

Endogenous Regime Changes in the Term Structure of Real Interest Rates*

Jørgen Haug[†] and Jacob S. Sagi[‡]

First Version: September 2004

This Version May 11, 2005

Abstract

We present a model that captures the tendency of real rates to switch between regimes of high versus low level and volatility, the general shape of the term structure in either regime, the relative frequency of the regimes, and the time varying risk premium associated with the yield curve. We do this by supplementing a pure endowment economy model with a simple constant returns to scale technology. The characteristics of the resulting equilibrium shift between those of a pure endowment and production economy. The shift induces endogenous regime switching in the real interest rate. Among the specifications we consider, combining a linear habit formation endowment economy with risk-free production appears to explain the broadest set of stylized facts.

JEL Classification: G12, G13

*We thank Jonathan Berk, Greg Duffee, Pierre Collin-Dufresne, Burton Hollifield, Joao Gomes, Ravi Jagannathan, Hayne Leland, Martin Lettau, and Steve Ross for valuable comments and suggestions. All mistakes are ours alone.

[†]Norwegian School of Economics and Business Administration, Department of Finance and Management Science, Helleveien 30, N-5045, Bergen, Norway. email: jorgen.haug@nhh.no, phone: (+47)-5595-9426

[‡]Haas School of Business, University of California at Berkeley, 545 Student Services Building, Berkeley, California, 94720-1900. email: sagi@haas.berkeley.edu, phone: (510)-642-3442

1 Introduction

There is mounting empirical evidence for regimes in the term structure of interest rates (Ang and Bekaert, 2004; Cecchetti et al., 1993; Dai et al., 2003; Evans, 2003; Garcia and Perron, 1996; Gray, 1996; Hamilton, 1988, among others). The stylized facts noted by this literature include the characterization of two regimes. The less frequent regime exhibits relatively higher mean and variance of the real short rate, and occurs roughly 20% of the time. The smooth (volatile) regime appears to feature an upward (downward) sloping term structure of interest rates, while the unconditional term structure is nearly flat. Finally, there is some evidence that the real interest-rate risk premium is time-varying and generally negative, although there are some states in which it can be positive. This paper offers a general equilibrium model — the first to our knowledge — in which these stylized facts are accommodated both qualitatively and quantitatively.

A likely source of regimes in interest rates is the government’s monetary policy or that of the Federal Reserve. Such an approach does not lend itself well to a general equilibrium analysis because it requires a model of the objectives of the government as well as the economic frictions leading to the relevance of its policies for the term structure. One would moreover expect that structural shifts in interest rate policy come about as a result of changes in more fundamental economic quantities than simply the whim of a policy maker.¹ This suggests that it may be useful to characterize the relationship between interest rate regimes and fundamental economic quantities even if one abstracts away from government policy. Following this line of thinking, this paper takes a more traditional approach to modeling interest regimes.

Our model can be understood as an economy in which a representative agent benefits from the output of several distinct production technologies. If the risk and return profile of the output from one technology is deemed relatively unattractive, capital is shifted away from it and towards alternative technologies. In the presence of large capital adjustment costs, the distribution of capital across technologies is ‘sticky’ and can be characterized by regimes in which some of the technologies are underutilized. These regimes generally exhibit distinct interest rate behavior since interest rates are linked to the marginal product of capital, which in turn depends on the distribution of capital.

¹Moreover, if one views a change in regime as linked to inversion in the slope of the *real* term structure, it is far from clear that government policy is the only rationale for regime switching.

Our main contribution is to demonstrate that (i) discontinuous regimes in real rates are natural even to an economy for which monetary policy is irrelevant;² (ii) such regimes can arise under very simple assumptions about the production technology; (iii) we are able to calibrate to a realistic and large set of stylized facts; (iv) the model links the onset of regimes to changes in economic fundamentals, and thus offers a rich set of implications and potential for empirical analyses. Current statistical models of regime switching in interest rates do not offer similar insights about causation. Our model may therefore serve to produce better econometric identifications of regimes, distinguish among the numerous and competing statistical specifications currently used to model interest rates, and increase one's confidence in the future reliability of what would otherwise be statistically specified economic models. Indeed, beyond fitting to stylized facts, we find support for the model when we perform simple tests of its predictions.

There are numerous ways in which the idea outlined above can be implemented. We study the simplest possible such setting: a single good endowment economy with a storage technology. To frame this as a story about shifting capital between technologies in the presence of adjustment costs, consider that ours is a setting with two independent technologies that produce the same consumption good. The first technology is analogous to the endowment process in an exchange economy, while the second is fully reversible and offers constant returns to scale (as in the model of Cox et. al., 1985a). Although the capital stock of the endowment technology cannot be changed, its output can be consumed or invested in the reversible technology. Capital adjustment costs are consequently infinite in the former while they are zero in the latter. This combination can give rise to two regimes: an 'endowment regime' corresponding to the case where no capital is invested in the reversible technology, and a 'production regime' corresponding to non-trivial investment. In the endowment regime the risk-free rate is determined solely by attributes of the endowment sector, while in the production regime it is determined by the yield on the reversible technology. In equilibrium, the economy and thus interest rates can shift endogenously between the two regimes.

While our model is too simple to be interpreted literally, it does possess the key features that will lead to interest rates regimes in a more realistic, albeit necessarily far more complicated setting.³

²At the very least, we provide an alternative or complementary channel for regime switching in interest rates.

³In particular, the 'sticky' zero capital boundary in the reversible technology appears unrealistic. We later argue, however, that one can trade the 'zero capital boundary' for any arbitrary minimal maintenance capital level without

One could, moreover, stretch the interpretation of our setup and view the reversible technology as proxying for the representative agent’s limited ability to ‘live beyond her means’ and smooth the consumption of aggregate output from labor capital; the irreversible ‘endowment’ sector therefore corresponds to the raw number of units of labor available in the economy, and its output can be viewed as aggregate labor income or GNP. Clearly, there is an increasing cost for the representative agent to ‘live beyond her means.’ We set this cost to be zero until such time as storage in our model is depleted, at which point the cost to consuming beyond the endowment is infinite. Under this interpretation, the ‘endowment regime’ is characterized by identical growth rates for consumption and labor income. By contrast, the ‘production regime’ exhibits consumption that is generally smoother than labor income. Thus our model directly relates heteroskedasticity in consumption growth rates to interest rate regimes — a prediction that is not apparent from statistical models of regime changes and for which we find some support in the data.

Another empirically testable relationship arises from the following consideration. If the economy is to switch among regimes in the steady state, the relative attractiveness of one technology vis á vis the other must be time varying (otherwise only one technology dominates in the steady state). Moreover, if it is prohibitively costly to instantaneously shift a finite stock of capital between the two technologies, then the distribution of capital will not fully and instantaneously respond to changes in the ‘relative attractiveness’ variable, and consequently will enter as a supplementary state variable. The need for two state variables can account for the observed time-variation in the ability of the slope of the term structure to forecast economic growth (Stock and Watson, 2003). Here too our model identifies consumption heteroskedasticity as an additional variable useful for forecasting consumption growth, and we do find support for this hypothesis in the data.

It is not the case that supplementing any endowment economy with production will be consistent with the listed stylized facts. If regime changes are to exist in a stationary equilibrium, the relative attractiveness of one technology vis á vis the other must be time varying as well. This is not the case, for instance, in a model where (i) the representative agent’s utility is time-separable and exhibits constant relative risk aversion, (ii) aggregate endowment evolves as geometric Brownian motion, and (iii) the yield on the reversible technology is time-invariant. Such a model will not lead

changing the model results.

to time variation in the relative ‘attractiveness’ of the endowment versus production technologies. Thus only one technology will dominate and there will be no regime changes in the steady state.

We qualitatively characterize endowment economies that can accommodate steady state regime switching when combined with a reversible constant returns to scale technology. Among the several parsimonious specifications we consider, only one has the potential to fit all the stylized facts. It combines an endowment economy with linear habit formation and random walk endowment growth (Campbell and Cochrane, 1999), with a reversible technology featuring constant risk-free returns. After calibrating the model to a set of stylized facts, it additionally predicts a weak positive correlation between consumption volatility and the level/volatility of real rates, as well as a stronger positive correlation between volatility and average consumption growth. We find evidence in support of our model when we test these predictions. Confirming the latter finding appears both novel and significant given that the most reliable forecasting variable for growth is the slope of the term structure (see Stock and Watson (2003) for a review).⁴

The next subsection reviews the literature. Section 2 outlines the qualitative aspects of an economy that combines the above mentioned features of production and endowment economies. Section 3 formally derives the model, solves an approximation of it, and characterizes the relevant macro-economic variables. Section 4 calibrates the model to fit the stylized facts and, beyond that, briefly assesses empirical predictions.

1.1 Related Work

While we are not aware of any other general equilibrium model that exhibits endogenous regimes in interest rates, there are a variety of important related papers. Of the papers already mentioned, Garcia and Perron (1996) consider regimes in the real rate, while Evans (2003) and Ang and Bekaert (2004) extend this to both real and nominal rates. Dai et al. (2003) is also a recent study of the nominal term structure.

⁴Consumption volatility is a proxy for the habit level in our model. We also largely confirm the model’s predicted relationship between habit level and interest rates and/or expected consumption growth using a constructed proxy of the habit level.

Evans (2003) uses market data on U.K. real and nominal bond prices to estimate a reduced form Cox-Ingersoll-Ross-type equilibrium model of the term structure with regimes. Evans identifies three distinct states for real and nominal U.K. data: (1) upward sloping real and nominal term structures, (2) downward sloping real and U-shaped nominal, and (3) upward sloping real and U-shaped nominal. This is interpreted to mean that there are two real regimes. He finds that the real term risk premium is negative when upward sloping (Regimes 1 and 3) and positive when downward sloping (Regime 2) (see his Table 2). Not only does the term premium change sign, but it is also positively correlated with rate volatility.

Ang and Bekaert (2004) are closely related to Evans (2003), but allow in addition time-varying conditional term premia and separate regimes for both the real rate and inflation. They filter the real rate to find two regimes; (1) upward sloping with low mean and low volatility, and (2) downward sloping with high mean and high volatility. The unconditional term structure is fairly flat at about 1.5%.

Dai et al. (2003) extend the single-factor regime switching models of Naik and Lee (1997); Landen (2000); Dai and Singleton (2002) to a multi-factor model. They allow for state dependent transition probabilities, and priced regime switching risk (as opposed to Bansal and Zhou, 2002). Estimating the model on nominal U.S. Treasury zero coupon bond prices, they document two regimes; (1) a persistent low mean, low volatility regime, and (2) a less persistent high mean, high volatility regime. Moreover, they find support for a hump in the term structure, but find it is pronounced only in the low volatility regime. State dependence of transition probabilities is argued to be important in capturing the relationship between the slope of the term structure and the business cycle.

One common feature of these models is that interest rates are modeled in reduced form. This characterizes most work on modeling the term structure after the early and influential general equilibrium study of the real term structure by Cox et al. (1985b) (based on the production-based asset pricing model of Cox et al. (1985a)). Some exceptions that relate to our particular approach, include Dunn and Singleton (1986) and Wachter (2004). Dunn and Singleton (1986) is an early attempt at relating an equilibrium (endowment economy) term structure model to real data. They consider whether a representative agent model with habit formation and durable goods can explain

the observed term structure. They do not find support for the model, however.

Wachter (2004) uses a pure endowment economy with external habit formation to explain key features of the nominal term structure of interest rates. It can account for aspects of the expectations puzzle (Campbell and Shiller, 1991), short- and long-term fluctuations in interest rates, as well as high equity premium and stock market volatility, due to external habit formation.

Despite being able to account for an impressive collection of features of the observed term structure, no general equilibrium model that we are aware of offers an explanation of the observed regimes in the term structure and their characteristics. We also note here that while our own work focuses on real interest rates, it also applies to nominal rates in so far as inflation does not have interesting dynamics.⁵

A related strand of literature focuses on the equilibrium term structure of commodity forward/future curves. Indeed, one can view the net convenience yield as an interest rate differential between real dollars and currency denominated in terms of the commodity. In this literature, capacity constraints and adjustment costs (i.e., investment irreversibility) play an important role in the onset and characterization of regimes. Routledge et al. (2000), for instance, model storage held by risk-neutral traders who optimally supply the commodity to the economy in high price states. The reversible technology plays a similar role in our model: just as zero-storage states drive the shape of the commodity forward curves in Routledge et al. (2000), in our model states in which no capital is invested or being invested in the reversible technology lead to a term structure distinct from that seen in other states. Recent papers by Kogan et al. (2003) and Cassasus et al. (2004) offer a related, though more detailed, explanation for the behavior of forward prices by considering an economy with irreversible investment. Here, regimes exist in capital investment policy due to non-trivial adjustment costs, capacity constraints and irreversibility. The underlying theme behind all these theoretical papers is that technological frictions lead to regimes in the shadow price of capital which, in turn, translate into regimes in the prices of commodity bonds (i.e., forward contracts). While our paper shares the underlying theme, the goal of these papers is to explain stylized facts peculiar to the term structure of commodity forwards (e.g., volatility structure, backwardation, etc.); by contrast, our goals and therefore modeling techniques focus exclusively on explaining

⁵One can easily supplement our real rate process with an exogenously specified inflation process.

attributes of the yield curve.

2 Qualitative Analysis

We consider an economy with a representative agent who receives endowments at dates $t, t + \Delta t, t + 2\Delta t, \dots$. The endowment, $y_t \Delta t$, is represented by a rate y_t received over a time interval of length Δt that we will shortly take to zero (our main results are stated in continuous time, but the intuition is more transparent in discrete time). y_t might be interpreted, for instance, as the return on labor capital: $y_t = r_t^y K_t$ where K_t is the stock of labor at date t and r_t^y is the return on this stock. In addition, part of the representative agent's wealth, Q_t , at date t is invested in a reversible risk-free technology that yields a return of η_t .⁶ The capital generating the endowment cannot be transferred to the reversible technology and vice versa (i.e., Q_t cannot be increased or decreased at the expense of K_t), thus we are in effect assuming that y_t is exogenously specified. Aside from risk-free growth at the rate of η_t , the amount of capital invested in the reversible technology can change only by investing a portion of the flow of endowment, say q_t . The quantity $q_t \Delta t$ can be no larger than $y_t \Delta t$ and no smaller than $-Q_t$, depending on whether capital is added or dismantled from the reversible technology.⁷ The aggregate rate of consumption, c_t , is given by $y_t - q_t$.

To understand how the level of Q_t is related to real rates, consider that if the instantaneous rate of risk-free borrowing, r_t is below η_t , an arbitrage opportunity exists at the level of the representative agent: borrow at the lower interest rate and invest at the higher yield of η_t . Thus an equilibrium imposes η_t as a lower bound for prevailing interest rates. On the other hand, one should not observe $Q_t > 0$ if short term interest rates are above η_t , since the representative agent could divest from the reversible technology, and invest the capital Q_t in higher yielding risk-free bonds. In other words, interest rates are at η_t or above, and are equal to η_t whenever $Q_t > 0$.

⁶We briefly argue later that assuming a risky reversible production technology does not qualitatively alter the results. It does, however, make the analysis more complicated and introduces an additional degree of freedom.

⁷The lower limit on investment in the technology can alternatively be set by requiring that Q_t doesn't fall below some arbitrary level of 'maintenance' capital. We point this out to emphasize that a 'stockout' need not correspond to zero capital investment, $Q_t = 0$. Rather, it corresponds to the point at which severe costs are imposed on divesting from an otherwise relatively reversible technology.

The remaining case, corresponding to $Q_t = 0$ with $r_t > \eta_t$, is economically admissible since no investment capital is available from the reversible technology; moreover, since there is no incentive to increase Q_t above zero, the economy is temporarily equivalent to that of pure endowment and exhibits the corresponding endowment economy short rate, hereafter denoted as r_t^E .⁸ In particular, this implies that an incentive to invest when $Q_t = 0$ exists only if $\eta_t > r_t^E$. The table below summarizes the relationship between the behavior of real interest rates (denoted by r_t), the level of capital in the reversible technology Q_t , and its yield η_t .

	$\eta_t < r_t^E$	$\eta_t \geq r_t^E$
$Q_t > 0$	Production Economy, $r_t = \eta_t$	Production Economy, $r_t = \eta_t$
$Q_t = 0$	Endowment Economy, $r_t = r_t^E$	Production Economy, $r_t = \eta_t$

The reversible technology is utilized in the ‘Production Economy’ regime while not in the ‘Endowment Economy’. The table illustrates the corresponding potential for ‘regime switching’ in real interest rates, where one regime corresponds to identifying real rates with the yield on the reversible technology, while the other regime corresponds to identifying real rates with those of the associated pure endowment economy. Recalling that interest rates are linked to the marginal product of capital, one can view the regimes in interest rates as arising from the distribution of capital among the technologies. While the regimes are particularly well defined in our model, the results would be qualitatively similar if one made the risk-free reversible technology risky or less reversible, the costs of shifting capital to or from the ‘endowment’ technology finite, or the point of irreversibility, $Q_t = 0$ an arbitrary level of minimal ‘maintenance capital’.

If η_t is always above r_t^E then capital in the reversible technology is never completely depleted. Likewise, if η_t is always below r_t^E then if capital is ever depleted, it is never subsequently replenished. Consideration of this argument leads immediately to our first result:⁹

Proposition 1. *A necessary condition for the economy described above to exhibit a stationary*

⁸The corresponding endowment economy is the one in which consumption is restricted to be equal to the endowment.

⁹We only include the more involved proofs to our results, and do so in a separate appendix.

equilibrium with regime switching is that the unconditional probability of each of the events, $\eta > r^E$ and $r^E > \eta$ is strictly positive.

In other words, to achieve a steady state equilibrium whose characteristics switch between those of endowment and production economies there must be a non-zero probability that η_t is greater than r_t^E , and vice versa. Note that the Proposition does not supply sufficient conditions for regime switching—these are generally model-dependent.¹⁰ An important corollary to this result is that one cannot achieve regime switching by introducing a constant yield technology into an endowment economy with constant interest rates.

Corollary to Proposition 1: *A necessary condition for a stationary economy to exhibit regime switching is that either η_t or r_t^E exhibit time-variation.*

In particular the stationary equilibrium of an economy with geometric random walk endowment combined with a constant yield risk-free technology generally leads to a trivial steady state which is either pure endowment (as in Rubinstein, 1976; Lucas, 1978; Breeden, 1979) or pure production (as in Brock, 1982; Cox et al., 1985a).¹¹ We are therefore led to consider assumptions that imply one or both of η_t or r_t^E exhibit time-variation. This can be achieved by positing that η_t itself is time varying, that the endowment growth rate is stochastic, or that preferences make r_t^E depend on an additional state variable. Consider in particular the following specifications:

Model	Preferences	Endowment Income	Reversible Technology
1	CRRA	Constant expected growth rate	AR(1) η_t
2	CRRA	AR(1) expected growth rate	Constant η_t
3	Exogenous proportional habit formation	Constant expected growth rate	Constant η_t
4	Exogenous linear habit formation	Constant expected growth rate	Constant η_t

CRRA preferences refer to time-separable and time homogeneous expected Bernoulli utility of consumption with constant relative risk aversion. Linear (proportional) habit formation is essentially the CRRA assumption with utility derived from the difference (ratio) between consumption

¹⁰For instance, if during states in which the economy depletes Q_t , it does so proportionately to Q_t , then a state in which $Q_t = 0$ will never occur in a stationary equilibrium where $Q_t > 0$ is an initial state.

¹¹This is also true for an Epstein and Zin (1989) representative agent economy.

and a weighted historical average of the aggregate endowment (see Abel, 1990; Campbell and Cochrane, 1999). The endowment is assumed to be a geometric random walk with an expected growth rate that is either constant (Models 1, 3 and 4) or mean reverting (Model 2).¹² Model 1 exhibits time variation in η_t while keeping r_t^E constant, whereas Models 2, 3 and 4 exhibit time variation in r_t^E while keeping η_t constant. Although there is a strong case to be made for combining some of these assumptions or introducing other specifications, we feel that it is sufficient to first consider these cases in isolation to illustrate the main insights. All of the models have the capacity to exhibit stationary regime changes when suitably parameterized. However, not all of them have the potential to be consistent with *all* the stylized facts we listed in the Introduction. We argue in Appendix B that Model 1 implies either that interest rates are highest in the smooth regime, or that the volatile regime arises too frequently. Moreover, while the smooth regime in Model 2 exhibits lower interest rates, reasonable parameter constraints appear to prevent it from occurring sufficiently frequent. Furthermore, the sign of the risk premium for interest rates in Models 1 and 2 is generally constant while it is always constant in Model 3. Risk premia for real bonds consequently do not change sign. Model 4, on the other hand, appears to be capable of simultaneously exhibiting all the stylized facts listed earlier. The model moreover has the advantage of simultaneously accommodating a high equity risk premium (as in Campbell and Cochrane, 1999) and keeping the volatility structure of interest rates realistic, which is not true of Models 1–3. Because of the relative complexity involved in analyzing any one of the models in detail, we focus on Model 4 and assume the economy can be represented by an agent with exogenous linear habit formation.¹³

Let s_t be the habit level relative to the agent’s endowment. In the pure endowment version of the economy we consider (Model 4), the short interest rate has the form

$$r_t^E = a_0 + \frac{b_0}{1 - s_t} - \frac{c_0}{(1 - s_t)^2}$$

where the coefficients b_0 and c_0 are positive. Since the variable s_t is mean-reverting, provided η is in the range of r_t^E , the economy will fluctuate between states in which η is greater and smaller than r_t^E . Note that if at some date $Q_t = 0$, the fact that η is in the range of r_t^E guarantees that

¹²A version of Model 2 is also studied by Deaton (1991). He focuses on the behavior of consumption and savings under liquidity constraints rather than the term structure of interest rates.

¹³We have detailed solutions for the other models, along the lines developed for Model 4—the more analytically involved of the four models. Solutions to the other models are available from us upon request.

$Q_{t'} > 0$ at some future date, $t' > t$.¹⁴ This is depicted in Figure 1 along with the constant risk-free yield, η . When $\eta > r_t^E$ or $Q_t > 0$ the economy must be in a regime, denoted ‘Regime P’, in which the reversible production is active and the actual rate is η (see the thick dashed line). If $r_t^E > \eta$ and no capital is invested in the reversible technology, the economy shifts to a pure endowment regime, referred to as ‘Regime E’ (i.e., Q_t is zero and there is no incentive to invest). In Regime E, the actual rate equals r_t^E (see the thick solid curve).

One can immediately deduce several qualitative implications from the graph. The real rate in Regime P is constant at η while it is time varying in Regime E. Moreover, since η is a lower bound for interest rates, the Regime P rates are low in both level and volatility relative to the Regime E rates. Ignoring the impact of possible risk adjustment, the model features a term structure that is upward sloping in the constant rate Regime P (since the short rates are as low as they can be), and potentially downward sloping in Regime E. As Figure 1 suggests, when the economy is in Regime E the short rate features a region where it increases and one where it decreases with s_t , the ratio of habits to endowment. In the increasing (decreasing) rate region of Regime E, a positive shock to the endowment, or equivalently consumption, is a negative shock to s_t and thus a *negative (positive)* shock to interest rates. This suggests that the risk premium for real rates can be negative or positive in Regime E (it must be zero in Regime P since the short rate is constant). The likelihood of exhibiting either sign depends on the unconditional distribution function of s_t .¹⁵

In the next two sections, we establish the existence of a stationary equilibrium with regime switching for Model 4 under reasonable parameter assumptions, and calibrate the model to exhibit the interesting and complex term structure dynamics outlined above and in the Introduction. In addition, the model makes two predictions concerning the heteroskedasticity of consumption growth and the forecasting power of the slope of the term structure, which we now turn to describe.

Intuitively, when the economy is in Regime P and capital in the reversible technology can be

¹⁴There is no guarantee, however, that capital invested in the reversible technology is ever depleted. Overall, the frequency of different regimes depends on the model parameters.

¹⁵Other studies of arithmetic habit formation are typically ‘one-sided’ in the sense that recessions are associated with large risk aversion and high habit levels. To accommodate the (infrequent) possibility of a positive risk premium for real rates we forego this monotonic relationship. However, it should be added that, given our parameter choices, a monotonic relationship between recessions and habit levels is the norm.

dismantled and consumed, aggregate consumption might be expected to be smoother relative to the volatility of the endowment (i.e., consumption in Regime E). When calibrated to the stylized facts, our model predicts a positive relationship between the level and volatility of real interest rates and the volatility of consumption growth. Alternatively, consumption volatility should be negatively related to any indicator of Regime P. We later check and find mixed support for these predictions in post-war US data (1957-2002).

Model 4 can also be shown to positively link expected consumption growth to two variables in Regime P: the efficiency of the reversible technology relative to r_t^E and the rate of investment, q_t . These, in turn, depend on the two state variables in the model (s_t and Q_t); thus to forecast consumption growth one needs proxies for these variables. The slope of the term structure is strongly regime dependent and therefore might be such a proxy. However, in the more frequent Regime P the short end of the term structure has little variability; given the low variability of the long end, the implication is that the slope of the term structure *within* the more frequent regime P does not vary much and is therefore not a good forecasting variable. Our model therefore predicts that the term spread is not an ideal indicator of future consumption growth. While the forecasting power of the latter for real activity is empirically well acknowledged (see Stock and Watson (2003) for a review), its effectiveness varies with subperiods. Our approach may therefore help to provide a theoretical rationale for the time varying forecasting power of the term spread. On the other hand, consumption volatility does vary within Regime P in our model and should therefore have additional forecasting power beyond what is offered by the term spread. We provide some confirmation of this later.

In the remainder of the paper we investigate Model 4 in detail, establish conditions under which the above stated results hold, calibrate the model and elaborate on the findings. In the last section we briefly investigate the degree to which the model's predictions with respect to consumption heteroskedasticity are consistent with US data in the post-war period.

3 The Model

We restrict ourselves to analyzing the case where there is a representative agent with time separable and time homogeneous Bernoulli utility with constant relative risk aversion over surplus consumption. In Appendix B we argue that the other specifications (Models 1–3) considered in Section 2 cannot be calibrated to all of the stylized facts. Specifically, the agent has inter-temporal utility for the consumption stream $\{c_{n\Delta t}\}_{n=0}^{\infty}$ given by

$$U = E_0 \left[\sum_{n=0}^{\infty} \frac{e^{-\beta n \Delta t}}{1-b} (c_{n\Delta t} - z_{n\Delta t})^{1-b} \Delta t \right]$$

where Δt is a time interval that will shortly be taken to zero, $\beta > 0$ and represents an exogenous habit level that corresponds to a smoothed history of the endowment stream. We define the level of habits

$$z_{n\Delta t} = \min \left\{ z_{-1} e^{-\alpha n \Delta t} + \lambda \sum_{j=0}^n e^{-\alpha(n-j)\Delta t} y_{j\Delta t} \Delta t, (1-\epsilon)y_{n\Delta t} \right\}$$

in terms of past endowments y_{τ} , $\tau \leq t$. This choice closely resembles that of Abel (1990) and Campbell and Cochrane (1999). While they study pure exchange economies for which $c_t = y_t$, this relationship does not hold exactly in the present economy where $c_t = y_t - q_t$. The distinction between c_t and y_t for computing z_t is at best of cursory interest, however, for the parameters we consider.¹⁶ Defining habits in terms of per capita consumption will not change our results, but will greatly increase the complexity of the model, as will defining habits in terms of individual consumption (as in Duesenberry, 1952; Pollak, 1970; Ryder and Heal, 1973; Sundaresan, 1989; Constantinides, 1990; Detemple and Zapatero, 1991). We moreover ensure that utility is well defined by assuming $\epsilon \in (0, 1)$. Again, this simplification does not affect our results for the parameters we consider, as the probability of $z_t \geq c_t$ is negligible. In addition to be analytically tractable, the specification we suggest avoids the problem of negative state prices (Chapman, 1998).

At time $t = n\Delta t$ the representative agent begins the period with a stock of capital, Q_t , invested in the reversible technology and an endowment of $y_t \Delta t$. She then invests $q_t \Delta t \in [-Q_t, y_t \Delta t]$ of the endowment and consumes the remainder, $c_t \Delta t = (y_t - q_t) \Delta t$, where q_t can be negative because

¹⁶For the parameters we choose, Figure 7 gives an indication of the equilibrium amount stored relative to labor income, q_t/y_t . The figure indicates that q_t , is negligible for the purpose of computing the level of habit z_t .

the technology is reversible. The stock of invested capital evolves into next period's stock as $Q_{t+\Delta t} = e^{\eta\Delta t}(Q_t + q_t\Delta t)$.

Theorem 1. *Consider the problem*

$$\sup_{\{q_{n\Delta t}\}_{n=0}^{\infty}} E_0 \left[\sum_{n=0}^{\infty} \frac{e^{-\beta n\Delta t}}{1-b} (y_{n\Delta t} - q_{n\Delta t} - z_{n\Delta t})^{1-b} \Delta t \right] \quad (1)$$

subject to

$$z_{n\Delta t} = \min \left\{ z_{-1} e^{-\alpha n\Delta t} + \lambda \sum_{j=0}^n e^{-\alpha(n-j)\Delta t} y_{j\Delta t} \Delta t, (1-\epsilon)y_{n\Delta t} \right\}$$

$$Q_{(n+1)\Delta t} = e^{\eta\Delta t}(Q_{n\Delta t} + q_{n\Delta t}\Delta t),$$

where the stochastic process $\{y_{n\Delta t}\}_{n=0}^{\infty}$ is a.s. bounded away from zero and defined relative to a probability space (Ω, \mathcal{F}, P) . Then for any $\beta > 0$ and $b > 1$, the optimization problem in (1) has a solution.

The first order (Kuhn-Tucker) conditions from the inter-temporal optimization problem easily imply the following result.

Proposition 2. *Assume the conditions in Theorem 1, let $t = n\Delta t$, and let r_t^E be the solution to $e^{-r_t^E \Delta t} = e^{-\beta\Delta t} E_t \left[\left(\frac{y_{t+\Delta t} - z_{t+\Delta t}}{y_t - z_t} \right)^{-b} \right]$. If $Q_t > 0$ or $r_t^E \leq \eta$ then*

$$e^{(\eta-\beta)\Delta t} E_t \left[\left(\frac{y_{t+\Delta t} - q_{t+\Delta t} - z_{t+\Delta t}}{y_t - q_t - z_t} \right)^{-b} \right] = 1 \quad (2)$$

where $y_{l\Delta t} > q_{l\Delta t} \geq -\frac{1}{\Delta t} Q_{l\Delta t}$ for any period, l .

To understand the intuition behind the proposition, first note that the risk-free rate in an otherwise identical pure endowment economy (i.e., absent the reversible technology) is r_t^E . Proposition 2 states that the risk-free return set by the inter-temporal marginal rate of substitution, $e^{-\beta\Delta t} \left(\frac{c_{t+\Delta t} - z_{t+\Delta t}}{c_t - z_t} \right)^{-b}$, must equal the return on the reversible technology whenever the technology is 'in use.' The latter is true when capital is not depleted (i.e., $Q_t > 0$), or when $Q_t = 0$ and $r_t^E < \eta$. Note that if $q_t = 0$ at $Q_t = 0$, then the interest rate is r_t^E ; if this is smaller than η then an arbitrage opportunity exists: one can borrow at a low risk-free rate and invest at the higher risk free yield.

Let $t = n\Delta t$. As it turns out, the Euler equation specified in Proposition 2 is sufficient to solve for the optimal investment policy, q_t , as a function of Q_t, y_t and z_t . In turn, this solution corresponds to the optimal consumption, c_t , and can be used to deduce the equilibrium real risk-free rate of return as well as other macroeconomic variables of interest. To obtain an intuition as to why Proposition 2 is all that is needed, consider the finite horizon case (the derivation of (2) is not affected by whether the horizon is infinite or not). When there is only a single period left and under the conditions stated in the proposition, the Euler equation becomes,

$$e^{(\eta-\beta)\Delta t} E_{T-\Delta t} \left[\left(\frac{y_T + e^{\eta\Delta t}(Q_{T-\Delta t} + q_{T-\Delta t}\Delta t) - z_T}{y_{T-\Delta t} - q_{T-\Delta t} - z_{T-\Delta t}} \right)^{-b} \right] = 1$$

where $q_T\Delta t$ was set equal to $-Q_T = -e^{\eta\Delta t}(Q_{T-\Delta t} + q_{T-\Delta t}\Delta t)$ (i.e., ignoring a bequest motive, it is optimal to completely deplete capital at date T). Assuming y_t is Markovian at all dates, this equation can in principle be inverted to solve for $q_{T-\Delta t}$ as a function of $Q_{T-\Delta t}, y_{T-\Delta t}$ and $z_{T-\Delta t}$. The solution can then be used via backward induction to solve for $q_{T-2\Delta t}$, and so on.

We note that if the reversible technology is risky, the analysis is similar though more involved. For instance, the Euler equation when $Q_t > 0$ in (2) becomes

$$e^{-\beta\Delta t} E_t \left[e^{\Delta v_t} \left(\frac{y_{t+\Delta t} - q_{t+\Delta t} - z_{t+\Delta t}}{y_t - q_t - z_t} \right)^{-b} \right] = 1$$

where Δv_t is the stochastic rate of return on the reversible technology between date t and $t + \Delta t$. By solving for q_t in this case one can also solve for the implied risk-free rate (by calculating the expected value of the inter-temporal marginal rate of substitution). The risk-free rate equals r_t^E when $q_t = 0 = Q_t$, and generally deviates from r_t^E otherwise, thereby exhibiting regimes (though perhaps not as pronounced as under our assumptions).

3.1 The Continuous Time Limit

We now let $\Delta t \rightarrow 0$, allowing us to derive a partial differential equation for the optimal policy. In this limit, our model assumptions are given by the following:

Assumption 1. *The endowment follows the diffusion process, $\frac{dy_t}{y_t} = \mu dt + \sigma dW_t^y$, the risk-free yield is constant at $\eta_t = \eta$, and $z_t \equiv \min \left\{ z_0 e^{-\alpha t} + \lambda \int_0^t e^{-\alpha(t-\tau)} y_\tau d\tau, (1-\epsilon)y_t \right\}$, for some $0 < \epsilon \ll 1$, $z_0 \in (0, 1-\epsilon)$ and $0 < \lambda < \alpha + \mu - \sigma^2 < \infty$.*

Under the assumption for α and λ , $s_t \equiv \frac{z_t}{y_t}$ follows a reflected and mean-reverting Itô process in $(0, 1 - \epsilon]$. When s_t is strictly inside that interval, $dz_t = (\lambda y_t - \alpha z_t)dt$ and

$$ds_t = (\alpha + \mu - \sigma^2) \left(\frac{\lambda}{\alpha + \mu - \sigma^2} - s_t \right) dt - \sigma s_t dW_t^y \quad (3)$$

Proposition 3. *For $s_t \in (0, 1 - \epsilon)$, the interest rate in the pure endowment economy is given by*

$$r^E(s_t) = \beta + b(\mu - \lambda) + b(\alpha + \mu - \lambda) \frac{s_t}{1 - s_t} - \frac{b(b+1)\sigma^2}{2(1 - s_t)^2}$$

The earlier discussion in Section 2 indicates that regime switching is only possible if $r^E(s) > \eta$ over a finite region of $s \in [0, 1 - \epsilon]$. Note that $r^E(s) - \eta \geq 0$ yields a quadratic inequality that is satisfied between its two roots. This requires the following parameter constraints:

Assumption 2. *The two roots to the equation $r^E(s) - \eta = 0$, denoted as s_- and s_+ , are real, distinct and $(0, 1 - \epsilon)$ contains at least one root.*

Without loss of generality, we assume $s_- < s_+$. A necessary condition for there to be at least one root in $(0, 1 - \epsilon)$ is that $\alpha + \mu - \lambda > \sigma^2(b+1)$.

The next proposition derives the partial differential equation governing the optimal investment policy and resulting from application of the first order conditions (the Euler equation and Kuhn-Tucker condition).

Proposition 4. *Set $s_t \equiv \frac{z_t}{y_t}$ and $x_t \equiv \frac{Q_t}{z_t}$, then $q(y_t, z_t, Q_t)$ is homogeneous of degree one in y_t . Moreover, letting $g_t = g(x_t, s_t) \equiv 1 - q(y_t, z_t, Q_t)/y_t - s_t$, for $s_t \in (0, 1 - \epsilon)$, g_t satisfies*

$$\frac{b+1}{2} \sigma^2 \frac{s^2 g_s^2}{g} = Ag + \left((\alpha + \eta - \frac{\lambda}{s})xs + 1 - s - g \right) \frac{g_x}{s} - (Cs - \lambda)g_s + \frac{\sigma^2 s^2}{2} g_{ss}$$

where $A \equiv (\beta + b\mu - b(b+1)\frac{\sigma^2}{2} - \eta)/b$, and $C \equiv \mu + \alpha - (b+1)\sigma^2$.

The variable, g_t , represents the surplus (net) consumption relative to the endowment, y_t . We suppress the t -dependence of g_t in the PDE for notational convenience (i.e., g_s, g_x and g_{ss} are derivatives of $g_t = g(x_t, s_t)$). To fix the boundary conditions note that $Q = 0$ iff $x = 0$ (since $z > 0$). The no-arbitrage boundary conditions at $Q = 0$ translate to $g(0, s_t) = 1 - s_t$ whenever $r_t^E \geq \eta$. Note also that as $x_t s_t = Q_t/y_t \rightarrow \infty$, $x_t \rightarrow \infty$ (since s_t is in $(0, 1)$). Moreover, as $\frac{Q_t}{y_t} \rightarrow \infty$,

the impact of y_t or s_t on the representative agent should be negligible, and it is sensible to require the optimal control, $q(y_t, z_t, Q_t)$, to be linear in Q_t as for a CRRA agent with wealth Q_t who can invest at a risk-free rate η and consume only from her investments. In other words, we require $g(x, s_t) \xrightarrow{x \rightarrow \infty} kxs$, where k is some constant. Substituting this into the differential equation in Proposition 4 in the limit $x \rightarrow \infty$ yields $k = A - M$ where $M = \mu - \eta - (b + 1)\sigma^2$.

The PDE for the optimal investment policy cannot be solved analytically and standard numerical approaches are difficult to implement due to the highly stiff nature of the solution as x_t approaches zero in the region $r^E(s_t) > \eta$. On the other hand, one may hope that by judiciously ignoring ‘small’ terms one can analytically approximate the solution.

Noting that when x_t is small, $g(x, s) \approx 1 - s$, we approximate $\frac{\sigma^2 s^2}{2} g_{ss}$ as zero and $\frac{b+1}{2} \sigma^2 \frac{s^2 g_s^2}{g}$ as $\frac{b+1}{2} \sigma^2 \frac{s^2}{1-s}$. This leads to the following:

$$Ag = \frac{b+1}{2} \sigma^2 \frac{s^2}{1-s} - (1 - s - g + (\alpha + \eta - \frac{\lambda}{s})xs) \frac{g_x}{s} + (Cs - \lambda)g_s \quad (4)$$

The validity of this approximation depends on the parameters and can be gauged on a case by case basis. We return to this when the model is calibrated.

Theorem 2. *Given Assumptions 1 and 2, and that $s_- < \frac{\lambda}{C} < s_+$, the solution to Eqn. (4) with boundary conditions $g(0, s_t) = 1 - s_t$ for $r^E(s) \geq \eta$ and $g(x, s_t) \xrightarrow{x \rightarrow \infty} (A - M)xs$ (with $M = \mu - \eta - (b + 1)\sigma^2$) is given by the solution to*

$$0 = xs - \frac{g}{A - M} + V_1(s) + V_2(s, g) \quad (5)$$

where the functions $V_1(s)$ and $V_2(s, g)$ are specified in Appendix A.

While the optimal investment policy is not trivial, it is straight forward to compute. More importantly, it can be used to simulate the evolution of the relative capital level state variable, x_t , and to calculate the volatility and expected rate of consumption growth, as well as the term structure of interest rates. We also have the following result:

Proposition 5. *If $A > M$, then $g(x, s)$ is increasing in x .*

Another necessary condition for regime switching in the steady state is that x_t is mean reverting. In particular, it must be that as $x \rightarrow \infty$ its expected growth rate is negative. This provides for

a final parameter constraint when calibrating the model. Note the boundary condition at $x \rightarrow \infty$ requires $A > M$ if it is the case that $g_x > 0$ in that limit. A more useful result in this regard can be obtained by examining the evolution of $\hat{x}_t = x_t s_t$:

Lemma 1.

$$d\hat{x}_t = (1 - g(x, s) - s + (\eta - \mu + \sigma^2)\hat{x}_t)dt - \hat{x}_t \sigma dW_t^y$$

Note that since s_t is mean-reverting, \hat{x}_t mean-reverts if and only if x_t does. For \hat{x}_t to be mean-reverting, its expected growth must be negative for large \hat{x}_t . From the Lemma, as $\hat{x} \rightarrow \infty$ it can be approximated by geometric motion with growth rate $M - A + \eta - \mu + \sigma^2$. If \hat{x}_t is initially large, then to ensure that the probability that it almost surely decreases one must require that $M - A + \eta - \mu + \frac{\sigma^2}{2} < 0$, or alternatively, $\beta + b\mu + \frac{b^2}{2}\sigma^2 > \eta$.

3.2 Evolution of Macroeconomic Variables

We now provide an analysis of some macroeconomic variables that depend on the investment policy. In particular, we focus on consumption, the maximal Sharpe ratio, and bond prices.

Consumption Growth

The evolution of aggregate consumption growth in the region described by $x = 0$ and $s \in [s_-, s_+]$ (see Assumption 2) coincides with that of the endowment (i.e., Regime E). Everywhere *outside* that region, the evolution is given by

$$\begin{aligned} \frac{dC_t}{C_t} &= \frac{d[y_t(g_t + s_t)]}{y_t(g_t + s_t)} = \frac{dy_t}{y_t} + \frac{dg_t + ds_t}{g_t + s_t} + \frac{[dg_t + ds_t][dy_t]}{y_t(g_t + s_t)} \\ &\approx \left(\mu + \frac{g_x}{(g_t + s_t)s_t} (1 - g_t - s_t + ((\alpha + \eta)s - \lambda)x_t) + \frac{g_s + 1}{g_t + s_t} (\lambda - (\alpha + \mu)s_t) + o(\sigma^2) \right) dt \\ &\quad + \left(1 - \frac{s_t(g_s + 1)}{g_t + s_t} \right) \sigma dW_t^y \\ &\approx \left(\mu + \frac{(1 - s_t - g_t)A + (\eta - r^E(s_t))(1 - s)/b}{g_t + s_t} + o(\sigma^2) \right) dt + \left(1 - \frac{s_t(g_s + 1)}{g_t + s_t} \right) \sigma dW_t^y \end{aligned}$$

The $o(\sigma^2)$ represents a term that has the same order of magnitude as the volatility of consumption squared and, anticipating it to be small, we ignore its contribution to consumption growth. In

optimizing, the representative agent considers the trade-offs between a strategy of high short term consumption, which depletes capital faster and therefore features lower future consumption growth, versus a more modest plan which allows capital to grow so that a higher growth rate of consumption can be enjoyed in the future. The expression for the expected consumption growth rate indicates that the key considerations are the relative amount invested today, $1 - s_t - g_t = q_t/y_t$, and the relative efficiency of the reversible technology, measured by $\eta - r^E(s_t)$. So long as $A > 0$ —which is typically true if σ is small—positive investment contributes to the future growth rate of consumption. Moreover, the higher the efficiency of the reversible technology, the more attractive it is for the representative agent to increase consumption growth by moderating current consumption. It is worth noting that the expected growth rate of consumption is not continuous across regimes. In particular, the regime change in interest rates also shows up as a regime change in the expected growth rate of consumption.

Consumption volatility is also a function of the optimal investment policy. Whenever $1 + g_s$ is positive, the reversible technology acts to smooth consumption. While not uniformly true, this is indeed the case for $x = 0$ and $s \in [0, s_-]$ since the reversible technology becomes increasingly less attractive as the endowment economy interest rate rises and $g_s = -1$ at s_- .

Maximal Sharpe Ratio

In Regime P, the market price of risk corresponds to the volatility of the pricing kernel,

$$\left(\frac{y_{t+\Delta t} - q_{t+\Delta t} - z_{t+\Delta t}}{y_t - q_t - z_t} \right)^{-b} = \left(\frac{y_{t+\Delta t} g_{t+\Delta t}}{y_t g_t} \right)^{-b}$$

Using Itô's Lemma, this is

$$\sigma_M \equiv b\sigma \left(1 - \frac{sg_s}{g} \right)$$

which continuously approaches the analogous expression in the endowment economy regime: $b\sigma \frac{1}{1-s_t}$. Following the argument made earlier, g_s might be expected to be negative. The market risk premium associated with a shock adW_t^y is $a\sigma_M$, thus σ_M can be interpreted as the instantaneous maximal Sharpe Ratio.

Interest Rates

The shape of the term structure generally depends on expectations for future interest rates, rate volatility and associated risk adjustments. Since the short rate is constant in Regime P, the shape of the term structure in the present economy is a function of the short rate dynamics in Regime E, and expectations about the onset of a new regime.

In Regime E, the short interest rate is not monotonic in s : increasing for $s \in [0, 1 - \frac{(b+1)\sigma^2}{\alpha+\mu-\lambda}]$ and decreasing thereafter. When increasing in s_t , the short rate is associated with a *negative* risk premium since shocks to the rate have the form $-\frac{dr^E}{ds}\sigma sdW_t^y$, and are negatively correlated with endowment shocks. The risk premium changes sign when r^E is decreasing in s . Thus, the risk premium associated with a given bond depends on the current regime, the state variable s_t , and the maturity of the bond. If, for instance, the distribution of s_t is such that most of the time r^E is increasing in s_t then one expects the risk premium for long maturity bonds to be positive always, while the risk premium for short maturity bonds might often be zero (i.e., in Regime P), occasionally positive, and sometimes negative. An application of the Feynman-Kac formula yields,

Lemma 2. *The date- t price of a zero-coupon bond maturing at date $t + \tau$, $P(\tau, t)$, solves the following differential equation:*

$$r_\tau P = -P_\tau + (1 - g_\tau - s_\tau + ((\alpha + \eta)s_\tau - \lambda)x_\tau)P_x + (\lambda - (\alpha + \mu)s_\tau + bs_\tau\sigma^2(1 - \frac{s_\tau g_s}{g}))P_s + \frac{\sigma^2 s_\tau^2}{2}P_{ss} \quad (6)$$

where $r_\tau = r^E(s_\tau)$ when $x = 0$ and $s \in [s_-, s_+]$, while $r_\tau = \eta$ otherwise. The PDE is solved subject to the boundary condition, $P(0, t) = 1$.

Notice that the risk premium for the bond, $bs_\tau\sigma^2(1 - \frac{s_\tau g_s}{g})P_s$, is generated through its dependence on s_t .

4 Analysis and Calibration

In this section, we calibrate the model to the stylized facts mentioned in the Introduction. We begin by summarizing the parameter restrictions from the previous section. These are given by

- i) $0 < \alpha, \lambda, \alpha + \mu - \lambda - \sigma^2 > 0$ (Assumption 1)

- ii) $\alpha + \mu - \lambda > (b + 1)\sigma^2$ (Assumption 2)
- iii) $\beta + b\mu + \frac{b^2}{2}\sigma^2 > \eta$ (Growth condition on x_t from Lemma 1)
- iv) $0 < s_- < \frac{\lambda}{C} < s_+$ (Theorem 2)

The first three constraints are trivial to satisfy in a realistic parameterization. For instance, one can associate y_t with labor income and $\alpha + \mu - \sigma^2$ with the length of a business cycle. Choosing μ and σ compatible with observed labor income growth rates, it is clear that (ii), and therefore (i), can be satisfied for reasonable values of b . The third inequality is also easily satisfied by setting $\beta \geq \eta$. While it is somewhat harder to see, it is a simple matter to show that for realistic levels of σ (e.g., $\sigma \sim 0.03$), one can generally find a β satisfying the last inequality.

In addition to the above constraints, we also require that $A > 0$, so that the expected consumption growth rate increases as the capital invested in the reversible technology increases, and attempt to achieve consistency with the following observations:

- i) Labor income growth is on average 2-3% per year with annual standard deviation of 3-4%.¹⁷
- ii) As with other habit formation models, one can interpret s_t as a business cycle variable and consistent with that, we require its mean reversion to correspond to a *time scale* of the order of a business cycle (≈ 5 years). Since the rate of mean reversion of s_t is $\alpha + \mu - \sigma^2$, this requirement sets $\alpha \approx 0.2$ (given the restrictions on μ and σ).
- iii) The maximal Sharpe ratio can vary by a factor of two over the business cycle. Given the relatively small volatility of s_t , this requires the median value of s_t to lie close to one and therefore constrains λ .
- iv) The low volatility regime in real interest rates appears to be more common in the post-war period and thus the high volatility state (Regime E) is less persistent than Regime P (Dai et

¹⁷One could legitimately associate y_t with GDP (i.e., total endowment growth). We prefer to think of it as labor income since y_t can have a very different growth rate from consumption. The numbers are estimates based on the labor income time-series from Martin Lettau's homepage. NIPA wage income or disposable income has a volatility that is smaller by a factor of two and closer to the volatility of aggregate consumption. GNP moments are similar to those of NIPA labor income data.

al., 2003). Moreover, in the low volatility regime the mean real short rate is low (1.4%) while in the high volatility regime it is about 1% higher (Ang and Bekaert, 2004). This fixes η to be close to 1% and requires one to set s_- sufficiently close to the median of s_t to ensure that Regime E is infrequent (with a frequency of about 20%).

v) The unconditional term structure is flat at between 1 and 2%.

We start with the following parameter values:

μ	0.025
σ	0.035
α	0.2
λ	0.185
η	0.014

Notice that η — the short rate in the more frequent Regime P — is set to be consistent with the observed mean rate in the smooth rate regime. The implied unconditional probability density of s_t is plotted in Figure 2.¹⁸ The first, median and 99th percentiles are, respectively, $s_{1\%} = 0.73$, $s_{50\%} = 0.82$, $s_{99\%} = 0.93$ with an unconditional mean and standard deviation of 0.82 and 0.043, respectively. This implies a potential variation by a factor of two in the market price of risk in the analogous endowment economy as s_t varies within 1.5 unconditional standard deviations of its median value. The probability that s exceeds 1 is 1.2×10^{-4} . Thus setting ϵ arbitrarily close to 1 allows us, to a good approximation, to ignore the fact that there is a reflective boundary near $s = 1$.

Only two parameters remain to be pinned down: b and β . The former can be used to calibrate to the observed maximal Sharpe ratio; for instance, given that at the median of s_t , the market price of risk in the analogous endowment economy is roughly $5b\sigma$, one can hope to generate a large Sharpe ratio by setting b between 3 and 6. The parameter β will then have to be adjusted so as to be consistent with the model requirements. After some experimentation, we settle on

¹⁸The unconditional probability density for s_t can be calculated from the Fokker-Planck counterpart to Eqn. (3) and is given by

$$\rho(s) \propto s^{-2\frac{\alpha+\mu}{\sigma^2}} e^{-2\frac{\lambda}{\sigma^2 s}}$$

$(b, \beta) = (3.3, 0.204)$. Figure 1 illustrates r^E along with s_-, s_+ and $s_{50\%}$ for this choice of parameters. The thick dashed line corresponds to the rate in Regime P (i.e., η). Note that β is chosen so that s_- is above the median s . This ensures that the economy is expected to be in Regime E strictly less than 50% of the time, since in addition to $s_t \in (s_-, s_+)$ (which happens less than 50% of the time) Regime E requires $Q_t = 0$. When in Regime E, short rates will be associated with a negative risk premium if $s_t \in (s_-, 0.868)$ and with a positive risk premium if $s_t \in (0.868, s_+)$. The unconditional likelihood of $s_t \in (0.868, s_+)$ is 9.05%, thus the likelihood of seeing shorter duration bonds exhibiting a negative premium is strictly less than this number. Finally, we note that $0.825 = s_- < \lambda/C = 0.842 < s_+ = 0.895$, so that the parameter choices are consistent with the model restrictions.

Before we proceed to compute key descriptive variables for the economy, consider the approximation error we make at the chosen parameter values. Figure 3 gives an indication of the error made in approximating the PDE for optimal investment. In the figure, we plot

$$\frac{\frac{\sigma^2 s^2}{2} g_{ss} + (b+1) \frac{\sigma^2 s^2}{2} \left(\frac{1}{1-s} - \frac{g_s^2}{g} \right)}{|Ag| + \frac{b+1}{2} \sigma^2 \frac{s^2}{1-s} + (|1-s-g| + |\alpha + \eta - \frac{\lambda}{s} |xs|) \left| \frac{g_x}{s} \right| + |(Cs - \lambda)g_s|}$$

which is the ratio of the terms ignored to the absolute value of terms depending on g in the approximate PDE. Over the relevant range of s_t and x_t the ignored terms constitute no more than a 10% correction to the PDE and the error is less than half of that for most of that range.

We next focus on the unconditional distribution of the relative capital stock, x_t , the investment policy (q_t/y_t), the term structure of interest rates, the behavior of the slope of the term structure, consumption volatility and expected growth rate, and the market price of risk (maximal Sharpe ratio). We begin by calculating the unconditional probability distribution of x in Regime P (see Figure 4). This is done by simulating 2500 paths for x_t , each 200 years in duration. The unconditional probability of being in Regime E is about 19.9%, consistent with the observations that the high volatility regime is less persistent (Dai et al., 2003). Conditional on being in Regime P, x resembles a Poisson distributed variable (the mean is close to the standard deviation and the median is approximately $\ln 2$ times the mean). Although the amount of invested capital is small relative to the habit level, we will also see shortly that the rates at which capital is depleted or replenished implies that it takes 2-3 years to completely deplete the median level of x_t if s_t is one standard deviation above its unconditional median. Intuitively, the ratcheting effect of habit

formation on relative risk aversion implies that even small changes in consumption growth have a significant utility impact.¹⁹ Figure 5 depicts the probability of finding the economy in Regime E given different values of s_t (i.e., $Pr(x = 0, q = 0|s)$). This also gives an indication as to how quickly capital can become depleted; specifically, $s = 0.9$ roughly corresponds to three *annual* standard deviations above the mean, reinforcing the intuition that capital can be depleted from its median level within 2-3 years given a series of negative shocks to the endowment (i.e., positive shocks to s_t). Figure 5 also gives an indication of the likelihood of negative risk-premium bonds: the probability of $s_t \in (0.868, s_+)$, i.e. 9.05%, times the likelihood that $Q_t = 0$ when $s_t \in (0.868, s_+)$, i.e., about 0.5. Thus one might expect, from this calibration, that once every 20 years, *real* bonds of duration 2-5 years may exhibit a negative risk premium. Figure 6 plots the distribution of x in Regime P conditional on s_t .

We calculate and plot the (relative) investment policy, $q_t/y_t = 1 - g_t - s_t$, in Figure 7. As established in Proposition 5, investment in the reversible technology decreases with the x_t . Moreover, investment decreases with the habit level. While intuitive, this is not generally true for $s > s_+$. However, this is not seen here since s_+ is above the range of s depicted.

By solving Eqn. (6) we calculate bond prices for zero coupon bonds with maturity varying from 1 to 20 years. Figure 9 depicts an estimate of the unconditional term structure, based on the simulated distribution of (x_t, s_t) , as well as the term structures conditional on each of the two regimes. The plots are within statistical tolerance of those estimated in Ang and Bekaert (2004). The unconditional term structure is nearly flat, while the Regime P unconditional term structure is slightly upward sloping and the Regime E curve steeper.

The next three figures plot the term structure of real interest rates for various values of x and for s roughly equal to its median, as well as one unconditional standard deviation above and below the median ($s = 0.78, 0.82, 0.86$). When the relative habit is at or below its unconditional mean, the term structure is sloping up. When capital invested in the reversible technology is high, the curve is flat in the short term, reflecting the low likelihood that regime change will take place in the short term. At the median values for s and x (Figure 11) the term structure is humped, and the

¹⁹Indeed, when b or σ is reduced, the investment policy tends to increase. For instance, in the case $(b, \beta) = (1, 0.01)$ both the median level of capital and the accompanying investment policy are roughly ten times higher.

hump grows more pronounced at lower levels of capital. When the relative habit is one standard deviation above its unconditional mean (Figure 12) the term structure is still sloping up for $Q_t > 0$, but is slightly *steeper*, reflecting the increased likelihood of Regime E. This trend continues as s increases. The Figure also illustrates a regime change: the slope of the term structure is negative and the short rate is high relative to its value in Regime P. Figure 13 gives an indication of how the risk premium affects the term structure by comparing bond yields when $Q_t = 0$ for $s = 0.85$ with $s = 0.88$. When $s = 0.85$ rates are on average expected to decrease faster than when $s = 0.88$ (since both are above the median value of s). Moreover, the instantaneous short rate at $s = 0.88$ is *higher* than when $s = 0.85$. Thus, if the risk premium was zero, the yield curve at $s = 0.85$ would lie below that of $s = 0.88$. The figure therefore indicates that the risk premium for bonds is sufficiently negative to shift the yield curve downward more than 50 basis points when $s = 0.88$ and $Q_t = 0$. Negative bond premia can thus significantly affect the term structure despite occurring only about 5% of the time.

We also calculate the slope of the term structure as the difference between the 20- and 1-year yields. This is plotted in Figure 8 for both cases. Most of the variation in the slope is at low x_t and high s_t . We will soon show that the consumption drift is monotonically increasing in s_t . Thus the slope of the term structure will forecast high growth in as much as it is negatively related to s_t . Since at low s_t the slope is uniformly high while this is only sometimes true at high s_t , there is a negative relationship between the two variables; this confirms two empirical observations summarized in Stock and Watson (2003): the slope of the term structure can forecast economic growth, but the forecasting power varies with time. A natural question is whether there is an alternative variable that might do a better job forecasting growth—at least in the context of our model. The most readily observable alternative, consistent with our model, is consumption volatility and we will shortly test its ability to forecast consumption growth.

Figure 14 plots the volatility of consumption growth relative to the endowment volatility. Consumption volatility decreases with x_t , increases with s_t , and is generally smoothed by the presence of the reversible technology. The magnitude of consumption volatility is entirely dependent on the magnitude of labor income. NIPA data indicates that consumption volatility is about 20% smaller than that of labor income or GNP. On the other hand, our calibration is based on a measure

of labor income constructed in Lettau and Ludvigson (2001) which has a much higher volatility than the data from NIPA. Consequently, our implied consumption growth volatility is also larger than post-war data on non-durables suggests. One can in principle, however, calibrate the model to NIPA data instead in an attempt to capture non-durable consumption growth dynamics better.²⁰ The expected growth rate of consumption is shown in Figure 15. It appears to be negatively related to consumption volatility and can be higher or lower than μ ; there is hardly any variation with respect to x_t save for the discontinuity during a regime shift, thus any variable that can proxy for the relative habit will be a good predictor of consumption growth. The range of variation in both consumption volatility and drift is about 20% and is restricted by the low levels of capital held for the purpose of modifying the endowment. In the case $(b, \beta) = (1, 0.01)$, where far more capital is typically held in the reversible technology, consumption volatility can be 40% that of the endowment and the expected growth rate of consumption can vary between 1.5% and 4.5%. Finally, the market price of risk, σ_M shown in Figure 16, resembles consumption volatility in shape. It is largest in Regime E and is firmly within the levels required to match the observed market Sharpe ratio.

In summary, the model appears to qualitatively and quantitatively capture the stylized facts listed in the Introduction: (i) regime changes where one regime is less persistent, and exhibits high mean and volatility (Regime E); (ii) the term structure is upward (downward) sloping in the smooth (volatile) regime. The unconditional term structure is relatively flat and seems to imply an unconditional negative risk premium for real interest rates (i.e., positive premium for bonds); and (iii) the risk premium for bonds is time varying and can be negative. In addition, the model appears capable of being in qualitative agreement with the observation that the slope of term structure has limited power to forecast consumption growth.

²⁰In addition, the autocorrelation of consumption growth, as calculated from a simulation of our model, is close to zero. While this is consistent with data from 1947-2004, it is not consistent with data from the 70's onward. Overall, ours does not seem to be a good model for capturing all of the peculiarities of non-durable consumption growth. On the other hand, there are many criticisms of using the non-durable consumption time-series to explain macro-asset pricing phenomena (see, for example, Vissing-Jørgensen (2002) and Aït-Sahalia, Parker and Yogo (2004) among other references).

4.1 Empirical Predictions

Figure 14 indicates that consumption volatility varies primarily with s_t . When in Regime E, consumption volatility is near its peak and this also coincides with a high level and volatility of real rates. The relationship is weak, however, since *within* each of the regimes consumption volatility and rates are uncorrelated. Moreover, since consumption growth rates also vary primarily with s_t , these should be well correlated with consumption volatility. Indeed, from Figures 14, 15 and 8 one might expect a proxy for s_t , such as consumption volatility, to be a better predictor of consumption growth than the slope of the term structure.

These observations are empirically testable and the purpose of this subsection is to examine the degree to which they are supported by post-war U.S. data. Table I lists each data series we use along with its source. The lower panel lists the variables constructed from the data series. Note that we calculate four different volatility measures for consumption growth. We do this because the volatility of this time-series appears non-stationary, decreasing by about a factor of two over the observation period. The *relative* volatility estimates, $\sigma_{5,ln,t}^C$ and $\sigma_{5,cn,t}^C$, are meant to somewhat account for this non-stationarity and provide a measure of the size of the standard deviation around the observation date relative to the standard deviation over the previous 5 years. While our model does not explicitly account for the possibility of secular trends in volatility, it is possible to do so: the qualitative results here do not change substantively if the endowment process itself exhibits a trend in its volatility. In particular, consumption volatility *relative* to that of the endowment process (Figure 14) will look qualitatively the same. Thus the correct volatility measure in such an economy would be the deviation from the trend, as is calculated in $\sigma_{5,ln,t}^C$ and $\sigma_{5,cn,t}^C$.

Table II.A presents the contemporaneous correlation between consumption volatility, real rate levels, real rate volatility, and the probability that the smooth rate regime prevailed as estimated by Ang and Bekaert (2004).²¹ Notice first that the regime probability variable has the correct (significant) relationship with both the rate level and volatility. Moreover, the data appears to indicate a significant positive relationship between the consumption volatility measures that are not trend-corrected and real rate volatility. A weaker positive relationship remains after trend-

²¹We thank Andrew Ang and Geert Bekaert for making their filtered time series of smooth regime probability available to us.

correction. The table also indicates that while a significant negative relationship exists between consumption volatility and the smooth regime likelihood (as predicted), this disappears with the trend correction. Finally, consumption volatility does not appear to have a significant relationship with the level of interest rates, despite the fact that the predicted relationship exists with respect to rate volatility. Overall, the data provides support for the hypothesis that consumption volatility exhibits a positive (though weak) correlation with the real rate level and volatility.

Table II.B presents contemporaneous correlations between the various measures in Table II.A and the constructed habit proxy. Note that the relationship between s_t and interest rate series is stronger than for the consumption volatility measures and consistent with the model predictions. The lack of correlation with the consumption volatility measures, on the other hand, suggests that the latter are not good proxies for s_t .

Table III presents correlations between lagged predictors and realized consumption growth. We include the nominal term spread, Lettau and Ludvigson's (2001) estimate of the consumption-wealth ratio (CAY), and the probability that the economy was in a smooth regime as estimated by Ang and Bekaert (2004). The latter is related to the real term spread and thus according to Harvey (1988) has predictive power for consumption growth. We use $\sigma_{5,ln,t}^C$ for consumption volatility to make the consumption volatility proxy stationary and also include the direct proxy for s_t . Here too the results are supportive of the model in the sense that the trend corrected volatility measure and the habit proxy appear to have forecasting power for consumption growth. The lower portion of Table III reports the results of regressions that further lend support to the proposition that consumption volatility and/or s_t have forecasting power for growth beyond that present in the slope of the term structure. This finding may be particularly useful since, according to Stock and Watson (2003), there are very few good predictors of growth.

5 Concluding Remarks

The regimes discussed here and in the literature have roughly the same duration as business cycles. As briefly noted in the empirical section above, the data does appear to indicate secular changes to the level of real rates and consumption volatility that are not entirely consistent with

a steady state stationary equilibrium. Our model is too simple to handle such complications, and it seems worthwhile to study a similar problem involving multiple technological sectors so as to introduce more than a single time scale. Related to that, a more realistic approach than provided here might explicitly model the irreversibility of capital investment along with time variation in production efficiency. This could be modeled, for instance, by allowing for the possibility that existing industries receive ‘shocks’ rendering them obsolete or less efficient as other technologies enter. Non-trivial adjustment costs and the irreversibility of investment might lead to regimes in economy-wide investment and, possibly, tighter empirical predictions.

Appendix A Proofs

Proof of Theorem 1:

Let $\omega \in \Omega$ be a generic element in Ω describing the state of the world. Define

$$H(y, Q, z) = \{q \in \mathbb{R} : -Q \leq q \leq y - z\} \quad (7)$$

$$\mathcal{H}(\omega) = \{(q_0, q_1, \dots) : q_n \in H(y_n, Q_n, z_n)(\omega)\}. \quad (8)$$

We can now rephrase the problem as $\sup_{q \in \mathcal{H}} U(q)$. In particular, we are interested in $\arg \max_{q \in \mathcal{H}} U(q)$. Clearly H is compact. Tychonoff's Theorem implies that \mathcal{H} is compact in the product topology. Note moreover that \mathcal{H} is non-empty: consider for instance the policy $\bar{q} \equiv (1 - \epsilon)y - z$. Since $u'(x) = x^{-b}$ is continuous, one can always find a topology-preserving metric that can be used to establish that U is upper semi-continuous on \mathcal{H} . By the Weierstraß maximum theorem U attains its maximum on \mathcal{H} (see for instance Luenberger, 1969). \square

Proof of Proposition 2:

Note that due to the Inada condition satisfied by $u(x)$ we can ignore the possibility that $q_t \Delta t = -Q_t$. For any finite horizon, the result is easy to derive. Note that concavity of the value function and convexity of the control space guarantees uniqueness of the solution for both finite and infinite horizon problems. Since $U(q)$ is bounded and \mathcal{H} is compact, the optimal finite horizon solution converges to the infinite horizon solution (by the Dominated Convergence Theorem). This is therefore also true for the sequence of Euler equations generated by taking the limit of the finite horizon problem to an infinite horizon. \square

Proof of Proposition 3:

This follows from setting $q_t = 0$ and substituting $\eta = r_t^E$ in the Euler Equation (Proposition 2), using Itô's Lemma and comparing terms of order dt . \square

Proof of Proposition 4:

Homogeneity of q follows from the Euler equation in Proposition 2 (Alternatively, see Duffie et al., 1997). The PDE can likewise be derived by using Itô's Lemma and comparing terms of order dt . \square

Proof of Theorem 2:

The function $V_1(s)$ and $V_2(g, s)$ in Eqn. (5) are given by,

$$V_1(s) = \frac{(b+1)\sigma^2(M-C-\lambda) - (A-M)(C-M-\lambda)}{M(A-M)(C-M)} - \frac{s}{C-M} \left(1 - \frac{(b+1)\sigma^2}{A-M}\right) - \frac{(b+1)\sigma^2}{(A-M)(C-\lambda)} \frac{\Phi_L\left(\frac{Cs-\lambda}{C-\lambda}, -\frac{M}{C}\right)}{C-\lambda}$$

$$V_2(g, s) = |Cs - \lambda|^{\frac{M}{C}} f \left(|Cs - \lambda|^{-\frac{A}{C}} \left\{ g + (b+1)\sigma^2 \left(\frac{\Phi_L\left(\frac{Cs-\lambda}{C-\lambda}, -\frac{A}{C}\right)}{C-\lambda} + \frac{\lambda + C - A(s+1)}{A(C-A)} \right) \right\} \right)$$

where

$$\Phi_L(z, \nu) = \sum_{n=0}^{\infty} \frac{z^n}{n + \nu}$$

and

$$f(y) = |Cs(y) - \lambda|^{-\frac{M}{C}} \left(\frac{1-s(y)}{A-M} + \frac{(b+1)\sigma^2}{(A-M)(C-\lambda)} \frac{\Phi_L\left(\frac{Cs(y)-\lambda}{C-\lambda}, -\frac{M}{C}\right)}{C-\lambda} + \frac{s(y)}{C-M} \left(1 - \frac{(b+1)\sigma^2}{A-M}\right) - \frac{(b+1)\sigma^2(M-C-\lambda) - (A-M)(C-M-\lambda)}{M(A-M)(C-M)} \right) \quad (9)$$

such that $s(y)$ solves

$$y = |Cs(y) - \lambda|^{-\frac{A}{C}} \left\{ 1 - s(y) + (b+1)\sigma^2 \left(\frac{\Phi_L\left(\frac{Cs(y)-\lambda}{C-\lambda}, -\frac{A}{C}\right)}{C-\lambda} + \frac{\lambda + C - A(s(y)+1)}{A(C-A)} \right) \right\}$$

with

$$\beta + b(\mu - \lambda) + b(\alpha + \mu - \lambda) \frac{s(y)}{1-s(y)} - \frac{b(b+1)\sigma^2}{2(1-s(y))^2} \geq \eta \quad (10)$$

$$(Cs - \lambda)(Cs(y) - \lambda) > 0 \quad (11)$$

To prove this, note first that Eqn. (4) is a quasi-linear PDE and, using the method of characteristics, has the general solution:

$$0 = F(g, s) + xs$$

where $F(g, s)$ solves the following linear PDE:

$$1 - s - g = \left(Ag - \frac{b+1}{2} \sigma^2 \frac{s^2}{1-s} \right) F_g + (Cs - \lambda) F_s - MF$$

where $M = \mu - \eta - \sigma^2(b+1)$. The general solution to the latter is given by

$$F(g, s) = -\frac{g}{A-M} + |Cs - \lambda|^{\frac{M}{C}} \int^s |Cw - \lambda|^{-\frac{M}{C}} \left((1-w) - \frac{(b+1)\sigma^2}{2(A-M)} \frac{w^2}{1-w} \right) (Cw - \lambda)^{-1} dw + |Cs - \lambda|^{\frac{M}{C}} f \left(g |Cs - \lambda|^{-\frac{A}{C}} + \frac{b+1}{2} \sigma^2 \int^s |Cw - \lambda|^{-\frac{A}{C}} \frac{w^2}{1-w} (Cw - \lambda)^{-1} dw \right)$$

where $f(\cdot)$ is an arbitrary differentiable function. Note that in the region $s \in (2\frac{\lambda}{C} - 1, 1)$ the Lerch Phi function, $\Phi_L(\frac{Cs-\lambda}{C-\lambda}, -\nu)$, is defined and that

$$\frac{d}{ds} \left\{ |Cs - \lambda|^{-\nu} \left[\frac{\Phi_L(\frac{Cs-\lambda}{C-\lambda}, -\nu)}{C-\lambda} + \frac{\lambda + C(1-\nu(s+1))}{C^2\nu(1-\nu)} \right] \right\} = |Cs - \lambda|^{-\nu} \frac{s^2}{1-s} (Cs - \lambda)^{-1}$$

This can be easily used to demonstrate that Eqn. (5) is the general solution to (4) in the region $s \in (2\frac{\lambda}{C} - 1, 1)$. It remains to establish the boundary conditions.

To see that the choice of f in Eqn. (9) satisfies the boundary conditions, note that $s(y) = s$ when $g = 1 - s$, and thus $F(1 - s, s) = 0$; thus as long as s satisfies $r^E(s) \geq \eta$, setting $x = 0$ gives $g = 1 - s$, as required by the boundary conditions. One now has to establish that (i) for any $g > 0$ there is always a unique solution for $s(y)$ when

$$y = |Cs - \lambda|^{-\frac{A}{C}} \left\{ g + \frac{b+1}{2} \sigma^2 \left(\frac{\Phi_L(\frac{Cs-\lambda}{C-\lambda}, -\frac{A}{C})}{C-\lambda} + \frac{\lambda + C - A(s+1)}{A(C-A)} \right) \right\}$$

and (iii) the boundary condition, $g > 0$ when $x = 0$ and $r^E(s) < \eta$ is satisfied, and (iii) that the solution to the PDE is unique.

Let s_- and s_+ be the roots of the equation

$$\beta + b(\mu - \lambda) + b(\alpha + \mu - \lambda) \frac{s(y)}{1-s(y)} - \frac{b(b+1)\sigma^2}{2(1-s(y))^2} = \eta$$

A necessary requirements for (i) and (ii) above is that s_- and s_+ are real and not equal, in which case one can assume that $s_- < s_+$, and $r^E(s) > \eta$ in the interval (s_-, s_+) . Another necessary condition is that the range of

$$y(s) = |Cs - \lambda|^{-\frac{A}{C}} \left\{ 1 - s + \frac{b+1}{2} \sigma^2 \left(\frac{\Phi_L(\frac{Cs-\lambda}{C-\lambda}, -\frac{A}{C})}{C-\lambda} + \frac{\lambda + C - A(s+1)}{A(C-A)} \right) \right\}$$

for $s \in [s_-, s_+]$ is *unbounded* above so that $s(y)$ has a solution for any $g > 0$ in the region $[s_-, s_+]$. The existence of real roots in the interior of $(0, 1)$ is guaranteed by the Theorem's hypothesis. Note that for $y(s)$ to be unbounded in $[s_-, s_+]$, it is necessary that $s_- < \frac{\lambda}{C} < s_+$ otherwise $y(s)$ is finite in $[s_-, s_+]$. Given this, note further that

$$\begin{aligned} \frac{d}{ds} y(s) &= |Cs - \lambda|^{-\frac{A}{C}} (Cs - \lambda)^{-1} \left(-A(1-s) - (Cs - \lambda) + \frac{\sigma^2 s^2}{1-s} \right) \\ &= -|Cs - \lambda|^{-\frac{A}{C}} (Cs - \lambda)^{-1} (1-s) (r^E(s) - \eta) \end{aligned}$$

So that if $s_- < \frac{\lambda}{C} < s_+$, $y(s)$ diverges to $+\infty$ as $s \rightarrow \frac{\lambda}{C}$, and a solution to $s(y)$ exists for any $g \geq 0$. To prove uniqueness, note that $\frac{d}{ds} y(s)$ is monotonic and thus invertible in each of the regions $(s_-, \frac{\lambda}{C})$ and $(\frac{\lambda}{C}, s_+)$. Thus the solution for $s(y)$ satisfying the constraints in (10) and (11) is unique.

To show that the solution satisfies the boundary conditions at $x \rightarrow \infty$, first note that g can only satisfy Eqn. (5) for arbitrary s as $x \rightarrow \infty$ if g also approaches infinity. In particular, this requires that $s(y) \xrightarrow{g \rightarrow \infty} \frac{\lambda}{C}$.

We calculate that

$$g_x = \frac{A - M}{1 - \left| \frac{Cs(y) - \lambda}{Cs - \lambda} \right|^{(A-M)/C}} \quad (12)$$

thus as $x \rightarrow \infty$, $g_x \rightarrow A - M$, which is our boundary condition.

Finally, we need to show that our solution is unique. To do this, it is sufficient to demonstrate that the solution satisfies a pair of boundary condition in a closed and connected region. In particular, we must extend the current boundary condition at $x = 0$ and $s \in [s_-, s_+]$ so that it bounds some closed and connected region. To establish this note that $y(s)$ is decreasing in $[0, s_-)$ and increasing in $[s_+, 1)$, thus in these regions $g(0, s)$ exists and is greater than $1 - s$; moreover, for any s one can find a unique and finite $x^*(s) \geq 0$ such that $g = 1 - s$ satisfies Eqn. (5) ($x^*(s) = 0$ by construction when $s \in [s_-, s_+]$). Thus the solution specified in Eqn. (5) and its extension to all of $s \in [0, 1 - \epsilon]$ satisfies the boundary condition $g = 1 - s$ at $x^*(s)$ and $g = (A - M)x$ at $x = \infty$, and is therefore unique. \square

Proof of Proposition 5:

Given Eqn. (12) in the last proof, it is sufficient to prove that $s(y) > s$ for $Cs - \lambda < 0$ and $s(y) < s$ for $Cs - \lambda > 0$. Consider $Cs - \lambda < 0$; since $s(y) \in [r_-, \lambda/C]$, if $s < r_-$ then we are done. If $s \in [r_-, \lambda/C]$, then note that $x = 0$, the arbitrage conditions require $g_x > 0$; thus $s(y) > s$ for x close to 0. However, the form of Eqn. (12) makes a sign change in g_x when s is in $(r_-, \lambda/C)$. Thus $g_x > 0$ for s below λ/C . A similar argument can be applied to the complementary region. \square

Appendix B

In this Appendix we argue that Models 1-3 in Section 2 are not compatible with the stylized facts outlined in the Introduction.

In Model 1, η_t is time varying while r^E is constant, thus the volatile rate regime corresponds to the case $Q_t > 0$ or $\eta_t > r^E$. Let $\bar{\eta}$ be the mean of η_t . If $r^E \leq \bar{\eta}$ then the volatile rate regime has a higher mean than r^E , consistent with the stylized facts. However, since η_t is an AR(1) process, the unconditional likelihood that $\eta_t > r^E$ is greater or equal to 50% in this case; and thus the volatile regime is at least as frequent as the smooth regime, in contradiction with the stylized facts. If, on the other hand, $r^E > \bar{\eta}$, then the volatile regime has a lower mean than r^E (the smooth regime mean), also contradicting the stylized facts.

In Model 2, the endowment process can be written in continuous time as $\frac{dy_t}{y_t} = \mu_t dt + \sigma_y dW_t^y$ with

$d\mu_t = \kappa(\bar{\mu} - \mu_t)dt + \sigma_\mu dW_t^\mu$. Setting $z_t = 0$ one can derive r_t^E as is done in Proposition 3 to yield

$$r_t^E = \beta + b\mu_t - \frac{b(b+1)}{2}\sigma_y^2$$

If $Q_t = 0$, the economy is in the smooth regime whenever $\eta > r_t^E$, or alternatively,

$$\frac{\beta - \eta}{b} + \bar{\mu} - \frac{b+1}{2}\sigma_y^2 < \bar{\mu} - \mu_t \quad (13)$$

The right hand side of the last inequality is a mean-zero Normal variate. Thus the smaller the left hand side of the inequality, the larger the likelihood of a smooth regime. One can make the left hand side small by choosing β and b appropriately. This, however, is constrained by a growth condition similar to the one following Lemma 1, indicating that Q_t/y_t mean reverts only if $b\bar{\mu} + \beta - \eta > \frac{b^2}{2}\sigma_y^2$. Moreover, if $b < 1$ then for utility to be finite one must have $\beta > (1-b)\bar{\mu}$.²² Assuming $\bar{\mu} \approx 0.025$, $\eta \approx 0.015$ and $\sigma_y < 0.05$ implies that either b must be unreasonably large — thereby implying unrealistic real rate volatility or a ‘knife edge’ model with $\sigma_\mu \ll \sigma_y$ — or the smooth regime at $Q_t = 0$ is characterized by $\bar{\mu} - \mu_t > a$ where a is no smaller than 0.01. For reasonable values of σ_μ this yields a frequency well under 50% when $Q_t = 0$. Note that even when $Q_t > 0$, the agent will deplete invested capital whenever $0.01 > \bar{\mu} - \mu_t$ (which is, again, most of the time). This precludes the possibility that the smooth regime can be credibly calibrated to dominate roughly 80% of the time.²³

Finally, Models 1-3 exhibit time varying real rates that are monotonic in their respective state variables.²⁴ Thus bond yields will also vary monotonically with these state variables. In the case of Model 3, this rules out a change of sign in the risk premium associated with the real rate. The same is generally true in the case of Models 1 and 2, although one may be able to find parameters under which it is possible to engineer the correlation between dW_t^y and dW_t^μ or dW_t^η so as to induce a sign change in the risk premium of bond yields.

References

Abel, Andrew B., “Asset Prices under Habit Formation and Catching Up with the Joneses,” *AEA Papers and Proceedings*, May 1990, 80, 38–42.

Aït-Sahalia, Yacine, Jonathan A. Parker and Motohiro Yogo, Ravi and Hao Zhou, “Luxury Goods and the Equity Premium,” *Journal of Finance*, 2004, 59, 2959–3004.

²²Theorem 2 assumes $b > 1$. This can be extended to $b > 0$ as long as appropriate parameter constraints are placed.

²³We’ve verified this qualitative argument numerically.

²⁴While it is slightly tedious to establish this, the proof is straight forward and we leave it to the reader. Note that in Model 3, utility is defined over y_t/z_t , where z_t is defined as in Model 4.

- Ang, Andrew and Geert Bekaert**, “The Term Structure of Real Rates and Expected Inflation,” April 2004. Manuscript, Columbia University.
- Bansal, Ravi and Hao Zhou**, “Term Structure of Interest Rates with Regime Shifts,” *Journal of Finance*, 2002, 57 (5), 1997–2044.
- Breedon, Douglas T.**, “An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities,” *Journal of Financial Economics*, 1979, 7, 265–296.
- Brock, William A.**, “Asset Prices in a Production Economy,” in “The Economics of Information and Uncertainty,” Chicago: University of Chicago Press, 1982, pp. 1–46.
- Campbell, John Y. and John H. Cochrane**, “By Force of Habit: A Consumption-based Explanation of Aggregate Stock Market Behavior,” *Journal of Political Economy*, 1999, 107 (2), 205–251.
- and **Robert J. Shiller**, “Yield Spreads and Interest Rate Movements: A Birds Eye View,” *Review of Economic Studies*, 1991, 58, 495–514.
- Cassasus, Jaime, Pierre Collin-Dufresne, and Bryan Routledge**, “Equilibrium Commodity Prices with Irreversible Investment and Non-linear Technologies,” April 2004.
- Cecchetti, S., P. Lam, and N. Mark**, “The Equity Premium and the Risk-free Rate: Matching the Moments,” *Journal of Monetary Economics*, 1993, 31, 21–45.
- Chapman, David A.**, “Habit Formation and Aggregate Consumption,” *Econometrica*, 1998, 66, 1223–1230.
- Constantinides, George M.**, “Habit Formation: A Resolution of the Equity Premium Puzzle,” *Journal of Political Economy*, 1990, 98 (3), 519–543.
- Cox, John C., Jonathan E. Ingersoll, Jr., and Stephen Ross**, “An Intertemporal General Equilibrium Model of Asset Prices,” *Econometrica*, 1985, 53, 363–384.
- , —, and —, “A Theory of the Term Structure of Interest Rates,” *Econometrica*, 1985, 53, 385–408.
- Dai, Qiang and Kenneth J. Singleton**, “Expectations Puzzles, Time-varying Risk Premia, and Affine Models of the Term Structure,” *Journal of Financial Economics*, 2002, 63, 415–441.
- , —, and **Wei Yang**, “Regime Shifts in a Dynamic Term Structure Model of U.S. Treasury Bond Yields,” October 2003. Manuscript, New York University and Stanford University.
- Deaton, Angus**, “Saving and Liquidity Constraints,” *Econometrica*, September 1991, 59 (5), 1221–1248.

- Detemple, Jerome B. and Fernando Zapatero**, “Asset Prices in an Exchange Economy with Habit Formation,” *Econometrica*, November 1991, *59* (6), 1633–1657.
- Duesenberry, James**, *Income, Saving, and the Theory of Consumer Behavior*, Harvard University Press, 1952.
- Duffie, Darrell, Wendell Fleming, H. Mete Soner, and Thaleia Zariphopoulou**, “Hedging in Incomplete Markets with HARA Utility,” *Journal of Economic Dynamics and Control*, 1997, *21*, 753–782.
- Dunn, Kenneth B. and Kenneth J. Singleton**, “Modeling the Term Structure of Interest Rates under Non-separable Utility and Durability of Goods,” *Journal of Financial Economics*, 1986, *17*, 27–55.
- Epstein, Larry G. and Stanley E. Zin**, “Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns I: A Theoretical Framework,” *Econometrica*, July 1989, *57* (4), 937–969.
- Evans, Martin D. D.**, “Real Risk, Inflation Risk, and the Term Structure,” *The Economic Journal*, 2003, *113*, 345–389.
- Garcia, R. and P. Perron**, “An Analysis of the Real Interest Rate Under Regime Shifts,” *Review of Economics and Statistics*, 1996, *78* (1), 111–125.
- Gray, Stephen F.**, “Modeling the Conditional Distribution of Interest Rates as a Regime Switching Process,” *Journal of Financial Economics*, September 1996, *42* (1), 27–62.
- Hamilton, James D.**, “Rational-Expectations Econometric Analysis of Changes in Regime: An Investigation of the Term Structure of Interest Rates,” *Journal of Economic Dynamics and Control*, 1988, *12*, 385–423.
- Harvey, Campbell**, “The Real Term Structure and Consumption Growth,” *Journal of Financial Economics*, December 1988, *22* (2), 305–333.
- Kogan, Leonid, Dmitry Livdan, and Amir Yaron**, “Futures Prices in a Production Economy with Investment Constraints,” November 2003.
- Landen, C.**, “Bond Pricing in a Hidden Markov Model of the Short Rate,” *Finance and Stochastics*, 2000, *4*, 371–389.
- Lettau, M. and S. Ludvigson**, “Consumption, Aggregate Wealth and Expected Stock Returns,” *Journal of Finance*, 2001, *56*, 815–849.
- Lucas, Robert E. Jr.**, “Asset Prices in an Exchange Economy,” *Econometrica*, November 1978, *46* (6), 1429–1445.

- Luenberger, David G.**, *Optimization by Vector Space Methods*, New York, New York: John Wiley & Sons, 1969.
- Naik, V. and M. H. Lee**, “Yield Curve Dynamics with Discrete Shifts in Economic Regimes: Theory and Estimation,” 1997. Working paper, University of British Columbia.
- Pollak, Robert A.**, “Habit Formation and Dynamic Demand Functions,” *Journal of Political Economy*, 1970, 78 (4), 745–763.
- Routledge, Bryan R., Duane J. Seppi, and Chester S. Spatt**, “Equilibrium Forward Curves for Commodities,” *Journal of Finance*, June 2000, 55 (3), 1297–1339.
- Rubinstein, Mark**, “The Valuation of Uncertain Income Streams and the Pricing of Options,” *Bell Journal of Economics*, Autumn 1976, 7 (2), 407–425.
- Ryder, Harl and Geoffrey Heal**, “Optimal Growth with Intertemporally Dependent Preferences,” *Review of Economic Studies*, January 1973, 40 (1), 1–33.
- Stock, James H. and Mark W. Watson**, “Forecasting Output and Inflation: The Role of Asset Prices,” *Journal of Economic Literature*, September 2003, 61, 788–829.
- Sundaresan, Suresh M.**, “Intertemporally Dependent Preferences and the Volatility of Consumption and Wealth,” *Review of Financial Studies*, 1989, 2, 73–89.
- Vissing-Jørgensen, Annette**, “Limited Asset Market Participation and the Elasticity of Inter-temporal Substitution,” *Journal of Political Economy*, 2002, 110, 825–853.
- Wachter, Jessica A.**, “A Consumption-based Model of the Term Structure of Interest Rates,” July 2004. Manuscript, University of Pennsylvania.

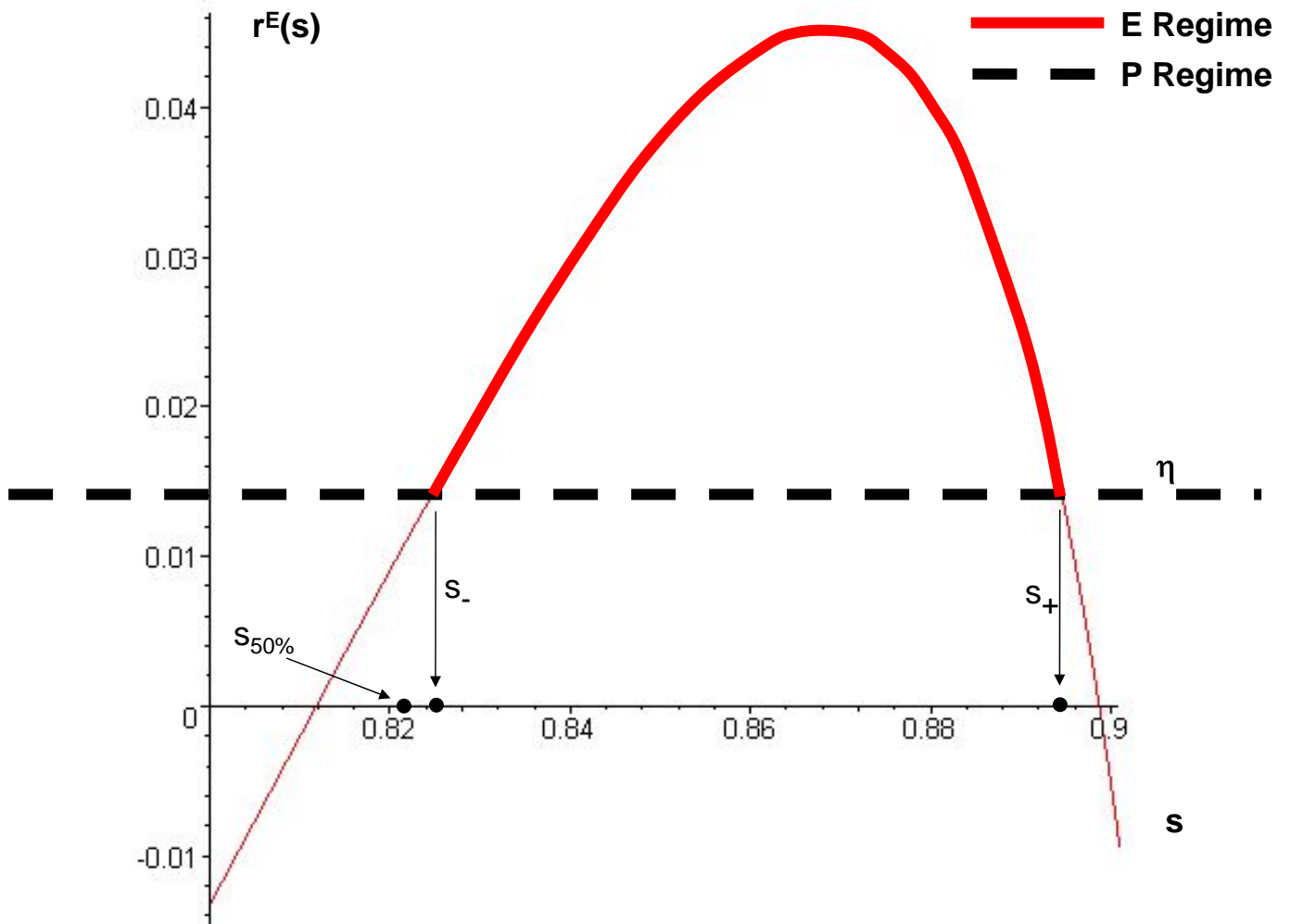


Figure 1: ‘Regime P’ and ‘Regime E’ rates. The dashed line corresponds to the yield on the reversible technology. The thick and thin solid curves correspond to the interest rate for the endowment economy, r_t^E , with linear habit formation. The true interest rate equals r_t^E only when both $Q_t = 0$ and $r_t^E \geq \eta$ (thick solid curve).

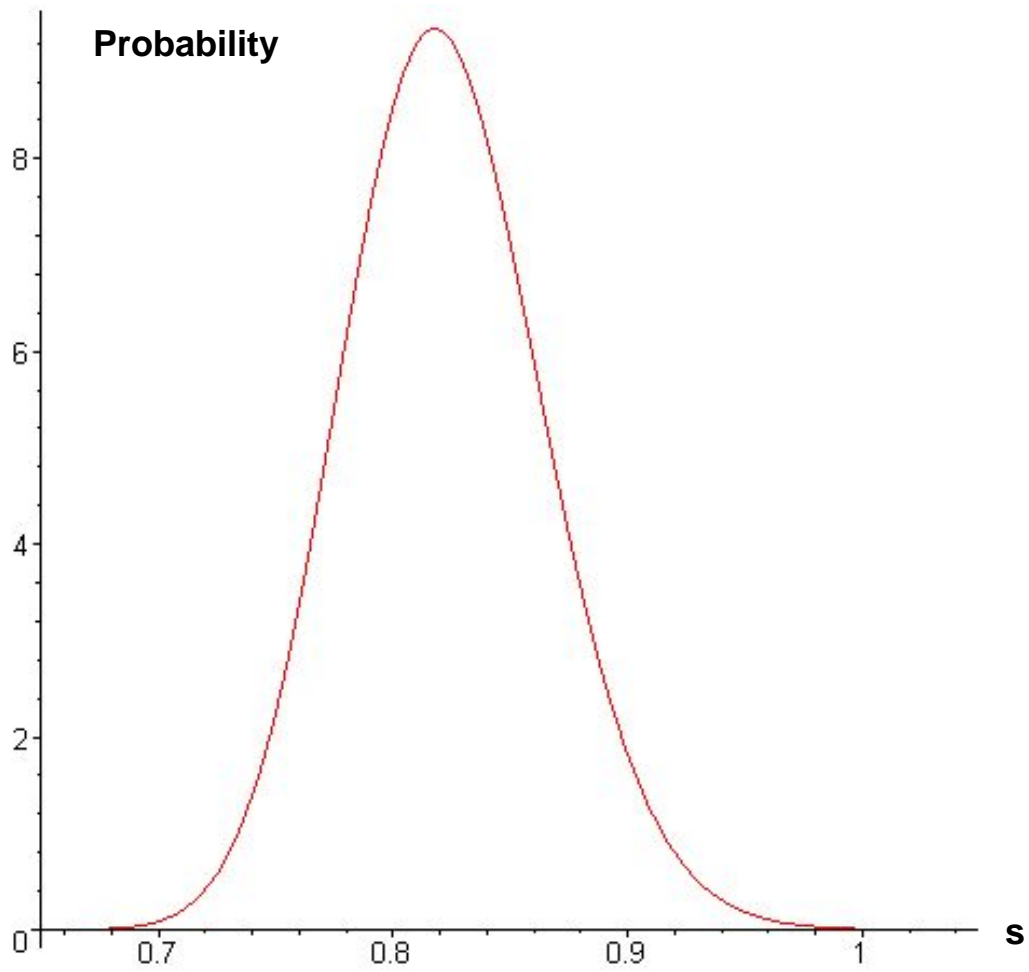


Figure 2: Distribution Function for s_t .

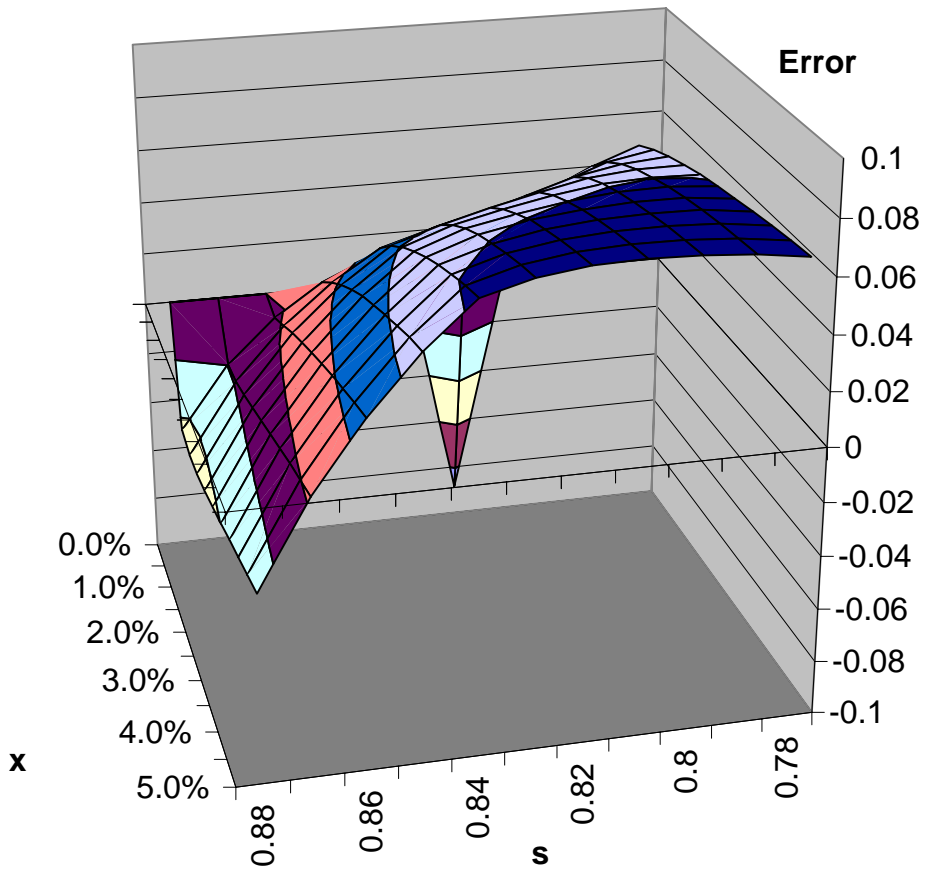


Figure 3: The relative magnitude of terms ignored in approximating the PDE for g_t .

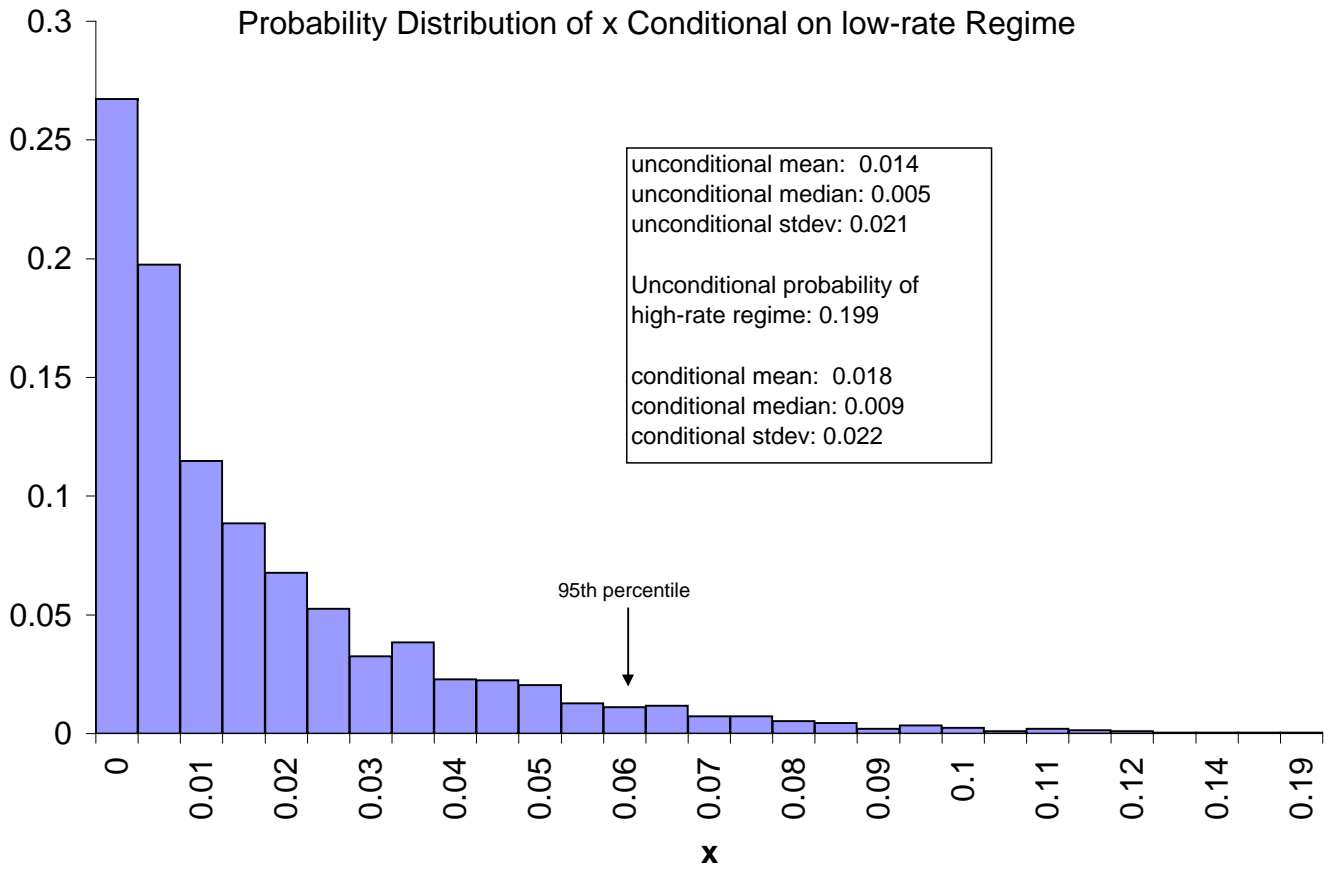


Figure 4: Unconditional distribution of x in Regime P

Probability of High-rate Regime

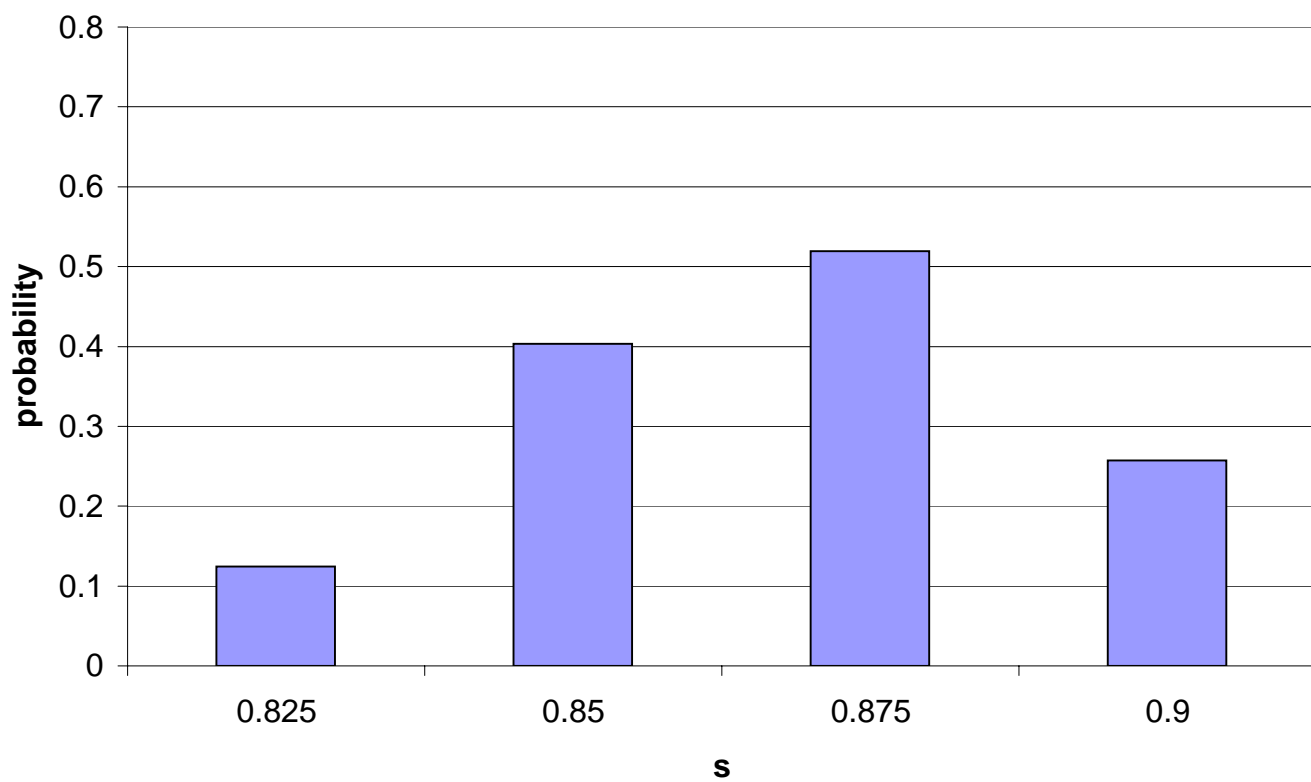


Figure 5: Conditional probability, $Pr(x = 0, q = 0|s)$, of finding the economy in Regime E.

Conditional Distribution of x (low-rate regime)

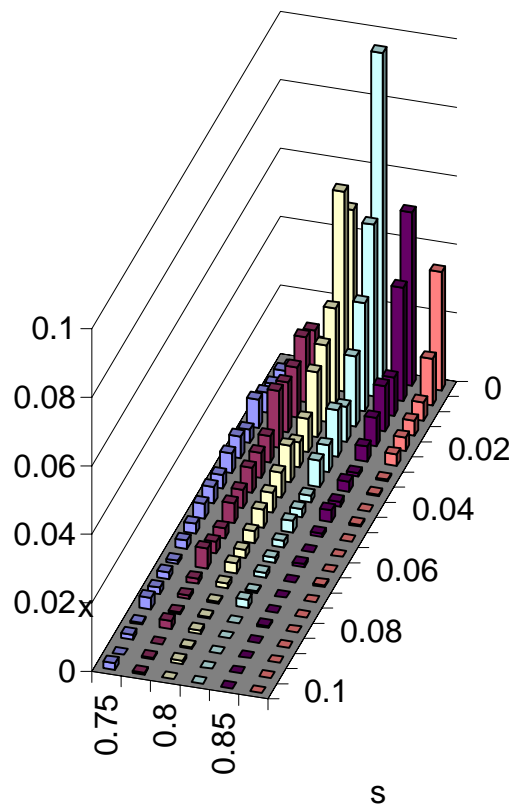


Figure 6: Joint distribution function in Regime P.

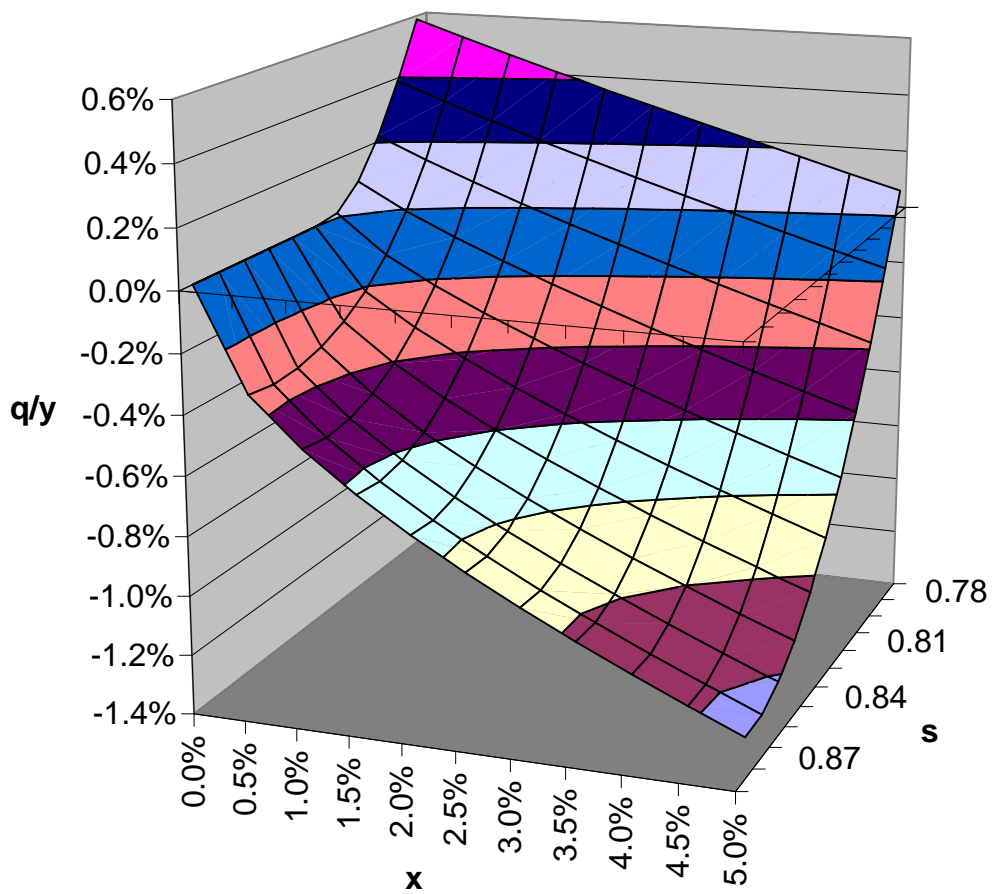


Figure 7: Investment Policy.

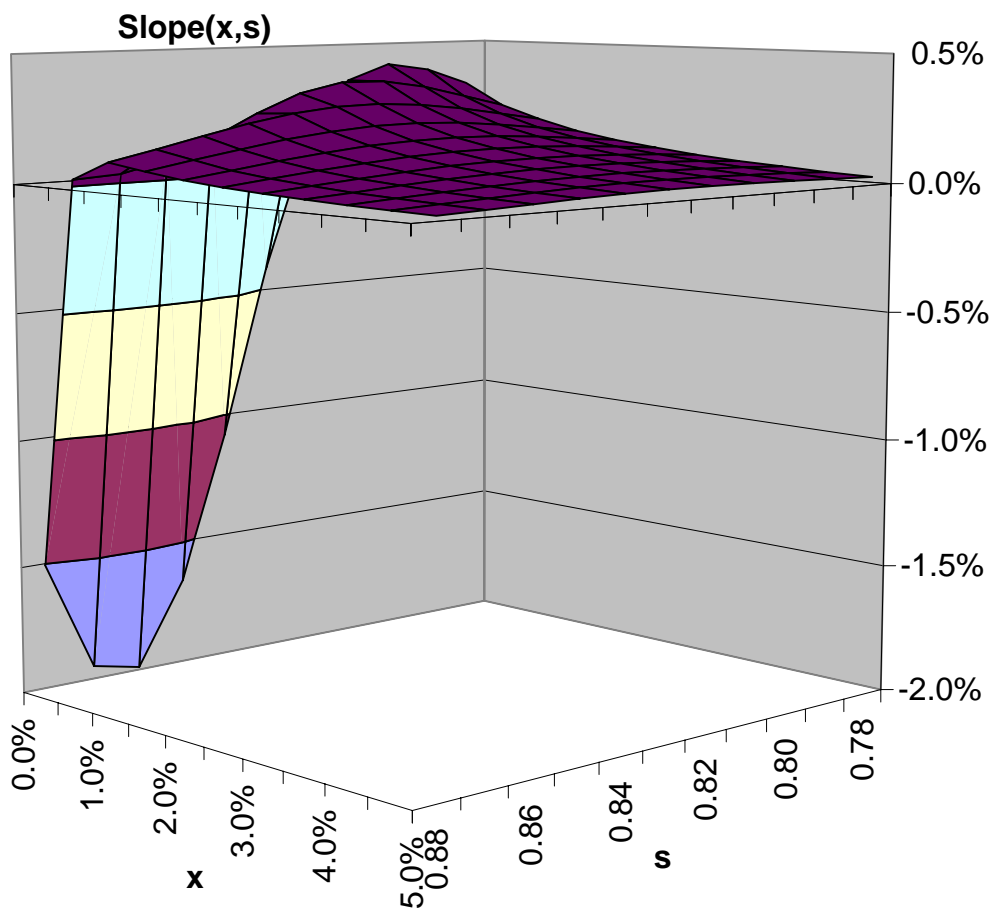


Figure 8: The slope of the term structure (difference between the yields on a 20 and one year zero coupon bonds).

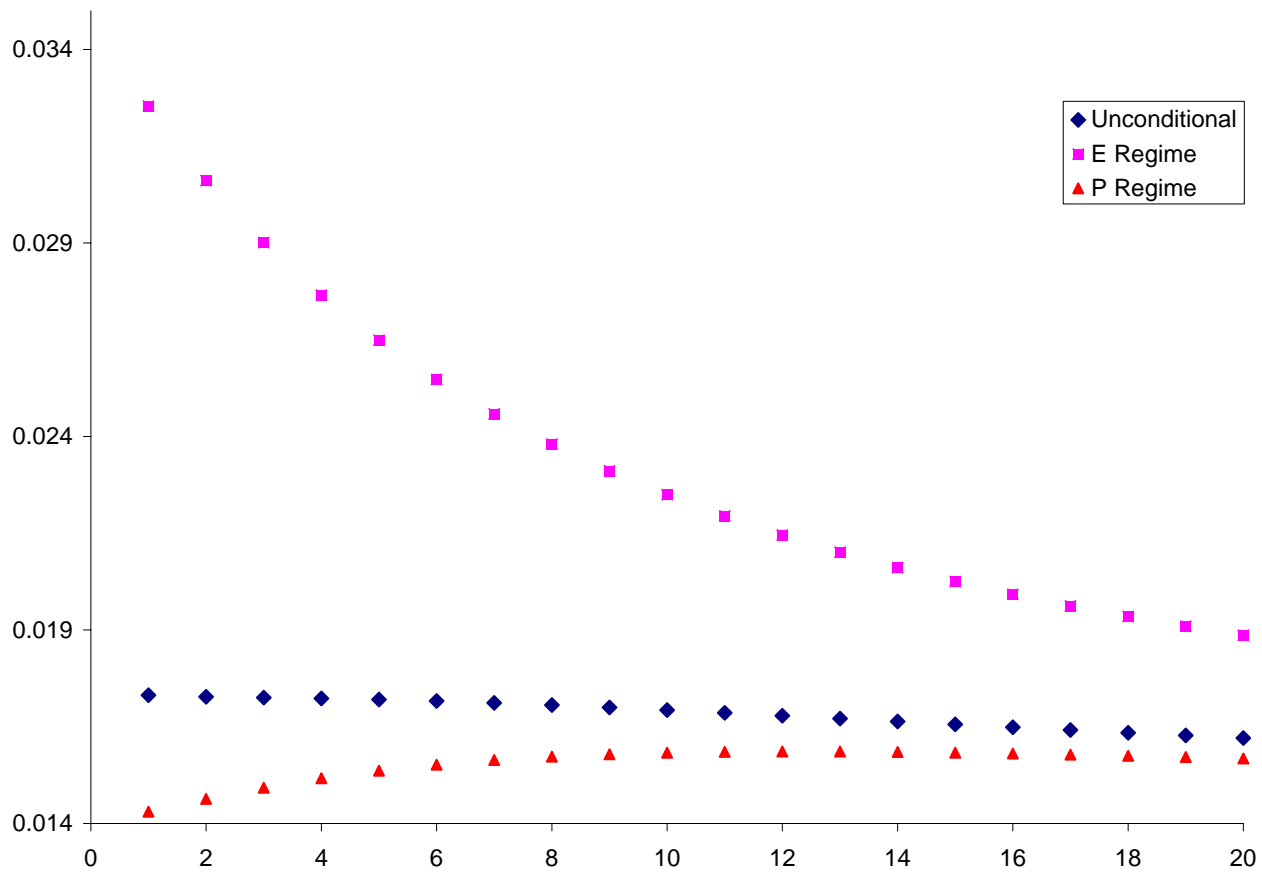


Figure 9: The average term structure of real rates.

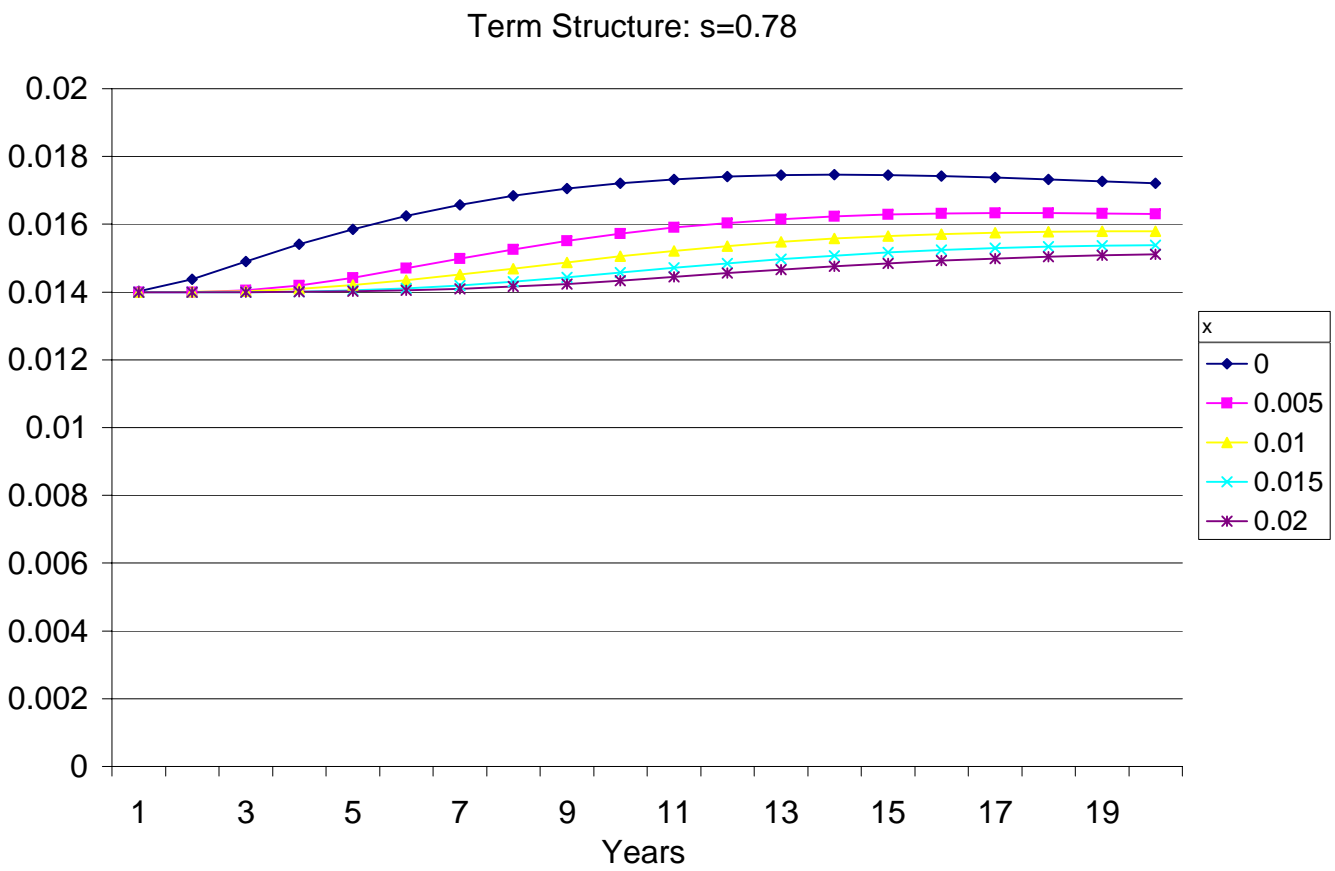


Figure 10: The term structure of real rates when $s = 0.78$ (one unconditional standard deviation below the median value of s).

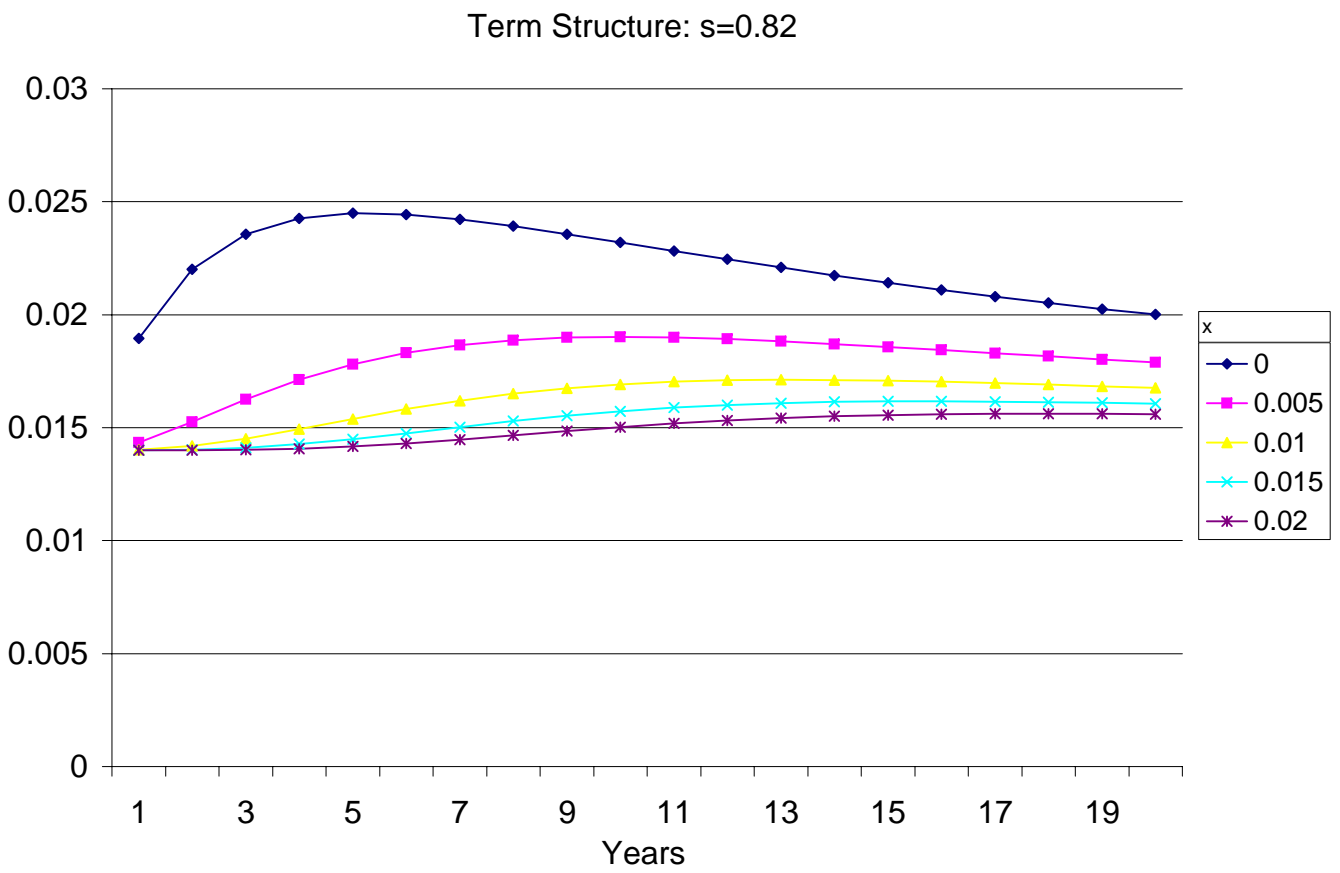


Figure 11: The term structure of real rates when $s = 0.82$ (at the median value of s).

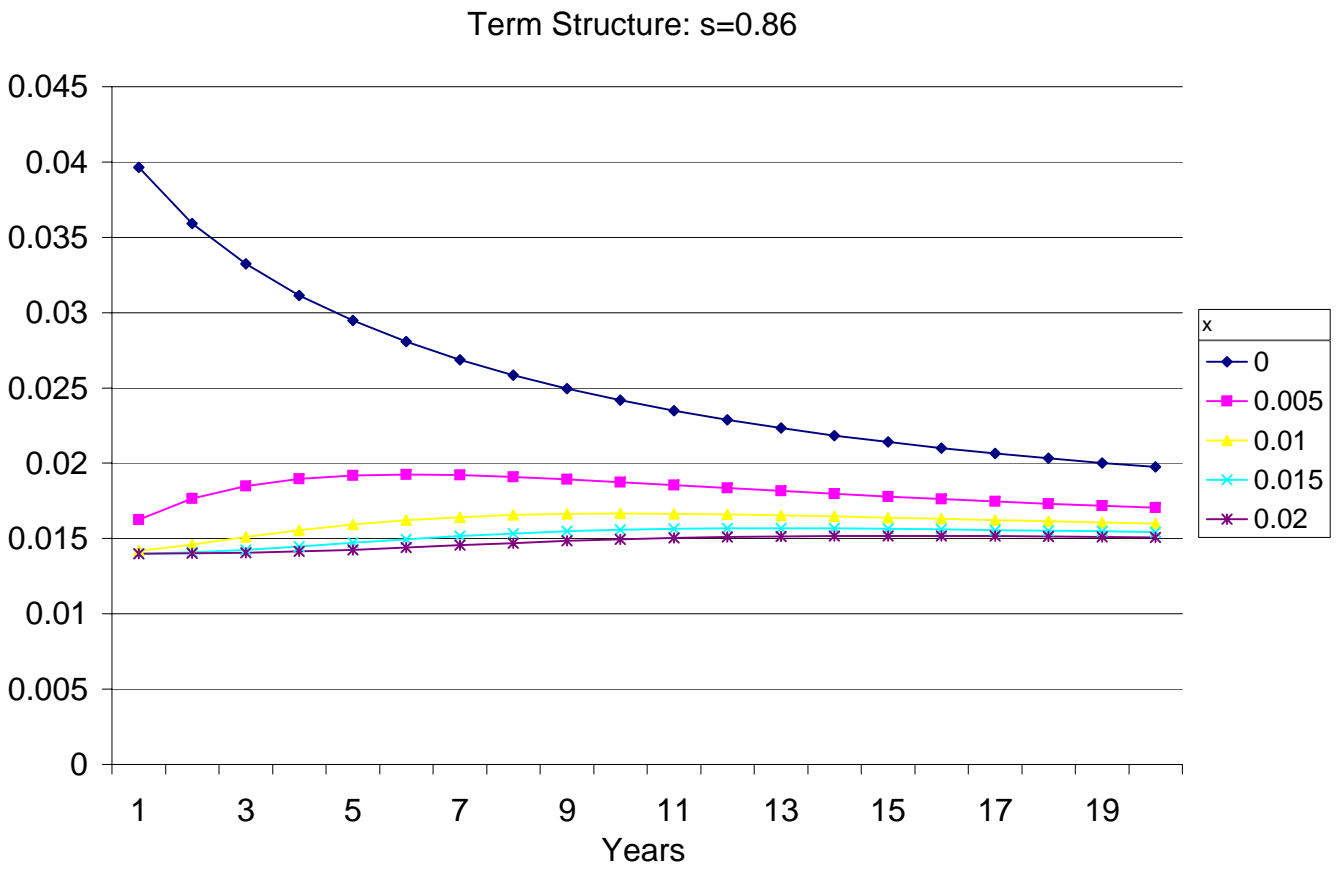


Figure 12: The term structure of real rates when $s = 0.78$ (one unconditional standard deviation above the median value of s).

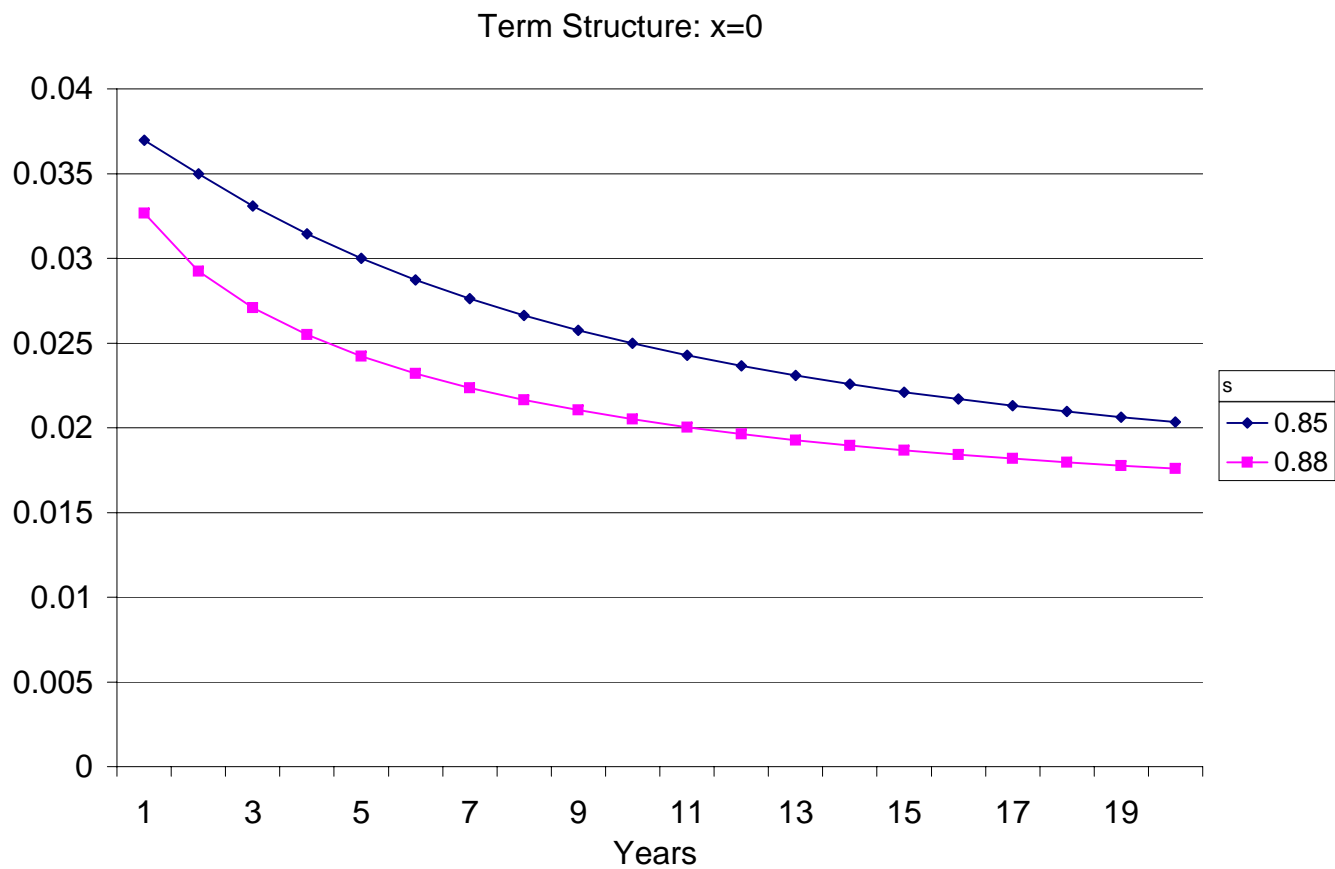


Figure 13: The term structure of real rates when $x = 0$ and $s = 0.85, 0.88$.

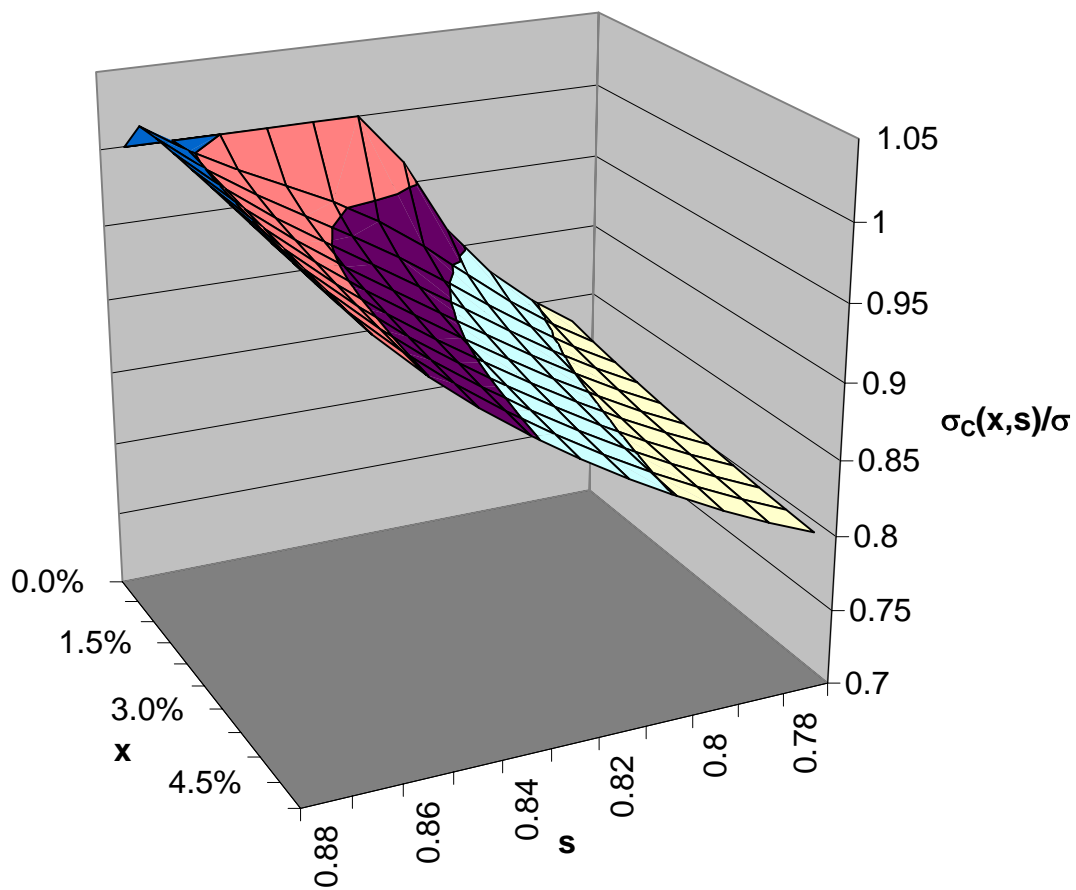


Figure 14: Consumption growth volatility as a fraction of endowment volatility.

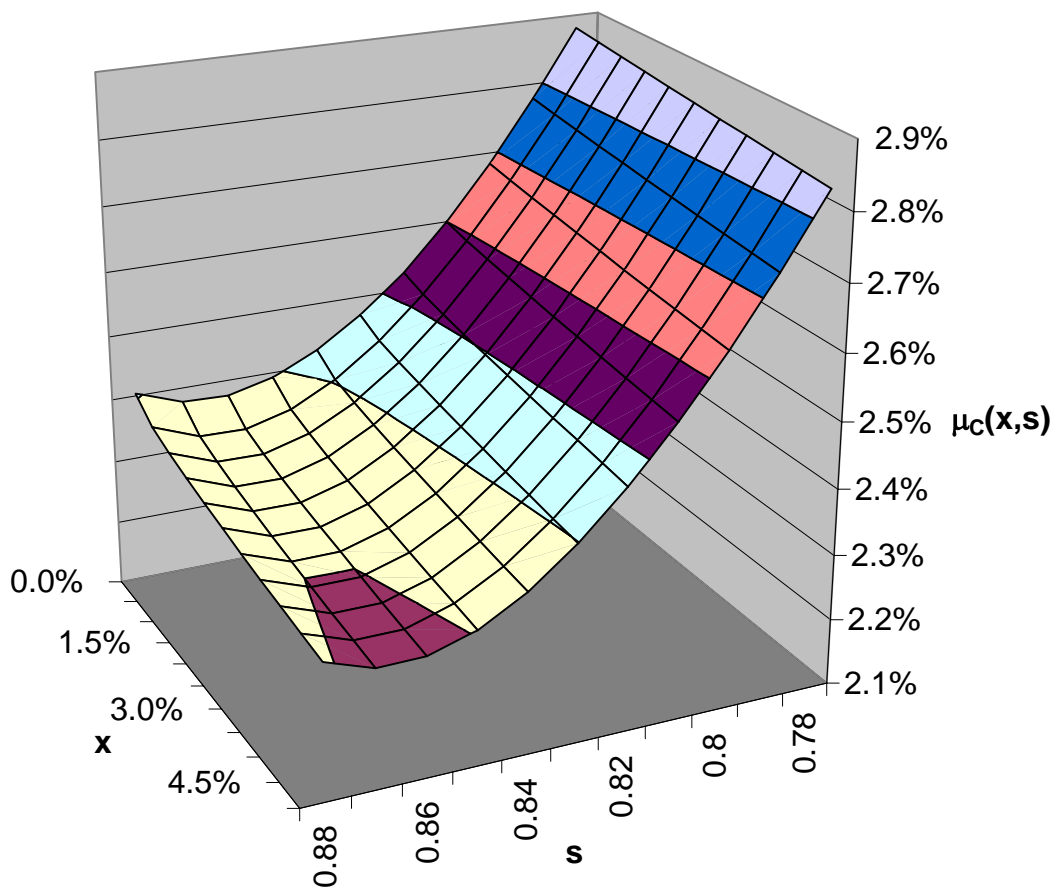


Figure 15: Expected consumption growth rate.

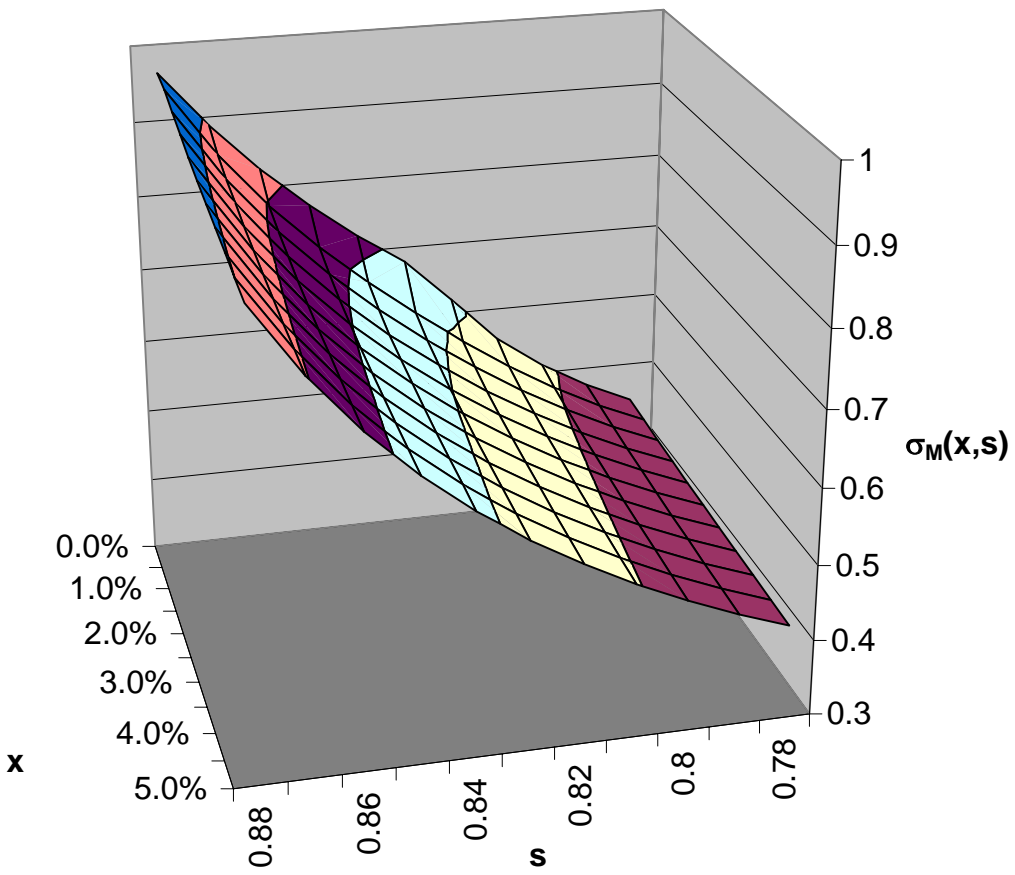


Figure 16: The market price of risk (maximum instantaneous Sharpe ratio).

Table I

Description of data used and data series constructed.

Series	Source
3-mo treasury rate (1947Q1:2004Q1)	Federal Reserve Economic Data (FRED® II)
CPI (1947Q1:2004Q1)	Federal Reserve Economic Data (FRED® II)
10-yr treasury yield (1947Q1 :2004Q1)	Federal Reserve Economic Data (FRED® II)
Consumption (1952Q2:2002Q3)	Martin Lettau's webpage
CAY (1952Q2:2002Q3)	Martin Lettau's webpage
Labor Income	Martin Lettau's webpage
Ang-Bekaert Regime Switching Probability (1952Q3 :2002Q3)	Courtesy of Andrew Ang and Geert Bekaert

Variable	Description
$r_{R,t}$	quarter-t real interest rate (nominal 3-mo treasury rate less AR(1) forecast of increase in CPI)
y_t	Date-t level of labor income
$\sigma_{5,c,t}^r$	Standard deviation of $r_{R,t-2}, r_{R,t-1}, r_{R,t}, r_{R,t+1}, r_{R,t+2}$
$\sigma_{5,l,t}^r$	Standard deviation of $r_{R,t-5}, r_{R,t-4}, r_{R,3}, r_{R,t-2}, r_{R,t-1}$
dc_t	Change in log-consumption growth from quarter t-1 to quarter t
$\sigma_{5,c,t}^c$	Standard deviation of $dc_{t-2}, dc_{t-1}, dc_t, dc_{t+1}, dc_{t+2}$
$\sigma_{5,cn,t}^c$	$\sigma_{5,c,t}^c$ divided by the standard deviation of $dc_{t-20}, \dots, dc_{t-1}$
$\sigma_{5,l,t}^c$	Standard deviation of $dc_{t-5}, dc_{t-4}, dc_3, dc_{t-2}, dc_{t-1}$
$\sigma_{5,ln,t}^c$	$\sigma_{5,l,t}^c$ divided by the standard deviation of $dc_{t-20}, \dots, dc_{t-1}$
cay_t	date-t value of Lettau and Ludvigson's CAY time-series
nts_t	Nominal term spread (10-year treasury yield minus 3-mo T-Bill rate)
prb_t	Ang-Bekaert estimate for the probability of being in the smooth real rate regime at quarter-t
s_t	20-quarter exponential smoothing of y_t using model parameters ($\alpha=0.2, \lambda=0.185$)

Table II

Correlations between measures of consumption volatility and interest rate level & volatility. Also shown are correlations with the likelihood of the smooth real rate regime as estimated by Ang and Bekaert (2004) as well as correlations with the constructed habit level.

A. 1957Q3:2002Q3

	$r_{R,t}$	$\sigma_{5,l,t}^r$	$\sigma_{5,c,t}^r$	prb_t
$\sigma_{5,l,t}^c$	0.01	0.34**	0.12*	-0.13*
$\sigma_{5,c,t}^c$	-0.05	0.13*	0.34**	-0.14*
$\sigma_{5,ln,t}^c$	-0.06	0.18**	-0.03	0.02
$\sigma_{5,cn,t}^c$	-0.07	-0.04	0.2**	-0.02
prb_t	-0.49**	-0.29**	-0.44**	

B.	$r_{R,t}$	$\sigma_{5,c,t}^r$	$\sigma_{5,l,t}^r$	$\sigma_{5,c,t}^c$	$\sigma_{5,l,t}^c$	$\sigma_{5,cn,t}^c$	$\sigma_{5,ln,t}^c$	prb_t	nts_t	cay_t
s_t	0.16*	0.25**	0.34**	0.05	0.08	0.00	0.01	0.19**	0.35**	0.53**

Table III

Correlations between consumption growth (dc_t) and lagged variables

1957Q3:2002Q3

	nts_{t-j}	$\sigma^c_{5,\ln,t-j}$	prb_{t-j}	cay_{t-j}	dc_{t-j}	s_{t-j}
j=1	0.19**	-0.12*	0.08	-0.24**	0.33**	-0.27**
j=4	0.22**	-0.15*	0.19**	-0.07	0.11	-0.09

Forecasting regressions of consumption growth (dc_t) 1957Q3:2002Q3. T-stats are in small fonts.

	nts_{t-j}	$\sigma^c_{5,\ln,t-j}$	prb_{t-j}	cay_{t-j}	dc_{t-j}	s_{t-j}	const	R^2
j=1	0.12	-0.002	0.001	-0.08	0.19	-0.06	0.10	0.21
	3.88	-1.47	0.75	-2.85	2.47	-2.00	4.65	
j=4	0.10	-0.002	0.002	-0.05	-0.01	0.07	0.05	0.10
	3.15	-1.96	2.08	-1.62	-0.18	3.01	2.62	