dz^*}{z^*} = \mu^*(z^*,x)dt+\sigma^*(z^*,x)dB_z. \tag{72}$$

Depending on how x is specified, the labor income can be interpreted either as an "external" or "internal" habit.

[18]When inflation risk is substantial, the hump may be masked by the inflation premium. This is consistent with the fact that a volatility hump is observed in the 80's and 90's in the U.S. treasury yields, when inflation was subdued, and is not observed in the 70's, when inflation was rampant.

5.1 *Labor Income as "External" Habit*

Suppose that the labor income process is given by

$$dx = \nu(x, z^*)dt + \psi(x, z^*)dB_x.$$

In this case, the labor income is not directly controlled by the investor's consumption policy. Rather, it is exogenously determined by the endowment process (more precisely the labor income process and the endowment process are jointly given exogenously). A change of variable $c = z + x$ reveals that the above model can be interpreted as a model of habit formation, with c being the aggregate consumption, and x being interpreted as an "external" habit. Asset pricing in this model is simple, because the presence of labor income merely enriches the exogenously specified probability distribution of the endowment process. In particular, the equilibrium pricing kernel is given by the utility gradient evaluated at the endowment: $m(t) = u_z(z(t)^*, t)$. Thus, the riskless rate is given by

$$r(z^*, x) = -\frac{u_{zt} + \mu^*(z^*, x)u_{zz} + \frac{1}{2}\sigma^*(z^*, x)^2 u_{zzz}}{u_z}. \tag{73}$$

By imposing a joint restriction on $\mu^*(z^*, x)$ and $\sigma^*(z^*, x)$, we can make the riskless rate constant or state-dependent in any desirable way. For arbitrary specifications of the endowment process, all endogenous variables, including the riskless rate, the Sharpe ratio, equity prices and returns, bond prices and returns, are functions of both state variables (z^*, x) in general.

A prominent example of this approach is Campbell and Cochrane [1999] (henceforth CCHABIT) where, consistent with the interpretation of x as an "external" consumption habit, z^* is interpreted as the *surplus consumption.* These authors specify the joint processes for the *log surplus consumption ratio*, $s \equiv \log \frac{z^*}{z^*+x}$, and the consumption process, $c \equiv z^* + x$, which is is equivalent to an exogenous specification of (z^*, x) after a simple change of variable. In Appendix G, I show that, in the continuous-time limit, CCHABIT is essentially given by

$$dS^* = h(\bar{S} - S^*)dt + (\mathcal{B} - S^*)vdB,$$

and

$$dx = (bz^* - \kappa x)dt + v(1 - \mathcal{B})(z^* + x)dB,$$

where $\bar{S}$, $\mathcal{B}$, h, b, κ are constants, and $S^* \equiv \frac{z^*}{z^*+x} = e^s$ can be referred to either as the *surplus consumption ratio* or as the *investor's share of output.*

5.2 *Labor Income as "Internal" Habit*

Now, consider the alternative specification for labor income:

$$dx = \nu(x, z)dt + \psi(x, z)dB_x.$$

In this case, the labor income process is directly controlled by the investor's consumption policy z. A change of variable $c = z + x$ reveals that the model can be interpreted as a model of habit formation, with x being interpreted as an "internal habit". Asset pricing is more involved, because a small positive amount of saving away from equilibrium produces three effects: (i) a reduction in current utility, captured by the utility gradient; (ii) an increase in future utility due to payoff from the saving; and (iii) a change in future endowment, because the current level of consumption z affects future levels of labor income (which depends on z), which in turn affects future levels of endowment (which depends on x). The first two effects are conventional. When the third effect is not present, an optimal trade-off between the first two effects leads to the utility gradient characterization of the pricing kernel. The presence of the third effect inserts a wedge between the pricing kernel and the utility gradient (Lemma 4 shows that, in the production setting, it inserts a wedge between the marginal utility of wealth and the marginal utility of consumption).

The production model SHABIT can be recast as an exchange model with the endowment process given by equation (72) and a labor income process given by equation (19). If z^* is given by equation (30), then the equilibrium riskless rate is given by equation (20), and the equilibrium return for the aggregate wealth $w(t) = m(t)^{-1}E_t\left[\int_t^\infty (z^*(s) + x(s))m(s)ds\right]$ is given by equations (21)–(22). To obtain these results from standard methods commonly employed in dealing with exchange economies is not a trivial matter. To see this, note that the utility gradient method of Duffie and Skiadas [1994] follows from the first order condition of the Saddle point problem:

$$\max_{\mathbb{Z}} \mathcal{L}(\mathbb{Z}),$$

where $\mathbb{Z}$ is the cumulative dividend process,

$$\mathcal{L}(\mathbb{Z}) = E_0\left[\int_0^\infty u(z(t),t)dt - \lambda\int_0^\infty (z(t) - z(t)^*)m(t)dt\right],$$

and $\lambda > 0$ is a constant at which the Saddle point problem has an optimal solution. The first order condition is obtained by setting the *directional* or *Gateaux* derivative of $\mathcal{L}$ to zero (see Duffie [1996]).

A standard assumption in the asset pricing literature is that the endowment process is not affected by the control variable $z(t)$. In this case, the first order condition states that that the utility gradient is equal to $\lambda m(t)$ (see Duffie and Epstein [1992] and Duffie and Skiadas [1994] on how to compute the utility gradient in the presence of an internal habit).

In the current case, $z^*(t)$ is affected by the process $x(t)$, which in turn is affected by the control policy $z(t)$. Thus, the directional derivative of $z^*(t)$ is not zero. This extra term inserts a wedge between the utility gradient (which is computed as usual) and $\lambda m(t)$, which in general depends on the entire history of the pricing kernel $m(t)$. In another word, the equilibrium pricing kernel in this case is determined by an integral equation, which is not trivial to solve.

5.3 *Models CHABIT, CCHABT, and SHABIT*

I conclude this section by contrasting the models CHABIT, CCHABIT, and SHABIT from the perspective of an endowment economy, thereby highlighting some key marginal contributions of this paper.

CHABIT can be recovered in the exchange setting by assuming that x is "internal", $\beta = 0$, $\mu^*(z,x) = \bar{\mu}^*$, $\sigma^*(w,x) = \bar{\sigma}^*$, where $\bar{\mu}^*$ and $\bar{\sigma}^*$ are constant (which is equivalent to assuming that $\epsilon = 0$). In SHABIT, both $\mu^*(z^*, x)$ and $\sigma^*(z^*, x)$ are state-dependent, induced by the state-dependence of the riskless rate. From the perspective of an endowment economy, the state-dependence of $\mu^*(z^*, x)$ and $\sigma^*(z^*, x)$ induces the state-dependence of the riskless rate and the predictability of expected returns. SHABIT captures several important aspects of the observed behavior of macroeconomy and asset markets that CHABIT (and even DHABIT) can not: (i) the dividend growth rate and the labor income growth rate are imperfectly correlated; (ii) expected bond returns and term spreads are positively correlated; (iii) labor income risk appears to play an important role in explaining cross-sectional behavior of equity returns.

Campbell and Cochrane [1999] focus mainly on a version of CCHABIT with $\mathcal{B} = 1$, in which case the consumption habit is deterministic, and the riskless rate is constant. Although the underlying habit process in CCHABIT is essentially the same as that in CHABIT, the expected returns in CCHABIT are state-dependent because aggregate consumption growth rate is i.i.d., whereas in CHABIT, the aggregate dividend growth rate is i.i.d..

In setting $\mathcal{B} \neq 1$, both the consumption habit and the riskless rate become stochastic in CCHABT. This version of the model, however, has some important limitations. First, assuming that their consumption habit can be equated with labor income, their model implies that the dividend growth rate and the labor income growth rate are perfectly correlated; (ii) the real interest rate is likely to be too volatile, because it depends on both z^* and x. The explicit dependence on z^* can produce unintended features in the term structure of interest rates; and (iii) in principle, the model can also generate non-zero correlations between the expected equity and bond returns and term spreads. However, because there is really one aggregate shock (the endowment shock and the labor income shock are perfectly correlated), the model may produce too much or too little predictability. It is not obvious how to relax the assumption of perfect correlation between the endowment shock and the labor income (habit) shock. Wachter [2001] extends CCHABIT by assuming that the expected consumption growth rate is itself a stochastic process and contains an independent shock.

An "external" version of SHABIT is obtained by assuming that the labor income process is determined by the endowment process rather than the consumption policy. That is,

$$dx = \left[bz^* - \kappa x + \frac{1}{2}\epsilon\beta^2 \right] dt + \beta z^* dB_x.$$

This model retains several key features of SHABIT, and is computationally tractable. While the "external" and "internal" versions of SHABIT share many features, there is an important distinction in their cross-sectional implications of asset returns. In the "external" version, all assets are priced by the dividend growth rate alone – even though there are two state variables (in the same spirit as the Breeden [1979]'s point that all assets are priced by the consumption growth rate alone in Merton [1973]'s model). In contrast, assets are priced by both the dividend and labor income growth rates as two separate risk factors. This distinction also appears in the Euler equation implied by each version of the model: the "internal" version is given by (23), whereas the "external" version is given by (24).

5.4 *A General Endowment Specification*

We can write down an exchange model with a more general specification of the endowment process. Let $\mathbb{D}$ be some suitably defined space of stochastic processes, with the property that for any $z \in \mathbb{D}$, there exists a unique solution to the following bi-variate stochastic differential equation:

$$\begin{aligned} dz^* &= \mu^*(z^*, x, z)dt + \sigma^*, x, z)dB_z, \\ dx &= \nu(z^*, x, z)dt + \psi(z^*, x, z)dB_x, \end{aligned}$$

where z^* is interpreted as the endowment to the investor, and x is the labor income to the worker. A limited participation model obtains if the investor solves

$$\max_{z \in \mathbb{D}} E_0 \left[\int_0^\infty u(z(t), t) dt \right].$$

This setup covers Detemple and Zapatero [1991]'s model (henceforth DZHABIT). These authors (implicitly) assume that the growth rate of the *aggregate* endowment $c^* = z^* + x$ is not affected by z (either directly or indirectly through x). Since x (being an internal habit) is affected by z, it must be the case that z^* is affected by z. This explains why DZHABIT is different from CHABIT (in which z^* is i.i.d.). The failure by Detemple and Zapatero [1991] to state clearly the implicit assumption mentioned above leads to a confusion over whether the state-price density in DZHABIT is even legitimate. Chapman [1998] points out correctly that if the growth rate of $z^* = c^* - x$ is i.i.d., as in CHABIT, then the state-price density in DZHABIT can be negative in some states of the world. But, the state-price density in DZHABIT is not supposed to hold in the first place if the growth rate of z^* is i.i.d.!

Incidentally, if the growth rate of c^* is i.i.d. as in CCHABIT, and the habit process is linear and deterministic as in CHABIT, then DZHABIT is essentially the "internal", continuous-time version of CCHABIT.[19]

6 Conclusion

In this paper, I have shown that habit formation is perhaps not what it is commonly perceived to be: an extension of preference specification for the representative agent. Rather, it captures a dynamic interaction between aggregate financial income and aggregate labor income. I have also shown that existing specifications of consumption habit can be extended to incorporate a stochastic shock, which is interpreted as the labor income shock. As a result of these two innovations, I have shown that a habit formation model can explain the equity premium, equity volatility, and riskfree rate puzzles simultaneously, and give an equilibrium justification for the predictability of equity and bond returns by dividend/price ratio and term spreads. To substantiate these claims, I present an extension of the Breeden-Lucas

[19] See Appendix G. While Campbell and Cochrane [1999] focus on the case of constant riskless rate by imposing a parametric restriction on the endowment process, Detemple and Zapatero [1991] focus on the general case with a stochastic riskless rate.

CCAPM by incorporating a particular form of heterogeneity assumption (related to Constantinides and Duffie [1996] and Heaton and Lucas [1996]) and a particular form of limited participation assumption (related to Mankiw and Zeldes [1991], Saito [1995], and Basak and Cuoco [1998]). The resulting model features a richer technological specification (from the perspective of a production economy) or a richer endowment specification (from the perspective of an exchange economy), but retains standard assumptions of constant relative risk aversion, complete markets, and frictionless trading from the perspective of the marginal investor.

The model can be readily extended to incorporate state-nonseparable preferences (see, e.g., Kreps and Porteus [1978], Epstein and Zin [1989], and Duffie and Epstein [1992]) and consumption durability (see, e.g., Dunn and Singleton [1986], Hindy and fu Huang [1993], and Heaton [1995]), and can be readily tested using standard econometric methods. These theoretical extensions and a rigorous empirical analysis of the model will be elaborated and presented in future work.

Appendix

A Investor's Problem in Terms of Aggregate Wealth

We begin by writing down the model in terms of financial capital $k(t)$ and labor income $x(t)$. We assume that the investor solves

$$\max_{z(t),a(t):t\geq 0} E_0\left[\int_0^\infty u(z(t),t;k(t),x(t))dt\right],$$

subject to

$$\begin{aligned} dk &= [a(\mu_e(k,x)-r)k+rk-z]\,dt + a\sigma_e(k,x)k\,dB_k, \\ dx &= \nu(x,z)\,dt + \psi(x,z)\,dB_x. \end{aligned}$$

Zero net supply of the riskless asset implies $a(t)=1$ in equilibrium.

Suppose that the above model has already been solved under the assumption of complete markets so that the value of human capital $n(t)=n(k,x)$ can be computed in equilibrium. For an arbitrary portfolio policy $a(t)=a(k,x)$ and a consumption policy $z(t)=z(k,x)$, we can define two new state variables: $\hat{w}=a(k,x)(k+n(k,x))+(1-a(k,x))k$ and $\hat{x}=a(k,x)x$, which can be interpreted, respectively, as the total amount of wealth controlled by the investor and the total amount of labor cost borne by the investor when he owns a fraction $a(t)$ of the risky production.

Ito's lemma implies that the dynamics of $\hat{w}$ and $\hat{x}$ can be generically written as

$$\begin{aligned} d\hat{w} &= [\alpha(\mu(\hat{w},\hat{x})-r(\hat{w},\hat{x}))\hat{w}+r(\hat{w},\hat{x})\hat{w}-\hat{z}-\hat{x}]\,dt+\alpha\hat{w}\sigma(\hat{w},\hat{x})dB_w, & (74)\\ d\hat{x} &= \hat{\nu}(\hat{x},\hat{z})dt+\hat{\psi}(\hat{x},\hat{z})dB_x, & (75) \end{aligned}$$

where $\alpha=\frac{a(k,x)(k+n(k,x)))}{\hat{w}}=\alpha(\hat{w},\hat{x})$ is the fraction of total wealth associated with the investment in the risky production and $\hat{z}=a(k,x)z(k,x)+(1-a(k,x))kr(k,x)=\hat{z}(\hat{w},\hat{x})$. α and $\hat{z}$ may be viewed as the portfolio and consumption policies, respectively, in terms of the new state variables $(\hat{w},\hat{x})$. Clearly, $\alpha=1$ if and only if $a=1$.

Replacing (k,x) by $(\hat{w},\hat{x})$, the expected utility can be rewritten as

$$E\left[\int_0^\infty \hat{u}(\hat{z}(t),t;\hat{w},\hat{x})dt\,\middle|\,\hat{w}(0)=\hat{w}_0,\hat{x}(0)=\hat{x}_0\right],$$

where w_0 solves $w_0=k_0+n(k_0,x_0)$ in terms of the initial values of the original state variables, k_0 and x_0. Equations (3)–(6) are recovered by assuming that $\hat{u}(\hat{z}(t),t;\hat{w},\hat{x})$ does not depend on the new state variables $(\hat{w},\hat{x})$, and dropping all hats in the new state variables $(\hat{w},\hat{x})$, the new control variable $\hat{z}$, and the utility function $\hat{u}$.

In equilibrium, $\alpha(t)=a(t)=1$. It follows that $\hat{z}=z$, and $\hat{u}(\hat{z},t;\hat{w},\hat{x})=\hat{u}(z,t;\hat{w},\hat{x})=u(z,t;k,x)$. Thus, $\hat{u}(\hat{z},t;\hat{w},\hat{x})$ is independent of $\hat{w}$ and $\hat{x}$ if and only if $u(z,t;k,x)$ is independent of k and x.

B The Worker's Outside Opportunity, Riskless Rate, and Reservation Value

We assume that the worker always has the option of abstaining from participating in risky production. In the event that a "strike" occurs, the state of the economy switches from (k, x) to $(0, x)$. Consumption of the current labor income represents an investment in the human capital, which generates a future stream of labor income. Therefore, we can think of the per capita labor income as the worker's endowment. In this case, the riskless rate must be equal to the the rate of return for the investment in human capital, and can be determined in equilibrium. To this end, let us consider the following exchange problem: the worker solves

$$\max_{\tilde{x}(t):t\geq 0} \int_0^\infty \tilde{u}(\tilde{x}, t)dt,$$

subject to the budget constraint:

$$\int_0^\infty \tilde{m}(t)\tilde{x}(t)dt \leq \int_0^\infty \tilde{m}(t)x(t)dt,$$

where $\tilde{m}(t)$ is the pricing kernel, and $x(t)$ is the endowment process (the amount of consumption that can be extracted from mother nature without the help of the risky technology).

Suppose that an equilibrium exists (with market clearing condition $\tilde{x} = x$. Then there exists a constant $\lambda > 0$, such that

$$\tilde{u}_x(x, t) = \lambda \tilde{m}(t).$$

For simplicity, assume $\tilde{u}(x, t) = e^{-\tilde{\rho}t}\tilde{u}(x)$. Furthermore, per Assumption 2, the endowment process is given by

$$dx = l_0(x)dt. \tag{76}$$

Thus, the equilibrium riskless rate (noting $d\tilde{m}(t) = -r(t)\tilde{m}(t)dt$) is given by

$$r(x, t) = -\frac{\tilde{u}_{xt}(x, t) + l_0(x)\tilde{u}_{xx}(x, t)}{\tilde{u}_x(x, t)} = \tilde{\rho} - l_0(x)\frac{\tilde{u}''(x)}{\tilde{u}'(x)} = r(x). \tag{77}$$

Thus, the riskless rate depends only x. Furthermore, the functional form is determined entirely by the worker's utility function $\tilde{u}$ and the nature of outside opportunity $(l_0(x))$.

From the investor's perspective, the riskless rate is exogenous, because the worker's behavior when pursuing an outside opportunity has nothing to do with the investor's problem.

In equilibrium, the value of human capital is given by

$$f(x) = \int_0^\infty e^{-\int_0^t r(x(s))ds}x(t)dt,$$

given $x_0 = x$. Alternatively, we can write

$$df(x) = [r(x)f(x) - x]\, dt. \tag{78}$$

Consistency between equations (76) and (78) requires that

$$r(x) = \frac{x + l_0(x)f'(x)}{f(x)}. \tag{79}$$

It follows immediately from equation (11) that

$$\tilde{\rho} + \frac{l_0(x)}{x}\tilde{A}(x) = r(x) = \frac{x + l_0(x)f'(x)}{f(x)}, \tag{80}$$

where $\tilde{A}(x) \equiv -\frac{x\tilde{u}''(x)}{\tilde{u}'(x)}$ is the Arrow-Pratt measure of relative risk aversion. Thus, the function $f(x)$ is intimately related to the utility specification.

The function $f(x)$ will be referred to as a "reservation value", because this is the least that the worker can get. The worker is willing to participate in the risky production, only if the value of human capital, $n(k, x)$, is higher than the reservation value. This is guaranteed by Assumption 5, which states that the worker should not only be compensated properly for bearing the labor income risk, but should also receive an extra incentive for participation. Obviously, when the worker participates in risky production, the expected return to human capital should be different from the riskless rate.

B.1 *Proof of Lemma 2*

To illustrate the close linkage between $f(x)$ and the utility function, let us assume that (i) $l_0 = -\kappa x$, where $\kappa > 0$ is a constant; (ii) $f(0) = 0$; (iii) for $x > 0$, $f'(x) > 0$ and $f''(x) < 0$, so that $A(x) \equiv -\frac{xf''(x)}{f'(x)} > 0$. Under these assumptions, we can show that $A'(x) \geq 0$ implies $r'(x) > 0$ and $\tilde{A}'(x) < 0$ for $\forall x > 0$. This is a stronger result than Lemma 2, thus its proof subsumes the proof for the lemma.

First, equation (80) implies that $r'(x) = -\kappa\tilde{A}'(x)$. Since $\kappa > 0$, it implies that if the riskless rate is increasing (decreasing) in x, the relative risk aversion is decreasing (increasing).

Next, differentiating equation (79) gives $r'(x) = \frac{F_1(x)+\kappa f'(x)F_2(x)}{f(x)^2}$, where $F_1(x) \equiv f(x) - xf'(x)$ and $F_2(x) \equiv f(x)A(x) - F_1(x)$. It is easy to check that (i) $F_1(0) = F_2(0) = 0$; (ii) $F_1'(x) = -f''(x) > 0$, thus $F_1(x) > 0$ for $x > 0$; and (iii) $F_2'(x) = f(x)A'(x) \geq 0$, thus $F_2(x) \geq 0$. It follows that $r'(x) > 0$ and consequently $\tilde{A}'(x) < 0$ for $x > 0$.

In the equity premium literature, a conventional wisdom is that a low elasticity of intertemporal substitution implies a low demand for saving, which implies a high riskless rate (see, e.g., Kocherlakota [1996]). This means that that $r'(x)$ and $\tilde{A}'(x)$ should have the same sign. The above proof gives a counter-example: when the endowment declines over time ($\kappa > 0$), $r'(x)$ and $\tilde{A}'(x)$ have the opposite sign. However, if the endowment is growing ($\kappa < 0$), then then $r'(x)$ and $\tilde{A}'(x)$ do have the same sign.

To reconcile the above result, it must be the case that the conventional wisdom holds only if the endowment is growing. When the endowment is declining, the worker has an urge to borrow rather than to save. A higher intertemporal elasticity of substitution implies a stronger urge to borrow and therefore a higher riskless rate.

C Proof of Lemma 3

Let $V(w,x,t)$ be the indirect utility function. The necessary optimality conditions for the existence of an equilibrium are given by the Hamilton-Jacobian-Bellman equation (see Merton [1969] and Merton [1971]):

$$0 = \max_{\alpha,z}\left[u(z,t) + V_t + \mathcal{A}^{\alpha,z}V\right], \tag{81}$$

where the *controlled* infinitesimal generator $\mathcal{A}^{\alpha,z}$ is given by

$$\begin{aligned}\mathcal{A}^{\alpha,z} &= \left\{\left[\alpha(\mu(w,x) - r(w,x)) + r(w,x)\right]w - z - x\right\}\partial_w + \frac{1}{2}\alpha^2\sigma(w,x)^2w^2\partial_{ww} \\ &+ \nu(x,z)\partial_x + \frac{1}{2}\psi(x,z)^2\partial_{xx} + w\alpha\sigma(w,x)\delta\psi(x,z)\partial_{xw}.\end{aligned}$$

Schematically,

$$\begin{aligned}\frac{\partial\mathcal{A}^{\alpha,z}}{\partial\alpha} &= (\mu(w,x) - r(w,x))w\partial_w + \sigma(w,x)^2\alpha w^2\partial_{ww} + w\sigma(w,x)\delta\psi(x,z)\partial_{xw}, \\ \frac{\partial\mathcal{A}^{\alpha,z}}{\partial z} &= -\frac{\partial}{\partial w} + \nu_z(x,z)\partial_x + \psi(x,z)\psi_z(x,z)\partial_{xx} + w\alpha\sigma(w,x)\delta\psi_z(x,z)\partial_{xw},\end{aligned}$$

from which we can read off the first-order conditions: $\frac{\partial\mathcal{A}^{\alpha,z}}{\partial\alpha}V = 0$, and $u_z + \frac{\partial\mathcal{A}^{\alpha,z}}{\partial z}V = 0$.

From the first-order conditions, we can solve for the optimal portfolio policy:

$$\alpha^* = A(w,x)^{-1}\sigma(w,x)^{-1}\left[\Lambda(w,x) - H(w,x)\delta\psi(x,z)\right], \tag{82}$$

where $\Lambda(w,x) \equiv \sigma(w,x)^{-1}(\mu(w,x) - r(w,x))$, $A(w,x) = -\frac{wV_{ww}}{V_w}$ is the relative risk aversion with respect to wealth, and $H(w,x) = -\frac{V_{wx}}{V_w}$.

Similarly, the optimal financial policy is given by

$$u_z = V_w - \nu_z(x,z)V_x - \psi(x,z)\psi_z(x,z)V_{xx} - w\alpha\sigma(w,x)\delta\psi_z(x,z)V_{xw}. \tag{83}$$

Evaluating the equation (81) at the optimal policies, differentiating both sides with respect to w, and noting that the optimal consumption policy z^* is a function of both w and x, we have

$$0 = u_z z_w^* + V_{wt} + \frac{\partial}{\partial w}\left[\mathcal{A}^{\alpha^*,z^*}V\right]. \tag{84}$$

Evaluating the last term explicitly and using equation (83), we have

$$\begin{aligned}\frac{\partial}{\partial w}\left[\mathcal{A}^{\alpha^*,z^*}V\right] &= \mathcal{A}^{\alpha^*,z^*}V_w - u_z z_w^* + \frac{\partial}{\partial w}\left\{\left[\alpha^*(\mu - r) + r\right]w - z^* - x\right\}V_w \\ &+ \frac{\partial}{\partial w}\left[w\alpha^*\sigma\delta\right]\psi V_{wx} + \frac{\partial}{\partial w}\left[\frac{\alpha^*\sigma^2\alpha^*w^2}{2}\right]V_{ww},\end{aligned} \tag{85}$$

Combining equations (84) and (85) yields

$$\frac{V_{wt} + \mathcal{A}^{\alpha^*,z^*}V_w}{V_w} + r = -wr_w - \mathcal{G}\Lambda_w,$$

where $\mathcal{G}(w,x) \equiv w\sigma'\alpha^* = wA^{-1}\left(\Lambda - H\delta\psi\right)$. Thus,

$$\frac{V_{wt} + \mathcal{A}^{\alpha^*,z^*}V_w}{V_w} = -r, \text{ if and only if, } wr_w + \mathcal{G}\Lambda_w = 0. \tag{86}$$

D Proof of Lemma 4

Let β be the share of aggregate wealth in the contingent claim. Then, the wealth constraint generalizes to

$$dw = [\alpha(\mu - r)w + \beta(\mu_P - r)w + rw - z - x - D]\, dt + \alpha w \sigma dB_w + \beta w \sigma_P dB_p.$$

The Hamilton-Jacobian-Bellman equation is generalized accordingly. It is easy to check that the first-order condition with respect to β is given by

$$(\mu_P - r)wV_w + \left[\beta w^2 \sigma_P^2 + \alpha w^2 \sigma \sigma_P \delta_{pw}\right] V_{ww} + w\sigma_P \psi \delta_{px} V_{wx} = 0, \tag{87}$$

and the first-order conditions with respect to α and z are the same as before under the market clearing condition $\beta = 0$. Evaluating equation (87) under the market clearing conditions $\alpha = 1$ and $\beta = 0$, we obtain

$$\mu_P - r = \sigma_P \left[\delta_{pw} A\sigma + \delta_{px} H\psi\right]. \tag{88}$$

Let $m(t)$ be a pricing kernel. By definition, any security with dividend $D(t)$ and ex-dividend price $P(t)$, $P(t)m(t) + \int_0^t D(s)m(s)ds$ must be a martingale. Let $B_x = \delta B_w + \sqrt{1-\delta^2} B_\perp$. Then, without loss of generality, we can write

$$\frac{dm(t)}{m(t)} = -rdt - \Lambda dB_w - \Lambda_\perp dB_\perp,$$

where the drift of $m(t)$ is determined by applying the the defining property of the pricing kernel to a real deposit account, with zero dividend, and price process $e^{\int_0^t r(s)ds}$; and the market price of risk for B_w is determined by applying the defining property to the aggregate wealth process (dividend $z + x$ and price w). According to equation (82), the equilibrium Sharpe ratio is given by

$$\Lambda = A\sigma + \delta H\psi. \tag{89}$$

Applying the defining property to the contingent claim, we have

$$\mu_P - r = -\mathrm{cov}\left(\frac{dP(t) + D(t)dt}{P(t)}, \frac{dm(t)}{m(t)}\right) = \sigma_P \left[\delta_{pw}\Lambda + \delta_{p\perp}\Lambda_\perp\right], \tag{90}$$

where $\delta_{p\perp} = \frac{\delta_{px} - \delta\delta_{pw}}{\sqrt{1-\delta^2}} \neq 0$. Comparing equations (88) and (90), and using equation (89), we conclude that $\Lambda_\perp = \sqrt{1-\delta^2} H\psi$. Thus, the pricing kernel is given by

$$\frac{dm(t)}{m(t)} = -rdt - \Lambda dB_w - \Lambda_\perp dB_\perp = -rdt - A\sigma dB_w - H\psi dB_x.$$

Lemma 3 and Ito's lemma imply that, in equilibrium,

$$\frac{dV_w}{V_w} = -rdt + \frac{wV_{ww}}{V_w}\sigma dB_w + \frac{V_{wx}}{V_w}\psi dB_x = -rdt - A\sigma dB_w - H\psi dB_x.$$

Thus, $m(t) = V_w$ (up to a scaling constant). It is clear that $m(t)$ is unique.

E Proof of Proposition 1

For simplicity, I temporarily restrict the representative investor from investment in the riskless asset. The more general case is addressed in Appendix F, where the optimal portfolio rule is also given.

The representative investor solves $\max_{z(t):0\leq t\leq\infty} E_0\left[\int_0^\infty e^{-\rho t}\frac{z(t)^\gamma}{\gamma}dt\right]$, subject to: (i) $dw = (\mu(w,x)w - z - x)dt + \sigma(w,x)wdB_w$, $w(0) = w_0$; and (ii) $dx = (l_0(x) + l_1(x)z + l_2(x)z^2)dt + l_3(x)zdB_x$, $x(0) = x_0$. This is a standard intertemporal optimization problem, with a bivariate state vector $(w(t), x(t))$, and stochastic control $z(t)$.

The structure of the problem suggests a conjecture as follows:

$$V(t,w,x) \equiv \max_{z(s):t\leq s\leq\infty} E\left[\int_t^\infty e^{-\rho s}\frac{z(s)^\gamma}{\gamma}ds\,\middle|\, w(t)=w, x(t)=x\right] = e^{-\rho t}J(w,x),$$

where $J(w,x)$ is the (time-invariant) indirect utility function. The Hamilton-Jacobi-Bellman equation then becomes

$$\begin{aligned}0 = \max_z\ & \frac{z^\gamma}{\gamma} - \rho J + [\mu(w,x)w - z - x]\,J_w + \frac{1}{2}\sigma(w,x)^2w^2J_{ww} \\ & + \left[l_0(x) + l_1(x)z + l_2(x)z^2\right]J_x + \frac{1}{2}l_3(x)^2z^2J_{xx} + \delta l_3(x)\sigma(w,x)wzJ_{wx}.\end{aligned} \tag{91}$$

Substituting equations (20), (21), and (22) into (91), we obtain

$$\begin{aligned}0 = \max_z\ & \frac{z^\gamma}{\gamma} - \rho J + \{[r(x) + \Lambda(x)g(x)]\,[w - f(x)] - z\}\,J_w + \frac{1}{2}g(x)^2\,[w - f(x)]^2\,J_{ww} \\ & + \left[l_1(x)z + l_2(x)z^2\right]J_x + \frac{1}{2}l_3(x)^2z^2J_{xx} + \delta l_3(x)g(x)\,[w - f(x)]\,zJ_{wx} \\ & + l_0(x)\,[f'(x)J_w + J_x]\,.\end{aligned} \tag{92}$$

Following the standard practice in dynamic programming, I guess a solution first, and then verify that it satisfies all necessary and sufficient optimality conditions. The structure of equation (92) suggests the following conjecture: The indirect utility function takes the form:

$$J(w,x) = \phi\frac{[w - f(x)]^\gamma}{\gamma}, \tag{93}$$

for some positive constant ϕ.[20] Under this conjecture, $J_x = -f'(x)J_w$, so that the last term (one of the inhomogeneous terms) in equation (92) vanishes. It follows immediately that the optimal financial policy is given by

$$z^* = h(x)[w - f(x)] \tag{94}$$

for some real function $h(x)$.

To verify that (93) and (94) constitute a solution, we need to check the *optimality conditions*:

[20] Since the felicity function has the same sign as γ, $J(w,x)$ must also have the same sign as γ. Consequently, we must have $\phi > 0$.

1. Under $z = z^*$, the first-order condition for z is satisfied;

2. Equation (91) is satisfied when the right-hand side is evaluated at $z = z^*$;

3. The objective function in (91) is concave at $z = z^*$;

4. The transversality condition is satisfied at $z = z^*$.

To verify that the solution is unique, all I need to show is that the objective function in (91) is globally concave in the admissible state space.

Before proceeding to check the optimality conditions, we note that

$$\begin{aligned}J_w &= \frac{\gamma}{w - f(x)}J, \quad J_{ww} = \frac{\gamma(\gamma-1)}{[w-f(x)]^2}J, \quad J_{wx} = \frac{\gamma(1-\gamma)f'(x)}{[w-f(x)]^2}J, \\ J_x &= -\frac{\gamma f'(x)}{w-f(x)}J, \quad J_{xx} = -\frac{\gamma f''(x)}{w-f(x)}J + \frac{\gamma(\gamma-1)f'(x)^2}{[w-f(x)]^2}J.\end{aligned} \tag{95}$$

Thus, equation (13) implies

$$l_2(x)J_x + \frac{1}{2}l_3(x)^2 J_{xx} = \frac{\gamma(\gamma-1)}{2}\frac{l_3(x)^2 f'(x)^2}{[w-f(x)]^2}J. \tag{96}$$

Thus, the right hand side of equation (92) is homogenous in $[w - f(x)]$.

First-Order Condition:

The first-order condition in the Hamilton-Jacobi-Bellman equation is given by

$$\begin{aligned}0 &= z^{\gamma-1} - J_w + l_1(x)J_x + 2z\left[l_2(x)J_x + \frac{1}{2}l_3(x)^2 J_{xx}\right] + \delta l_3(x)\sigma(w,x)wJ_{wx} \\ &= z^{\gamma-1} - J_w + l_1(x)J_x + \gamma(\gamma-1)l_3(x)^2 f'(x)^2 \frac{z}{[w-f(x)]^2}J + \delta l_3(x)g(x)(w-f(x))J_{wx},\end{aligned}$$

where the second equality holds because of equations (96) and (22). Substituting (93) and (94) into this equation, we obtain equation (31).

Bellman Optimality Principle:

Evaluating the right-hand side of equation (92) at $z = z^*$, and making use of equations (96), we obtain

$$\begin{aligned}0 = &\frac{h(x)^\gamma}{\gamma\phi} - \frac{\rho}{\gamma} + [\Lambda(x)g(x) + r(x)] - h(x)(1 + l_1(x)f'(x)) - \frac{1-\gamma}{2}g(x)^2 \\ &- \frac{1-\gamma}{2}l_3(x)^2 h(x)^2 f'(x)^2 + (1-\gamma)\delta l_3(x)g(x)h(x)f'(x).\end{aligned} \tag{97}$$

The terms associated with $[w - f(x)]$ drop out due to the homothetic structure. Equation (32) follows immediately from equations (31) and (97).

Global Concavity:

The second-order derivative of the objective function in (91) is given by

$$(\gamma-1)z^{\gamma-2}+2\left[l_2(x)J_x+\frac{1}{2}l_3(x)^2J_{xx}\right]=-(1-\gamma)\left[z^{\gamma-2}+\phi l_3(x)^2f'(x)^2[w-f(x)]^{\gamma-2}\right]<0,$$

where the inequality comes from the fact that $\gamma<0$ and $\phi>0$. Thus, the objective function is globally concave. It follows that z^* is not only optimal, but also unique.

Transversality Condition:

Applying Ito's lemma, it is straightforward to show that, under the optimal policy,

$$\frac{d(w-f(x))}{w-f(x)}=\mu_v dt+\sigma_{vw}dB_w+\sigma_{v\perp}dB_\perp, \tag{98}$$

where $B_x=\delta B_w+\sqrt{1-\delta^2}B_\perp$, $B_\perp$ is orthogonal to B_w, and

$$\begin{aligned}\sigma_{vw}&=g(x)-\delta l_3(x)h(x)f'(x),\quad \sigma_{v\perp}=-\sqrt{1-\delta^2}l_3(x)h(x)f'(x),\\ \mu_v&=\Lambda(x)g(x)+r(x)-(1+l_1(x)f'(x))h(x)=\frac{\rho}{\gamma}-\frac{h(x)^\gamma}{\gamma\phi}+\frac{1-\gamma}{2}\left(\sigma_{vw}^2+\sigma_{v\perp}^2\right),\end{aligned} \tag{99}$$

where the last equality comes from equation (33).

Applying Ito's lemma to $|V(t)|=e^{-\rho t}|J(w(t),x(t))|=\phi e^{-\rho t}\frac{(w(t)-f(x(t)))^\gamma}{|\gamma|}$ yields

$$\begin{aligned}\frac{d|V(t)|}{|V(t)|}&=\left[-\rho+\gamma\mu_v-\frac{\gamma(1-\gamma)}{2}\left(\sigma_{vw}^2+\sigma_{v\perp}^2\right)\right]dt+\gamma\sigma_{vw}dB_w+\gamma\sigma_{v\perp}dB_\perp\\ &=-\frac{h(x)^\gamma}{\phi}dt+\gamma\sigma_{vw}dB_w+\gamma\sigma_{v\perp}dB_\perp.\end{aligned}$$

Since $\phi>0$, $|V(t)|$ is a supermartingale. Since $w_0-f(x_0)>0$, we have $\lim_{t\to\infty}E_0[|V(t)|]\le |V(0)|=|J(w_0,x_0)|<\infty$.

F Proof of Proposition 2

Let $\alpha(t)$ be the share of aggregate wealth invested in the risky technology, and $1-\alpha(t)$ be the riskless asset. The representative investor solves $\max_{z(t),\alpha(t):0\le t\le\infty}E_0\left[\int_0^\infty e^{-\rho t}\frac{z(t)^\gamma}{\gamma}dt\right]$, subject to (i) $dw=[(\alpha(\mu(w,x)-r(x))+r(x))w-z-x]\,dt+\alpha\sigma(w,x)wdB_w$, $w(0)=w_0$; and (ii) $dx=(l_0(x)+l_1(x)z+l_2(x)z^2)dt+l_3(x)zdB_x$, $x(0)=x_0$.

The Hamilton-Jacobi-Bellman equation is given by

$$\begin{aligned}0=\max_{z,\alpha}\ &\frac{z^\gamma}{\gamma}-\rho J+[\alpha(\mu(w,x)-r)w+rw-z-x]\,J_w+\frac{1}{2}\alpha^2\sigma(w,x)^2w^2J_{ww}\\ &+(l_0(x)+l_1(x)z+l_2(x)z^2)J_x+\frac{1}{2}l_3(x)^2z^2J_{xx}+\delta l_3(x)\alpha\sigma(w,x)wzJ_{wx}.\end{aligned} \tag{100}$$

The first-order condition with respect to α is given by

$$0=(\mu(w,x)-r(x))wJ_w+\alpha^*\sigma(w,x)^2w^2J_{ww}+\delta l_3(x)\sigma(w,x)wzJ_{wx}.$$

Solving, we obtain the optimal portfolio allocation rule:

$$\alpha^* = \frac{\Lambda(x) + (1-\gamma)\delta l_3(x)h(x)f'(x)}{(1-\gamma)\sigma(w,x)} \frac{w - f(x)}{w}. \tag{101}$$

It is easy to show that, if equations (21)–(22) hold, then the the indirect utility function and the optimal financial policy z^* remain the same as in Appendix E. It is also easy to check that global concavity obtains, so the optimal policies (α^*, z^*) constitute a unique solution to the model. Evaluating equation (100) at the optimal policies imposes a restriction on $h(x)$, $g(x)$, and $\Lambda(x)$.

What remains to be shown that when $\mu(w,x)$ and $\sigma(w,x)$ are given by equations (21)–(22) under a suitable choice of $g(x)$ and $\Lambda(x)$, the market for the riskless asset clears. This is straightforward. Setting $\alpha^* = 1$, equation (101) and equation (22) imply equation (33). Moreover, at $\alpha^* = 1$, equation (100) evaluated at z^* reduces to equation (97). Equation (34) follows by combining equations (31) and (97). Thus, the unknown function $g(x)$ is jointly determined with $h(x)$ by equations (31) and (34), whereas the unknown function $\Lambda(x)$ is determined by equation (33) once $g(x)$ is determined.

F.1 *Proof of Corollary 1*

In equilibrium,

$$\begin{aligned}
dw &= (\mu(w,x)w - z - x)dt + \sigma(w,x)w dB_w \\
&= [(\Lambda(x)g(x) + r(x)) - h(x)]\,[w - f(x)]dt + (rf(x) - x)dt + g(x)[w - f(x)]dB_w, \\
df(x) &= \left[f'(x)(l_0(x) + l_1(x)z + l_2(x)z^2) + f''(x)\frac{l_3(x)^2}{2}z^2\right]dt + f'(x)l_3(x)z dB_x \\
&= f'(x)\left[(l_0(x) + l_1(x)z)dt + l_3(x)z dB_x\right].
\end{aligned}$$

Taking the difference of the two equations, and making use of equation (11), we obtain

$$\begin{aligned}
\frac{d[w - f(x)]}{w - f(x)} &= [\Lambda(x)g(x) + r(x) - (1 + l_1(x)f'(x))h(x)]\,dt + g(x)dB_w \\
&\quad - l_3(x)h(x)f'(x)dB_x.
\end{aligned}$$

Using equations (31) and (34), we can show that

$$\begin{aligned}
&\Lambda(x)g(x) + r(x) - (1 + l_1(x)f'(x))h(x) \\
= &\frac{r(x) - \rho}{1-\gamma} + \frac{2-\gamma}{2(1-\gamma)^2}\Lambda(x)^2 + \frac{2-\gamma}{2}(1-\delta^2)l_3(x)^2h(x)^2f'(x)^2.
\end{aligned}$$

F.2 *Derivation of Equations (44) and (45)*

In equilibrium, the state process can be written as

$$\begin{aligned}
dw &= [\mu(w,x)w - z^* - x]\,dt + \sigma(w,x)w dB_w, \\
dx &= \mu_x dt + \sigma_x(\delta dB_w + \sqrt{1-\delta^2}dB_\perp),
\end{aligned}$$

where $\mu(w,x) = r(x) + \Lambda(x)\sigma(w,x)$, $\sigma(w,x) = g(x)\frac{w-f(x)}{w}$, $\mu_x = l_0(x) + l_1(x)z^* + l_2(x)z^{*2}$, $\sigma_x = l_3(x)z^*$, $z^* = h(x)(w - f(x))$. Using Ito's lemma, I can write

$$dk = \mu_k dt + \left(\sigma_w w \frac{\partial k}{\partial w} + \delta\sigma_x \frac{\partial k}{\partial x}\right) dB_w + \sqrt{1-\delta^2}\sigma_x \frac{\partial k}{\partial x} dB_\perp,$$

where

$$\mu_k = \left[\mu(w,x)w - z^* - x\right] \frac{\partial k}{\partial w} + \frac{\sigma(w,x)^2 w^2}{2}\frac{\partial^2 k}{\partial w^2} + \mu_x \frac{\partial k}{\partial x} + \frac{\sigma_x^2}{2}\frac{\partial^2 k}{\partial x^2} + \delta\sigma_x\sigma(w,x)w\frac{\partial^2 k}{\partial w \partial x}.$$

Since $k(t)m(t) + \int_0^t z^*(s)m(s)ds$ must be a martingale, we must have

$$\mu_k - r(x)k + z^* = \Lambda(x)\sigma(w,x)w\frac{\partial k}{\partial w} + \left[\delta\Lambda(x) + \sqrt{1-\delta^2}\Lambda_\perp(x)\right]\sigma_x\frac{\partial k}{\partial x}.$$

Equation (44) follows immediately by combining the last two equations. Equation (45) is derived similarly.

G Continuous-Time Limit of Campbell and Cochrane

In the continuous-time limit, the consumption and log surplus consumption ratio in Campbell and Cochrane [1999] can be written as

$$\frac{dC}{C} = gdt + vdB, \tag{102}$$

$$ds = h(\bar{s} - s)dt + \lambda(s)vdB, \tag{103}$$

where C is the aggregate consumption, $s = \log S$ is the log surplus consumption ratio, $S = \frac{C-X}{C}$ is the surplus consumption ratio, and X is the consumption habit. g and v are assumed to be constant. $\lambda(s) = \frac{\mathcal{B}}{S}\sqrt{1-2(\bar{s}-s)} - 1$, if $s \le s_{max}$, and $\lambda(s) = 0$ if $s \ge s_{max}$, where $\bar{S}$ is the steady-state mean of S, $\bar{s}$ is the steady state mean of s, and $s_{max} = \bar{s} - 0.5$.

Note that $\lambda(s)$ may be viewed as the linearization of $\frac{\mathcal{B}}{S} - 1$:

$$\frac{1}{S} \approx \frac{1}{\bar{S}}(1 - (s - \bar{s})) \approx \frac{1}{\bar{S}}\sqrt{1 - 2(s - \bar{s})},$$

where $\bar{S} = e^{\bar{s} - \frac{1}{2}var(s)}$. Thus, I will replace $\lambda(s)$ by $\frac{\mathcal{B}}{S} - 1$,

Furthermore, $\bar{s} - s$ may be viewed as a linearization of $\frac{\bar{S}-S}{S}$ and accordingly equation (103) may be viewed as a linearization of (ignoring a Jensen's term proportional to v^2)

$$dS = h(\bar{S} - S)dt + (\mathcal{B} - S)vdB. \tag{104}$$

If we take equations (102) and (104) as the primitive assumptions, and apply Ito's lemma to $X = C(1-S)$, we obtain

$$dX = \left[gX + bC - aX\right]dt + v(1-\mathcal{B})CdB, \tag{105}$$

where $a = h + v^2$ and $b = h(1 - \bar{S}) + v^2(1 - \mathcal{B}) = a - h\bar{S} - v^2\mathcal{B}$.

Define $c = Ce^{-gt}$, $x = Xe^{-gt}$, and $z = Ze^{-gt}$, we obtain

$$\begin{aligned} \frac{dc}{c} &= vdB, \\ dx &= (bc - ax)dt + v(1 - \mathcal{B})cdB. \end{aligned}$$

It is hard to believe that taking the continuous-time limit and applying the "reverse approximation" to equation (103) make any material difference. Therefore, when $\mathcal{B} = 1$, the habit process implied by CC is essentially identical to that of CHABIT.

H Proof of Corollary 3

When $\hat{l}(x) = \hat{l}_3(x) = 1$, $l_1(x) = b$, $l_3(x) = \beta$.

First, in the limit $\beta \to 0$, equation (31) becomes $\frac{h(x)^{\gamma-1}}{\phi} = 1 + bf'(x)$. Taking the derivative with respect to x gives $(\gamma - 1)\frac{h(x)^{\gamma-2}}{\phi}h'(x) = bf''(x) < 0$. Since $\gamma - 1 < 0$, it follows that $h'(x) > 0$.

Next, in the limit $\beta \to 0$, equations (33) and (34) imply

$$2\Lambda(x)\Lambda'(x) = -2(1-\gamma)\left[r'(x) + (1-\gamma)\frac{h(x)^{\gamma-1}}{\phi}h'(x)\right] < 0.$$

Since $\Lambda(x) \geq 0$, $\Lambda'(x) < 0$.

Finally, in the limit $\beta \to 0$, $g(x) = \frac{\Lambda(x)}{1-\gamma}$. Thus, $g'(x) < 0$.

References

Abel, A. B. (1990). Asset prices under habit formation and catching up with the Joneses. *American Economic Review Papers and Proceedings 80*, 38–42.

Basak, S. and D. Cuoco (1998). An equilibrium model with restricted stock market participation. *Review of Financial Studies 11*, 309–341.

Breeden, D. (1979). An intertemporal asset pricing model with stochastic consumption and investment opportunities. *Journal of Financial Economics 7*, 265–296.

Campbell, J. and J. Cochrane (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy 107*(2), 205–251.

Campbell, J. Y. (1999). *Handbook of Macroeconomics*, Volume 1, Chapter 19. North-Holland, Amsterdam.

Chapman, D. A. (1998, September). Notes and comments: Habit formation and aggregate consumption. *Econometrica 66*(5), 1223–1230.

Constantinides, G. (1990). Habit formation: A resolution of the equity premium puzzle. *Journal of Political Economy 98*(3), 519–543.

Constantinides, G. and J. D. Duffie (1996). Asset pricing with heterogeneous consumers. *Journal of Political Economy 104*, 219–240.

Dai, Q. and K. J. Singleton (2001). Expectations puzzles, time-varying risk premia, and affine models of the term structure. Accepted for publication, Journal of Financial Economics.

Detemple, J. B. and F. Zapatero (1991, November). Asset prices in an exchange economy with habit formation. *Econometrica 59*(6), 1633–1657.

Duffie, D. (1996). *Dynamic Asset Pricing Theory, 2nd edition.* Princeton University Press, Princeton, N.J.

Duffie, J. D. and L. Epstein (1992). Stochastic differential utility. *Econometrica 60*(2), 353–394.

Duffie, J. D. and C. Skiadas (1994). Continuous-time security pricing: A utility gradient approach. *Journal of Mathematical Economics 23*, 107–131.

Dunn, K. B. and K. J. Singleton (1986). Modeling the term structure of interest rates under non-separable utility and durability of goods. *Journal of Financial Economics*, 27–55.

Epstein, L. G. and S. E. Zin (1989). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica 57*, 937–969.

Fama, E. (1984). The information in the term structure. *Journal of Financial Economics 13*, 509–528.

Fama, E. and R. Bliss (1987). The information in long-maturity forward rates. *American Economic Review 77*(4), 680–692.

Fama, E. and K. French (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics 25*, 23–49.

Hansen, L. and K. Singleton (1982). Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica 50*, 1269–1286.

Hansen, L. P. and R. Jagannathan (1991). Implications of security market data for models of dynamic economies. *Journal of Political Economy 99*(2), 225–262.

Heaton, J. (1995, May). An empirical investigation of asset pricing with temporically dependent preference specifications. *Econometrica*.

Heaton, J. and D. Lucas (1996). Evaluating the effects of incomplete markets on risk sharing and asset pricing. *Journal of Political Economy 104*(3), 443–487.

Hindy, A. and C. fu Huang (1993). Optimal consumption and portfolio rules with durability and local substitution. *Econometrica 61*(1), 84–121.

Ibbotson, R. G. and G. P. Brinson (1987). *Gaining he Performance Advantage: Investment Markets*. McGraw-Hill, Inc., New York.

Klein, L. R. (1950). *Economic Fluctuations in the United States, 1921–1941, Cowles Commission for Research in Economics, Monograph No. 11.* John Wiley & Sons, New York.

Kocherlakota, N. R. (1996). The equity premium: It's still a puzzle. *Journal of Political Literature 34*(1), 42–71.

Kreps, D. and E. Porteus (1978). Temporal resolution of uncertainty and dynamic choice theory. *Econometrica 46*, 185–200.

Kugler, P. (1997). Central bank policy reaction and expectations hypothesis of the term structure. *International Journal of Financial Economics 2*, 217–224.

Lucas, R. (1978). Asset prices in an exchange economy. *Econometrica 46*, 1429–1445.

Mankiw, N. G. and S. P. Zeldes (1991). The consumption of stockholders and nonstockholders. *Journal of Financial Economics 29*, 97–112.

McCallum, B. T. (1994). Monetary policy and the term structure of interest rates. NBER working paper No. 4938.

Mehra, R. and E. C. Prescott (1985, March). The equity premium: A puzzle. *Journal of Monetary Economics 15*, 145–161.

Merton, R. C. (1969). Lifetime portfolio selection under uncertainty: The continuous-time case. *Review of Economics and Statistics 51*, 247–257.

Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory 3*, 373–413.

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica 41*(5), 867–887.

Saito, M. (1995). Limited market participation and asset pricing. Manuscript, University of British Columbia.

Sundaresan, S. M. (1989). Intertemporally dependent preferences and the volatility of consumption and wealth. *Review of Financial Studies 2*, 73–88.

Wachter, J. (2001). Habit formation and returns on bonds and stocks. Working paper, NYU.

Weil, P. (1989). The equity premium puzzle and the riskfree rate puzzle. *Journal of Monetary Economics 24*, 401–421.