

From Equity Premium Puzzle to Expectations Puzzle: A General Equilibrium Production Economy with Stochastic Habit Formation*

Qiang Dai

First Draft: May 22, 2000

Abstract

This paper develops a general equilibrium model for a representative agent, production economy with stochastic internal habit formation. The model describes a scale-independent economy, with a unique stochastic investment opportunity set. Local correlation between the stochastic interest rate and time-varying market price of risk can be determined endogenously and leads to correct predictions on the sign and magnitude of several major empirical puzzles in both equity and bond markets.

In the empirical part of the paper, we calibrate our model, simultaneously, to the equity premium puzzle, the riskfree rate puzzle, and the expectations puzzle, and show that the three puzzles are completely resolved under reasonable parameter values.

Thus, we establish, conclusively, the inextricable link between the equity and bond markets, both theoretically and empirically.

Our model subsumes the internal habit formation models of Sundaresan (1989) and Constantinides (1990), and, perhaps somewhat surprisingly, the external habit formation model of Campbell and Cochrane (1999).

*The author is at the Stern School of Business, New York University, and can be reached at 212-998-0358 or qdai@stern.nyu.edu. I would like to thank Yakov Amihud, Stephen Brown, John Campbell, Darrell Duffie, Ned Elton, Marti Gruber, Alex Shapiro, Costas Skiadas, Suresh Sundaresan, and Robert Whitelaw, and especially Ken Singleton, Jun Liu, and Anthony Lynch for their helpful comments. All remaining errors are mine.

Contents

1	Introduction	2
2	Stochastic Habit Formation Model with Incomplete Market	7
2.1	Production Technology	7
2.2	Consumption Habit and Time-Nonseparable Preference	8
2.3	Detrending and Rescaling the Economy	9
2.4	Consumption Choice and Portfolio Selection	11
2.5	Admissible State Space and Transversality Condition	13
2.6	Equilibrium Consideration	14
3	Equilibrium Properties	15
3.1	Correlation Between Riskless Rate and Sharpe Ratio	15
3.2	A Consumption Based Real Riskless Rate Model	16
3.3	Surplus Wealth and Consumption Processes	18
3.4	Surplus Consumption Ratio	19
3.5	Steady State Means	21
4	Further Refinements	22
4.1	Stochastic Habit Formation Model with Complete Market	23
4.2	A Nominal Economy with Real Inflation Risk	25
4.3	An Additional Source of Riskless Rate Volatility	27
5	Explaining Equity Market Puzzles	28
5.1	Equity Premium Puzzle	28
5.2	Stock Market Volatility Puzzle	29
5.3	Riskfree Rate Puzzle	31
5.4	Calibrating to the Equity Return and Consumption Data	31
5.4.1	Sample Observations	32
5.4.2	Calibration	32
6	Term Structure of Interest Rates	35
6.1	Linear Approximation of Term Structure Dynamics	38
6.2	Inflation Consideration	40
6.3	Expectations Puzzle	41
7	Reconciling Equity Market and Expectations Puzzles	43
7.1	An Identification Problem	45
7.2	Aggregate Stock Market \neq Aggregate Wealth	46
7.3	One More Loose End	47
8	Comparative Statics Under Calibrated Parameter Values	48
8.1	Exact Solution under Calibrated Parameter Values	49
8.2	Economic Properties Seen Through Simulation	49
8.3	Discussion and Future Research Directions	51
9	Conclusion	52

1 Introduction

Over the past two decades or so, two asset pricing puzzles have been documented and stand prominently as two major road blocks toward a full understanding of the financial markets. The *equity premium puzzle* was documented by Mehra and Prescott (1985), who show that the equity premium is too large to be explained by traditional asset pricing models with time-separable preferences at reasonable values of the risk aversion parameter. The tension between empirical observations and traditional asset pricing models was visualized elegantly by Hansen and Jagannathan (1991) in terms of the mean-variance bounds on model implied pricing kernels. The *expectations puzzle* was documented by Campbell and Shiller (1991), who show that the expectations hypothesis, which implies that the slope of the yield curve should forecast short term changes of long term yields, is inconsistent with empirical observations.

In addition to the empirical regularities associated with the individual asset markets, Fama and French (1989) also documented a possible connection between the two markets: *predictability of the equity premium* by the term spread. Like the equity premium and expectations puzzles, this predictability result appears to be fairly robust. An extensive literature has subsequently elaborated on the predictability of equity return and volatility by additional variables, and on the implications of return and volatility predictability for investment decisions.

Among many theories put forward to explain the equity premium puzzle,¹ habit formation models have emerged as a strong candidate. Sundaresan (1989), Constantinides (1990), Abel (1990), and Campbell and Cochrane (1999) show that a representative agent habit formation model is capable of explaining the equity premium puzzle with reasonable values of the agent's coefficient of relative risk aversion. Habit formation models achieve this by breaking time-separability, thereby inserting a wedge between the relative risk aversion and intertemporal rate of substitution. All of these models, however, are completely silent on the expectations puzzle and on the long-term predictability of equity premium by term spread – all of them focus on an economy with constant riskless rate and constant risk premium.

The literature on expectations puzzle, on the other hand, has accumulated strong evidences that the risk premium on bond returns are time-varying and predictable by the term spread (see Fama (1976), Fama (1984b), Fama (1984a), Fama and Bliss (1987), and Backus et al. (1997).) Recently, Dai and Singleton (2000) show that the expectations puzzle can be explained if the forward premium is a linear function of the forward-spot spread. Furthermore, the requisite behavior of the forward premium can be generated by a one-factor Gaussian model with market price of risk being a linear function of the short rate with positive linear coefficient (the short rate mean reverts faster under the risk neutral measure than under the physical measure.)

The primary objective of this paper is to construct an economic theory capable of explaining key empirical regularities in both the equity and bond markets, *simultaneously*. We demonstrate conclusively that this ambitious and seemingly difficult goal can be achieved with a general equilibrium model with stochastic habit formation. Two key ingredients to the empirical success of our theory, to be developed fully in the paper, are: (1) time-nonseparable

¹See Campbell (2000) and Sundaresan (2000) for the most recent surveys of the literature.

preference induced by habit formation; and (2) stochastic evolution of habit formation. The first ingredient was captured, to a large extent, by the internal habit formation model of Sundaresan (1989) and Constantinides (1990). The second ingredient was implicitly present in Campbell and Cochrane (1999), but its theoretical and empirical implications were underdeveloped due to the exogenous nature of their external habit formation model and their deliberate focus on an economy with constant riskless rate and risk premium. Our theory integrates these two ingredients tightly in a general equilibrium model of a representative agent production economy. We derive the theoretical and empirical implications of our model by characterizing explicitly the equilibrium solution of our model and calibrating the equilibrium solution to both the equity market puzzles and the expectations puzzle.

In the rest of this introduction, we trace out, in detail, the motivations for key modeling assumptions, and give a preview of key theoretical and empirical results.

Unlike previous studies of habit formation models, we explicitly require that a non-trivial term structure dynamics must be an integral part of the model. To this end, we begin by laying out important features of the term structure dynamics our model must capture:

1. Mean reversion of riskless rate.
2. Local interest rate risk, or local uncertainty.
3. Non-zero correlation between interest rate shock and the market shock.
4. Non-zero and state-dependent market price of interest rate risk.

In any model with habit formation, there are at least two controlled state variables: the aggregate wealth and the consumption habit. If the riskless rate depends on wealth, then the riskless rate will be too volatile.² Thus, we seek to construct a model in which the riskless rate will depend only on the consumption habit. This implies that, at the minimum, the habit process must have the following features:

1. By definition, it must depend on the history of past consumption. This is how a standard power utility function is modified to accommodate a time-nonseparable preference in habit formation models.

In an *external* habit formation model, an agent's habit does not depend on its own consumption history, but rather on the consumption history of other agents (as in Abel (1990)'s "catching up with the Joneses" theory) or the history of aggregate consumption (as in Campbell and Cochrane (1999)). An external habit formation model usually starts with the consumption and portfolio selection problem faced by an individual agent whose optimal decisions do not affect the external habit. In the end, when the model is taken to the aggregate data, some form of aggregation must occur. However, it is not always clear in this literature how the aggregation process works.

In an *internal* habit formation model, an agent's habit depends on its own past consumption history. This can be true for either an individual agent or a representative agent. Thus, the aggregation problem does not exist. Dunn and Singleton (1986) were

²See Campbell (2000).

among the first to study the link between aggregate consumption and interest rate behavior using a time-nonseparable utility induced by service flow of durable goods, which can be interpreted as an internal habit formation model. Sundaresan (1989) and Constantinides (1990) were among the first to characterize an internal habit formation model in a continuous-time setting, and to link the model to the equity premium puzzle. Abel (1990)'s model also allows for internal habit formation under suitable parameterizations. However, his model is an exchange economy in discrete-time, and the consumption habit enters the utility function multiplicatively. The model we develop in this paper is a production economy, in continuous-time, and the consumption habit enters additively.

Since interest rates are intrinsically aggregate variables, it seems sensible to assume that the habit formation process for a representative agent should be *internal*, rather than *external*. Thus, the habit formation process in our model will be internal. We develop the model in a continuous-time setting, because it allows us to characterize the equilibrium solution explicitly in closed form.

2. The habit process must be mean-reverting, for two reasons. First, when the surplus consumption (defined as consumption minus habit) approaches zero, we do not expect the habit to rise. Indeed, as past high consumption levels recede further into history, we expect that the habit level to come down, unless surplus consumption increases again. Second, historical evidences suggest that the interest rates are stationary. In order to induce a stationary interest rate, the habit process must itself be mean-reverting. The mean reversion feature is absent in Sundaresan (1989). Constantinides (1990) adds this extension.
3. The habit process must have an unexpected shock, for otherwise, the interest rate dynamics would be trivial: even if it is state-dependent, it would not have any local risk, which is counter-factual.
4. The habit process must have a non-zero correlation with the production shock. We will see that the price of interest rate derives primarily from the correlation between the habit shock and the production shock.
5. The habit shock must vanish at zero surplus consumption. Otherwise, the habit process can move above the consumption with a positive probability. Thus, the habit process is controlled by the consumption process, not only in its conditional mean, but also in its conditional volatility.

The last three features are associated with the stochastic nature of the habit formation process, and are absent in both Sundaresan (1989) and Constantinides (1990). Thus, our model will be characterized as a model of *stochastic habit formation*.

The external habit formation model of Campbell and Cochrane (1999) implies a habit process that is stochastic in nature. However, the specific form of the implied habit process is not consistent with the last two features we require. It turns out, as we will see in Section 3.4, that the exogenous specifications of the log surplus consumption ratio and consumption processes in Campbell and Cochrane (1999) can be derived, as an approximation, from the

endogenous processes in the exact solution to our model, and under an additional restriction that the habit shock and the production shock be perfectly correlated. Thus, in the context of our model, the inconsistency between the implied habit process from Campbell and Cochrane (1999) and the stochastic habit formation process in this paper comes from two features of Campbell and Cochrane (1999): (1) implicitly, the habit shock and the production shock are perfectly correlated; and (2) the approximation does not respect the requirement that the conditional volatility of the habit process should vanish at zero surplus consumption.

With a standard power utility function, modified by habit formation, we expect the model to have a homothetic solution, corresponding to a scale-independent economy.³ Unfortunately, the only scale-independent economy possible with the above naive specifications of stochastic habit formation is an economy with constant riskless rate and Sharpe ratio – essentially the same economy studied by Constantinides (1990). An important innovation of the paper is to recognize that, in order to accommodate a scale-independent economy with *stochastic* riskless rate and *time-varying* risk premium, the expected habit growth rate must be *nonlinear* in the surplus consumption. The specific form of this nonlinear term is related to the specific form of the conditional volatility of the habit shock. We will see that, a sensible specification of the conditional volatility of the habit shock that supports a homothetic solution is linear in the surplus consumption. Consequently, the nonlinear term in the drift will have to be quadratic in the surplus consumption. This feature is absent in all previous habit formation models. A more detailed discussion of the motivation for this innovation and its economic interpretation is best explained during the process of solving our model.

At an abstract level, our model is a subset of the general equilibrium framework of Cox et al. (1985). However, our model contains an important refinement. In Cox et al. (1985), the production technologies can be specified almost arbitrarily. The optimality conditions give rise to an endogenous relationship between the riskless rate dynamics and the assumed production technologies. Conversely, any sensible term structure model can be embedded in a general equilibrium model by reverse engineering the requisite production technologies. This is not the case in our model. The time-nonseparable preference induced by stochastic habit process imposes a very strong restriction on the admissible form of the investment opportunity set. In fact, there is only one unique stochastic investment opportunity set in our model that gives rise to a homothetic solution.

Furthermore, this stochastic investment opportunity set has *realistic* properties: under admissible parameter values, the riskless rate is monotonically increasing with the detrended habit process, while the Sharpe ratio of the production technology is monotonically decreasing with the detrended habit process. As a result, a robust, *negative*, local correlation between the riskless rate and the Sharpe ratio emerges endogenously from our model. The sign of this correlation is consistent with many important empirical regularities in both the equity and bond markets. For an example, it leads directly to the prediction that both the equity premium and the expected excess return on zero coupon bonds are positively correlated with the term spread (which is negatively correlated with a mean-reverting riskless rate).

The uniqueness of the stochastic investment opportunity set and consequently the unique-

³Campbell (2000) gives an excellent discussion on why a scale-independent economy, or a homothetic solution, is desirable in order for a model to be consistent with historical experiences.

ness of the equilibrium solution to our model would spell its instant demise if empirical predictions of the model are counter-factual. Fortunately, this is not the case. We have already mentioned that the sign of the local correlation between the riskless rate and the Sharpe ratio gives the correct qualitative predictions on return predictability. We will also show that under very reasonable parameter values (say with a curvature of 2 for the power utility function and a surplus consumption ratio of 5%), the model is able to fit exactly the sample volatility of the consumption growth rate (hence the equity premium puzzle), the sample mean of the real riskless rate (hence the riskfree rate puzzle), and the expectations puzzle, *simultaneously*. The uniqueness of the solution therefore works in our favor: it shows that the empirical regularities in both the equity and bond markets share the same economic origin – time-nonseparable preference induced by stochastic habit formation with a unique investment opportunity set in a scale-independent economy.

Thanks to the explicit form of the equilibrium solution, the empirical implications of our model are both richer and sharper, compared to previous models of habit formation. For an example, while our model is able to explain three of the four puzzles in the equity and bond markets almost perfectly, our analysis also reveals that the stock market volatility puzzle of Campbell (1999) *can not* be explained by our model – once it is required to explain the equity premium puzzle, the riskfree rate puzzle, and the expectations puzzle. However, if we do not require our model to explain the expectations puzzle, then all three equity market puzzles can be explained exactly and simultaneously.

Thus, we subsume the results of Constantinides (1990) and Campbell and Cochrane (1999), both calibrate their models to fit the volatility ratio of the consumption growth rate and the equity return in the absence term structure restrictions.⁴ But we go beyond their work by showing that the equity return volatility can be explained by the habit formation model only when the equity and the bond markets are disconnected.

The rest of the paper is organized as follows. We begin with an incomplete market setup of our model and its solution in Section 2, followed by a complete characterization of key theoretical properties of the model in Section 3, including the local correlation between the riskless rate and the Sharpe ratio, the equilibrium log surplus consumption process, and general properties of the riskless rate dynamics. We complete the theoretical portion of the paper by presenting several important refinements, in Section 4, to our basic setup, including the completion of market using a console bond, and the addition of inflation risk.

In the empirical part of the paper, we start by calibrating the model to the equity market puzzles without imposing moment restrictions implied by bond returns (that is, the theoretical link we have established is artificially broken). We show that the equity premium puzzle, the stock market volatility puzzle, and the riskfree rate puzzles are simultaneously resolved in our model under reasonable parameter values, especially the risk aversion parameter γ . We then show, in Section 6 and Section 7, that the stock market volatility puzzle and the expectations puzzle are mutually incompatible. This incompatibility stems from the implicit assumption in the calibration exercise that the production technology, or the aggregate wealth, consists of the stock market alone. Once we break this assumption, the high volatility for equity return is no longer a puzzle. We show that our model resolves the

⁴Term structure restrictions are necessarily absent in these models because they focus on economies with constant riskless rate and constant risk premium.

remaining three puzzles: the equity premium, the riskfree rate, and the expectations puzzles, simultaneously. Before the conclusion in Section 9, we present key comparative statistics and simulated time-series of important economic variables, computed from the exact solution to our model using parameters calibrated to the asset market puzzles.

2 Stochastic Habit Formation Model with Incomplete Market

In this section, we introduce the model of stochastic habit formation in an incomplete market setting. All key economic intuitions are developed in this basic setup. Key properties of the model are analyzed in Section 3. In Section 5, we show that the model explains all important puzzles in the equity market. In linking the model to the fixed-income markets, however, we need several important refinements to the model, which is taken up in Section 4. In particular, we will introduce a console bond to complete the market.

2.1 Production Technology

We assume that a representative agent has a wealth stock $W(t)$ at time t denominated in units of the consumption good, investing a fraction $\alpha(t)$ of the capital in a risky production technology

$$\frac{dS(t) + D(t)dt}{S(t)} = \mu_S dt + \sigma_S dB_s(t), \quad (1)$$

and $1 - \alpha(t)$ of the capital at riskless rate r . $D(t)$ is the aggregate output.

It is important to point out that we need not think of the production technology as the stock market, *alone*. In this paper, we are not interested in how an individual investor allocates his/her wealth between the stock market and the riskless asset. Rather, we are interested in a representative agent's problem of how to strike an optimal balance between consumption and growth, at the aggregate level. The growth opportunities can come, not only from the stock market, but also from investments in human capital, national infrastructure, public or privately held real estate, and privately held corporations. Wealth capitalized in the public stock market represents only a relatively small fraction of the total wealth of a nation-state. For some countries, such as China, the wealth capitalized in the publicly traded stock market is almost negligible.

However, the price discovery function of an efficient stock market serves to determine the market price of risk for the return to aggregate wealth. Thus, in examining empirical implications of our model, we will assume that the market price of risk gleaned from the stock market is the same as that of the market price of risk for the aggregate wealth process. Consequently, if the return to wealth is less volatile, which is most likely the case due to diversification effect, the risk premium in the expected return to wealth will be smaller than the equity premium.

At a theoretical level, this distinction is not so critical. Indeed, most theoretical asset pricing models do not make this distinction. However, at an empirical level, the distinction

is crucial in refining the notion of “stock market volatility puzzle” of Campbell (1999), and in reconciling the difference between the consumption/wealth ratio and the dividend/price ratio for equity. This issue will be revisited and elaborated in the empirical part of the paper.

2.2 Consumption Habit and Time-Nonseparable Preference

The representative agent consumes at rate $C(t)$. She also develops a consumption habit $X(t)$, which enters her utility function in the manner:

$$U_0 = E_0 \int_0^\infty e^{-\hat{\rho}t} \frac{(C(t) - X(t))^\gamma}{\gamma} dt, \quad (2)$$

where $\gamma < 1$, $\gamma \neq 0$, is the consumption curvature of the utility function, and $\hat{\rho}$ is the subjective discount rate.

The agent forms optimal consumption and portfolio policies by solving

$$\max_{\{C(t), \alpha(t): 0 \leq t \leq \infty\}} U_0, \quad (3)$$

subject to the budget constraint:

$$dW(t) = \{[(\mu_S - r)\alpha(t) + R]W(t) - C(t) + Y(t)\} dt + \sigma_S \alpha(t)W(t)dB_s(t), \quad (4)$$

The *point of departure* of this paper from previous work, such as Sundaresan (1989) and Constantinides (1990), is to assume that the habit formation process is *stochastic*:

$$dX(t) = \left[\left(b_X(C(t) - X(t)) + \frac{\epsilon_X \beta^2}{2} (C(t) - X(t))^2 \right) - (a_X - b_X)X(t) \right] dt + \beta(C(t) - X(t))dB_x(t), \quad (5)$$

where $\delta = \text{corr}(dB_x, dB)$, $\delta\beta < 0$, $\epsilon_X > 0$, $a > b > 0$.

Thus, the habit process receives an unexpected shock, or has a local uncertainty, which is correlated with the production shock. The stochastic nature of habit formation is a key innovation, because it gives rise to a non-trivial term structure dynamics that is also empirically relevant.

At a technical level, the inclusion of the quadratic term in the drift of the habit process is necessary in order for the economy to have a homothetic solution. At an intuitive level, the quadratic term introduces a nonlinear aspect of complementarity between adjacent consumptions: most recent past consumptions will be weighted more than consumptions in the distant past, relative to exponential weighting associated with linear and deterministic habit formation models of Sundaresan (1989) and Constantinides (1990).

The volatility of the habit shock is increasing linearly with the surplus consumption. It is critical that the volatility decreases to zero as the surplus consumption goes to zero. Otherwise, there would be a positive probability that the habit may move above the consumption level – this is forbidden in our utility specification. When the volatility of the habit shock is not linear in the surplus consumption, the model may not have a homothetic solution.

The target level must also vanish with the surplus consumption. This is because, at zero surplus consumption, there is no reason for the habit to grow. In fact, since we expect the habit to be weighted sum of past consumption, once the consumption *drops* to habit level, the habit can only be expected to decrease as past consumption level recedes further into the more distant past and receives less and less weights. Any constant in the drift will rule out a homothetic solution to the model.

If $\epsilon_X < 0$, the habit process would be counter-intuitive: for sufficiently high consumption surplus, the habit would be expected to decline. Thus we rule out this case.

When $\epsilon_X = 0$, the model is well-defined, but gives rise to a trivial term structure of interest rates. In particular, if β is also zero, then the model reverts back to Constantinides (1990). If $\beta \neq 0$, but $\epsilon_X = 0$, the solution to the model is qualitatively similar to Constantinides (1990): constant riskless rate and constant Sharpe ratio. Again, the term structure of interest rates is trivial.

We are certainly not claiming that this is a unique specification of habit formation. An important justification for this specification is that it leads to a model that is not only *well-defined*, in the sense that a non-trivial solution exists and is unique, but also *sensible*, in the sense that the predictions of the model accord well with the empirical observations.

The model is well-defined for arbitrary sign of $\delta\beta$. However, the model will be empirically irrelevant for the term structure of interest rates if $\delta\beta \geq 0$. First, if β were zero, then interest rate would have to be constant, or locally riskless. Second, if δ were zero, then interest rate risk, even if present, would not be priced. Third, if $\delta\beta > 0$, then expected return on long term bonds will be negatively correlated with the equity risk premium, which is counter-factual. In particular, it will lead to the wrong direction in which the expectations hypothesis is violated.

The sign of δ is arbitrary, because the economically meaningful sign is the correlation between the habit shock and the production shock, which is determined by the product $\delta\beta$. We choose $\delta < 0$, because, as will become more clear later, this is consistent with the usual convention in the term structure literature. Given the choice of this convention, negative correlation between the habit shock and the production shock implies, or is implied, by the sign restriction $\beta > 0$.

It is natural to impose the restriction $a_X > b_X > 0$, as does Constantinides (1990).

The appearance of β in the quadratic term does not entail any loss of generality: as long as both ϵ_X and β are not zero, it is a matter of parameterization. This particular choice of parameterization makes the notations much simpler.

2.3 Detrending and Rescaling the Economy

We expect that in a non-trivial economy, aggregate output, consumption, habit, and wealth are expected to grow. We assume, and verify later, that the model allows a deterministic time-trend for all of the flow and stock variables, at a rate $G(t)$, the permanent growth rate, so that after detrending, all variables have a steady state. The growth rate can be state-dependent. However, the model does not have any theoretical restriction on the form of the state-dependency. For simplicity, we will assume that it is simply a constant, G .⁵

⁵Later, we will see that it is natural to think of G as the growth rate of the economy in its steady state.

A dimensional analysis reveals that the parameter $\epsilon_X > 0$ has the dimension of inverse consumption. Since the consumption level is assumed to have a positive permanent trend, ϵ_X must have a negative permanent trend. Thus, we write:

$$\epsilon_X = \epsilon e^{-\int_0^t G(s)ds}, \quad (6)$$

where ϵ is a constant.

Define the following *detrended and rescaled* variables

$$\begin{aligned} w(t) &= \epsilon e^{-\int_0^t G(s)ds} W(t), \\ s(t) &= \epsilon e^{-\int_0^t G(s)ds} S(t), \\ d(t) &= \epsilon e^{-\int_0^t G(s)ds} D(t), \\ c(t) &= \epsilon e^{-\int_0^t G(s)ds} C(t), \\ x(t) &= \epsilon e^{-\int_0^t G(s)ds} X(t). \end{aligned} \quad (7)$$

Equations (7) are nothing more than a change of variables. The transformations themselves do not have any economic content whatsoever. Economic content is added by assuming, and later verifying, that the detrended processes have a steady state distribution.⁶ This allows us to model the riskless rate and the market price of risk, which are typically thought of as stationary processes, as a function of the state of the economy $x(t)$ and $w(t)$.

It is straight-forward to verify that

$$\frac{ds(t) + d(t)dt}{s(t)} = \mu_s dt + \sigma_s dB_s(t). \quad (8)$$

where $\mu_s = \mu_S - G$, $r = R - G$, $\sigma_s = \sigma_S$. The Sharpe ratio of the risky technology, $\Lambda = \frac{\mu_S - R}{\sigma_S} = \frac{\mu_s - r}{\sigma_s}$, is not affected.

Furthermore, the detrended and rescaled habit process is given by

$$dx(t) = \mu_x dt + \sigma_x dB_x(t), \quad (9)$$

where,

$$\begin{aligned} \mu_x &= \left(b(c(t) - x(t)) + \frac{\beta^2}{2}(c(t) - x(t))^2 \right) - (a - b)x(t), \\ \sigma_x &= \beta(c(t) - x(t)), \end{aligned} \quad (10)$$

and $a = a_X + G$, $b = b_X$. The parametric restrictions on the original habit process are carried over: $\delta < 0$, $\beta > 0$, $\delta < 0$, $a > b > 0$.⁷

⁶Parametric restrictions under which the model allows a steady state economy is discussed in Section 3.

⁷For empirically relevant cases, we would expect the permanent growth rate to be positive. Thus $a > b$ is implied by $a_X > b_X$. However, it is not necessary to require that $G > 0$ on theoretical grounds. The model is applicable also for an economy with a steady decline, in which case $a > b$ is a stronger assumption than $a_X > b_X$. In any case, we will insist that $a > b$, in addition to the assumption $a_X > b_X$.

2.4 Consumption Choice and Portfolio Selection

In terms of detrended variables, the agent's problem becomes

$$\max_{\{c(t), \alpha(t): 0 \leq t \leq \infty\}} E_0 \int_0^\infty e^{-\rho t} \frac{[c(t) - x(t)]^\gamma}{\gamma} dt, \quad (11)$$

subject to the constraint:

$$dw(t) = \{[(\mu_s - r)\alpha(t) + r]w(t) - c(t)\} dt + \sigma_s \alpha(t) w(t) dB_s(t), \quad (12)$$

where $\rho = \hat{\rho} - G\gamma$.

Note that the parameter ϵ has dropped out completely from the rescaled economy. ϵ^{-1} may be interpreted as the scale of an economy under consideration. Our model is applicable to an infinite class of economies that differ only in its scale ϵ^{-1} , and permanent growth rate G .

The portfolio selection problem is a standard one, with an important twist to be explained below. Since we are working in an incomplete market setting, we use the dynamic programming approach. Assuming, and later verifying through consistency check, that the indirect utility function takes the following form: $V(w(t), x(t)) = e^{-\rho t} J(w(t), x(t))$, we arrive at the Hamilton-Jacobi-Bellman equation:

$$\begin{aligned} 0 = & \max_{\alpha(t), c(t)} \frac{[c(t) - x(t)]^\gamma}{\gamma} - \rho J \\ & + \{[(\mu_s - r)\alpha(t) + r]w(t) - c(t)\} J_w + \frac{1}{2} \sigma_s^2 \alpha(t)^2 w(t)^2 J_{ww} \\ & + \left[bc(t) - ax(t) + \frac{\beta^2}{2} (c(t) - x(t))^2 \right] J_x + \frac{\beta^2}{2} (c(t) - x(t))^2 J_{xx} \\ & + \delta \sigma_s \alpha(t) w(t) \beta (c(t) - x(t)) J_{wx} \end{aligned} \quad (13)$$

Furthermore, we assume, and later verify with consistency check, that the time-invariant indirect utility function takes the following form:

$$J(w, x) = \phi \frac{(w - f(x))^\gamma}{\gamma}, \quad (14)$$

which is concave in wealth. We call $f(x)$ the *habit level of wealth*, and $w - f(x)$ the *surplus wealth*. Nonlinear complementarity of consumption is directly linked to nonlinearity of $f(x)$.

Since $J(w, x)$ is concave in wealth, the objective function of the optimization problem (13) is quadratic and concave in the portfolio policy α . Consequently there is always a unique solution for the portfolio policy except for degenerate cases, which we rule out by assuming that the utility function is not linear. The first order condition for optimal portfolio policy is given by

$$(\mu_s - r)w J_w + \sigma_s^2 \alpha^* w^2 J_{ww} + \delta \sigma_s \beta (c^* - x) w J_{wx} = 0. \quad (15)$$

The first order condition for optimal consumption policy is given by

$$(c^* - x)^{\gamma-1} - J_w + [b + \beta^2 (c^* - x)] J_x + \beta^2 (c^* - x) J_{xx} + \delta \sigma_s \beta \alpha^* w J_{wx} = 0. \quad (16)$$

From equation (14), we have

$$J_w = \frac{\gamma}{w - f(x)} J, \quad (17)$$

$$J_{ww} = \frac{\gamma(\gamma - 1)}{(w - f(x))^2} J, \quad (18)$$

$$J_x = -\frac{\gamma f'(x)}{w - f(x)} J \quad (19)$$

$$J_{xx} = -\frac{\gamma f''(x)}{w - f(x)} J + \frac{\gamma(\gamma - 1) f'(x)^2}{(w - f(x))^2} J \quad (20)$$

$$J_{wx} = \frac{\gamma(1 - \gamma) f'(x)}{(w - f(x))^2} J \quad (21)$$

Now comes the twist. If the second order derivative of $f(x)$ does not vanish, the two terms in J_{xx} are not homogeneous in surplus wealth, $w - f(x)$. This could potentially rule out a homothetic solution to the model. But this is exactly where the quadratic term in the expected habit growth comes to the rescue:⁸ if the habit level of wealth $f(x)$ takes the following form:

$$f(x) = \eta (1 - e^{-x}), \quad (22)$$

where η is an arbitrary positive constant, then the offending term in J_{xx} and the quadratic term in the drift of the habit process cancel exactly. We are then left with a homothetic solution:

$$c^* = x + h(x)(w - f(x)), \quad (23)$$

$$\alpha^* = \frac{g(x)}{\sigma_s} \frac{w - f(x)}{w}, \quad (24)$$

provided that the production technologies take the following form:⁹

$$r(x) = \frac{1 - (a - b)f'(x)}{f(x)/x}, \quad (25)$$

$$\Lambda(x) = (1 - \gamma) [g(x) - \delta\beta h(x)f'(x)]. \quad (26)$$

⁸When $f(x)$ is linear, the quadratic term itself becomes a problem, and the model will not have a solution. The model can have a solution with linear $f(x)$, however, if the quadratic term is removed, by setting ϵx to zero in equation (5). In that case, the variables will not be scaled by ϵ , and the model still goes through, except that the homothetic solution to the model again implies a trivial term structure of interest rates.

⁹Although conceptually, the agent takes the production technologies as given in forming her asset demand, it would have been difficult to guess, *a priori*, what the production technologies should look like as functions of the *controlled* state variable, x . We *solve* for the *requisite* production technologies by guessing first the form of the indirect utility function. The riskless rate and the Sharpe ratio, and subsequently, the condition volatility of the production return, are all *endogenously* determined and tightly linked together. This makes perfect sense in a general equilibrium framework, and this is what makes the empirical predictions of this model so tight.

The state-dependent functions $h(x)$ and $g(x)$ are determined by substituting the optimal policies into equations (13) and (16):

$$\begin{aligned} \phi^{-1}h(x)^{\gamma-1} &= (1 + bf'(x)) + (1 - \gamma)\beta^2h(x)f'(x)^2 \\ &\quad - \delta\beta(1 - \gamma)g(x)f'(x) \end{aligned} \quad (27)$$

and

$$p_2(x)g(x)^2 + p_1(x)g(x) + p_0(x) = 0, \quad (28)$$

where

$$\begin{aligned} p_2(x) &= \frac{1 - \gamma}{2}, \quad p_1(x) = -\frac{1 - \gamma}{\gamma}\delta\beta h(x)f'(x), \\ p_0(x) &= r(x) - \frac{\rho}{\gamma} + \frac{1 - \gamma}{\gamma}(1 + bf'(x))h(x) \\ &\quad + \frac{(1 - \gamma)(2 - \gamma)}{2\gamma}\beta^2h(x)^2f'(x)^2. \end{aligned} \quad (29)$$

Consumption choice and portfolio selection with stochastic investment opportunity set has been examined by many authors, starting from the classical papers of Merton (1969) and Merton (1971). Recent work have focused on the effects of stochastic interest rates and time-varying risk premium. The literature is huge, and we can only mention a few. Kim and Omberg (1996) solves the portfolio selection problem with a constant riskless rate and a time-varying equity premium that follows a homoskedastic AR(1) process. Liu (2000) derives explicitly optimal consumption and portfolio policies with affine interest rates and risk premium. Campbell and Viceira (2000) solves (approximately) the optimal consumption choice and portfolio selection problem with the presence of both indexed and nominal bonds (characterized by a two-factor Gaussian model), with Epstein-Zin utility (Epstein and Zin (1989)). Lynch (1999) also examines the portfolio selection problem with the presence of both time-varying real interest rates and risk premium. A common feature of these work and most, if not all, other work in the literature with realistic features of return predictability and stochastic interest rates is that the stochastic investment opportunity set is *exogenously* specified. In contrast, we show that both the riskless rate and the Sharpe ratio in our model take a unique functional form, dictated by the requirement that the solution be homothetic.

2.5 Admissible State Space and Transversality Condition

Since $f(x)$ is known, the functions $h(x)$ and $g(x)$ can be obtained jointly, for each given state x , by solving the equations (27) and (28) simultaneously. The solution does not necessarily exists at all habit levels. Since equation (28) is quadratic in $g(x)$, a solution can exist only if

$$\Delta(x) = p_1(x)^2 - 4p_0p_2 \geq 0. \quad (30)$$

We will show later that, for empirically relevant parameter values, $\Delta(x)$ will be decreasing in x , and eventually becomes zero at a critical value x_{max} :

$$\Delta(x_{max}) = 0. \quad (31)$$

Thus, there is a limit, in this economy, on the equilibrium habit level, beyond which the economy can not be sustained. At x_{max} , the Sharpe ratio is given by

$$\Lambda(x_{max}) = \frac{(1 - \gamma)^2}{\gamma} \delta \beta h(x_{max}) f'(x_{max}). \quad (32)$$

The admissible state space for the habit is therefore

$$0 < x < x_{max} \quad (33)$$

When the volatility of habit shock is small, i.e., $\beta \ll 1$, and/or the correlation between the habit shock and the production shock is small, i.e., $\delta \ll 1$, the risk premium at the critical habit level x_{max} is close to zero. This is because when the habit level rises, the agent becomes increasingly more aggressive in investing in the production shock – as long as the risk premium is positive, and increasingly less aggressive in investing in the riskless asset (less lending or even more borrowing) to sustain a consumption level that is necessarily higher than the habit level. We will see in more detail in Section 3 and Section 8, the risk premium is driven down and the riskless rate is driven up by a rising habit level.

In general, if the steady state of the habit process is below x_{max} , habit will be pulled back as it approaches x_{max} . However, there is always a theoretical possibility that a big, though rare, habit shock can send the habit level above the critical value x_{max} . This can be handled by pasting another model on top of our model that deals with this situation. The habit process (9) only applies when $x \leq x_{max}$. However, this does not change the nature of the economy in the admissible region of the current model, because the behavior of the economy is determined by the local dynamics of the habit process. By assuming that the habit shock is relatively weak, and the steady state of the habit is a finite distance away from x_{max} , we do not need to worry about this theoretical possibility in exploring the empirical implications of our model. We will give a quantification of this argument after we have calibrated the model to empirical data.

The transversality condition, $\lim_{t \rightarrow \infty} E_0 [e^{-\rho t} J(w(t), x(t))] = 0$, is satisfied, if

$$\rho \geq \gamma r(x) + \frac{\gamma}{2(1 - \gamma)} \Lambda(x)^2 + \frac{\gamma(1 - \gamma)}{2} (1 - \delta^2) \beta^2 h(x)^2 f'(x)^2. \quad (34)$$

2.6 Equilibrium Consideration

Our model can be closed as a general equilibrium model by imposing the market clearing condition:

$$c^* = d, \quad (35)$$

$$\alpha^* = 1. \quad (36)$$

Equations (24) and (36) tie down the last piece of the production technology: the conditional volatility the total return to wealth:

$$\sigma_s = g(x) \frac{w - f(x)}{w}. \quad (37)$$

Since σ_s does not depend on either of the two stochastic controls, its state-dependency does not affect the consumption choice and portfolio selection problem.

The market clearing conditions imply that $w = s$ in equilibrium.

In Section 8, we will see that $g(x)$ decreases with the habit level x , and the habit level is positively correlated with the surplus consumption ratio. Thus, given the wealth level, the conditional volatility of the return to the dividend or consumption claim is decreasing with surplus consumption ratio. This is consistent with the finding of Campbell and Cochrane (1999). This is not an coincidence – we show in the next section that the exogenous assumptions of Campbell and Cochrane (1999) on the surplus consumption ratio process arise endogenously in our model.

3 Equilibrium Properties

In this section, we characterize formally three key equilibrium properties of our model.

First, we analyze the local correlation between the riskless rate and the Sharpe ratio. This correlation lies at the center of many predictability results documented in the literature, including predictability of equity returns and bond returns. In order to induce a correlation between the riskless rate and the risk premium, both the riskless rate and the risk premium are necessarily time-varying.

Second, we characterize the dynamics of the real riskless rate in the equilibrium.

Third, we characterize the log surplus consumption ratio and the optimal consumption growth. These are primitive processes in the external habit formation model of Campbell and Cochrane (1999). We show that their exogenous assumptions arise endogenously as an approximation to the equilibrium solution of our model. This allows our model to make contact with a wealth of empirical regularities analyzed by Campbell and Cochrane (1999). Furthermore, our model potentially offers richer and stronger predictions by virtue of the fact that the exact solution is a lot richer than what Campbell and Cochrane (1999) assume.

3.1 Correlation Between Riskless Rate and Sharpe Ratio

To simplify notations, let us define:

$$\begin{aligned}
 r_0 &= r(0), & \eta &= \frac{1}{r_0 + a - b}, \\
 r &= r(x), & f &= f(x), \\
 r_x &= r'(x), & f_x &= f'(x) = \eta e^{-x}, \\
 h^0 &= \lim_{\beta \rightarrow 0} h(x), & h_x^0 &= \lim_{\beta \rightarrow 0} h'(x), \\
 g^0 &= \lim_{\beta \rightarrow 0} g(x), & g_x^0 &= \lim_{\beta \rightarrow 0} g'(x), \\
 \Lambda^0 &= \lim_{\beta \rightarrow 0} \Lambda(x) = (1 - \gamma)g^0, & \Lambda_x^0 &= \lim_{\beta \rightarrow 0} \Lambda'(x) = (1 - \gamma)g_x^0.
 \end{aligned}$$

Our model then implies that¹⁰

$$f_x = \eta e^{-x} > 0, \quad (38)$$

$$r_x = r \left[\frac{(a-b)f_x}{1-(a-b)f_x} + \frac{f-xf_x}{xf} \right] > 0, \quad (39)$$

$$g^0 = \frac{\Lambda^0}{1-\gamma} > 0, \quad (40)$$

$$h^0 = \frac{\rho - \gamma r - \frac{1}{2}\gamma(1-\gamma)(g^0)^2}{(1-\gamma)(1+bf_x)} > 0, \quad (41)$$

$$\phi = \frac{(h^0)^{\gamma-1}}{1+bf_x} > 0, \quad (42)$$

$$h_x^0 = \frac{bh^0 f_x}{1-\gamma} > 0, \quad (43)$$

$$g_x^0 = -\frac{1}{(1-\gamma)g^0} [r_x + (1-\gamma)\phi^{-1}(h^0)^{\gamma-1}h_x^0] < 0, \quad (44)$$

$$\Lambda_x^0 = (1-\gamma)g_x^0 < 0. \quad (45)$$

Two of the most important signs are

$$r_x = r'(x) > 0, \quad \Lambda_x = \Lambda'(x) < 0. \quad (46)$$

That is, when the habit shock is not too volatile, the riskless rate is *positively* correlated with the habit level and the Sharpe ratio is *negatively* correlated with the habit level.

Although we have only shown that the result is unambiguously true in the limit $\beta \rightarrow 0$, we expect it to be valid for small β as well, by continuity. If the riskless rate is mean-reverting (which we confirm later,) we are led directly to the conclusion that both the equity risk premium and the expected excess return on long term bonds are *positively* correlated with the term spread. The equity premium predictability result confirms Fama and French (1989), and the bond return predictability result confirms Fama (1976), Fama (1984b), Fama (1984a), and Fama and Bliss (1987).

3.2 A Consumption Based Real Riskless Rate Model

In this subsection, we characterize the dynamics of the real riskless rate implied by our model. This description is complete, even in the incomplete market setting. However, we do not yet have a term structure model, because the market price of interest rate risk is yet to be determined.

¹⁰ We show details for the two most important signs: $r_x > 0$ and $\Lambda_x^0 < 0$ – in the limit $\beta \rightarrow 0$. To show $r_x > 0$, note that when $x > 0$, $f(x) > 0$, $f'(x) > 0$, $f(x) - xf'(x) > 0$, and $xf(x) + xf'(x) - f(x) > 0$. Thus, $r'(x) = \frac{(f(x)-xf'(x))+(a-b)f'(x)(xf(x)+xf'(x)-f(x))}{f(x)^2} > 0$. To show $h_x^0 > 0$, note that as $\beta \rightarrow 0$, equation (27) becomes $\phi^{-1}h(x)^{\gamma-1} = 1 + bf'(x)$. It follows that $h'(x) = \frac{bh(x)f'(x)}{1-\gamma} > 0$. To show $g_x^0 < 0$, note that as $\beta \rightarrow 0$, equation (28) simplifies to $g(x)^2 = -\frac{2}{1-\gamma} \left[r(x) - \frac{\rho}{\gamma} + \frac{1-\gamma}{\gamma} \phi^{-1}h(x)^\gamma \right]$. It follows that $g'(x) = -\frac{1}{(1-\gamma)g(x)} [r'(x) + (1-\gamma)\phi^{-1}h(x)^{\gamma-1}h'(x)] < 0$. It follows that $\Lambda'(x) = (1-\gamma)g'(x) < 0$.

Using Ito's lemma and equation (25), we can derive the dynamics of the riskless rate (the instantaneous short rate):

$$r(x) = \frac{1 - (a - b)f'(x)}{f(x)/x}, a > b, \quad (47)$$

$$f(x) = \eta(1 - e^{-x}), \eta > 0, \quad (48)$$

$$dx(t) = \kappa(\theta_x - x)dt + \sigma_x dB_x(t), \quad (49)$$

where

$$\begin{aligned} \kappa &= a - b > 0 \\ \theta_x &= \frac{1}{\kappa} \left[bz(t) + \frac{\beta^2}{2} z(t)^2 \right] \\ \sigma_x &= \beta z(t), \\ z(t) &\equiv c(t) - x(t), \\ c - x &= h(x)(w - f(x)). \end{aligned}$$

Thus, the riskless rate is a nonlinear transformation of a mean-reverting process with stochastic central tendency and stochastic volatility, both of which are driven by the surplus consumption, $z = c - x$.

The riskless rate in the actual economy is given by $R = r + G$. R is non-negative whenever it is non-negative at zero habit level. This is equivalent to the parametric restriction:

$$0 < \eta \leq \frac{1}{a - G - b} = \frac{1}{a_X - b_X}, \quad (50)$$

Augmented by the endogenously determined stochastic process (56), we have a two-factor model for the real riskless rate. Since one of the state variables is effectively the consumption process, and the habit is effectively observed from the riskless rate through a one-to-one mapping, the model can be characterized as a two-factor *consumption-based real interest rate model*, or CRIM. The model has several very appealing properties.

1. It is a structural model derived from fundamental economic theory, and has the hope of being a realistic description of the data.
2. The riskless rate is strictly positive, if the riskless rate at zero habit level is strictly positive. This is because the riskless rate is an increasing function of the habit.
3. Although the surplus consumption $z = c - x$ is endogenously determined by the model, it is effectively observable, since the consumption habit can be backed out from the interest rate itself. Viewed this way, CRIM is completely specified in terms of macroeconomic observables. With a suitable estimation technique for non-Markovian models, we don't even need to specify a dynamics for the surplus consumption in order to estimate the model. Of course, a more powerful test is also available if we use the endogenous restrictions imposed by our model on the dynamics of the surplus consumption process.

4. CRIM gives a structure interpretation for term structure models with central tendency and stochastic volatility (see Balduzzi et al. (1996) and Dai and Singleton (1999)).
5. Finally, the expectations puzzle can be explained by CRIM. We make it transparent by showing that in a suitable limit, CRIM leads to a dynamic term structure model proposed recently by Dai and Singleton (2000) to explain the expectations puzzle.

The implication for the expectations puzzle will be the focal point of our analysis of the empirical implications of our generalized habit formation model. Thus we need to examine the pricing of interest rate risk in the bond market.

There are two sources of interest rate risk: the habit shock that enters directly due to the direct dependence of the riskless rate on the consumption habit; and the production shock that enters indirectly through the conditional mean and conditional volatility of the consumption habit. In general, we can decompose the habit shock into two components: one parallel to the production shock and the other orthogonal to the production shock:

$$B_x = \delta B_s + \sqrt{1 - \delta^2} B_\perp, \quad (51)$$

where B_\perp is a standard Brownian motion orthogonal to B_s . We already know what the market price for the production shock is: Λ . To the extent that the habit shock is correlated with the production shock with correlation coefficient δ , it will be priced, and the market price for the component of the habit shock parallel to the production shock is simply $\delta\Lambda$. What about the market price of risk associated with B_\perp ? Since bonds, if exist, are exposed to the risk associated with B_\perp , they will serve as a natural hedge against the habit shock. Therefore the risk associated with B_\perp should command a premium in the equilibrium. To determine this premium, we need to extend our model by introducing bonds into the agent's investment opportunity set. This completes the market. This extension will be worked out in Section 4.

In Detemple and Zapatero (1991), stochastic interest rate is introduced through the fascility of an exogenously specified endowment process in a Lucas (1978)–type exchange economy. In Campbell and Cochrane (1999), stochastic interest rate is introduced through exogenously specified consumption and surplus consumption processes, in an effectively Lucas-type exchange economy. In contrast, the interest rate dynamics and the market prices of interest rate are endogenously determined in our model, as is the full pricing kernel (see Section 4.)

3.3 Surplus Wealth and Consumption Processes

It is straight forward to show that, in equilibrium, the growth rate of the surplus wealth, $\omega = w - f(x)$, is determined completely by the habit process:

$$\frac{d(w - f(x))}{w - f(x)} = \mu_\omega dt + \sigma_{\omega s}(x) dB + \sigma_{\omega x}(x) dB_x, \quad (52)$$

where

$$\mu_\omega(x) = \frac{r(x) - \rho}{1 - \gamma} + \frac{2 - \gamma}{2(1 - \gamma)^2} \Lambda(x)^2 + \frac{2 - \gamma}{2} (1 - \delta^2) \beta^2 h(x)^2 f'(x)^2, \quad (53)$$

$$\sigma_{\omega s}(x) = g(x), \quad (54)$$

$$\sigma_{\omega x}(x) = -\beta h(x) f'(x). \quad (55)$$

Similarly, from equations (23) and (52), the growth rate of the surplus consumption process, $z = c^* - x$, is completely determined by the habit and the surplus consumption processes:

$$\frac{dz}{z} = \mu_z(z, x)dt + \sigma_{zs}(z, x)dB_s + \sigma_{zx}(z, x)dB_x, \quad (56)$$

where

$$\mu_z = \mu_\omega + \frac{h'(x)}{h(x)} (\mu_x + \sigma_{\omega x}\sigma_x + \delta\sigma_{\omega s}\sigma_x) + \frac{1}{2} \frac{h''(x)}{h(x)} \sigma_x^2, \quad (57)$$

$$\sigma_{zs} = \sigma_{\omega s}(x), \quad (58)$$

$$\sigma_{zx}(z, x) = \sigma_{\omega x} + \frac{h'(x)}{h(x)} \sigma_x. \quad (59)$$

The above results imply that both the surplus wealth and the surplus consumption will be strictly positive, as long as they are initially so. The positivity of the surplus consumption, however, does not guarantee that the habit is strictly positive. At $x = 0$, the drift and the diffusion of the habit process are $bz + \frac{1}{2}z^2$ and βz , respectively. The fact that the diffusion does not vanish at $x = 0$ implies that there is always a theoretical possibility that the habit could be negative. The possibility of consumption substitution does not invalidate our solution. The probability of the habit becoming and staying negative is small, because the drift is positive. Thus our model allows local substitution of consumption on an infrequent basis.

3.4 Surplus Consumption Ratio

We are now ready to make a formal connection to Campbell and Cochrane (1999), who specify, *exogenously*, the stochastic processes governing the surplus consumption ratio and the consumption processes. In what follows, we show that some key aspects of their assumptions arise, *endogenously*, from our model.

Following Campbell and Cochrane (1999), we define the log surplus consumption ratio by

$$\xi = \log \frac{c - x}{c} = \log \frac{C - X}{C}. \quad (60)$$

It is easy to see that the surplus consumption ratio is the same for both the original economy and our detrended, rescaled economy.

From equations (9), (52), and (56), we obtain the dynamics of the log surplus consumption ratio:

$$d\xi = \mu_\xi dt + \sigma_{\xi s} dB_s + \sigma_{\xi x} dB_x, \quad (61)$$

where

$$\mu_\xi = \frac{x\mu_z}{x(z+x)} - \frac{(z+x)^2 - z^2}{2z^2(z+x)^2}\sigma_z^2 - \frac{\mu_x}{z+x} + \frac{\sigma_x^2}{2(z+x)^2}, \quad (62)$$

$$\sigma_{\xi s} = \frac{xg(x)}{z+x}, \quad (63)$$

$$\sigma_{\xi x} = \left[\left(\frac{xh'(x)}{h(x)} - 1 \right) \frac{\sigma_x}{z+x} - \frac{\beta x h(x) f'(x)}{z+x} \right]. \quad (64)$$

Similarly, we can obtain the dynamics of the consumption process:

$$\frac{dc}{c} = \mu_c dt + \sigma_{cs} dB_s + \sigma_{cx} dB_x, \quad (65)$$

where

$$\mu_c = \frac{z\mu_z}{z+x}, \quad (66)$$

$$\sigma_{cs} = \frac{c-x}{c}\sigma_{zs} = \frac{zg(x)}{z+x}, \quad (67)$$

$$\sigma_{cx} = \frac{c-x}{c}\sigma_{zx} = \frac{c-x}{c} \left(\frac{h'(x)}{h(x)}\sigma_x - \beta h(x)f'(x) \right). \quad (68)$$

If we ignore the habit shock, we immediately obtain the *sensitivity function*, defined as the ratio of the shock to the log surplus consumption ratio and the shock to the consumption growth rate:

$$\zeta = \frac{\sigma_{\xi s}}{\sigma_{cs}} = \frac{x}{z} = e^{-\xi} - 1. \quad (69)$$

Let $\bar{\xi}$ be the steady state mean¹¹ of ξ . Linearizing the right hand side of equation (69), we have

$$\begin{aligned} \zeta &\approx e^{-\bar{\xi}} \left(1 - (\xi - \bar{\xi}) + \frac{1}{2} \text{var}(\xi - \bar{\xi}) \right) - 1 \\ &\approx \frac{e^{\text{var}(\xi - \bar{\xi})}}{\Xi} (1 - (\xi - \bar{\xi})) - 1 \approx \frac{e^{\text{var}(\xi - \bar{\xi})}}{\Xi} \sqrt{1 - 2(\xi - \bar{\xi})} - 1, \end{aligned} \quad (70)$$

where $\Xi = e^{\bar{\xi} + \frac{1}{2}\text{var}(\xi)}$ is the steady state mean of e^ξ .

When $\text{var}(\xi - \bar{\xi}) = 0$, that is, local uncertainty of the log surplus consumption ratio at the steady state mean is zero, the right hand side of equation (70) is exactly the same as the sensitivity function specified by Campbell and Cochrane (1999)! Thus we recover the constant riskless rate case of Campbell and Cochrane (1999) in the limit $\beta \rightarrow 0$.

In our model, stochastic interest rate is induced by local fluctuations of habit near its steady state mean. In the context of Campbell and Cochrane (1999), we expect that stochastic interest rate would be induced by local fluctuations of the log surplus consumption ratio

¹¹See next subsection for conditions under which the steady mean exists.

near its steady state mean. This is accomplished by letting $var(\xi - \bar{\xi}) \neq 0$. This is exactly equivalent to modifying the steady state mean in the sensitivity function of Campbell and Cochrane (1999).

From next subsection and Section 8, we will see that under suitable parameter values, the state variables of the economy will be mean reverting around their steady state means. Consequently, the log surplus consumption ratio will also be mean-reverting around its steady state mean.

Thus, we recover all key assumptions that define the external habit formation model of Campbell and Cochrane (1999), which can now be seen as a linearization of the optimal solution to our model around the steady state of the economy.

Because our model is solved exactly in equilibrium, it puts a a great deal more structure on a number of important volatilities associated with the equity market puzzles. For an example, ignoring the habit shock, the conditional volatility of the consumption growth rate is $g(x) = \frac{\Lambda(x)}{1-\gamma}$ multiplied by the surplus consumption ratio, $\frac{c-x}{c}$, while the conditional volatility of the production return is $g(x)$ multiplied by the surplus wealth ratio, $\frac{w-f(x)}{w}$. In Section 4.1, we will also see that the conditional volatility of the endogenously determined pricing kernel is given by $\Lambda(x)$. Thus, the equity premium puzzle boils down to the relative magnitude of surplus consumption ratio and the preference parameter $1 - \gamma$, while the volatility puzzle boils down to the relative magnitude of the surplus consumption ratio and the surplus wealth ratio. These endogenous relationship will be exploited in Section 5 to give a simultaneous resolution of all major equity market puzzles.

3.5 Steady State Means

Combining equations (9) and (52), we have a bi-variate Markovian system for the surplus wealth and the habit processes, the two controlled state variables of the economy. A complete characterization of the dynamic properties of this Markovian system is beyond the scope of this paper. For our purpose, it is sufficient to show the existence of steady state means for the surplus wealth and the habit processes, under empirically relevant parameter values. This will be done in Section 8. To prepare for that, we characterize conditions under which the steady state means of the state variables exist.

We define the steady state means $\bar{\omega}$ and \bar{x} as the long run values of the surplus wealth process ω and the habit process x when both the production shock and the habit shock are turned off. In another word, at $\omega = \bar{\omega}$ and $x = \bar{x}$, the drift of the surplus consumption and the habit processes are zero:

$$\mu_{\omega}(\bar{x}) = 0, \tag{71}$$

$$\mu_x(\bar{z}, \bar{x}) = 0, \tag{72}$$

where

$$\bar{z} = h(\bar{x})\bar{\omega} \tag{73}$$

is the steady state mean of the surplus consumption.¹²

¹²We do not define \bar{z} as the value of z at which the drift of z vanishes. This is convenient for our purpose.

Equations (71) and (72) must be solved jointly (though in general we must solve them numerically.) The steady state means depend only on the parameters of the detrended, rescaled economy. From equation (72), the steady state mean of the surplus consumption and the steady state mean of the habit are related by

$$\bar{x} = \frac{1}{a-b} (b\bar{z} + \frac{1}{2}\beta^2\bar{z}^2). \quad (76)$$

A useful corollary of equation (76) is that, in the limit $\beta \rightarrow 0$, we have

$$\frac{\bar{z}}{\bar{x}} = \frac{a-b}{b}, \quad (77)$$

$$\frac{\bar{c}}{\bar{x}} = \frac{a}{b}. \quad (78)$$

In the rest of the paper, we will assume that parameter values are chosen so that

$$0 < \bar{x} < x_{max}, \quad \bar{z} > 0, \quad (79)$$

which we will refer to as *steady state conditions*.¹³

When x is only a small deviation, $\hat{x} = x - \bar{x}$, away from the steady state mean, we have the following approximation:

$$r(x) = r(\bar{x}) + r'(\bar{x})\hat{x} + o(\hat{x}) + o(\beta), \quad (80)$$

$$\Lambda(x) = \Lambda(\bar{x}) + \Lambda'(\bar{x})\hat{x} + o(\hat{x}) + o(\beta). \quad (81)$$

4 Further Refinements

The basic model with incomplete market analyzed in the preceding sections is adequate for studying the equity market in isolation. One main objective of this paper, however, is to explain the empirical regularities of the equity market and the bond market simultaneously. A careful treatment of the bond market requires some important extensions to the basic model.

Due to the presence of the Jensen terms, the drift of the surplus consumption will not be zero, in general, when evaluated at $\bar{z} = h(\bar{x})\bar{\omega}$, and \bar{x} :

$$\mu_z(\bar{z}, \bar{x}) = \mu_\omega(\bar{x}) + \frac{1}{2} \frac{h''(\bar{x})}{h(\bar{x})} \beta^2 \bar{z}^2 + \frac{h'(\bar{x})}{h(\bar{x})} (\delta g(\bar{x}) - \beta h(\bar{x}) f'(\bar{x})) \beta \bar{z}, \quad (74)$$

$$= \frac{1}{2} \frac{h''(\bar{x})}{h(\bar{x})} \beta^2 \bar{z}^2 + \frac{h'(\bar{x})}{h(\bar{x})} (\delta g(\bar{x}) - \beta h(\bar{x}) f'(\bar{x})) \beta \bar{z}. \quad (75)$$

Note, however, that as $\beta \rightarrow 0$, the Jensen terms vanish. In this limit, both the surplus consumption and the surplus wealth have zero drift at the steady state means.

¹³Economies that violate the steady state conditions are not ruled out by our model. However, we must take a stand on whether the U.S. economy has been in a steady state or not, when we bring the model to bear with empirical observations. It seems sensible, and certainly within the economic mainstream, to assume that the U.S. economy has been in a steady state – after detrending and rescaling, as usual.

First, we need to determine the market price of risk for the interest rate risk. The model with incomplete market only determines the price of the interest rate shock parallel to the production shock. In order to determine equilibrium price of interest rate risk orthogonal to the production shock, we need to extend our model to a complete market setting, where bonds are brought into the agent's investment opportunity set explicitly. This issue does not arise in deterministic habit formation models of Sundaresan (1989), Constantinides (1990), or Detemple and Zapatero (1991), which have only a single source of uncertainty, or in Campbell and Cochrane (1999), in which the habit shock and the production shock are perfectly correlated – effectively there is only one source of uncertainty as well.

A second extension is to add expected inflation and inflation risk to our model. This is necessitated by the fact that the literature on expectations theory has traditionally focused on empirical regularities on nominal yields.

A third extension is more subtle: we add another source of volatility to the riskless rate, in addition to the habit shock. The reason we need to explicitly decompose the volatility of the riskless rate into two components, one related to the habit shock, and one does not, is because attributing the volatility of the riskless rate entirely to habit is both counter-factual and destabilizing: the habit shock would be too volatile for the economy to stay near the steady state mean.

4.1 Stochastic Habit Formation Model with Complete Market

Consistent with the assumption that the agent is infinitely lived, we introduce a console bond into the agent's investment opportunity set. Let ψ be the constant rate of coupon flow, and p be the ex-coupon price of the console bond. Without loss of generality, its stochastic development can be written as

$$\frac{dp + \psi dt}{p} = \mu_p dt + \sigma_p \left(\delta_p dB_s + \sqrt{1 - \delta_p^2} dB_\perp \right). \quad (82)$$

We now assume that the agents invest α in the production technology, α_p in the console bond, and the rest in the riskless technology. The budget constraint becomes

$$\begin{aligned} dw = & \{[(\mu_s - r)\alpha + (\mu_p - r)\alpha_p + r] w - c\} dt + \alpha \sigma_s w dB_s \\ & + \alpha_p \sigma_p w \left(\delta_p dB_s + \sqrt{1 - \delta_p^2} dB_\perp \right). \end{aligned} \quad (83)$$

Thus, the Hamilton-Jacobian-Bellman equation now becomes

$$\begin{aligned} 0 = & \max_{c, \alpha, \alpha_p} \frac{(c - x)^\gamma}{\gamma} - \rho J + \{[(\mu_s - r)\alpha + (\mu_p - r)\alpha_p + r] w - c\} J_w \\ & + \frac{1}{2} [(\alpha \sigma_s + \alpha_p \delta_p \sigma_p)^2 + \alpha_p^2 (1 - \delta_p^2) \sigma_p^2] w^2 J_{ww} \\ & + \left[(\alpha \sigma_s + \alpha_p \delta_p \sigma_p) \delta \sigma_x + \alpha_p \sqrt{1 - \delta_p^2} \sigma_p \sqrt{1 - \delta^2} \sigma_x \right] w J_{wx} \end{aligned} \quad (84)$$

The first order conditions for α and α_p are:

$$0 = (\mu_s - r) w J_w + (\alpha \sigma_s + \alpha_p \delta_p \sigma_p) \sigma_s w^2 J_{ww} + \delta \sigma_s \sigma_x w J_{wx}, \quad (85)$$

$$0 = (\mu_p - r) w J_w + (\alpha \sigma_s + \alpha_p \delta_p \sigma_p) \delta_p \sigma_p w^2 J_{ww} + \left(\delta_p \delta + \sqrt{1 - \delta_p^2} \sqrt{1 - \delta^2} \right) \sigma_p \sigma_x w J_{wx} \quad (86)$$

Solving, we obtain

$$\alpha = \frac{g - \delta_p g_p}{\sigma_s(1 - \delta_p^2)}, \quad (87)$$

$$\alpha_p = \frac{g_p - \delta_p g}{\sigma_p(1 - \delta_p^2)}, \text{ where,} \quad (88)$$

$$g = \frac{\Lambda}{1 - \gamma} + \delta \beta h(x) f'(x), \quad (89)$$

$$g_p = \frac{\Lambda_p}{1 - \gamma} + \left(\delta_p \delta + \sqrt{1 - \delta_p^2} \sqrt{1 - \delta^2} \right) \beta h(x) f'(x), \quad (90)$$

where $\Lambda_p = \frac{\mu_p - r}{\sigma_p}$ is the Sharpe ratio of the console bond.

To determine the equilibrium risk premium on the habit shock, we assume that the console bond is in zero supply, that is, $\alpha_p = 0$ in equilibrium. This implies

$$g_p = \delta_p g, \quad (91)$$

or equivalently

$$\Lambda_p = \delta_p \Lambda - \sqrt{1 - \delta_p^2} \sqrt{1 - \delta^2} \beta h(x) f'(x). \quad (92)$$

In general, the stochastic development of the pricing kernel, or the state price $m(t)$, of the model can be written as

$$\frac{dm(t)}{m(t)} = -r dt - \Lambda dB_s - \Lambda_\perp dB_\perp, \quad (93)$$

where the drift of $(-r)$ comes from the fact that a bank account deflated by the state price must be a Martingale. Furthermore, the diffusion function associated with the production shock is $(-\Lambda)$ because the equilibrium wealth process must be the price of the consumption stream, therefore the *deflated gain process*, $w(t)m(t) + \int_0^t c(s)m(s)ds$, must be a Martingale.¹⁴

Since p is the price of the coupon stream ψ , $p(t)m(t) + \int_0^t \psi m(s)ds$ must also be a Martingale. This implies that the Sharpe ratio of the console bond must be

$$\Lambda_p = \delta_p \Lambda + \sqrt{1 - \delta_p^2} \Lambda_\perp. \quad (94)$$

Equation (94) may be viewed as a version of the consumption based capital asset pricing model (CCAPM) applied to the bond market.

Comparing equations (92) and (94), we conclude that

$$\Lambda_\perp = -(1 - \gamma) \sqrt{1 - \delta^2} \beta h(x) f'(x). \quad (95)$$

This is the equilibrium price of risk associated with the random shock B_\perp .

Once we impose the market clearing condition for the console bond, the rest of the model is exactly the same as before. That is, the functions $r(x)$, $\Lambda(x)$, $f(x)$, $h(x)$, and $g(x)$, will

¹⁴See Duffie (1996).

all be the same as before. The only effect of introducing a zero supply console bond into the agent's investment opportunity set is to allow us to determine the equilibrium price for the habit shock.

The pricing kernel is now completely determined and is unique in the presence of a bond market:

$$\frac{dm(t)}{m(t)} = -r(x)dt - \Lambda(x)dB_s + \sqrt{1 - \delta^2}\beta h(x)f'(x)dB_\perp, \quad (96)$$

This pricing kernel is strictly positive. As pointed out by Chapman (1998), this is not the case for endowment economies such as Detemple and Zapatero (1991). Furthermore, Chapman (1998)'s critique on the production economy with deterministic habit formation (Constantinides (1990)) does not apply to our model, since the investment opportunity set in our model is both *stochastic* and *realistic*.

4.2 A Nominal Economy with Real Inflation Risk

The model we developed so far describes a real economy. In order to make contact with the empirical literature on expectations theory, which has historically only dealt with nominal returns, we need to bring inflation into the picture.

We start by setting up the represent agent's problem in nominal terms

$$\max_{\{C^\S(t), \alpha(t): 0 \leq t \leq \infty\}} E_0 \int_0^\infty e^{-\rho^\S t} \frac{(C^\S - X^\S)^\gamma}{\gamma}, \quad (97)$$

subject to the budget constraint:

$$dW^\S(t) = \{[(\mu_S^\S - R^\S)\alpha(t) + R^\S]W^\S(t) - C^\S(t)\}dt + \sigma_S^\S \alpha(t)W^\S(t)dB_s(t), \quad (98)$$

where

$$dX^\S(t) = \left[b_X(C^\S(t) - X^\S(t)) + \frac{\epsilon_X^\S \beta_X^2}{2}(C^\S(t) - X^\S(t))^2 - (a_X - b_X)X^\S(t) \right] dt + \beta_X(C^\S(t) - X^\S(t))dB_x(t). \quad (99)$$

We assume that the price level, $\Pi = \frac{W^\S}{W}$, has the following dynamics:¹⁵

$$d \log \Pi = \pi dt + \sigma_\pi dB_\pi, \quad \Pi(0) \equiv 1, \quad (100)$$

where π is the expected inflation, and dB_π is the unexpected inflation shock, $\delta_{\pi s} = \text{corr}(dB_\pi, dB_s)$, $\delta_{\pi x} = \text{corr}(dB_\pi, dB_x)$.

Again, a dimension analysis reveals that it is appropriate to assume

$$\epsilon_X^\S = \epsilon e^{-Gt} \Pi^{-1}, \quad (101)$$

¹⁵The conditional volatility is assumed to be proportional to the surplus consumption, so that the nominal model allows a homothetic solution. Both π and β_π can potentially be state-dependent.

where ϵ is a constant. Define the *detrended*, *rescaled*, and *deflated* processes by

$$\begin{aligned} w(t) &= \epsilon e^{-Gt} \Pi^{-1}(t) W(t), \\ s(t) &= \epsilon e^{-Gt} \Pi^{-1}(t) S(t), \\ d(t) &= \epsilon e^{-Gt} \Pi^{-1}(t) D(t), \\ c(t) &= \epsilon e^{-Gt} \Pi^{-1}(t) C(t), \\ x(t) &= \epsilon e^{-Gt} \Pi^{-1}(t) X(t). \end{aligned} \tag{102}$$

In terms of the lower case variables, the model can be written as

$$\max_{\{c(t), \alpha(t): 0 \leq t \leq \infty\}} E_0 \int_0^\infty e^{-\rho t} \frac{[c(t) - x(t)]^\gamma}{\gamma} dt, \tag{103}$$

subject to the constraint:

$$\begin{aligned} dw(t) &= \{[(\mu_s - r)\alpha(t) + r] w(t) - c(t)\} dt \\ &+ \sigma_s \alpha(t) w(t) dB_s(t) - \sigma_\pi w dB_\pi(t), \end{aligned} \tag{104}$$

where $\rho = \rho^s - \gamma G - \gamma \log \Pi$, $\mu_s = \mu_S - G - \pi$, and $r = R^s - G - \pi$.

The habit process, x , is now governed by

$$\begin{aligned} dx &= \left[\left(b(c - x) + \frac{\beta_x^2}{2} (c - x)^2 \right) - (a - b)x \right] dt \\ &+ \beta_x (c - x) dB_x(t) - \sigma_\pi x dB_\pi(t), \end{aligned} \tag{105}$$

where $a = a_X + G + \pi$, $b = b_X$, and $\beta_x = \beta_X$.

Note that the expected inflation, π , drops out of the real problem. Thus, no matter whether the inflation risk has real effect, the real model does not impose any restriction on the expected inflation.

The only difference between the nominal problem (103), (104), and (105) and the real problem (11), (12), and (9) is the appearance of an additional inflation shock in the wealth process (104) and the habit process (105). If inflation shock is absent, that is, if inflation is completely predictable, then inflation does not have any real effect – we recover exactly the real economy described in previous sections.

When the inflation shock is non-trivially correlated with the production and habit shocks, inflation has important real effects. With a suitable choice of the conditional volatility, σ_π , for the inflation shock, e.g.,

$$\sigma_\pi = \beta_\pi \frac{c(t) - x(t)}{x(t)}, \tag{106}$$

we can solve the problem exactly.

Under the specification¹⁶ (106), the solution to the model has the same form of the indirect utility function as before, i.e., equation (14). However, the curvature of the habit

¹⁶There are alternative specifications of σ_π that allow well-defined solutions. The question of which alternative is empirically more relevant will be deferred to future research. For our purpose, it suffices to say that it does not have any material effect on the expectations puzzle, which has more to do with the state-dependency of the expected inflation, π , rather than the exact specification of the inflation shock.

level of wealth, $f(x)$, is modified:

$$f(x) = \eta (1 - e^{-qx}), \quad (107)$$

where

$$q = \frac{\beta_x^2}{\beta_\pi^2 - 2\delta_{\pi x}\beta_\pi\beta_x + \beta_x^2}. \quad (108)$$

The riskless rate is still determined by (25).

The optimal consumption and portfolio policies now take the form

$$c^* - x = h(x, w)(w - f(x)), \quad (109)$$

$$\alpha^* = \frac{g(x, w)}{\sigma_s} \frac{w - f(x)}{w}, \quad (110)$$

where the functions $h(x, w)$ and $g(x, w)$ depend on both controlled state variables. Accordingly, the Sharpe ratio of the production technology now depends on both habit and wealth. Furthermore, a hedging demand for the production technology is induced by the correlation between the inflation shock and the production shock, $\delta_{\pi s}$.

Using this model, we can study the real effects of monetary policy in an equilibrium setting. The model also gives rise to a three-factor model for the term structure of nominal interest rates: production, habit, and inflation. Empirical implications of this model will be examined elsewhere.

4.3 An Additional Source of Riskless Rate Volatility

To introduce an additional source of risk to the riskless rate, we can add an additional term, $y \frac{w}{f'(x)}$, to the conditional mean of the habit process:

$$dx(t) = \mu_x dt + \sigma_x dB_x(t), \quad (111)$$

where,

$$\begin{aligned} \mu_x &= \left(y \frac{w}{f'(x)} + b(c(t) - x(t)) + \frac{\beta^2}{2}(c(t) - x(t))^2 \right) - (a - b)x(t), \\ \sigma_x &= \beta(c(t) - x(t)), \end{aligned} \quad (112)$$

The rest of the model is kept the same.

It is easy to show that, as long as y does not depend on the stochastic controls, c and α , the solution to the extended model is the same as before except that the riskless rate acquires an additional term:

$$\begin{aligned} \hat{r} &= r + y, \\ r &= \frac{1 - (a - b)f(x)}{f(x)/x}. \end{aligned} \quad (113)$$

Thus, we have effectively added another factor to the riskless rate. Our model does not impose any restriction on the specific form of y . For illustrative purposes, let us suppose that y is a slowly mean-reverting process with zero long run mean:

$$dy = -\kappa_y dt + \sigma_y dB_y, \quad (114)$$

and B_y is independent of either B_x or B_s .

It follows that the steady state mean of \hat{r} is the same as that of r . The assumption that B_y is totally idiosyncratic implies that all local correlations are not affected by the presence of this additional factor.

Furthermore, the unconditional variance of \hat{r} is given by

$$\text{var}(\hat{r}) = \text{var}(r) + \text{var}(y) \quad (115)$$

Under an one-factor Gaussian approximation of r , $\text{var}(r) = \frac{\sigma_r^2}{2\kappa} = \frac{\beta^2 \bar{r}_x^2 \bar{z}^2}{2\kappa}$. Thus, a small β can be easily accommodated with the observed volatility $\text{var}(\hat{r})$.

5 Explaining Equity Market Puzzles

In this section, we explore implications of our model to empirical puzzles associated with Lucas-style asset pricing model with traditional time-separable utility: the *equity premium puzzle* of Mehra and Prescott (1985), the *stock market volatility puzzle* of Campbell (1999), and the *riskfree rate puzzle* of Weil (1989). The latter two are closely related to the equity premium puzzle. We will refer to these three puzzles collectively as *equity market puzzles*.

The analysis begins with a general discussion of all three puzzles, and culminates in a calibration of model parameters under which the equity market puzzles are explicitly resolved.

Throughout this section, we will work under the limit $\beta \rightarrow 0$. Thus extensions discussed in the last section are irrelevant. All theoretical relations come from the incomplete market model. We will discuss implications for the bond market toward the end of the section.

5.1 Equity Premium Puzzle

We characterize the equity premium puzzle in terms of the volatility ratio of the consumption growth rate and the pricing kernel rate along the line of Hansen and Jagannathan (1991) and Campbell (2000).

From equation (23), the conditional volatility of the consumption growth rate is given by

$$\sqrt{\sigma_{cs}^2 + \sigma_{cx}^2 + \delta \sigma_{cs} \sigma_{cx}},$$

where $\sigma_{cs} = \frac{c-x}{c} g(x)$, and $\sigma_{cx} \propto \beta$. From equation (96), the conditional volatility of the pricing kernel is given by

$$\sqrt{\Lambda^2 + (1 - \delta^2) \beta^2 h(x)^2 f'(x)^2}.$$

In the limit $\beta \rightarrow 0$, the two volatilities simplify to $\sigma_{cs} = \frac{c-x}{c}g(x) = \frac{c-x}{c} \frac{\Lambda}{1-\gamma}$ and Λ , respectively. Thus,

$$\frac{\text{std}_t\left(\frac{dc}{c}\right)}{\text{std}_t\left(\frac{dm}{m}\right)} = \frac{1}{1-\gamma} \frac{c-x}{c}. \quad (116)$$

Evaluated at the steady state means, we have

$$\frac{\text{std}_t\left(\frac{dc}{c}\right)}{\text{std}_t\left(\frac{dm}{m}\right)} = \frac{1}{1-\gamma} \frac{a-b}{a}. \quad (117)$$

Note that the parameters a and b play a crucial role here. In particular, the mean reversion coefficient $a-b$ of the habit process is intimately tied to the equity premium puzzle. Later, we will show that $a-b$ also emerges as the mean reversion coefficient for the riskless rate, and plays an equally important role in explaining the expectations puzzle.

It is also important to note that the equity premium puzzle does not identify a and b individually: given γ , and the empirical observation of the left hand side of equation (116), only the ratio $\frac{a}{b}$ is identified. The six groups of parameters reported by Constantinides (1990) are characterized by different pairs of (a_X, b_X) . It turns out that these different pairs of (a_X, b_X) imply roughly the same ratio of a and b : $\frac{b}{a} \approx 0.8$. If we take $1-\gamma = 2.2$, as in Constantinides (1990), the ratio of conditional volatilities of the consumption growth rate and the pricing kernel is about 9%, a quite reasonable value.

Without habit, the power utility function implies that

$$\frac{\text{std}_t\left(\frac{dc}{c}\right)}{\text{std}_t\left(\frac{dm}{m}\right)} = \frac{1}{1-\gamma}. \quad (118)$$

The empirically observed low volatility of consumption growth rate relative to the lower bound of the volatility of the pricing kernel implies that $1-\gamma$ must be very large. Table 1 implies $1-\gamma = 24.1$ – hence the equity premium puzzle.

5.2 Stock Market Volatility Puzzle

In equilibrium, the conditional volatility of the wealth growth rate $\frac{dw}{w}$ is

$$\text{std}_t\left(\frac{dw}{w}\right) = \sigma_s = \frac{w-f(x)}{w}g(x).$$

and the conditional volatility of the consumption growth rate is, in the limit $\beta \rightarrow 0$,

$$\text{std}_t\left(\frac{dc}{c}\right) = \sigma_{cs} = \frac{c-x}{c}g(x).$$

Thus, in the limit $\beta \rightarrow 0$,

$$\frac{\text{std}_t\left(\frac{dc}{c}\right)}{\text{std}_t\left(\frac{dw}{w}\right)} = 1 - \frac{x-h(x)f(x)}{c}. \quad (119)$$

We will see later that, at the steady state mean, $x = \bar{x}$, and in the limit $\beta \rightarrow 0$,

$$x > h(x)f(x).$$

Both $h(x)$ and $f(x)$ are concave, so that as x increases, the wedge widens.

To understand the intuition, let us link the above volatility ratio to the relative risk aversion, RRA , and the elasticity of consumption substitution, ECS .

Following Constantinides (1990), we define the relative risk aversion coefficient as the scaled curvature of the indirect utility function with respect to wealth:

$$RRA = -w \frac{J_{ww}}{J_w} = (1 - \gamma) \frac{w}{w - f(x)}, \quad (120)$$

and the elasticity of consumption substitution as the inverse of the scaled curvature of the utility function with respect to consumption

$$ECS^{-1} = -c \frac{u_{cc}}{u_c} = (1 - \gamma) \frac{c}{c - x}, \quad (121)$$

where $u = \frac{(c-x)^\gamma}{\gamma}$. Thus both the RRA and ECS are time-varying and can be substantially larger than $1 - \gamma$.

From the above, we obtain

$$ECS \cdot RRA = \frac{c - x}{c} \frac{w}{w - f(x)} = 1 - \frac{x - h(x)f(x)}{c} = \frac{\text{std}_t \left(\frac{dc}{c} \right)}{\text{std}_t \left(\frac{dw}{w} \right)}. \quad (122)$$

Thus, the low volatility of the consumption growth rate relative to the volatility of the total returns to wealth is equivalent to the product of RRA and ECS being substantially smaller than 1. In the time-separable power utility models, $RRA = ECS^{-1} = (1 - \gamma)^{-1}$. Thus, $ECS \cdot RRA = 1$ – hence the stock market volatility puzzle. The presence of habit inserts a wedge between RRA and ECS^{-1} .

In calibrating this puzzle to the data, we need to be very careful about the distinction between the “volatility of equity return” and the “volatility of wealth return”. The equity return and its volatility are directly observed and easily measured, while the return to wealth and its volatility are not. It is therefore tempting to treat the stock market volatility as the same as the volatility of wealth growth rate. We will see that this appears innocuous if the bond market is artificially delinked from the equity market. However, as soon as we bring in moment restrictions implied by the expectations puzzle, the problem of equating the volatility of equity return with the volatility of wealth return becomes transparent.

In the end, we will argue that the stock market volatility puzzle need not be a puzzle in our model, once we make the distinction between the production technology and the stock market.

The equity premium puzzle is immune to this problem: under the plausible assumption that the market price of risk (Sharpe ratio) is the same for both the equity market and the aggregate wealth, the equity premium puzzle is not affected by the difference between the volatility of the stock market and the volatility in the aggregate wealth.

5.3 Riskfree Rate Puzzle

The riskfree rate puzzle is resolved in the same way it is resolved in Constantinides (1990) and Campbell and Cochrane (1999): in order to explain the equity premium puzzle, we no longer require that $1 - \gamma$ be a large number. The riskfree rate, or riskless rate, can therefore be consistent with the empirically observed mean value of about 1% with a reasonable value of the subjective discount rate, ρ .

Furthermore, the riskless rate is only weakly correlated with the production shock: $\delta\beta \ll 1$. Thus, our model is consistent with the fact that the riskless rate does not vary significantly across time and across countries with consumption shocks.

5.4 Calibrating to the Equity Return and Consumption Data

A natural strategy for calibrating the model parameters is to separate them into two groups. The first group consists of $(G, \epsilon, \rho, \gamma, a, b, \bar{x})$, which defines the model in the limit $\beta \rightarrow 0$. All properties of economy at its steady state mean are determined by these parameters. The second group consists of (β, δ) , both of which are associated with the stochastic nature of the habit process. Interest rate dynamics is not trivial only if $\beta \neq 0$.

Throughout this subsection, we will assume that $\beta = 0$. Thus, the second group of parameters and therefore the term structure dynamics are irrelevant. Accordingly, we will calibrate the first group of parameters using the equity and consumption data only (except for the steady state mean of the riskless rate, which is needed to compute the risk premium and Sharpe ratio). We will, however, discuss implications of the calibrated values for the first group of parameters for the interest rate dynamics, as some important aspects of the term structure dynamics are also determined by parameters in the first group.

In order to make contact with empirical observations, we must make the necessary assumption that the U.S. economy has been fluctuating around its steady state mean, so that the sample means of the investment opportunity set: $r(x)$ and $\Lambda(x)$, should correspond to theoretical values evaluated at the steady-state mean¹⁷ of the consumption habit, \bar{x} .

To assist the discussion, let us define the following auxiliary variables.

$$\begin{aligned}
 \bar{r} &= r(\bar{x}), & \bar{r}_x &= r'(\bar{x}), \\
 \bar{f} &= f(\bar{x}) = \eta(1 - e^{-\bar{x}}), & \bar{f}_x &= f'(\bar{x}) = \eta e^{-\bar{x}}, \\
 \bar{h} &= \lim_{\beta \rightarrow 0} h(\bar{x}) = h^0(\bar{x}), & \bar{h}_x &= \lim_{\beta \rightarrow 0} h'(\bar{x}) = h_x^0(\bar{x}), \\
 \bar{g} &= \lim_{\beta \rightarrow 0} g(\bar{x}) = g^0(\bar{x}), & \bar{g}_x &= \lim_{\beta \rightarrow 0} g'(\bar{x}) = g_x^0(\bar{x}), \\
 \bar{\Lambda} &= \lim_{\beta \rightarrow 0} \Lambda(\bar{x}) = (1 - \gamma)\bar{g}, & \bar{\Lambda}_x &= \lim_{\beta \rightarrow 0} \Lambda'(\bar{x}) = (1 - \gamma)\bar{g}_x.
 \end{aligned}$$

¹⁷Throughout this section, we will ignore all Jensen terms, justified by the assumption that β is small.

5.4.1 Sample Observations

Among the variables defined above, only \bar{r} and $\bar{\Lambda}$ are directly observed (see Table 1):

$$\bar{r} = \bar{R} - G = 0.94\% - 1.89\% = -0.95\%, \quad (123)$$

$$\bar{\Lambda} = \frac{\bar{\mu}_S - \bar{R}}{\bar{\sigma}_S} = \frac{7.85\% - 0.94\%}{15.92\%} = 0.4341. \quad (124)$$

The rest are determined by the the model endogenously, given the first group of parameters. All relevant sample observations are summarized in the following table:

Table 1: Observables

The Sharpe ratios for bonds and volatility of real riskless rate are taken from Campbell and Viceira (2000), and the rest are taken from Campbell and Cochrane (1999), for the post-war sample period. Except for the sample means $\bar{R} = \bar{r} + G$, and $\mu_S = \mu_s + G$, all other observed variables are scale and trend free – that is, they are the same in both the actual economy and the rescaled, detrended economy. The Sharpe ratio for nominal bonds is the average of the Sharpe ratios for 1-year (0.246), 3-year (0.141), and 10-year (0.08) bonds. The volatility of the pricing kernel is the same as the Sharpe ratio on equity.

Steady-state Means		
Growth rate	G	1.89%
Riskless rate	\bar{R}	0.94%
Expected equity return	$\bar{\mu}_S$	7.85%

Volatilities		
Real riskless rate	$\Sigma_r = \frac{\sigma_r}{\sqrt{2\kappa}}$	1.01%
Stock return	$\sqrt{\text{var}(ds/s)/dt}$	15.92%
Consumption growth	$\sqrt{\text{var}(dc/c)/dt}$	1.50%
Pricing kernel	$\sqrt{\text{var}(dm/m)/dt}$	43.41%

Sharpe Ratios		
Sharpe Ratio for indexed bonds	$\bar{\Lambda}_p$	0.374
Sharpe Ratio for nominal bonds	$\bar{\Lambda}_p^s$	0.156

5.4.2 Calibration

We require that the parameters of our model satisfy the following specific restrictions:

- (I) Equations (71) and (72) are satisfied at the steady state means; and the steady-state conditions are satisfied;

- (II) \bar{r} and $\bar{\Lambda}$ are matched exactly;
- (III) Equity premium and stock volatility puzzles are resolved explicitly.
- (IV) R_0 is close to zero.
- (V) Admissibility conditions:

$$\begin{aligned} a + G &> b, \quad \gamma < 1, \quad \Delta(\bar{x}) \geq 0, \\ \rho - \gamma\bar{r} - (1 - \gamma)(1 + b\bar{f}_x)\bar{h} &> 0. \end{aligned} \tag{125}$$

(I) and (II) are endogenous restrictions derived from the equilibrium solution, (III) is an empirical restriction associated with the equity premium and stock market volatility puzzles, (IV) is a plausibility restriction related to the riskfree rate puzzle (see below), and (V) are parametric restrictions that ensure the model is well-defined at or near the steady state mean.¹⁸

Restriction (I) is satisfied by setting ρ and \bar{z} according to

$$\rho = \bar{r} + \frac{2 - \gamma}{2(1 - \gamma)}\bar{\Lambda}^2, \tag{126}$$

$$\bar{z} = \frac{a - b}{b}\bar{x}. \tag{127}$$

The latter implies that $\bar{c}/\bar{x} = a/b$.

Restriction (II) is satisfied by setting r_0 and \bar{h} according to

$$r_0 = \left(\frac{\bar{r}}{\bar{x}} - (a - b) \right) (1 - e^{-\bar{x}}), \tag{128}$$

$$\bar{h} = \frac{\rho - \gamma\bar{r} - \frac{1}{2}\frac{\gamma}{1-\gamma}\bar{\Lambda}^2}{(1 - \gamma)(1 + b\bar{f}_x)}. \tag{129}$$

Finally, to explain the equity market puzzles, that is, the restriction (III), we construct unconditional sample counterparts to equations (116) and (119), evaluated at the steady state means:

$$\textbf{Premium:} \quad \frac{1}{1 - \gamma} \frac{a - b}{a} \approx \frac{\sqrt{\text{var}(\frac{dc}{c})/dt}}{\bar{\Lambda}} = \frac{0.015}{0.4341} = 0.03455, \tag{130}$$

$$\textbf{Volatility:} \quad 1 - \frac{b}{a}(1 - \bar{h}\bar{f}/\bar{x}) \approx \frac{\sqrt{\text{var}(\frac{dc}{c})/dt}}{\sqrt{\text{var}(\frac{dw}{w})/dt}} = \frac{0.015}{0.1592} = 0.09428. \tag{131}$$

¹⁸ $a_X > b_X$ ensures that the habit process does not explode. $\gamma < 1$ ensures that the investor is risk averse. $\Delta(\bar{x}) \geq 0$ ensures that risk premium exists. Since $\Delta(x)$ decreases with x , this condition is equivalent to the steady-state mean condition: $\bar{x} < x_{max}$. $\rho - \gamma\bar{r} - (1 - \gamma)(1 + b\bar{f}_x)\bar{h} > 0$ ensures that the transversality condition is satisfied. By continuity, the transversality condition is satisfied near the steady-state mean as well.

Note that an important assumption is made in deriving the parametric restriction (131): we are assuming that the volatility of the equity return is the same as the volatility of growth rate of the aggregate wealth. In another word, we are assuming that the production technology is identified with the stock market. There are serious problems in this assumption. First, this assumption implies that the stock dividend is the same as the aggregate consumption, which is clearly counter-factual.¹⁹ Second, in Section 7, we will show that this assumption is incompatible with the expectations puzzle. To best illustrate the point, we proceed by following the tradition of the literature and maintaining this assumption for the time being.

Restriction (IV) is added to ensure that the observed riskfree rate is associated with a plausible value of the riskless rate at zero habit level. It turns out that the ability of the model to explain the equity premium and the stock market volatility puzzles, and most qualitative properties of the model are not sensitive to the choice of γ and \bar{x} , as long as a and b are suitably chosen given γ and \bar{x} , even if we do not impose the restriction (IV). However, some combinations of \bar{x} , a , and b may lead to implausible values of R_0 (for example, -6%), *given* the observed steady state value of $\bar{R} = 0.94\%$ (hence the link to the riskfree rate puzzle). To formalize (IV) in the calibration, we require that

$$r_0 + G = 0, \quad (132)$$

which effectively ties down \bar{x} . We could fix $r_0 + G$ to any plausible values other than 0 but less than $\bar{r} + G = 0.94\%$, the behavior of the economy would be qualitatively the same.

It is easy to see that once (I) and (II) are satisfied, there are four parameters left to fit the equity market puzzles. γ , \bar{x} , a , and b , with three empirical restrictions (130), (131), and (132). Thus, one of the four parameters is unidentified. This indeterminacy is a welcome relief: it means that we can pick freely any reasonable value of γ .

In choosing the parameters to fit the equity market puzzles, we also need to impose the constraints $a_X = a + G > b_X = b$ and $\gamma < 1$ explicitly. The presence of these constraints imply that γ need not be completely unidentified: for some values of γ , the constraints may be binding.

Thus, we adopt the following basic strategy for model calibration: search over all possible combinations of the four parameters, (γ, \bar{x}, a, b) , by minimizing the least square criterion function:

$$L_1 = k_1 \left(\frac{1}{1-\gamma} \frac{a-b}{a} - 0.03455 \right)^2 + k_2 \left(1 - \frac{b}{a} (1 - \bar{h}\bar{f}/\bar{x}) - 0.09428 \right)^2 + k_3 (r_0 + G)^2, \quad (133)$$

where $k_i > 0$, $i = 1, 2, 3$, are fixed constants, subject to the constraints: $a + G > b$ and $\gamma < 1$.

Although the equity market puzzles are often stated in terms of the volatility ratios, they can be stated directly in terms of the volatilities themselves. Since the volatility of the pricing kernel is given, fitting the equity premium puzzle is the same as fitting the volatility of the consumption growth rate. Once the volatility of the consumption growth rate is fit,

¹⁹Campbell and Cochrane (1999) show that the growth rates of the aggregate stock dividend and the aggregate consumption are very weakly correlated, and the former has a much higher volatility than the latter.

fitting the stock market volatility puzzle is the same as fitting the volatility of the equity, or wealth return. Thus, an alternative and equivalent criterion function is

$$L_2 = k_1 \left(\frac{\bar{c} - \bar{x}}{\bar{c}} \bar{g} - 1.5\% \right)^2 + k_2 \left(\frac{\bar{w} - \bar{f}}{\bar{w}} \bar{g} - 15.92\% \right)^2 + k_3 (r_0 + G)^2, \quad (134)$$

where $\bar{g} = \frac{\bar{\lambda}}{1-\gamma}$.

The calibration result is reported in column EM:I of Table 2. We use the criterion function L_2 , with weights k_i , $i = 1, 2, 3$, to ensure different parts are scaled properly, or emphasized as we choose. The results are almost identical if we used L_1 as our criterion function.

Comparing rows 5–6 (for volatility ratios) or 7–8 (for volatilities) of column 2 and column EM:I, we see that the equity market puzzles are fitted almost exactly!

The optimal value for γ is -0.565 , which conforms with the popular notion that the risk aversion parameter should not be too large. Column EM:II is identical to Column EM:I, except that γ is fixed at -1 , a value comparable to that used by Constantinides (1990) and Campbell and Cochrane (1999). Comparing rows 5–6 or 7–8 of column 2 and column EM:II, we see that the fit the equity market puzzles are nearly perfect as well, indicating that the model is not sensitive to the choice of γ , within some range. In both EM:I and EM:II, the riskless rate at zero habit, R_0 , is practically zero.

The surplus consumption ratio is 5.45% when $\gamma = -0.565$ in column EM:I, and 6.95% when $\gamma = -1$ in column EM:II. These calibrated steady state values for the surplus consumption ratio are remarkably close to the steady state value of 5.7% calibrated by Campbell and Cochrane (1999).

If empirical regularities in the equity market were all we are concerned about, or if we believed that empirical regularities in the equity market and those in the bond market are not related, then the story ends here – we have completely resolved all three puzzles in the equity market, simultaneously!

Is it too good to be true? In some sense, it is. By ignoring moment restrictions from the term structure of interest rates, the model gets, unfairly perhaps, a free hand, in explaining all empirical regularities of the equity market, even though it is not supposed to explain some of them. One important example is the equity return volatility. The above calibration exercise is based on the assumption that the stock market volatility is the same as the conditional volatility of the production technology. This is an over-simplification. We will refine the notion of stock market volatility after we show how to bring in moment restrictions from the term structure of interest rates, in particular, from the expectations puzzle.

6 Term Structure of Interest Rates

In Section 3.2, we derived a two-factor model for the term structure of real interest rates, as part of the equilibrium solution of our stochastic habit formation model with incomplete market. In Section 4.1, we show how to complete the market by introducing a console bond, and derive the equilibrium risk premium for both sources of interest rate risk under the assumption that the console bond has a zero net supply. A complete term structure model is therefore endogenously determined in the complete market setting.

Table 2: Calibrating to Equity Market Puzzles

Parameter	Sample	EM:I	EM/EH:I	EM:II	EM/EH:II
γ		-0.5652	0.9396	-1.0000	-1.0000
\bar{x}		0.0686	0.4375	0.0441	0.4215
a		2.7084	14.4369	3.2184	0.4487
b		2.5618	14.4068	2.9960	0.4176
Premium: $\frac{1}{1-\gamma} \frac{a-b}{a}$	0.03455	0.0345	0.0345	0.0345	0.0346
Volatility: $1 - \frac{b}{a}(1 - \bar{h}\bar{f}/\bar{x})$	0.09428	0.0943	0.2721	0.0942	0.2938
$\sqrt{\text{var} \left(\frac{dc}{c} \right) / dt} = \frac{\bar{c}-\bar{x}}{\bar{c}} \bar{g}$	1.5%	1.50%	1.50%	1.49%	1.50%
$\sqrt{\text{var} \left(\frac{ds}{s} \right) / dt} = \frac{\bar{w}-\bar{f}}{\bar{w}} \bar{g}$	15.92%	15.90%	5.51%	15.91%	5.12%
$\frac{\bar{c}-\bar{x}}{\bar{c}}$		5.41%	0.20%	6.91%	6.93%
\bar{w}		1.2184	30.3872	0.7958	35.5445
\bar{f}		0.5197	30.1542	0.2122	27.1545
\bar{h}		0.0056	0.0039	0.0056	0.0037
\bar{g}		0.2773	7.1903	0.2170	0.2170
ϕ		168.72	0.0017	2101.2	3152.1
ρ		0.1449	1.6455	0.1318	0.1318
R_0		-0.0000	0.0005	-0.0000	0.0004
κ		0.1465	0.0301	0.2224	0.0311
$\lambda_1^{\$}$		0.0123	0.0195	0.0151	0.0203
\bar{r}_x		0.1368	0.0196	0.2128	0.0206
$\bar{\Lambda}_x$		-0.8731	-0.4349	-1.3449	-0.4680
λ_1		0.0297	0.0469	0.0363	0.0488

Parameters (γ, \bar{x}, a, b) are free parameters. The rest of the parameters are determined from the model endogenously. Parameters in Column EM:I are estimated using least square criterion L_2 subject to the admissibility conditions (125). Column EM/EH:I differ from Column EM:I by adding an additional requirement to fit the expectations puzzle. In Columns EM:II and EM/EH:II, γ is fixed at -1 , with (\bar{x}, a, b) chosen by the constrained least square fit.

$$\lambda_1 = -\sigma_r \frac{\Lambda_p \bar{\Lambda}_x}{\bar{\Lambda} \bar{r}_x}, \lambda_1^{\$} = \frac{\bar{\Lambda}_p^{\$}}{\bar{\Lambda}_p} \lambda_1.$$

If we take this model seriously, then empirical implications of our model is still widely open. In this section, we characterize the properties of the term structure dynamics and its implications for the expectations puzzle. We show that the theoretical link between the equity and bond markets is broken at an empirical level if we take as given parameters fitted to explain all three equity market puzzles.

To elaborate on the theoretical link between the two asset markets, we note that some parameters of our model are essential for explaining both equity market and bond market puzzles.

One such parameter is $\kappa = a - b$. It is the mean reversion parameter of the consumption habit process. We have seen that, at the steady state mean, the volatility of the consumption growth rate is proportional to κ . This parameter will also emerge as the mean reversion coefficient of the riskless rate. We will see that it plays a crucial role in explaining the expectations puzzle.

Another parameter is λ_1 , reported in the last row of Table 2. This parameter is the sensitivity of the market price of interest rate risk with respect to the real short rate, and determines the strength of time-varying risk premium. A related parameter, reported in row 16, is $\lambda_1^{\$}$, the sensitivity of the market price of interest rate risk with respect to the nominal short rate. We will give a full explanation of how λ_1 and $\lambda_1^{\$}$ are defined and computed in this section. As a preview, it is sufficient to mention that they are proportional to the ratio of $\bar{\Lambda}_x$ and \bar{r}_x , reported in rows 17 and 18, respectively, which can be computed from the model, even without the presence of a term structure dynamics. As discussed in Section 3.1, the ratio $\frac{\bar{\Lambda}_x}{\bar{r}_x}$ captures the fundamental local correlation between the Sharpe ratio and the riskless rate. We can see in Table 2 (and to be explained later) that the unambiguous negative sign of the ratio leads to an unambiguous positive sign for both λ_1 and $\lambda_1^{\$}$, which leads to the correct sign for the predictability of expected bond excess returns by term spread, which is ultimately related to the expectations puzzle.

To explore implications of this term structure model for the expectations puzzle, we will specialize to a one-factor approximation of the model. The approximation is justified on several grounds. First, the secondary factor, namely the surplus consumption, does not enter the riskless rate and the market price of risk directly. Rather, it enters bond prices only indirectly by controlling the conditional mean and conditional volatility of the riskless rate and the Sharpe ratio. Thus, its influence in bond yields is only gradually felt in the sense that the term structure of the factor loading with respect to the log surplus consumption, that is, the derivative of the bond yields with respect to the log surplus consumption, starts out at exactly zero and increases gradually. It will eventually be pulled toward zero by mean reversion.²⁰ In contrast, the term structure of the factor loading of bond yields with respect to the primary factor, namely the riskless rate, starts out at 1, by definition, and then declines gradually as maturity of the bond increases. Thus, the yield loading of the primary factor easily dominates that of the secondary factor, especially when the habit shock is relatively weak (i.e., $\beta \ll 1$). Second, Dai and Singleton (2000) make a strong case that a one-factor Gaussian model with linear market price of risk is capable of generating the requisite time-varying risk premium to explain the expectations puzzle. We will see that if we ignore the dependency of bond yields on the log surplus consumption as a zeroth order approximation, our exact two-factor term structure model reduces to exactly a one-factor Gaussian model with linear market price of risk! Third, the one-factor approximation not only substantially reduces the computational burden, but also help bring out the economic intuition of the expectations puzzle much more clearly. Working with the exact solution requires the characterization of a nonlinear bi-variate Markovian system of $r(x)$ and $c - x$, and the economic intuition will be masked by non-essential nonlinear terms. Finally, the one-factor approximation is ex post justified by the sensibility of the empirical results.

In the rest of the section, we begin by deriving the one-factor approximation to the real

²⁰See Dai and Singleton (1999).

riskless rate dynamics and the market price of risk. Expected inflation is added to generate a one-factor model of the nominal term structure, which is necessary in order to compare directly with the empirical results of Campbell and Shiller (1991). We will then apply the analysis of Dai and Singleton (2000) to link the resulting dynamic model to the expectations puzzle. Finally, we take parameters other than λ_1 and $\lambda_1^\$$ in column EM:I of Table 2 as given, and calibrate the term structure dynamics using sample moments of real and nominal bond returns. This allows us to calculate the values of λ_1 and $\lambda_1^\$$. We show that under these parameter values, the expectations puzzle is only partially explained, substantiating the claim at the end of Section 5. In Section 7, we re-calibrate our complete model to both the equity market and expectations puzzles, and establish the link, not only at the theoretical level, but also at the empirical level, the inextricable link between the equity and bond markets.

6.1 Linear Approximation of Term Structure Dynamics

We begin by pricing a zero coupon bond under the exact solution. Let p_T and $y_T = -\frac{1}{T} \log P_T$ be the price and yield, respectively, of a zero coupon bond with par of \$1 and maturity of T years. Under the exact solution, p_T or y_T is a function of the two state variables r and z .²¹ We define the loadings of y_T on the state variables by

$$b_r(T) = \frac{\partial y_T}{\partial r}, \quad b_z(T) = \frac{\partial y_T}{\partial \log z}. \quad (135)$$

Since $y_0 \equiv r$, it is necessarily the case that $b_r(0) = 1$ and $b_z(0) = 0$.

Using Ito's lemma, the stochastic development of p_T can be written as²²

$$\frac{dp_T}{p_T} = \mu_p dt + \frac{\partial \log p(r, z)}{\partial \log z} (\sigma_{zs}(x) dB_s + \sigma_{zx}(x) dB_x) + \frac{\partial \log p(r, z)}{\partial r} \sigma_r(x) dB_x \quad (136)$$

$$= \mu_p dt - \sigma_p \left(\delta_p dB_s + \sqrt{1 - \delta_p^2} dB_\perp \right), \quad (137)$$

where

$$\begin{aligned} \delta_p &= -\frac{\sigma_{ps}}{\sigma_p}, & \sigma_p &= \sqrt{\sigma_{ps}^2 + \sigma_{p\perp}^2}, \\ \sigma_{ps} &= -[b_r \delta \sigma_r + b_z (\sigma_{zs} + \delta \sigma_{zx})] T, & \sigma_{p\perp} &= -[b_r \sigma_r + b_z \sigma_{zx}] \sqrt{1 - \delta^2} T, \\ \sigma_r &= r'(x) \sigma_x, & \mu_p &= \frac{1}{p(r, z)} \mathcal{L} p(r, z), \end{aligned} \quad (138)$$

with σ_{zs} and σ_{zx} given by equations (58)–(59), and \mathcal{L} being the infinitesimal generate associated with the bi-variate Markovian process (r, z) .

²¹The primitive state variables are x and w . However, in equilibrium, with mild regularity conditions, there is a one-to-one mapping between the pair (x, w) and the pair (r, z) . Thus, we can think of (r, z) as the state-variables.

²²The negative sign in front of σ_p is a convention.

The partial differential equation for zero coupon bond prices can be determined by noting that $p_T m$ must be a Martingale, which implies (see equation 96)

$$\Lambda_p \equiv \frac{\mu_p - r}{\sigma_p} = -\delta_p \Lambda - \sqrt{1 - \delta_p^2} \Lambda_\perp, \quad (139)$$

which is the exact pricing equation for zero coupon bonds.

If $|b_z| \ll |b_r|$, we can approximate the bond price by ignoring its dependency on z . This is achieved by setting b_z to zero, and replacing z in equation (139) by its steady state mean, \bar{z} . This is equivalent to the following approximations to the riskless rate and the market price of risk:²³

$$dr_t \approx \kappa(\theta_r - r_t)dt + \sigma_r dB_x(t), \quad (140)$$

$$\lambda \approx -\sigma_r \Lambda_p = \sigma_r(\delta \Lambda + \sqrt{1 - \delta^2} \Lambda_\perp) \approx \sigma_r \delta \Lambda, \quad (141)$$

where $\kappa = a - b > 0$, $\theta_r = \bar{r} + \kappa^{-1}(b\bar{z} + \frac{1}{2}\beta^2\bar{z}^2)\bar{r}_x$, and $\sigma_r = \beta\bar{z}\bar{r}_x$.

The additional approximation on the right hand side of equation (141) is justified by the assumption that β is small, and δ is not too small (we will see that $\delta = -0.8616$ from sample moments). When β is small, the habit process will stay near the steady state mean. Thus, we can approximate the market price of risk further by linearizing it around the state state mean of x :

$$\lambda \approx \lambda_0 + \lambda_1 r, \quad (142)$$

where $\lambda_0 = \sigma_r(\bar{\Lambda} - \frac{\bar{r}}{\bar{r}_x}\bar{\Lambda}_x) < 0$, and

$$\lambda_1 = \delta\sigma_r \frac{\bar{\Lambda}_x}{\bar{r}_x} > 0. \quad (143)$$

Thus, near the steady state, the term structure of interest rates are characterized by a Gaussian short rate with linear market price of risk. Furthermore, the market price of interest rate risk is *positively* correlated with the short rate. This is exactly what Dai and Singleton (2000) use to explain the expectations puzzle. $\lambda_0 < 0$ is consistent with the fact that the historical means of the expected returns on zero coupon bonds, or the historical mean of the forward-spot spreads, are positive.²⁴

²³The market price of risk can be defined in terms of the effect of a measure change on the Brownian motion: under the risk neutral measure, the Brownian motion associated with the habit shock becomes $d\tilde{B}_x = dB_x + \frac{\lambda}{\sigma_r}dt$, where λ is defined as the market price of risk (associated with B_x). When $b_z = 0$, we have $\delta_p = \delta$. Consequently,

$$\frac{dp_T}{p_T} = \mu_p dt - \sigma_p dB_x = (\mu_p + \frac{\sigma_p}{\sigma_r} \lambda)dt - \sigma_p d\tilde{B}_x.$$

Since the instantaneous return of the bond must be r under the risk neutral measure, we have

$$\lambda = -\sigma_r \Lambda_p.$$

²⁴See Dai and Singleton (2000).

Under the same approximation, δ can be estimated using equation (139) and sample observations of the Sharpe ratios:

$$\delta \approx \frac{\bar{\Lambda}_p}{\bar{\Lambda}} = -0.8616.$$

Since the sample volatility of the riskless rate is 1.01%, the conditional volatility of the riskless rate is, using the estimate of $\kappa = 0.1465$ from column EM:I of Table 2,

$$\sigma_r = 1.01\% \times \sqrt{2 \times 0.1465} = 0.5467\%.$$

Using the values of $\bar{r}_x = 0.1333$ and $\bar{\Lambda}_x = -0.8602$, we finally obtain

$$\lambda_1 = -0.8616 \times 0.5467\% \times \frac{-0.860}{0.133} = 0.03.$$

6.2 Inflation Consideration

We can easily modify our setup to accommodate expected inflation. In Section 4.2, we show that whether or not inflation is neutral, the real model does not impose any restriction on the expected inflation. In particular, the expected inflation may be state-dependent, but the form of the state-dependency can be completely arbitrary. Near the steady-state, the expected inflation can be linearized, in the manner:

$$\pi = \bar{\pi} + \bar{\pi}_x \hat{x}, \quad (144)$$

where the variables (and below) with a hat represent the deviations from their steady state means. $\bar{\pi}$ can be determined by the unconditional mean of the expected inflation, $\bar{\pi}_x$ can be determined by the correlation between the expected inflation and the real or nominal short rate.

Suppose that the expected inflation depends only on the habit process.²⁵ Expanding the nominal rate $r^{\$} = r + \pi$ around its steady state mean, we have

$$\hat{r}^{\$} = (\bar{r}_x + \bar{\pi}_x) \hat{x} = \frac{\bar{r}_x + \bar{\pi}_x}{\bar{r}_x} \hat{r}. \quad (145)$$

It follows that

$$\frac{\hat{r}^{\$}}{\hat{r}} = \frac{\sigma_r^{\$}}{\sigma_r}. \quad (146)$$

The same no-arbitrage argument we used to relate the market price of risk to the Sharpe ratio of the production technology goes through unmodified in nominal terms. Thus, we have

$$\begin{aligned} \lambda^{\$} &\approx -\sigma_r^{\$} \Lambda_p^{\$} = \sigma_r^{\$} \frac{\Lambda_p^{\$}}{\Lambda_p} \delta \Lambda_s \\ &= \sigma_r^{\$} \delta \frac{\sigma_p}{\sigma_p^{\$}} (\bar{\Lambda} + \bar{\Lambda}_x x + \frac{G}{\sigma_s}) = \sigma_r^{\$} \delta \frac{\bar{\Lambda}_p^{\$}}{\bar{\Lambda}_p} (\bar{\Lambda} + \bar{\Lambda}_x x + \frac{G}{\sigma_s}), \end{aligned} \quad (147)$$

²⁵If it depends explicitly on the wealth process, it would be a lot more volatile than we actually observe.

where we have used the fact that while the expected excess return on the bond is time-varying, the volatility is not in a Gaussian model.

It follows that the sensitivity of the nominal market price of risk to the nominal short rate is given by

$$\lambda_1^{\$} = \frac{\hat{\lambda}^{\$}}{\hat{r}^{\$}} = \frac{\hat{\lambda}^{\$} \hat{x} \hat{r}}{\hat{x} \hat{r} \hat{r}^{\$}} = \frac{\bar{\Lambda}_p^{\$}}{\bar{\Lambda}_p} \lambda_1. \quad (148)$$

Using the values of $\bar{\Lambda}_p^{\$} = 0.156$ and $\bar{\Lambda}_p = 0.374$, in Table 1, we get

$$\lambda_1^{\$} = \frac{0.156}{0.374} \times 0.03 = 0.0125.$$

Since the expected inflation is assumed to be a function of the consumption habit, the mean reversion coefficient for the nominal short rate is the same as that of the real short rate:

$$\kappa^{\$} = \kappa = 0.143.$$

Dai and Singleton (2000) show that, within the one-factor Gaussian model with linear market price of risk, $\kappa^{\$}$ and $\lambda_1^{\$}$ determine completely the slope coefficients in the long-rate regression of Campbell and Shiller (1991). In the next subsection, we define precisely what we mean by the expectations puzzle, and summarize the key results of Dai and Singleton (2000).

6.3 Expectations Puzzle

We now show what happens if we take the model (140) and (142) seriously, reinterpreted as a model for the nominal interest rates, in the manner discussed in the previous subsection.

Results presented in this subsection are taken directly from Dai and Singleton (2000). All variables in this subsection are nominal.

Define

- P_t^n as the price of an n -period zero coupon bond;
- $R_t^n = -\frac{1}{n} \log P_t^n$ as the n -period zero coupon bond yield;
- $f_t^n = -\log P_t^{n+1} P_t^n$ as the one-period forward rate settled n periods hence;
- $p_t^n = f_t^n - E_t[R_{t+n}^1]$ as forward premium;
- $c_t^n = R_t^n - \frac{n}{\sum_{i=0}^{n-1}} E_t[R_{t+i}^1]$ the yield premium; and
- $D_{t+1}^n = \log \frac{P_{t+1}^{n-1}}{P_t^n} - r_t$ as one-period holding period return on an n -period zero coupon bond.

The one-factor Gaussian model with linear market price of risk (equations (142) and (142)) implies that the forward premium can be predicted by the forward-spot spread:

$$p_t^n = \delta_n + \alpha_n(f_t^n - R_t^1), \quad (149)$$

where

$$0 < \alpha_n = \frac{e^{-\kappa^s n} - e^{-(\kappa^s + \lambda_1^s)n}}{1 - e^{-(\kappa^s + \lambda_1^s)n}} < 1, \quad (150)$$

where $\kappa^s = \kappa$ and $\lambda_1^s = \frac{\bar{\Lambda}_P^s}{\Lambda_P} \lambda_1$.

It follows that the expected return on a n -period zero coupon bond is predictable by the forward spreads²⁶

$$e_t^n \equiv E_t[D_{t+1}^n] = E_t[D_{t+1}^{*n}], \quad (152)$$

where

$$D_{t+1}^{*n} = -(n-1)(c_{t+1}^{n-1} - c_t^{n-1}) + p_t^{n-1}, \quad (153)$$

$$c_t^n = \frac{1}{n} \sum_{i=0}^{n-1} p_t^i. \quad (154)$$

The long-rate regression of Campbell and Shiller (1991) must now be generalized to

$$E_t \left[R_{t+1}^{n-1} - R_t^n + \frac{1}{n-1} D_{t+1}^{*n} \right] = \frac{1}{n-1} (R_t^n - R_t^1). \quad (155)$$

The expectations hypothesis corresponds to the case where D_{t+1}^{*n} , or equivalently, p_t^n , are constants for all n .

D_{t+1}^{*n} , the *premium part* of the excess return, is completely determined from our model, given model parameters: it is completely determined by p_t^n , which in turn is completely determined by the observed forward-spot spread, $f_t^n - R_t^1$, and model parameters (through δ_n and α_n).

Given model parameters, so that D_{t+1}^{*n} is known, equation (155) forms the basis of a linear regression, with the slope coefficient given by

$$d_n = \frac{\text{cov}(R_{t+1}^{n-1} - R_t^n + D_{t+1}^{*n}/(n-1), (R_t^n - R_t^1)/(n-1))}{\text{var}((R_t^n - R_t^1)/(n-1))}, \quad (156)$$

²⁶By definition, we have

$$D_{t+1}^n = D_{t+1}^{*n} + \sum_{i=1}^{n-1} (E_t R_{t+i}^1 - E_{t+1} R_{t+i}^1), \quad (151)$$

where D_{t+1}^{*n} is given by (153).

Under the expectations hypothesis, the slope coefficient is given by

$$d_n^0 = \frac{\text{cov}(R_{t+1}^{n-1} - R_t^n, (R_t^n - R_t^1)/(n-1))}{\text{var}((R_t^n - R_t^1)/(n-1))}, \quad (157)$$

When the riskless rate is mean-reverting, i.e., $\kappa^s > 0$, and the market price of risk is positively correlated with the riskless rate, i.e., $\lambda_1^s > 0$, D_{t+1}^{*n} is positively correlated with the yield spread $R_{t+1}^{n-1} - R_t^1$, for large enough n . Thus, d_n^0 is always *lower* than d_n at long maturities. If our model is a good approximation of the dynamics under the expectations puzzle, we expect d_n^0 to be close to 1, and the difference, $d_n - d_n^0$, accounts for the expectations puzzle.

We do not need a model to compute d_n^0 . To compute d_n , we need to know α_n , but not δ_n . Since α_n is completely determined by κ^s and λ_1^s , so is d_n .

Figure 1 plots the slope coefficients from the long rate regression of Campbell and Shiller (1991) under the expectations hypothesis (circles), and under the term structure dynamics implied by two sets of parameter values. The dashed line corresponds to the set of parameter values found in column EM:I of Table 2: $\kappa^s = 0.1465$, $\lambda_1^s = 0.0123$. The solid line corresponds to the set of parameter values found in column EM/EH:II of Table 2: $\kappa^s = 0.0311$, $\lambda_1^s = 0.0203$.

We can see that the dashed line is above the circles across the entire maturity range from 1 month to 120 months, indicating a correction in the right direction. However, it is still negative and increasing in magnitude with maturity – the expectations puzzle is not resolved. This is a direct cost of calibrating our model to the equity market puzzles without the discipline of moment restrictions from the term structure of interest rates. In the next section, we will show how to resolve this issue and allow the model to explain equity market and the expectations puzzles simultaneously.

7 Reconciling Equity Market and Expectations Puzzles

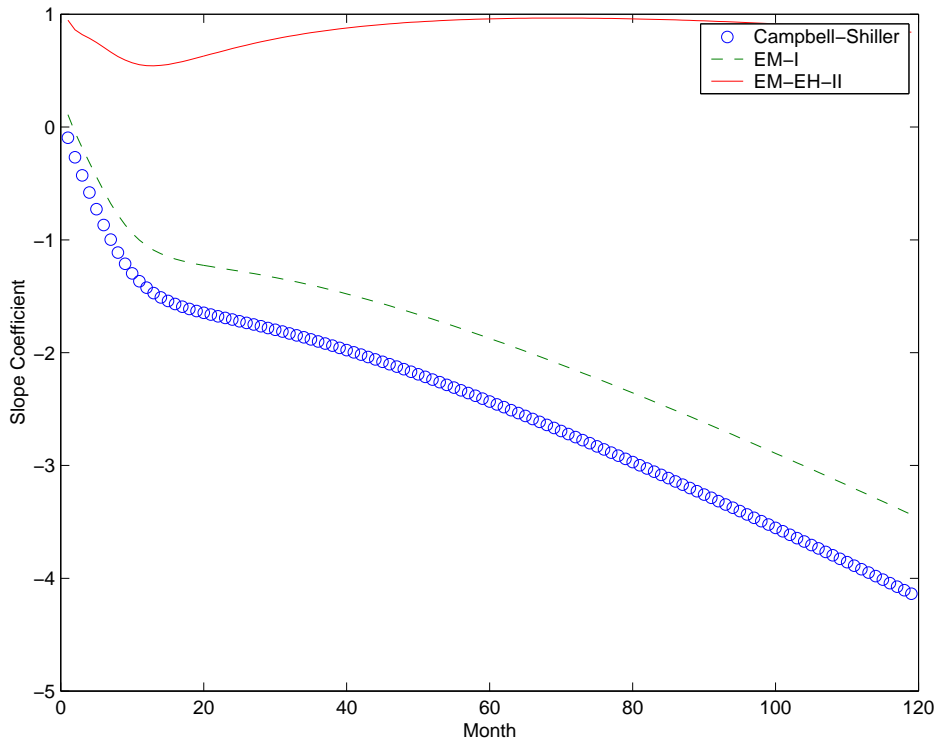
Parameters reported in columns EM:I and EM:II of Table 2 are calibrated *sequentially*: we first calibrate (γ, \bar{x}, a, b) to the equity market puzzles, which are then held fixed when we calibrate the rest of the parameters responsible for term structure dynamics. This procedure inherently tilts the model in favor of the equity market puzzles.

To put the equity and bond market on an equal footing, as they are in our theory, we need to re-design our calibration strategy to explicitly include the objective of explaining the expectations puzzle into our fitting criterion. To achieve this, we can modify the least square criterion function L_2 (or L_1 in similar fashion) to

$$L_2 = k_1 \left(\frac{\bar{c} - \bar{x}}{\bar{c}} \bar{g} - 1.5\% \right)^2 + k_2 \left(\frac{\bar{w} - \bar{f}}{\bar{w}} \bar{g} - 15.92\% \right)^2 + k_3 (r_0 + G)^2 + k_4 (\kappa^s - 1.48\lambda_1^s)^2, \quad (158)$$

where we use a simple rule of thumb that the expectations puzzle is resolved when the ratio of κ^s and λ_1^s is roughly 1.48. In a more rigorous treatment, the last term would be replaced

Figure 1: Expectations Regressions



The circles are the slope coefficients from the long-rate regression of Campbell and Shiller (1991), which is the same as fixing D_{t+1}^{*n} at a constant in equation (155). The dashed line and the solid line plot the slope coefficients from the regression (155), where D_{t+1}^{*n} is obtained from our model, with κ^s and λ_1^s given in columns EM:I and EM/EH:II, respectively, of Table 2.

by a properly defined measure of deviations from the expectations puzzle. It turns out that our simple rule of thumb works so well that the more rigorous treatment won't do any better.

Column EM/EH:I of Table 2 report the least square values of the parameters (γ, \bar{x}, a, b) that minimizes the criterion function L_3 , subject to the constraints (125). Note that additional information needed to compute the last piece of the criterion function L_3 is provided by the sample moments in Table 1. Thus there is no new parameters to calibrate. Since the last term depends on the free parameters, its inclusion in the criterion function brings discipline to the model with moment restrictions from the bond market. Column EM/EH:II is the same as column EM/EH:I, except that γ is fixed at -1 , a value comparable to that used by Constantinides (1990) and Campbell and Cochrane (1999).

The model under the two parameterizations in columns EM/EH:I and EM/EH:II represent two different economies ((accordingly, we will use EM/EH:I and EM/EH:II to designate these two economies). In both economies, the equity premium puzzle, the riskfree rate puzzle, and the expectations puzzle are explicitly resolved. First, the volatility of the consumption growth rate is matched almost perfectly in both economies. Thus, the equity premium puzzle is resolved. Second, the steady state value of the riskfree rate is fixed at the sample mean in

calibrating the model parameters, and the riskless rate at zero habit is fixed at a plausible value of 0%. Thus, the riskfree rate puzzle is explicitly resolved. Finally, the expectations puzzle is resolved by showing that, after accounting for the time-varying risk premium, the slope coefficients in the long rate regression of Campbell and Shiller (1991) are restored to 1. The solid line in Figure 1 shows the slope coefficients from the regression (155), using parameters from column EM/EH:II. Except for a small “smile” near the short end of the maturity spectrum, the slope coefficients are essentially restored to 1.²⁷

Thus, we have established conclusively that the equity premium puzzle, the riskfree rate puzzle, and the expectations puzzle can be explained simultaneously, and almost perfectly, in our model.

Let us now look more closely under the hood to find out how the reconciliation between the equity premium and expectations puzzles is achieved.

7.1 An Identification Problem

Although the properties of the equity market appear to be largely unaffected by the existence of a bond market, the latter plays an important role in identifying a key property of the economy: the mean reversion coefficient of the consumption habit.

Recall that the resolution of the equity premium puzzle boils down to a proper parameterization of the model so that (130) holds.

Without the habit process, $b = 0$, equation (130) immediately puts the model in trouble: the requisite value of $\gamma = 28.94$ is perceived to be too high. The presence of a habit process introduces two additional free parameters, a and b . This gives the model additional degrees of freedom to explain the right hand side of equation (130) under reasonable values of γ .

As we have noted earlier, given γ , equation (130) identifies only the ratio of a and b , and but not a and b separately. This presents an identification problem.

In the context of our model, the riskfree rate puzzle requires that the value of R_0 be reasonable. In general, we expect that $0 \leq R_0 \leq \bar{R}$. If we knew the precise value of γ and \bar{x} , then restricting R_0 to a reasonable value, say 0%, would help identify a and b . Unfortunately, the steady state mean of the consumption habit, \bar{x} , is not directly observed. In fact, the role of the riskfree rate puzzle in this paper is precisely to help identify \bar{x} , given other model parameters. Thus, the riskfree rate puzzle does not provide any discriminatory power for possible values of a and b .

The identification issue can be resolved in two ways.

First, we can identify a and b by adding restrictions implied by the expectations puzzle. Dai and Singleton (2000) show that when both κ and $\lambda_1^{\$}$ are small, moment restrictions associated with the expectations puzzle identifies the ratio of κ and $\lambda_1^{\$}$. The theoretical model developed in this paper imposes structural restrictions between κ and $\lambda_1^{\$}$ so that the individual values of κ and $\lambda_1^{\$}$ are identified, given other model parameters. Thus, given γ , the equity premium puzzle, which identifies $\frac{a}{b}$, and the expectations puzzle, which identifies

²⁷Small adjustment of the parameter values, by changing the relative strength of $\kappa^{\$}$ and $\lambda_1^{\$}$ can give a slightly better fit for the short end, at the expense of over-shooting the long end slightly. But such finesse is not essential, and may not even be necessary. The structure in the short end may be due to something entirely un-related to time-varying risk premium, e.g., liquidity effect.

$a - b$, jointly identify a and b . Columns EM/EH:I and EM/EH:II of Table 2 show that the ability of the model to explain the equity premium and expectations puzzles is insensitive to the value of γ . However, the calibrated parameter values of a and b , as well as implied subjective discount rate ρ , are.

Second, we can identify a and b by forcing the model to explain the stock market volatility puzzle, which boils down to the parametric restriction (131). Note that the left hand side of equation (131) depends on individual values of a and b , not just their ratio. Columns EM:I and EM:II of Table 2 show that individual values of a and b are indeed identified by both the equity premium and stock market volatility puzzles, and are relatively insensitive to the parameter value of γ .

Although both the stock market volatility puzzle and the expectations puzzle are capable of identifying a and b , these two puzzles appear to be mutually incompatible. The requisite parameterizations under which these two puzzles are resolved correspond to very different economic behaviors. In the next subsection, we elaborate on the differences and give a plausible explanation.

7.2 Aggregate Stock Market \neq Aggregate Wealth

The economy EM:II resolves the stock market volatility puzzle, but not the expectations puzzle, and the economy EM/EH:II resolves the expectations puzzle, but not the stock market volatility puzzle. This dilemma can be resolved by noting that if we actually observe the volatility of the return to aggregate wealth (which is a portfolio of aggregate stock market, human capital, national infrastructure, public or privately held real estate properties and lands, farms, privately owned enterprises, etc) and it is about a third of the volatility of the equity return, then the volatility puzzle is resolved: the volatility of the consumption growth rate is still a lot less volatile than the volatility of the wealth growth rate, but they are in line with the equilibrium pricing relationship between the two (wealth is the price of the consumption stream). The volatility of the equity return can be a lot higher than the volatility of the wealth return, and therefore the equity premium will be also a lot higher than the risk premium in the expected return to wealth. However, this does not necessarily constitute a puzzle either because it is in line with the equilibrium pricing relationship: the equity return is more volatile because the dividend stream is a lot more volatile than the consumption stream.

To substantiate the above discussion, let us compare some key properties of the two economies.²⁸

First, the steady state mean of the consumption/wealth ratio in EM:II is 5.96%, which is close to the historical mean of dividend/price ratio, while the consumption/wealth ratio in EM/EH:II is 1.27%. This is not surprising because in calibrating EM:II, we essentially assumed the production technology consists of the equity market alone. We break this assumption in calibrating EM/EH:II, thereby breaking the equality between the dividend/price ratio and the consumption/wealth ratio.

²⁸EM:I and EM:II are similar. We compare EM:II and EM/EH:II, because it will be shown that EM/EH:II is a realistic representation of the observed economy, and EM:II has the same γ as EM/EH:II.

Second, the steady state mean of the equity premium in EM:II is about 6.91%, which is close to the historical mean of equity premium, while the steady state mean of the wealth premium in EM/EH:II is about 2.22%. Note that we deliberately make a distinction between “equity premium” and “wealth premium”. This is appropriate: with the same market price of risk, different amount of volatility leads to different amount of risk premium.

In both cases, the total return finances exactly the consumption: $6.91\% + \bar{r} = 5.96\%$ and $2.22\% + \bar{r} = 1.27\%$.

The reason that EM/EH:II has about one third of the consumption/wealth ratio and one third of the equity return of EM:II is directly linked to the fact that EM/EH:II is accounting for only about one third of the equity return volatility.

Third, the absolute level of consumption and wealth in EM:II are 0.0474 and 0.7959 respectively, while those in EM/EH:II are a lot higher at 0.4529 and 35.5446, respectively. We may interpret the ratio of $\frac{0.7959}{35.5446} = 2.24\%$ as the wealth capitalized in the equity market as a percentage of aggregate wealth.

Finally, we have one more powerful evidence that suggests EM/EH:II is more realistic than EM:II or EM:I by examining the admissible region of the real riskless rate. From the calibrated parameters, we can compute the maximum level of consumption habit, and therefore the corresponding maximum level of real riskless rate. In EM/EH:II, the maximum level of real riskless rate is $R_{max} = r(x_{max}) + G = 1.97\%$, which is almost the same as (only 8 basis points higher than) the permanent growth rate $G = 1.89\%$. On the other hand, in EM:I and EM:II, the maximum level of real riskless rate is 4.41%, way above the permanent growth rate. It makes perfect sense, *although we did not explicitly impose the restriction*, that the maximum level of real riskless rate should be comparable to G . Because we calibrate the model so that the lowest level of real riskless rate is 0, the admissible region of the riskless rate in EM/EH:II is almost precisely $[0, G]$.

This feature of the economy EM/EH:II is very appealing: it gives an intuitive interpretation for the endogenous restriction our model imposes on the admissible state space of the habit process: the maximum habit level represents the technological limit of the economy. Higher habit levels can not be sustained by the natural limit on the rate of wealth production and accumulation. We must all live within our means!

We conclude from this discussion that EM/EH:II is a more realistic representation of the observed U.S. economy than either EM:I or EM:II.

7.3 One More Loose End

One loose end remains but can be easily resolved. EM/EH:I and EM/EH:II appear to give similar performance in terms of matching the equity premium puzzle, the riskfree rate puzzle, and the expectations puzzle. Yet the two economies have very different properties. Which one is a more realistic representation of the observed economy?

We come to an immediate conclusion by examining the implied dynamics of the state variables. Figure 8 plots the conditional mean and the conditional volatility of the habit and surplus wealth processes. Notice that the conditional mean of the habit is increasing with the habit level, suggesting a self-destructive behavior. *Static* properties of the economy (when it is fixed at the steady state mean) are consistent with the empirical puzzles. However, the

economy is *dynamically* unstable.

We can also examine the properties of the model when γ takes other values between -1 and 1. In general, if γ is too close to 1, then economy becomes unstable: the agent must have a very high subjective discount rate to satisfy the transversality condition. When γ is close to 0 or negative, the properties of the economy are qualitatively similar: all fit the equity premium puzzle, riskfree rate puzzle, and the expectations puzzle equally well. Furthermore, the maximum values of R_{max} are, G and G minus 4 basis points, when γ is 0.1 and -0.1, respectively. Since the equity premium puzzle is explained by fitting $\frac{1}{1-\gamma} \frac{c-x}{c}$, a different γ implies a different value of surplus consumption ratio. At $\gamma = 0.1$, the surplus consumption ratio is about 3.12%. At $\gamma = -0.1$, the surplus consumption ratio is about 3.46%. All appear reasonable, and are roughly on the same order of magnitude as the value of 6.94% when $\gamma = -1$. Since $\gamma = -1$ is comparable to what Constantinides (1990) and Campbell and Cochrane (1999) chose, we will examine the properties of the economy under this value.

8 Comparative Statics Under Calibrated Parameter Values

In all of the calibration exercises, we have relied on an assumption that the parameter β , which controls the magnitude of the habit shock, is small. Three questions arise naturally. First, is β really small? Second, suppose that β is small, how good are the approximations we used in the calibration exercise? Third, what does the *exact solution* to our model look like, if $\beta \neq 0$?

The first question is not trivial. If we take our model in Section 2, or its term structure implication in Section 3.2 literally, then the conditional volatility of the riskless rate comes entirely from the habit shock. This implies that, to a good approximation,

$$\beta \approx \frac{\sigma_r}{\bar{z}\bar{r}_x}.$$

Using observed moments from Table 1 and parameters in column EM/EH-II of Table 2, we have $\beta = 3.8568$. This is much too big. In a refinement in Section 4.3, we show that β can be small, if the volatility of the riskless rate can come from additional idiosyncratic shocks. We do not attempt, nor there is a need to, decompose the volatility of the riskless rate into the component associated with the habit shock, and the volatility associated with the additional idiosyncratic shocks. This is because the market price of interest rate risk, as we defined it, is scaled by the conditional volatility of the riskless rate. To see this more clearly, we note that equations (138), (139) and (143) imply, for small β ,

$$\lambda_1 = \delta\sigma_r \frac{\bar{\Lambda}_x}{\bar{r}_x}, = \sigma_r \frac{\Lambda_p}{\Lambda} \frac{\bar{\Lambda}_x}{\bar{r}_x}.$$

Since Λ_p is inversely proportional to the volatility of the bond yields, and therefore the volatility of the riskless rate, λ_1 is actually independent of the short rate volatility.

Therefore, our model, refined by the extension in Section 4.3, does not put any specific restriction on the size of β . However, it can be easily shown that when β is too large, the

economy is not stable, because the habit process will frequently reach the maximum level, x_{max} , and the economy will break down.

In what follows, we will pick a reasonable value of β , say 0.05,²⁹ to compliment the parameters calibrated in column EM/EH:II of Table 2. This gives us a complete parametric specification of the model. We will examine the exact solution under this parameterization, and confirm that all key economic properties relevant for our empirical study are not sensitive to the parameter value of β , as long as it is small. We will also simulate the economy and show graphically that the time-series properties of the economy are very realistic.

8.1 Exact Solution under Calibrated Parameter Values

The exact solution can be computed, without making the assumption that β be small, even though the parameter value of β we actually use is relatively small. We solve equations (27) and (28) numerically under the parameter values in column EM/EH:II of Table 2 and $\beta = 0.05$. Once $h(x)$ and $g(x)$ are determined, the exact solution to our model can be plotted.

Figure 2 plots the riskless rate, the Sharpe ratio, $h(x)$ and $g(x)$, as functions of the habit level x . We can see that for the entire admissible state space $0 < x < x_{max} = 0.9407$, the riskless rate is monotonically increasing, from the minimum value of 0% to the maximum value close to G . The Sharpe ratio is monotonically decreasing. Thus, they are negatively correlated everywhere.

Figure 3 plots the drift and diffusions of the bi-variate Markovian process $(x, w - f(x))$. We can see that both state variables are mean reverting. The conditional volatility of the habit process increases with habit level, because with the surplus wealth fixed, a rising habit level increases the surplus consumption ($h(x)$ is increasing in x), which drives up the volatility. The drift of the surplus wealth is positive when $x < \bar{x}$ and negative when $x > \bar{x}$, where \bar{x} is about 0.4215.³⁰ The conditional volatility of the surplus wealth is primarily determined by the habit shock, and the sensitivity to the habit shock is decreasing as the habit level rises.

We can see that the parameters we calibrated satisfy all of the necessary conditions in order for the model to be well-defined.

8.2 Economic Properties Seen Through Simulation

We will examine the time-series properties of the economy through simulation. First, we simulate daily observations for two thousand years the bi-variate Markovian processes $(\omega = w - f(x), x)$ using equations (9) and (52). A first-order Euler scheme is used, and the economy

²⁹At $\beta = 0.05$, the standardized distance from the steady state mean to the maximum habit level, evaluated at the steady state values, is $d = \frac{x_{max} - \bar{x}}{\beta \bar{x}} = \frac{0.9407 - 0.4215}{0.05 \times 0.0314} = 330.59$. If the habit process were a pure Brownian motion, then the probability of the habit process reaching the maximum level from its the steady state mean in $T = 200$ years would be (see Karatzas and Shreve (1988)): $\sqrt{\frac{2}{\pi}} \int_{\frac{d}{\sqrt{T}}}^{\infty} e^{-\frac{x^2}{2}} dx$, which is practically 0. Since the habit process is actually mean reverting, the actual probability will be even less. Obviously, the probability will increase if β increases.

³⁰ $\bar{x} = 0.4215$ at $\beta = 0$. When $\beta \neq 0$, \bar{x} will be different. When β is small, \bar{x} is almost unaffected.

is started from the steady state means. All other time-series can be constructed from the simulated state variables.

For each of the Figures 4–7, the top half plots 200 years of history, during which business cycles are evident. The bottom half plots 2000 years of history, during which the rise and fall of fortune of a nation-state is graphically displayed.

Figure 4 plots the habit process (x), the surplus wealth ($w - f(x)$), the riskless rate $r(x)$, and the Sharpe ratio $\Lambda(x)$, for a typical realization of random shocks. We can see that the habit level is positively correlated with the surplus wealth, or wealth.³¹ Thus, the riskless rate is *pro-cyclic*, and the Sharpe ratio is *counter-cyclic*. The wealth process and the habit process are not perfectly correlated. Thus, business cycles is a noisy predictor of the wealth premium.

Figure 5 plots the equilibrium consumption and wealth processes, the surplus consumption and wealth processes, the relative risk aversion with respect to wealth, and the inverse of the elasticity of consumption substitution. Note that the consumption process is much smoother locally than the wealth process. RRA and ECS^{-1} are highly correlated, and generally counter-cyclic: the agent becomes more aggressive and less willing to save as the economy expands. Unlike time-separable models, RRA is much smaller than ECS^{-1} in our economy, which is consistent with the equity premium puzzle.

Figure 6 plots the log surplus consumption ratio, the sensitivities of the log surplus consumption ratio to the consumption, production and habit shocks. We can see that the log surplus consumption is mean-reverting around its steady state mean ($\bar{\xi} \approx -2.67$). The consumption shock sensitivity is strictly positive throughout the sample, and is *negatively* correlated with the log surplus consumption ratio. The production shock sensitivity shares the same property, as the consumption shock is primarily driven by the production shock. The habit shock sensitivity is negative and small. The exogenous assumptions of Campbell and Cochrane (1999) are born out clearly in these pictures.

Figure 7 plots the wealth premium, $\mu_s - r = \Lambda\sigma_s$, the consumption-wealth ratio, $\frac{c}{w}$, the riskless rate, r , and the wealth level, w .

The simulated economy reveals a number of important properties of our model.

1. From time to time, a combination of low levels consumption and habit and favorable production shocks sent the economy into a vigorous expansion that last for decades, followed by a general gradual decline that last for decades or more. The wealth gains accumulated through those expansions are gradually withered away by an elevated level of consumption and habit, because the positive productions shocks do not persist.

Within an “epochal” expansion and contraction, which corresponds to the rise and fall of a nation-state, there are smaller “booms” and “busts”, corresponding to the business cycles one actually experiences in a span of a life time.

There was a point at which the habit reached and moved slightly above³² the maximum level, because the wealth was increasing at a very fast pace. The economy quickly

³¹Both visually and from sample correlations. The sample correlation between the habit and the surplus wealth is about 0.2261, and the sample correlation between the habit and wealth levels is 0.6143.

³²In a simulation, the habit level can go slight above the maximum level, because of the way we treat the boundary: we set the risk premium at the lowest possible level, instead of rendering the process invalid, whenever the habit rises above the maximum level. Since this rarely happens, it does not affect the steady-

corrected itself, as it always does, because the highest level of habit can not be sustained in expectation. Except for this instance, the economy stayed within the admissible state space through the entire 2000 years of history.

There was also a point at which the wealth became almost zero ($w_{min} = 0.1412$), due to a series of unfavorable production shocks. It takes a few centuries for the economy to pull itself back, during which time consumption level is very low, and the volatilities for consumption growth and equity return are very high. The agent becomes extremely risk averse, and the propensity to save sends the riskless rate down to the lowest level, and equity or wealth premium to the highest level.

2. Even though the habit shock and the production shock are negatively correlated locally, the habit process is pro-cyclic, i.e., the sample correlation between the habit level and the wealth level is *positive*! This is due to the fact that the habit process is mean-reverting to a target level that rises and falls with the surplus consumption level, which in turn rises and falls with the surplus wealth. This feedback effect in the conditional mean introduces a positive correlation between the habit and the wealth processes. The positive correlation overcomes the negative local correlation over long horizons.

This observation has important ramifications for the equity return predictability literature. It rationalizes the observation that the predictability results are stronger over longer horizons. The extant literature has typically focused on the level dependency of expected return on say, dividend yield.

3. The bi-variate Markov processes, (ω, w) have a steady state distribution. For a long history of two thousand years, the chance of the habit process hitting the maximum admissible level is still very small. It occurred only once.
4. The sensitivity function of the log surplus consumption rate with respect to the consumption shock is positive, and is negatively correlated with the log surplus consumption itself. Furthermore, the log surplus consumption ratio is positively correlated with either the habit process or the wealth process over long horizons. Therefore we confirm Campbell and Cochrane (1999) in their association of high level of log surplus consumption ratio with a “boom”, and low level of log surplus consumption ratio with a “bust”.
5. The wealth premium is predictable by its own lag, consumption-wealth ratio, and riskless rate (or term spread), and is correlated with business cycles, confirming empirical observations documented over the last two decades.

8.3 Discussion and Future Research Directions

Table 3 summarizes correlations among key economic variables, computed from the simulated sample. Clearly, both wealth premium and its conditional volatility are predictable by a variety of variables, including their own lags at long horizons (the monthly autocorrelations

state properties of the economy.

of the wealth premium and its conditional volatility are 0.979 and 0.976, respectively.) To understand the correlation structure, we need to keep in mind that there are two state variables in the economy. The strength of the correlation between any two economic variables is, to a large extent, determined by their relative loadings on the two state variables.

The table generally confirms a large body of empirical literature that documents the predictability of equity return and volatility (if we assume that the wealth return and the equity return behave similarly, aside from the fact that they have a different scale in terms expected premium and conditional volatility.)

For examples, Campbell and Shiller (1988), Fama and French (1988), and Hodrick (1992), among others, documented predictability of equity premium by the dividend-price ratio. Poterba and Summers (1988) documented the autocorrelation in equity premium. The predictability of equity premium by term spread was documented by Campbell (1987), Fama and French (1989), and Keim and Stambaugh (1986), among others. Furthermore, using ARCH model of Engle (1982) and GARCH model of Bollerslev (1986), French et al. (1987), among others, documented the predictability of return volatility by the equity premium. A high correlation between the equity premium and its conditional volatility reported in the table confirms their findings. This suggests that our model gives a theoretical grounding for the ARCH/GARCH modeling approach for return volatility. It makes perfect sense: the conditional volatility of equity return in our model is autoregressive, and is driven by two state-variables.

Our model can provide more precise comments on this literature, for both aggregate markets and individual stocks, by characterizing an empirical model of dividend processes, and use the pricing kernel (96) to determine equity prices and returns.

In addition, our model gives theoretical predictions on the relationship between aggregate consumption and investment opportunities. Dunn and Singleton (1986) and Kandel and Stambaugh (1990), among others, investigated this relationship. Additional research along this line can provide additional tests for our model.

There is also a huge literature that address consumption choice and portfolio selection problems faced by an individual investor with finite investment horizon, with or without endowment and labor income. These issues can also be addressed within our theoretical framework.

All of the economic properties gleaned from the simulation exercise can be characterized formally by using the exact solution to our model. We have done this in explaining the equity market puzzles and the expectations puzzle, which are the focus of this paper. Analytical characterizations of additional empirical regularities discussed in this section will be deferred to future research.

9 Conclusion

In a sharp contrast to most work in existing asset pricing and portfolio selection literature, the stochastic investment opportunity set and consequently the pricing kernel in our model are endogenously determined, rather than exogenously specified. As such, our model provides a firm theoretical framework for re-examining many empirical regularities in both equity and bond markets, and their relationship with the macro-economy.

Table 3: Correlations in the Simulated Sample

The correlations are computed from daily data with a simulation length of 2000 years.

	$\mu_s - r$	w	$w - f(x)$	c	x	$\log \frac{c-x}{c}$	r	$\frac{w}{c}$	σ_{cs}	σ_s
$\mu_s - r$	1.0000	-0.2240	0.1755	-0.3663	-0.4071	0.8479	-0.4073	0.9569	0.9557	0.9807
w	-0.2240	1.0000	0.8054	0.9634	0.9392	0.2700	0.9412	-0.3188	0.0001	-0.0529
$w - f(x)$	0.1755	0.8054	1.0000	0.6339	0.5657	0.6542	0.5670	0.1886	0.4106	0.3477
c	-0.3663	0.9634	0.6339	1.0000	0.9963	0.0825	0.9961	-0.4918	-0.1761	-0.2211
x	-0.4071	0.9392	0.5657	0.9963	1.0000	0.0181	0.9998	-0.5433	-0.2295	-0.2707
$\log \frac{c-x}{c}$	0.8479	0.2700	0.6542	0.0825	0.0181	1.0000	0.0180	0.8250	0.9536	0.9211
r	-0.4073	0.9412	0.5670	0.9961	0.9998	0.0180	1.0000	-0.5441	-0.2288	-0.2697
$\frac{w}{c}$	0.9569	-0.3188	0.1886	-0.4918	-0.5433	0.8250	-0.5441	1.0000	0.9358	0.9323
σ_{cs}	0.9557	0.0001	0.4106	-0.1761	-0.2295	0.9536	-0.2288	0.9358	1.0000	0.9816
σ_s	0.9807	-0.0529	0.3477	-0.2211	-0.2707	0.9211	-0.2697	0.9323	0.9816	1.0000

In this paper, we show conclusively that two of most important puzzles in asset pricing literature: the equity premium puzzle in the equity literature, and the expectations puzzle in the term structure literature, share the same economic origin and are completely resolved within a single theoretical framework.

The model also provides a wealth of endogenous relationships between time-varying aggregate consumption, time-varying expected asset returns, and time-varying volatilities and correlations, beyond those associated with major asset market puzzles. Empirical investigations of these relationships will be deferred to future research.

The paper raises the bar for alternative models explaining the equity premium puzzle. For an example, models based on recursive (state-nonseparable) utility (see Kreps and Porteus (1978), Epstein and Zin (1991), Weil (1989), Duffie and Epstein (1992), among others) have also enjoyed considerable success in explaining the equity premium puzzle and the riskfree rate puzzle. We must now ask: Can these models accommodate a general equilibrium economy with realistic term structure dynamics and time-varying risk premium? Can they simultaneously explain key empirical regularities in both equity and bond returns?

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.
- Backus, D., Foresi, S., Mozumdar, A., and Wu, L. (1997). Predictable changes in yields and forward rates. Working Paper, New York University.
- Balduzzi, P., Das, S. R., Foresi, S., and Sundaram, R. K. (1996). A simple approach to three factor affine term structure models. *Journal of Fixed Income*, 6:43–53.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31:307–327.

- Campbell, J. and Cochrane, J. (1999). By force of habit: A consumption-based explanation of aggregate stock market behavior. *Journal of Political Economy*.
- Campbell, J. Y. (1987). Stock returns and the term structure. *Journal of Financial Economics*, 18:373–399.
- Campbell, J. Y. (1999). *Asset prices, consumption, and the business cycle*, in *Handbook of Macroeconomics*, volume 1. North-Holland, Amsterdam.
- Campbell, J. Y. (2000). Asset pricing at the millennium. *Journal of Finance*, forthcoming.
- Campbell, J. Y. and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies*, 1:195–228.
- Campbell, J. Y. and Shiller, R. J. (1991). Yield spreads and interest rate movements: A bird's eye view. *Review of Economic Studies*, 58:495–514.
- Campbell, J. Y. and Viceira, L. M. (2000). Who should buy long-term bonds? *American Economic Review*, forthcoming.
- Chapman, D. A. (1998). 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.
- Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1985). An intertemporal general equilibrium model of asset prices. *Econometrica*, 53:363–384.
- Dai, Q. and Singleton, K. (1999). Specification analysis of affine term structure models. *Journal of Finance*, forthcoming.
- Dai, Q. and Singleton, K. J. (2000). Expectations puzzles, time-varying risk premium, and dynamics of the term structure. Working paper.
- Detemple, J. B. and Zapatero, F. (1991). 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, D. and Epstein, L. (1992). Stochastic differential utility and asset pricing. *Econometrica*, 60:353–394.
- Dunn, K. B. and Singleton, K. J. (1986). Modeling the term structure of interest rates under non-separable utility and durability of goods. *Journal of Financial Economics*, pages 27–55.
- Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica*, 50(4):987–1006.

- Epstein, L. G. and Zin, S. E. (1989). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica*, 57:937–969.
- Epstein, L. G. and Zin, S. E. (1991). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis. *Journal of Political Economy*, 99:263–286.
- Fama, E. (1976). Forward rates as predictors of future spot rates. *Journal of Financial Economics*, 3:361–377.
- Fama, E. (1984a). The information in the term structure. *Journal of Financial Economics*, 13:509–528.
- Fama, E. (1984b). Term premiums in bond returns. *Journal of Financial Economics*, 13:529–546.
- Fama, E. and French, K. (1989). Business conditions and expected returns on stocks and bonds. *Journal of Financial Economics*, 25:23–49.
- Fama, E. F. and Bliss, R. R. (1987). The information in long-maturity forward rates. *American Economic Review*, 77(4):680–692.
- Fama, E. F. and French, K. R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics*, 22:3–27.
- French, K. R., Schwert, G. W., and Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of Financial Economics*, 19:3–29.
- Hansen, L. P. and Jagannathan, R. (1991). Implications of security market data for models of dynamic economies. *Journal of Political Economy*, 99(2).
- Hodrick, R. J. (1992). Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *Review of Financial Studies*, 5:357–386.
- Kandel, S. and Stambaugh, R. F. (1990). Expectations and volatility of consumption and asset returns. *Review of Financial Studies*, 3(2):207–232.
- Karatzas, I. and Shreve, S. (1988). *Brownian Motion and Stochastic Calculus*. New York: Springer-Verlag.
- Keim, D. and Stambaugh, R. (1986). Predicting returns in stocks and bond markets. *Journal of Financial Economics*, 17:357–390.
- Kim, T. S. and Omberg, E. (1996). Dynamic nonmyopic portfolio behavior. *Review of Financial Studies*, 9:141–161.
- Kreps, D. M. and Porteus, E. L. (1978). Temporal resolution of uncertainty and dynamic choice theory. *Econometrica*, 46:185–200.
- Liu, J. (2000). Portfolio selection in stochastic environments. Working paper, UCLA.

- Lucas, R. (1978). Asset prices in an exchange economy. *Econometrica*, 46:1429–1445.
- Lynch, A. W. (1999). Portfolio choice and equity characteristics: Characterizing the hedging demands induced by return predictability. Working paper, New York University.
- Mehra, R. and Prescott, E. C. (1985). 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.
- Poterba, J. and Summers, L. H. (1988). Mean reversion in stock returns: Evidence and implications. *Journal of Financial Economics*, 22:27–60.
- Sundaresan, S. M. (1989). Intertemporally dependent preferences and the volatility of consumption and wealth. *Review of Financial Studies*, 2:73–88.
- Sundaresan, S. M. (2000). Continuous-time methods in finance: A review and an assessment. *Journal of Finance*, forthcoming.
- Weil, P. (1989). The equity premium puzzle and the riskfree rate puzzle. *Journal of Monetary Economics*, 24:401–421.

Figure 2: Exact Solution for Economy EM/EH:II

$h(x)$ and $g(x)$ are jointly solved from equations (27) and (28) under the following parameter values taken from column EM/EH:II of Table 2, and $\beta = 0.05$. $r(x)$ and $\Lambda(x)$ are computed from (25) and (26), respectively.

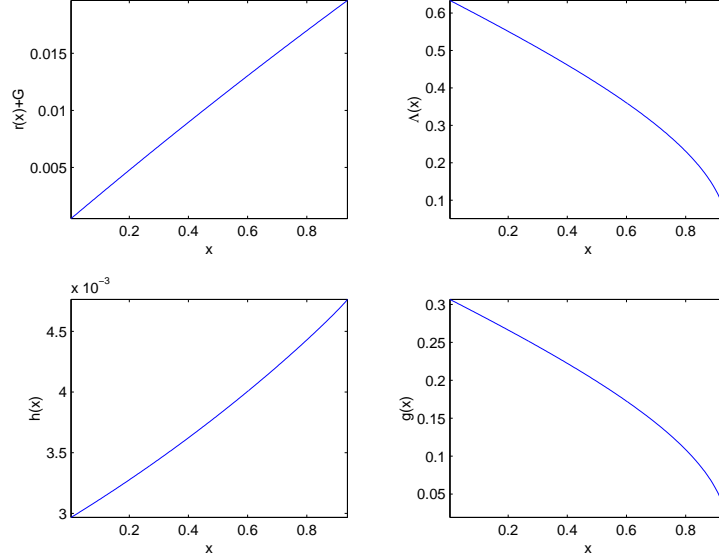


Figure 3: Equilibrium Dynamics

μ_x and σ_x are computed from equation (10). When x varies, the surplus consumption varies according to $c - x = h(x)(\bar{w} - f(\bar{x}))$. μ_ω , $\sigma_{\omega s}$, and $\sigma_{\omega x}$ are computed from equations (53) – (55).

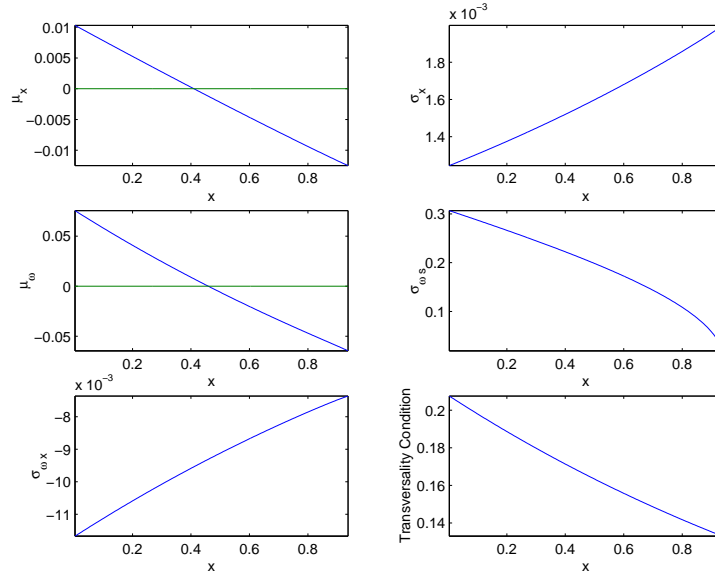


Figure 4: Equilibrium State Variables and Investment Opportunity Set x and $w - f(x)$ are simulated from equations (9) and (52). $r(x)$ and $\Lambda(x)$ are computed from equations (25) and (26), respectively, using the simulated habit process.

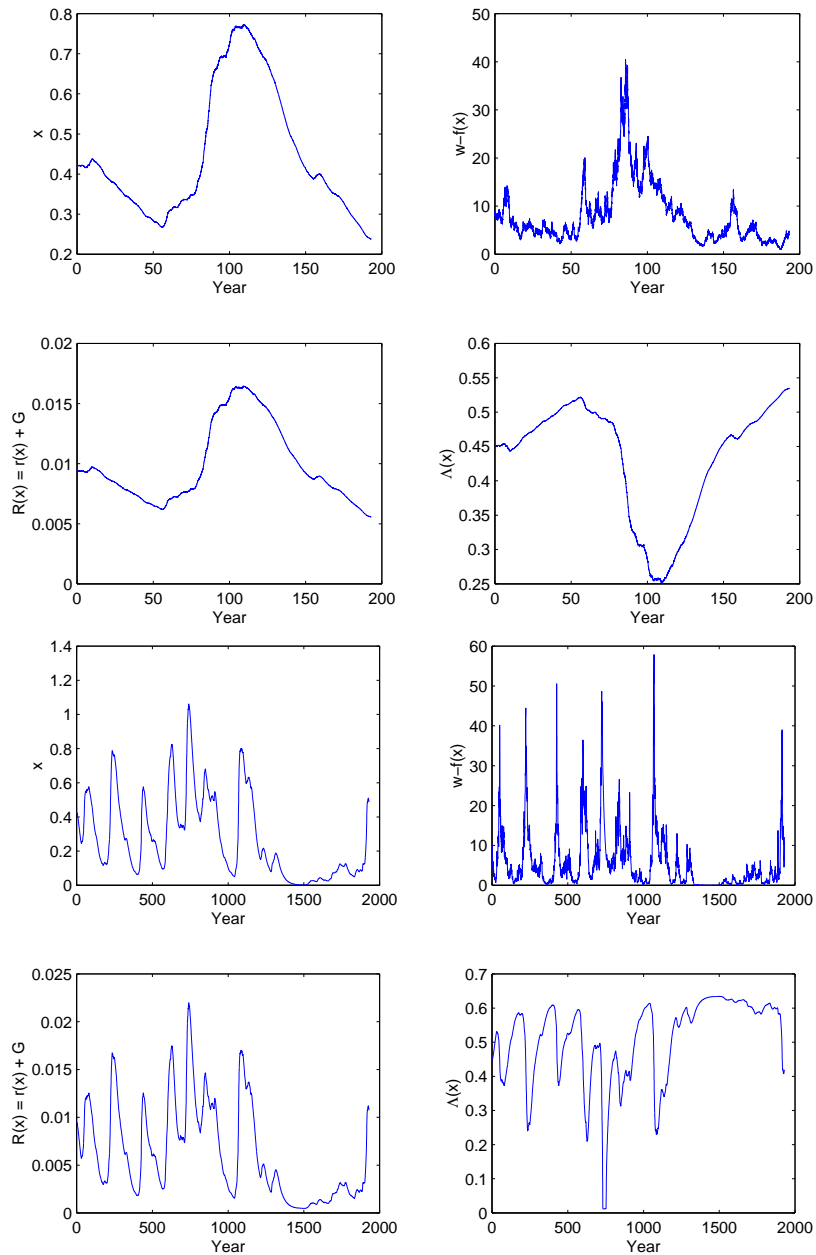


Figure 5: Equilibrium Consumption and Wealth

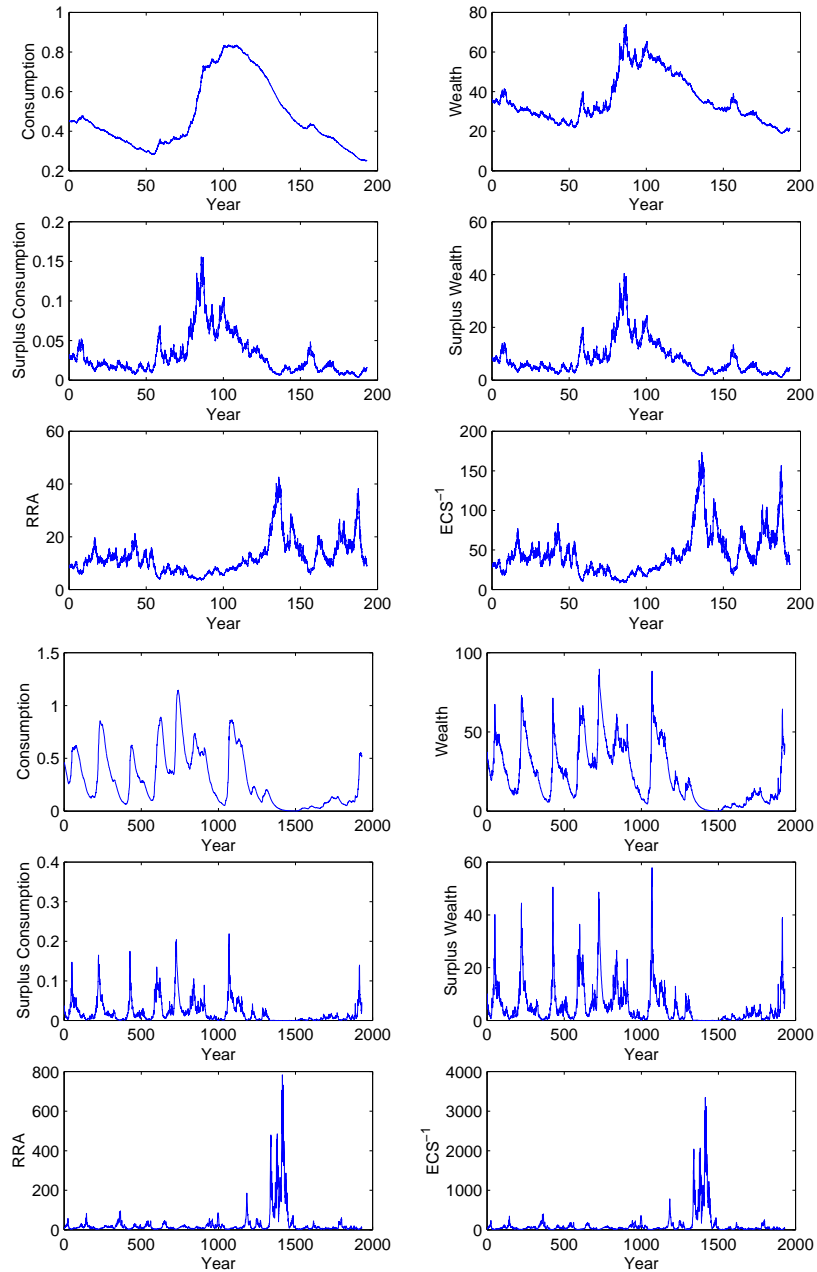


Figure 6: Surplus Consumption Ratio

The log surplus consumption ratio is simulated from equation (61) and the sensitivities are simulated from equations (69), (63) and (64), respectively.

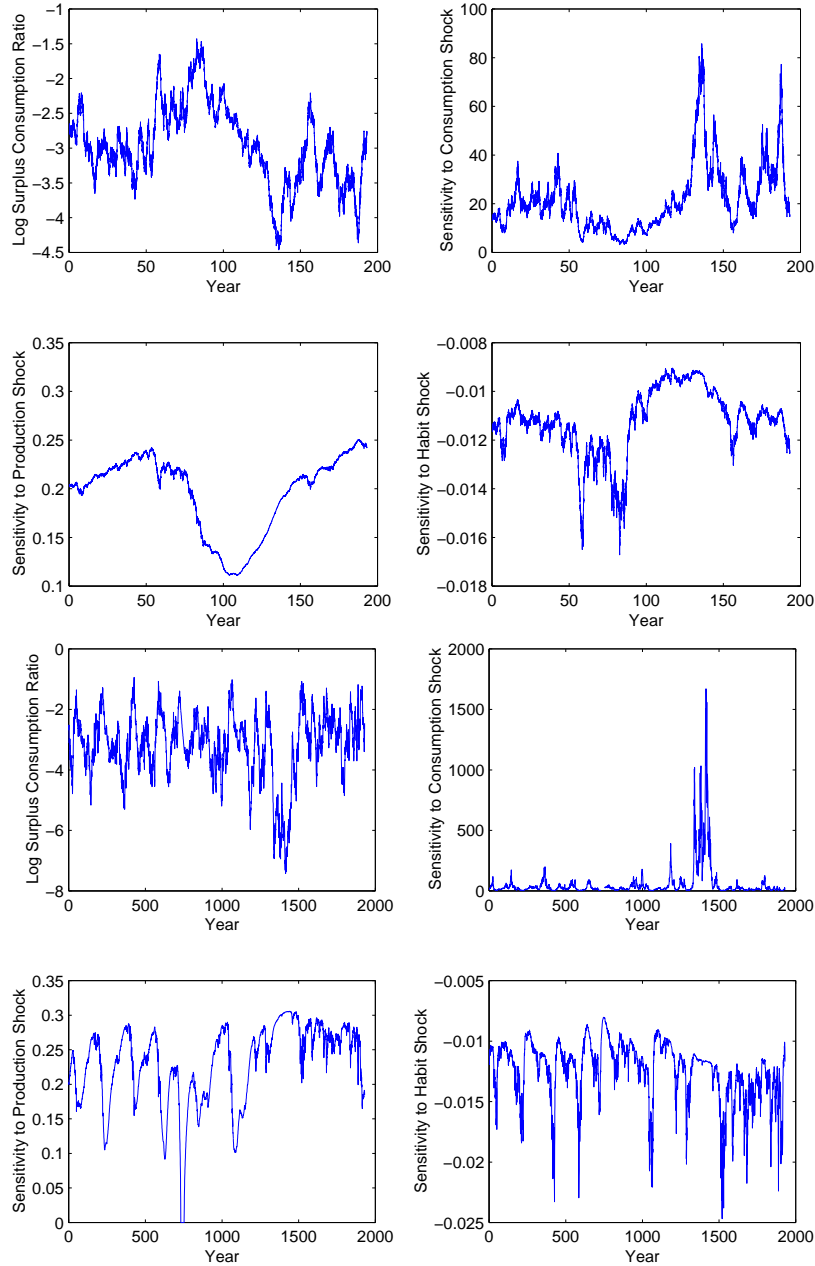


Figure 7: Wealth Returns and Volatilities
 The wealth premium is obtained by multiplying the Sharpe ratio by the endogenously determined conditional volatility (37).

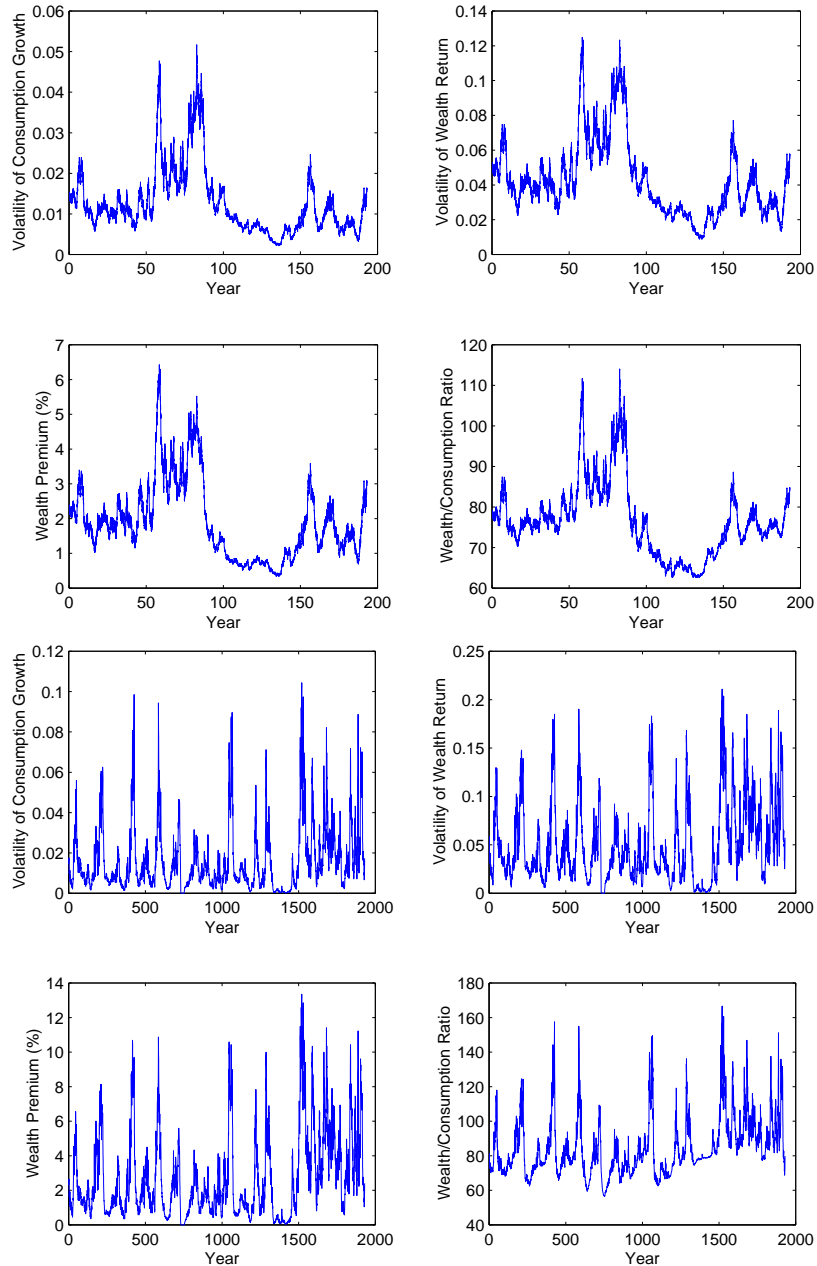


Figure 8: Dynamics Implied by EM/EH-I

