universität
wien

MAGISTERARBEIT

Titel der Magisterarbeit

„The Equity Premium Puzzle and Habit Formation“

Verfasser

Bastian Offermann

angestrebter akademischer Grad

Magister der Sozial- und Wirtschaftswissenschaften (Mag.rer.soc.oec.)

Wien, im Oktober 2011

Studienkennzahl lt. Studienblatt:	A 066 913
Studienrichtung lt. Studienblatt:	Magisterstudium Volkswirtschaftslehre UG2002
Betreuerin / Betreuer:	Univ.-Prof. Dipl. Ing. Dr. Gerhard Sorger

The Equity Premium Puzzle and Habit Formation

Bastian Offermann

October 6, 2011

Abstract

The contribution of Prescott and Mehra (1985) to the asset pricing literature triggered an enormous amount of research addressing the so called *equity premium puzzle*. In the following thesis I will briefly review the origins of asset pricing by presenting the seminal paper of Lucas (1978) and deriving simple closed-form asset pricing equations. In the second part of the thesis a brief review of the equity premium puzzle based on Prescott and Mehra (1985) and Hansen and Jagannathan (1991) will be given. Finally, I will discuss the approach of Campbell and Cochrane (1999) which tries to resolve the puzzle by introducing an alternative class of utility function that accounts for what is known as *habit formation*.

Keywords: *Equity Premium Puzzle, Habit Formation, CCAPM*
JEL Classification: D53, E44, G12, G13

Contents

1 Introduction

The following thesis provides an overview of the so called equity premium puzzle first published in Prescott and Mehra (1985). Prescott and Mehra published their paper in response to seminal works by Lucas (1978) and Breeden (1979). The equity premium observed empirically deviates significantly from its pendant predicted by consumption-based equilibrium models like the Lucas tree model and the CCAPM. A large number of explanations exist to adress the problem. I split the thesis into 3 parts. Part 1 serves as a basic introduction to general equilibrium asset pricing and introduces a simple contingent claims equilibrium as well as a more realistic setting in the form of a security market equilibrium. The major part of this section will describe the fundamental tree model as in Lucas (1978) and the closely related model in Breeden (1979). In part 2 the equity premium puzzle will be introduced as in Prescott and Mehra (1985). This section will be split into 2 subparts, namely a theory part to review the basic idea of the Presocott/Mehra paper and an empirical part to illustrate the most important findings from real-world data. Empirics will consist mainly of readily available data and results as in Campbell (1999). The final part then introduces the concept of habit formation which is one of the most powerful modeling frameworks to explain the puzzle. I will discuss the external habit version of Campbell and Cochrane (1999).

2 General Equilibrium Asset Pricing

In general equilibrium theory economies are viewed as systems in which the equilibrium values of the main variables of interest - consumption, production, asset prices - are determined simultaneously. Economic agents interact with each other through anonymous markets. Markets are assumed to be *perfectly competitive*, *i.e.* agents act as *price takers* and cannot influence them individually. One of the earliest works on general equilibrium theory was published by Walras (1874). Debreu (1959) extended Walras' work and provided conditions that guarantee existence of an equilibrium. With the two *Fundamental Theorems of Welfare* Debreu (1959) provided conditions (*local-non satiation* and *transitivity* of preferences) under which a *competitive equilibrium* leads to a *Pareto efficient* allocation (first theorem) and how an efficient allocation may be sustainable by a competitive equilibrium (*convexity* of production and consumption set; second theorem). Hirshleifer (1965), Hirshleifer (1966) and Radner (1972) were the first papers to provide an integrated approach by explicitly taking financial markets into account in their models. This literature was in turn extended by Merton (1973), Lucas (1978) and Breeden (1979). The Lucas *tree model* and the Consumption-based Capital Asset Pricing Model (*CCAPM*) are two of the major building blocks of what is called *equilibrium asset pricing theory.*

2.1 A Contingent Claim Economy

In the following section we will describe a simple one-period pure endowment economy and show how equilibria are determined in such settings, *i.e.* what prices and allocations ensure utility maximization for all agents as well as market clearing. As stated initially, only a one-period economy with a finite number of possible future states s is considered. Agents in such an economy negotiate contracts with each other that yield a certain payoff contingent on the realized future state s and may trade claims to such contracts for all possible states s. Such *contingent claims* are also often referred to as *Arrow-Debreu securities* (see Danthine and Donaldson (2005), p. 147). Arrow-Debreu securities are characterzized by the fact that they yield a payoff of 1 if a certain state s occurs and 0 otherwise. The realization of s is unknown at $t = 0$. Only the probability distribution over the set of possible future states is known. The realized state becomes known as soon as all contracts have been negotiated and trading takes place.

2.1.1 Model and Assumptions

In the following a formal definition of the underlying model is described as in Altug and Labadie (2008).

- There are finitely many consumers, $\{1, 2, \ldots, I\}$.
- Each possible future state s is assigned a probability $\pi_s \in (0, 1)$ such that

$$\sum_{s=1}^{S} \pi_s = 1.$$

  The set of states and the probabilities associated with each state s are known by all consumers.
- There is one commodity only.
- Consumption vectors for individual agents are of length S

$$c^i \equiv \{c_1^i, \ldots, c_S^i\}.$$

  The vector components $c_s^i \in \mathbb{R}_+$ denote the consumption of the commodity of agent i in state s. The commodity space is therefore $\mathbb{R}_+^S$ and is finite-dimensional since we assumed a finite number of states and one single commodity respectively.
- For each agent i there is an endowment vector of length S of the form

$$\omega^i \equiv \{\omega_1^i, \ldots, \omega_S^i\}.$$

- Each agent i has a utility function $u_i : \mathbb{R}_+^S \to \mathbb{R}$. The function u_i is assumed to be separable with respect to *states*.

$$u_i(c^i) = \sum_{s=1}^{S} \pi_s \cdot U_i(c_s^i).^1$$

[1] This implies that preferences are additively separable across states s. Here total utility u_i of a con-

- Finally, we can summarize a *contingent claim economy* as a list $\{(u_i, \omega_i) : i = 1, \ldots, I\}$. Each agent i is fully characterized by the tuple (u_i, ω_i) (see Lengwiler (2006), p. 24).

Having introduced the model setup we will proceed with some important definitions.

- A vector $\mathbf{C} \equiv (c^1, \ldots, c^I)$ with $\mathbf{C} \in \mathbb{R}_+^{S \times I}$ is called an allocation.
- An allocation $\mathbf{C}$ is called *feasible* or *attainable* if

$$\sum_{i=1}^{I} (c_s^i - \omega_s^i) \leq 0$$

  for all $s = 1, \ldots, S$. Thus, all consumers together can at most consume their total endowment with the consumption good in state s.

- An allocation $\mathbf{C}$ is called *Pareto optimal* if no other feasible allocation $\overline{\mathbf{C}} \equiv (\bar{c}^1, \ldots, \bar{c}^I)$ exists such that

$$u_i(\bar{c}^i) \geq u_i(c^i)$$

  for all $i = 1, \ldots, I$ and

$$u_i(\bar{c}^i) > u_i(c^i)$$

  for some i.

2.1.2 Contingent Claim Equilibrium

In what follows Altug and Labadie (2008) define a so called *complete contingent claims equilibrium* (CCE). Before the realization of a particular state s agents trade contingent claims. A contingent claim may state that an agent i transfers a certain amount of the commodity of his endowment ω^i to agent k given a certain state s is realized. A total of S contingent claims are traded in that economy (one claim for each of the S possible states). Let $p \in \mathbb{R}_+^S$ be a price vector of the form $p \equiv (p_1, \ldots, p_S)$ and $p_s \in \mathbb{R}_+$ be the price of a claim to one unit of the commodity contingent on state s. A mapping $p : \mathbb{R}_+^S \to \mathbb{R}_+$ is called a *price system* and assigns a cost to a commodity c and a value to an arbitrary endowment ω (see also Debreu (1959)). Thus,

$$p \cdot c^T \equiv \sum_{s=1}^{S} p_s \cdot c_s = p_1 \cdot c_1 + \ldots + p_S \cdot c_S.$$

So, the cost of a bundle is the *inner product* of the price vector p and commodity bundle c. A *competitive equilibrium* then is a pair (p, C) with price vector $p \geq \underline{0}$ and C is feasible[2] such that c^i solves the following constrained optimization problem for all $i = 1, \ldots, I$:

$$\max_{c^i} \; u_i(c^i)$$

sumption vector c^i can be represented as a weighted sum of utilities U_i of state contingent consumption c_s^i.

[2] $\underline{0}$ denotes the null vector of length S.

subject to

$$p \cdot {c^i}^T \leq p \cdot {\omega^i}^T.$$

2.1.3 Deriving a Contingent Claim Equilibrium

Altug and Labadie (2008) compute an equilibrium vector of prices by means of simple Lagrange multipliers. Let c_s^i denote the consumption of agent i in state s. Each agent i chooses a consumption vector $c^i \in \mathbb{R}_+^S$ by solving the following optimization problem:

$$\max_{\{c_s^i\}_{s=1}^S} \sum_{s=1}^{S} \pi_s \cdot U_i(c_s^i)$$

subject to

$$\sum_{s=1}^{S} p_s \cdot (\omega_s^i - c_s^i) \geq 0. \tag{1}$$

The utility function $U_i(.)$ is assumed to satisfy the following properties *(Inada conditions)*. Let $U_i : \mathbb{R}_+ \rightarrow \mathbb{R}$ be a concave, increasing C^2 function with

$$\lim_{c_s^i \to 0} \left(\frac{\partial U_i(c_s^i)}{\partial c_s^i} \right) = \lim_{c_s^i \to 0} U_i'(c_s^i) = +\infty$$

and

$$\lim_{c_s^i \to \infty} \left(\frac{\partial U_i(c_s^i)}{\partial c_s^i} \right) = \lim_{c_s^i \to 0} U_i'(c_s^i) = 0.$$

By the *Kuhn-Tucker Theorem* we can find a multiplier $\lambda_i \geq 0$ and a consumption vector c^i such that

$$u_i(c^i) + \lambda_i \cdot (p \cdot {\omega^i}^T - p \cdot {c^i}^T) \tag{2}$$

is maximized with respect to c^i. Equation (2) can be rephrased as

$$\mathcal{L} = \sum_{s=1}^{S} \pi_s \cdot U_i(c_s^i) + \lambda_i \cdot \left(\sum_{s=1}^{S} p_s \cdot \omega_s^i - p_s \cdot c_s^i \right)$$

being the Lagrangian to be optimized with respect to c^i. The *first-order conditions* (FOC) thus are

$$\pi_s \cdot U_i'(c_s^i) - \lambda_i \cdot p_s = 0 \tag{3}$$

for all s and i. Define

$$g_i(.) \equiv (U_i')^{-1}(.) \quad .$$

Rearranging equation (3) we get

$$U_i'(c_s^i) = \frac{p_s \cdot \lambda_i}{\pi_s}. \tag{4}$$

One can show that, by the *Implicit Function Theorem (IFT)*, the following holds:

$$c_s^i = g_i\left(\frac{p_s \cdot \lambda_i}{\pi_s}\right)$$

for all s and i. To obtain the equilibrium allocation we can substitute above equation into the original budget constraint (1)

$$\sum_{s=1}^{S} p_s \cdot \left[\omega_s^i - g_i\left(\frac{p_s \cdot \lambda_i}{\pi_s}\right)\right] = 0.$$

Since prices, endowments and state probabilities are given we obtain a function of λ_i for all $i = 1, \ldots, I$. The *left-hand side (lhs)* is strictly increasing in λ_i. This can easily be derived from equation (4). When λ_i increases marginal utility $U_i'(.)$ also increases. Since $U_i(.)$ is concave, marginal utility only increases when consumption decreases. Hence, if λ_i increases $g_i(.)$ must decrease. Thus, by the *IFT*, one can express λ_i as a function of prices. So,

$$\lambda_i = \zeta_i(p), \ \ \forall i.$$

Finally,

$$\sum_{i=1}^{I} g_i\left(\frac{\zeta_i(p) \cdot p_s}{\pi_s}\right) = \sum_{i=1}^{I} \omega_s^i$$

can be solved for the price vector p for all states $s = 1, \ldots, S$.

2.2 An Asset Economy

The model introduced in the previous section has a few drawbacks. Contingent claim markets in the narrower sense of its definition do not exist for most commodities. What is observable for actual economies are *(physical) goods* as well as *financial markets* (see Lengwiler (2006), p. 37). The following section is therefore supposed to extend the previous simple general equilibrium model by introducing the notion of an asset market to derive a *security market equilibrium*. The claims traded on such financial markets are claims to random payoffs denominated in monetary terms rather than commodities as shown in the previous section.

2.2.1 Model and Assumptions

As in the *contingent claim economy* security trading takes place before the state s is realized. Also, trading in the commodity markets only takes place after security trading has ended. Let us assume the following:

- There are finitely many securities $(1, \ldots, N)$.
- Each security n yields a certain (absolute) payoff $x_{n,s}$ for all states $s = 1, \ldots, S$. The vector of security payoffs in state s thus is $x_s \equiv (x_{1,s}, \ldots, x_{N,s})$.
- Let X be an $N \times S$ payoff matrix of the form

$$X \equiv \begin{pmatrix} x_{1,1} & \ldots & x_{1,S} \\ \vdots & \ddots & \vdots \\ x_{N,1} & \ldots & x_{N,S} \end{pmatrix}$$

.

- Let $q = (q_1, \ldots, q_N)$ denote the vector of security prices.
- Let $p_s \in \mathbb{R}_+$ denote the price of the commodity in state s and $p \in \mathbb{R}_+^S$ the price vector of the commodity for states $s = 1, \ldots, S$ (analogous definition to previous section 2.1.1).

A *portfolio* is a vector $\theta^i \equiv (\theta_1^i, \ldots, \theta_N^i) \in \mathbb{R}^N$. The vector components θ_n^i can be interpreted as the number of shares of security n held by agent i. Hence, the inner product $\theta^i \cdot q^T$ equals the *portfolio value*. Since $\theta_n \in \mathbb{R}$ we allow for negative portfolio weights which is called *short-selling* (see Hull (2008) for details).
An agent's objective is to choose a portfolio $\theta^i \in \mathbb{R}^N$ and a consumption vector $c^i \in \mathbb{R}_+^S$ by solving the following constrained optimization problem for given prices (q, p):

$$\max_{c^i, \theta^i} \; u_i(c^i) = \sum_{s=1}^{S} \pi_s \cdot U_i(c_s^i)$$

subject to

$$\theta^i \cdot q^T \leq 0$$

$$p_s \cdot {c_s^i}^T \leq p_s \cdot {\omega_s^i}^T + \theta^i \cdot {x_s}^T.$$

The first constraint simply states that an agent cannot generate positive net wealth by buying and selling securities, *i.e.* he can only invest the proceeds earned from short-selling. Hence, every purchase of a security must be financed by selling another security. This is also known as a *self-financing portfolio*. His initial endowment at $t = 0$ is therefore assumed to be 0. The major difference to the prior section's model is the fact that agents cannot trade arbitrary contingent claims for every possible state of the world. In contrast agents trade in *spot markets* which are markets for physical commodities which are *not* contingent on any future state s. Hence, in the present framework, an agent faces a different budget constraint for each state $s = 1, \ldots, S$. His consumption in state s is constrained by his endowment and the payoff generated from his investment portfolio.

2.2.2 Security Market Equilibrium

A *security market equilibrium* (SME) is a list

$$((\theta^1, c^1), \dots, (\theta^I, c^I), (q, p))$$

such that (θ^i, c^i) solves the optimization problem for all $i = 1, \dots, I$ and markets clear, *i.e.*

$$\sum_{i=1}^{I} \theta_n^i = 0, \ \ \forall n \tag{5}$$

and

$$\sum_{i=1}^{I} (c_s^i - \omega_s^i) = 0, \ \ \forall s \tag{6}$$

for given prices (q, p). The second condition again states that - in the aggregate - consumption equals supply in each state s. The first condition refers to market clearing for financial assets. Every asset that is bought by some investor has to be issued by another one first. Aggregating over all long- and short-positions portfolio holdings must sum to 0, *i.e.* securities are *"in zero net supply"* (see Lengwiler (2006), p. 50).

2.2.3 Deriving a Security Market Equilibrium

Altug and Labadie (2008) normalize the price of the single commodity to $p_s = 1$. Then each consumer solves

$$\max_{c^i, \theta^i} \ \sum_{s=1}^{S} \pi_s \cdot U_i(c_s^i)$$

subject to

$$\sum_{n=1}^{N} \theta_n^i \cdot q_n \leq 0$$

$$c_s^i \leq \omega_s^i + \sum_{n=1}^{N} \theta_n^i \cdot x_{n,s}, \ \ \forall s. \tag{7}$$

Hence, the Lagrangian equals

$$\mathcal{L} = \sum_{s=1}^{S} \pi_s \cdot U_i(c_s^i) - \mu^i \cdot \sum_{n=1}^{N} \theta_n^i \cdot q_n + \sum_{s=1}^{S} \Big(\lambda_s^i \cdot (\omega_s^i + \sum_{n=1}^{N} \theta_n^i \cdot x_{n,s} - c_s^i) \Big).$$

The *first-order conditions* with respect to the decision variables c_s^i and θ^i are

$$\pi_s \cdot U_i'(c_s^i) = \lambda_s^i \tag{8}$$

$$\mu^i \cdot q_n = \sum_{s=1}^{S} \lambda_s^i \cdot x_{n,s}, \ \ \forall n \tag{9}$$

for all $s = 1, \ldots, S$ and $i = 1, \ldots, I$. The corresponding market clearing conditions are those as in equation (5) and (6). Equilibrium allocations and prices can then be derived by first substituting equation (8) into (9) to obtain

$$\mu^i \cdot q_n = \sum_{s=1}^{S} \pi_s \cdot U'(c_s^i) \cdot x_{n,s}, \ \ \forall n. \tag{10}$$

Goods market clearing suggests that

$$\omega_s = c_s^i + \sum_{\substack{j=1 \\ j \neq i}}^{I} c_s^j.$$

We also have,

$$\omega_s = \omega_s^i + \sum_{\substack{j=1 \\ j \neq i}}^{I} \omega_s^j.$$

Hence, combining both, we can express c_s^i as

$$c_s^i = \omega_s^i + \sum_{\substack{j=1 \\ j \neq i}}^{I} \underbrace{(\omega_s^j - c_s^j)}_{\equiv A^j}$$

with

$$A^j \equiv - \sum_{n=1}^{N} \theta_n^j \cdot x_{n,s}.$$

The latter identity can simply be derived from the budget constraint (7). Substituting c_s^i into equation (10) we obtain

$$q_n = \sum_{s=1}^{S} \frac{\pi_s \cdot U_i'(\omega_s^i - \sum_{\substack{j=1 \\ j \neq i}}^{I} \sum_{n=1}^{N} \theta_n^j \cdot x_{n,s}) \cdot x_{n,s}}{\mu_i}.$$

The *right-hand side (rhs)* is strictly increasing in θ_n^j by the same argument provided in the prior section (see section 2.1.3). By the *Implicit Function Theorem* we can express portfolio holdings θ_n^i as a function of $\mu \equiv (\mu_1, \ldots, \mu_I)$ and $q \equiv (q_1, \ldots, q_N)$.

$$\theta_n^i = g_n^i(\mu, q), \ \ \forall i$$

From here we can simply use the security market budget constraint $\sum_{n=1}^{N} g_n^i(\mu, q) \cdot q_n = 0$ to solve for μ^i as a function of security prices q, *i.e.* $\mu^i = h_i(q)$. The market clearing condition for securities $\sum_{i=1}^{I} g_n^i(h(q), q) = 0$ eventually allows us to solve for q with $h(q) \equiv$

$(h_1(q), \ldots, h_I(q))$. A major point to consider is the difference between the contingent claims prices, which are fixed in advance, and security prices in a security market equilibrium. Since trading takes place before the state s is realized security prices reflect expectations regarding future payoffs, *i.e.*

$$q_n = \sum_{s=1}^{S} \frac{\pi_s \cdot U_i'(c_s^i)}{\mu^i} \cdot x_{n,s}.$$

Each payoff $x_{n,s}$ is discounted by a factor of $U_i'(c_s^i)/\mu^i$ and weighted by the probability of state s occuring. Finally, one might wonder how both models - the contingent claim and security market equilibrium - are related. In the contingent claim equilibrium agents maximize utility subject to a single budget constraint. In contrast to that, in a security market equilibrium agents face multiple budget constraints, *i.e.* one for each state. A natural question that arises in this context is under what conditions both (consumption) allocations coincide? Arrow (1964) shows that a contingent claims equilibrium can be attained by a security market equilibrium if the number of states S equals the number of securities N. A brief summary of these results can be found in Altug and Labadie (2008).

2.3 A Representative Agent

So far we have allowed for *ex ante* heterogeneity among the members of a set of individual agents. As demonstrated in the previous sections one would have to solve the optimization problem for all agents simultaneously. For a population of multiple agents it may therefore seem plausible to derive equilibrium allocations and prices by means of *one representative agent*. This idea can simply *"be justified by the fact that, in a competitive equilibrium with complete securities markets there is an especially intuitive sense of a representative agent: one whose utility function is a weighted average of the utilities of the various agents in the economy"* (see Danthine and Donaldson (2005), p. 162). Representative agents are a powerful tool to deal with the complex issue of aggregation especially when there is heterogeneity among agents in a population. The circumstances under which a representative agents exists are evaluated in this section.

2.3.1 Constructing a Representative Agent

Suppose we are in a one-period contingent claim economy with agents $i = 1, \dots, I$ choosing state contingent consumption c_s^i for period $t = 1$ (same setup as before). There is only one perishable consumption good and $c_s^i, p_s \in \mathbb{R}_+$. The procedure that I am presenting here is based on a central result which is closely related to the *second theorem of welfare*. One can show that for every *pareto optimal* allocation there exists a set of non-negative numbers $\{\lambda_i\}_{i=1}^I$ such that the same allocation can be achieved by a central planner who maximizes a linear combination of individual utility functions using $\{\lambda_i\}_{i=1}^I$ as weights (see Varian (1992), pp. 329 and Huang and Litzenberger (1988) ch. 5). We will derive conditions under which the central planner and individual optimization problems deliver the same equilibrium. The central planner maximizes

$$\max \sum_{i=1}^{I} \lambda_i \cdot \left[\sum_{s=1}^{S} \pi_s \cdot U_i(c_s^i) \right]$$

where π_s denotes the probability of state s occuring and U_i denotes the utility function of agent i over c_s^i. Assuming that utility is strictly increasing, the weights λ_i are strictly positive, *i.e.*

$$\lambda_i > 0, \ \forall i.$$

Since individuals may at most consume the available aggregate consumption level one has to impose the following budget constraint:

$$\sum_{i=1}^{I} c_s^i = C_s, \ \forall s$$

where C_s denotes aggregate consumption in state s. Taking this constraint into account one can set up the Lagrangian

$$\max_{c_s^i} \mathcal{L} = \sum_{i=1}^{I} \lambda_i \cdot \left[\sum_{s=1}^{S} \pi_s \cdot U_i(c_s^i) \right] + \sum_{s=1}^{S} \phi_s \cdot \left[C_s - \sum_{i=1}^{I} c_s^i \right]. \tag{11}$$

The FOC then is

$$\frac{\partial \mathcal{L}}{\partial c_s^i} = \lambda_i \pi_s \cdot \frac{\partial U_i(c_s^i)}{\partial c_s^i} - \phi_s = 0, \ \ \forall i, \ \ \forall s. \tag{12}$$

Accordingly, the optimization problem of an individual agent i then is

$$\max \sum_{s=1}^{S} \pi_s \cdot U_i(c_s^i)$$

subject to

$$\sum_{s=1}^{S} p_s \cdot c_s^i = \sum_{s=1}^{S} p_s \cdot \omega_s^i$$

where p_s denotes the price of a claim on one unit of consumption in state s and ω_s^i denotes the endowment of agent i in state s. The Lagrangian is

$$\max_{c_s^i} \mathcal{L} = \sum_{s=1}^{S} \pi_s \cdot U_i(c_s^i) + \psi_i \cdot \left[\sum_{s=1}^{S} p_s \omega_s^i - p_s c_s^i \right] \tag{13}$$

with $\psi_i > 0, \ \ \forall i$. The corresponding FOC then is

$$\frac{\partial \mathcal{L}}{\partial c_s^i} = \pi_s \cdot \frac{\partial U_i(c_s^i)}{\partial c_s^i} - \psi_i \cdot p_s = 0, \ \ \forall s. \tag{14}$$

By setting $\phi_s = p_s$ and $\lambda_i = \psi_i^{-1}$ the optimality conditions for a single agent (see equation (14)) and a central planner (see equation (12)) are equivalent. Henceforth, a central planner who wants to achieve the same pareto optimal allocation as if households optimized individually would have to assign a weight of ψ_i^{-1} to individual i. This result has an intuitive economic interpretation. From equation (13) we get that the parameter ψ_i is the Lagrangian multiplier with respect to the budget constraint. Furthermore, since such multipliers are interpreted as shadow prices, an agent i's weight λ_i is equal to the inverse of this shadow price obviously. From equation (14) one can spot that the product of shadow price ψ_i and the price of a claim in state s equals the marginal utility of consumption contingent on state s, *i.e.* an agent i is considered important when λ_i is large which is equivalent to ψ_i being small (and hence marginal utility in state s being small). Since marginal utility is small when consumption is high, a high weight λ_i corresponds to an "important" agent which makes sense intuitively. This derivation, however, only works when markets are complete. A market is considered complete when a complete set of state contingent claims exist. Prices of such claims are referred to as *state prices*. Why is market completeness required? In order to value endowments one needs a state price vector of consumption claims. Although such a vector exists in both, complete and incomplete markets, it is only *unique* in complete markets. Hence, different state prices imply different representative agents leading the idea of *one* representative agent *ad absurdum*.

Now that we have derived conditions under which a representative agent exists one can construct the representative agent as follows: Define

$$U_r(x) \equiv \max_{\{x_i\}_{i=1}^{I}} \sum_{i=1}^{I} \lambda_i \cdot U_i(x_i) \tag{15}$$

subject to

$$\sum_{i=1}^{I} x_i = x$$

with $\lambda_i = \psi_i^{-1}$ being the weight that the central planner assigns to an agent i. Taking the first derivative of this function with respect to aggregate consumption C_s yields

$$\frac{\partial U_r(C_s)}{\partial C_s} = U_r'(C_s) = \sum_{i=1}^{I} \lambda_i \cdot U_i'(c_s^i) \cdot \frac{\partial c_s^i}{\partial C_s} = \sum_{i=1}^{I} \underbrace{\lambda_i \cdot \psi_i}_{=1} \cdot \frac{p_s}{\pi_s} \cdot \frac{\partial c_s^i}{\partial C_s} = \frac{p_s}{\pi_s}$$

with

$$\sum_{i=1}^{I} \frac{\partial c_s^i}{\partial C_s} = 1.$$

The product $\lambda_i \cdot \psi_i$ is trivially equal to 1 since $\lambda_i = \psi_i^{-1}$. Similarly, the partial derivatives of c_s^i with respect to C_s adds up to 1 trivially. The reason is intuitive. When aggregate consumption increases by one unit each agent i retains his share in that additional unit. Thus, the sum of these shares must add up to a total of 1. Finally, to show the existence of a representative agent, let $C_s, \ \forall s$ denote the representative agents initial endowment in state s. Let π_s denote its subjective probability of state s and let utility be denoted by $U_r(C_s)$. To show the existence of the representative agent we will derive that state prices of consumption claims equal p_s. Since we are in a representative agent economy no trading occurs, *i.e.* prices must be such that the representative agent never wants to trade. In equilibrium the representative agent's marginal utility of consumption contingent on state s $\pi_s \cdot U_r'(C_s)$ must be equal to the state price p_s of the consumption claim. The agent will "buy" claims as long as $p_s < \pi_s \cdot U_r'(C_s)$ and "sell" them as long as $p_s > \pi_s \cdot U_r'(C_s)$. From $U_r'(C_s) = p_s/\pi_s$ we trivially get

$$\pi_s \cdot U_r'(C_s) = \pi_s \cdot \frac{p_s}{\pi_s} = p_s.$$

Hence, the representative agent's marginal utility equals state prices and therefore our representative agent exists.

2.4 The Lucas Model and CCAPM

The following section will introduce the *consumption-based capital asset pricing model* based on the papers of Lucas (1978) and Breeden (1979). The setup comprises a simple pure endowment economy and a representative agent who dynamically optimizes equity portfolio- as well as consumption allocations. We will therefore use a *recursive equilibrium approach.* Recursive methods first appeared in the works of Wald (1945), Bellman (1957) and Kalman (1960) and provide the necessary tools to study and analyze dynamic economic systems. A comprehensive summary of the latest methods is the book of Ljungqvist and Sargent (2004).

2.4.1 Model and Assumptions

As in the previous chapters we study a *pure endowment economy* with a representative agent, a single, perishable consumption good and n distinct production units. Our aim is to derive fundamental pricing functions for risky assets. With the *equity premium puzzle* in mind our focus will therefore be on the pricing of equity securities.

We will begin by describing the stochastic nature of the output process. In each period n exogenous shocks s_t affect the output process y_t where $y_t = \sum_{i=1}^n y_{i,t}$. The component $y_{i,t}$ refers to the output of unit i in period t. Let $s_t \in S \subset \mathbb{R}^n$ where S is assumed to be compact[3]. The shocks s_t follow a *Markov process* with *transition function* ϕ. The transition function is a mapping $\phi : S \times S \to [0, 1]$ such that

$$\phi(s, s') \equiv \text{Prob}(s_{t+1} \leq s' | s_t = s)$$

and can be interpreted as assigning probabilities to certain shocks s' occuring in period $t+1$ given that s occured in the prior period. It possesses the *Feller property (assumption a)*, *i.e.* for any bounded and continuous function $h : S \to \mathbb{R}$ the term $\int h(s') \ d\phi(s, s')$ is continuous in s (see Lucas (1978), p. 1431). The process generated by the transition function ϕ has a stationary distribution Φ (see mathematical appendix A.1 for details). Output is modelled as a function of the shocks

$$y_t \equiv y(s_t)$$

which is *invariant* with respect to time. Since the shock process s_t follows a first-order Markov process the output process is of the same kind. Furthermore, we assume that the output process takes only positive values in a compact set. Let

$$\mathcal{Y} \equiv [\underline{y}, \overline{y}]$$

with $\underline{y} > 0$ being a lower bound and $\overline{y} < \infty$ an upper bound. Hence, the mapping $y : S \to \mathcal{Y}$ is continuous and "*bounded away from zero*" *(assumption b)* (see Altug and Labadie (2008),

[3]In a real vector space this is equivalent to being closed and bounded by the *Heine-Borel Theorem.*

p. 163)[4]. Let $\{c_t\}_{t=0}^{\infty}$ be a sequence of consumptions. Preferences of the representative agent are given by

$$\mathbf{E}_0 \left\{ \sum_{t=0}^{\infty} \beta^t \cdot U(c_t) \right\} \tag{16}$$

where $\beta \in (0,1)$ is a discount factor and $\mathbf{E}_0$ is an expected value operator conditional on information avaiable at $t = 0$. Preferences are assumed to be *additively separable* with respect to consumption across time. Let $U : \mathbb{R}_+ \to \mathbb{R}_+$ denote a strictly concave and strictly increasing C^2 function representing preferences over consumption with Inada conditions $U(0) = 0$ and $\lim_{c \to 0} U'(c) = \infty$ *(assumption c)*. The latter assumption makes sure that we obtain an interior solution of the maximization problem (see Danthine and Donaldson (2005), p. 166).

The shares traded in our economy are claims on the output process. Let $q_t = (q_{1,t}, \ldots, q_{n,t}) \in \mathbb{R}_+^n$ denote the price vector of a share *ex dividend*, *i.e.* after dividends have been paid. By $z_t = (z_{1,t}, \ldots, z_{n,t}) \in \mathbb{R}_+^n$ we denote the *beginning-of-period* share holdings. Each outstanding equity share is assumed to be perfectly divisible. The budget constraint faced by the representative agent in each period t equals

$$c_t + q_t \cdot z_{t+1} \leq (y_t + q_t) \cdot z_t \tag{17}$$

for all $t = 0, 1, \ldots$. Period t consumption c_t can then be interpreted as the difference between total output y_t[5] plus the total value of financial assets $q_t \cdot z_t$ held in period t and the value of financial assets $q_t \cdot z_{t+1}$ carried forward to period $t+1$ such that "the *rhs* finances the *lhs*". The initial share holdings z_0 are taken as given. Trading in financial markets in any period t only takes place after the corresponding realized output y_t has been observed for that period. The agent chooses sequences for consumption and equity so as to maximize equation (16) subject to the budget constraint (17) and the following

$$c_t \geq 0$$

$$0 \leq z_{t+1} \leq \overline{z}, \;\; \overline{z} \gg \mathbf{e}$$

for all t where $\mathbf{e} \equiv (1, \ldots 1)$ is an n-element vector of ones. Sequences for equity prices are taken as given. Market clearing holds when

$$c_t = y_t$$

$$z_{t+1} = \mathbf{e}$$

[4]Lucas (1978) expresses this condition as $\phi(y, 0) = 0$, *i.e.* for a given y the next period's realization of the process can never equal 0.

[5]Total output y_t equals total dividends paid. There are no retained earnings.

for all t. The first condition simply states that the whole output is consumed *(the consumption good is perishable and cannot be stored).* The second one implies security market clearing, *i.e.* all shares are held by the representative agent.

2.4.2 Recursive Competitive Equilibrium

In the following we will show the existence of a *recursive competitive equilibrium*, *i.e.* we will establish a functional relationship between asset prices and the exogenously determined production shocks that affect output. As stated previously, total output equals total consumption in an equilibrium allocation. For any equilibrium allocation the agent's expected utility must therefore be finite (bounded). Consider the following:

Lemma 2.1 *Under assumptions a–c, for any consumption sequence* $\{c_t\}_{t=0}^{\infty}$ *with* $c_t \leq y_t$*, we have*

$$\mathbf{E}_0 \left\{ \sum_{t=0}^{\infty} \beta^t \cdot U(c_t) \right\} \leq \mathcal{B} < \infty.$$

Proof In the prior section we made sure that output takes values in a compact set by defining $\mathcal{Y} \equiv [\underline{y}, \overline{y}]$. Thus, consumption can be chosen such that $c_t \in [0, \overline{y}]$ with $y_t \leq \overline{y}$. Thus, under the continuous mapping $U(.)$ the image of a compact domain is again compact. Hence, total utility is bounded, *i.e.*

$$\mathcal{B} \equiv \sum_{t=0}^{\infty} \beta^t \cdot U(\overline{y}) = \frac{1}{1-\beta} \cdot U(\overline{y}) < \infty.$$

□

The agent's problem may then be formulated as a *stationary dynamic programming problem.* The relevant *state variables* are the number of shares z and the exogenous shock parameter s. The aim is to maximize expected utility subject to a budget constraint and market-clearing conditions. Let v denote the value function. The price function

$$q : S \to \mathbb{R}_+^n$$

is assumed to be given by the agent. The dynamic optimization problem can then be formulated recursively as

$$v(z, s) = \max_{c, z'} \left\{ U(c) + \beta \cdot \int v(z', s') \ d\phi(s, s') \right\} \tag{18}$$

subject to

$$c + q \cdot z' \leq (y + q) \cdot z \tag{19}$$

$$c \geq 0, \ \ z' \in Z. \tag{20}$$

where $z' \in Z \Leftrightarrow 0 \leq z'_i \leq \overline{z}_i$ for all $i = 1, \ldots, n$. Variables with a prime superscript denote future states and those without prime denote present states[6]. The relevant choice variables are current state consumption c and the future allocation of financial assets z'. A recursive competitive equilibrium can then be defined as follows:

Definition A recursive competitive equilibrium is a price function $q : S \to \mathbb{R}^n_+$ and a value function $v : Z \times S \to \mathbb{R}_+$ such that (i) given $q(s)$, $v(z, s)$ solves the agent's optimization problem and (ii) markets clear.

Since we are in a representative agent economy one has to introduce the notion of a so called *no trade equilibrium*. In a multi-agent economy one obtains an equilibrium when supply equals demand, *i.e.* for a given price some agents are willing to sell exactly what others are willing to buy. A representative agent does not have any trading counterparts. Hence, a no trade equilibrium is characerized in such a way that supply equals demand and both are at the same time equal to zero. At the equilibrium price the agent is thus willing to own all shares outstanding. *"Therefore the essential question being asked is: What prices must securities assume so that the amount the representative agent must hold (for all markets to clear) exactly equals what he wants to hold"* (see Danthine and Donaldson (2005), p. 154).

In what follows, a *proof of the existence* of the value function $v(.)$ will be given. The relevant mathematical preliminaries can be found in the appendix (see section A.3). Let the price function $q(s)$ be given. Let $\mathcal{S} \equiv Z \times S$. $\mathcal{S}$ is then a compact set being the cartesian product of compact sets. Let $\mathcal{C}(\mathcal{S})$ be the space of bounded, continuous functions $v : \mathcal{S} \to \mathbb{R}_+$ equipped with the supremums norm

$$\|u\| \equiv \sup_{z,s \in \mathcal{S}} |u(z, s)|, \ \ \forall u \in \mathcal{C}(\mathcal{S})$$

so that we have, in fact, a complete metric space such that every Cauchy sequence (of functions) converges to an element in that space (see Ljungqvist and Sargent (2004), p. 926).

Theorem 2.2 *For any given continuous price function $q(s)$ there exists a unique, bounded, continuous and nonnegative solution $v^* \in \mathcal{C}(\mathcal{S})$ to the functional equation defined by* (18). *The function v^* is concave and increasing in z.*

Proof Let $v \in \mathcal{C}(\mathcal{S})$ and define an operator T

$$Tv(z, s) = \max_{c,z'} \left\{ U(c) + \beta \cdot \int v(z', s') \ d\phi(s, s') \right\}$$

subject to (19) and (20). The sets $\mathcal{Y}$, S and Z are compact by assumption. The utility function U is also continuous by assumption as is v since $v \in \mathcal{C}(\mathcal{S})$. Hence, maximizing a

[6] The recursive nature of the problem allows us to drop the time indices, since the optimization problem is reduced to a two-period framework *"today and tomorrow"*.

continuous real-valued function v over a compact set $\mathcal{S}$ yields a maximum. Furthermore, both, U and v, are bounded. Hence, Tv is bounded and it is continuous as the sum of continuous functions is again continuous. T is then a mapping, $T : \mathcal{C}(\mathcal{S}) \rightarrow \mathcal{C}(\mathcal{S})$, which maps the space of continuous, bounded functions into itself. To show that T is indeed a *contraction mapping* one has to prove the *monotonicity* and *discounting properties (Blackwell conditions)* (see A.5). Let $u, w \in \mathcal{C}(\mathcal{S})$ with $u \geq w, \ \forall z, s \in \mathcal{S}$. Hence,

$$\begin{aligned} Tu &= \max_{c,z'} \left\{ U(c) + \beta \cdot \int u(z', s') \ d\phi(s, s') \right\} \\ &\geq \max_{c,z'} \left\{ U(c) + \beta \cdot \int w(z', s') \ d\phi(s, s') \right\} = Tw \end{aligned} \tag{21}$$

establishing the argument. The discounting property is easily verfied as follows: Let k be an arbitrary constant. Then

$$\begin{aligned} T(v + k) &= \max_{c,z'} \left\{ U(c) + \beta \cdot \int [v(z', s') + k] \ d\phi(s, s') \right\} \\ &= \max_{c,z'} \left\{ U(c) + \beta \cdot \int v(z', s') \ d\phi(s, s') \right\} + \beta \cdot k \\ &= Tv + \beta \cdot k \end{aligned}$$

and T satisfies the Blackwell conditions. Having shown that T is, in fact, a contraction mapping and considering that $\mathcal{C}(\mathcal{S})$ is a complete, normed, linear space then T has a unique fixed point and $\lim_{n\to\infty} T^n v_0 = v^*$ for any $v_0 \in \mathcal{C}(\mathcal{S})$ by the *contraction mapping theorem* (see A.2).

What remains to be shown is that v^* is increasing and concave. Let $\mathcal{C}'(\mathcal{S}) \subset \mathcal{C}(\mathcal{S})$ be the subspace of continuous, bounded, increasing and concave real-valued functions equipped with the supremums norm. $\mathcal{C}'(\mathcal{S})$ is a closed, complete, normed, linear space. Let $w \in \mathcal{C}'(\mathcal{S})$. Since w is increasing in z we have $w(z_1, s) < w(z_2, s)$ for $z_1 < z_2$[7]. Since T satisfies the Blackwell conditions this implies $Tw(z_1, s) < Tw(z_2, s)$. Finally, one has to show that T preserves concavity. Let $z_0, z_1 \in Z$ be arbitrary share allocations and $c_0, c_1 \in \mathcal{Y}$ be arbitrary consumptions. Let $\alpha \in [0, 1]$, $z_\alpha = \alpha \cdot z_0 + (1 - \alpha) \cdot z_1$ and $c_\alpha = \alpha \cdot c_0 + (1 - \alpha) \cdot c_1$. Since (c_i, z_i') is a feasible allocation for $i = 0, 1$ any convex combination of them must also be feasible. Hence, (c_α, z_α') satisfies (17). $Tw(z_i, s)$ is attained at (c_i, z_i') for $i = 0, 1$. Then

[7] The subscripts denote an index i for a whole vector z and do not refer to the *i-th* element of vector z.

$$\begin{aligned}
Tw(z_\alpha, s) &\geq U(c_\alpha) + \beta \cdot \int w(z'_\alpha, s')\ d\phi(s,s') \\
&\geq \alpha \cdot U(c_0) + (1-\alpha) \cdot U(c_1) + \alpha \cdot \beta \cdot \int w(z'_0, s')\ d\phi(s,s') \\
&+ (1-\alpha) \cdot \beta \cdot \int w(z'_1, s')\ d\phi(s,s') \\
&= \alpha \cdot \Big[U(c_0) + \beta \cdot \int w(z'_0, s')\ d\phi(s,s')\Big] \\
&+ (1-\alpha) \cdot \Big[U(c_1) + \beta \cdot \int w(z'_1, s')\ d\phi(s,s')\Big] \\
&\geq \alpha \cdot Tw(z_0, s) + (1-\alpha) \cdot Tw(z_1, s).
\end{aligned}$$

Since U and w are concave by assumption the second inequality must hold. $Tw(z_i, s)$ is attained at (c_i, z'_i) for $i = 0, 1$. Hence, we can factor out α and $(1-\alpha)$ respectively to obtain the third inequality. Eventually, putting together the fact that T is a contraction mapping on $\mathcal{C}(\mathcal{S})$ and $T(\mathcal{C}'(\mathcal{S})) \subseteq \mathcal{C}'(\mathcal{S})$ we must have $v^* \in \mathcal{C}'(\mathcal{S})$ by corollary A.3 (see Lucas (1978), p. 1432–1433). □

Hence, a solution to the consumers constrained optimization problem exists for a given price function q. So far we proved the existence of a value function v given the respective pricing function q. In the following, we will show the *existence of the pricing function* q for equities given that v exists. A proof regarding the differentiability of v is shown in Lucas (1978) (p. 1433–1434) and will be omitted here. The overall approach to follow is very similar to the one previously taken for v. Using the objective function (18), the budget constraint (19) and a Lagrange multiplier $\lambda(s)$ one can set up the Lagrangian

$$\begin{aligned}
\mathcal{L} = U(c) + \beta \cdot \int v(z', s')\ d\phi(s,s') \\
+ \int \lambda(s) \cdot \Big(\mathbf{y} \cdot z - c + q \cdot z - q \cdot z'\Big)\ d\phi(s,s').
\end{aligned}$$

where $\mathbf{y} = (y_1, \ldots, y_n)$[8] for convenience. The FOC with respect to c and z' are then

$$\frac{\partial \mathcal{L}}{\partial c} = U'(c) - \lambda(s) = 0$$

and

$$\frac{\partial \mathcal{L}}{\partial z'_i} = \beta \cdot \int \frac{\partial v(z', s')}{\partial z'_i}\ d\phi(s,s') - \lambda(s) \cdot q_i = 0,\ \ \forall i$$

The *envelope conditions*, which are the partial derivatives of the Lagrangian with respect to the parameter z, equal

[8] y_i denotes the output of unit i.

$$\frac{\partial \mathcal{L}}{\partial z_i} = \lambda(s) \cdot (y_i + q_i) = \frac{\partial v(z,s)}{\partial z_i}, \ \forall i.$$

The latter equality holds by Lucas (1978) (see proposition 2, p. 1433–1434). Let $c^*(z,s)$ and $z^*(z,s)$ denote the equilibrium solutions of the policy functions of the Bellman equation (18). From the market clearing conditions one then gets in equilibrium $c^*(1,s) = y(s)$ (total output is consumed) and $z^*(1,s) = \mathbf{e}$ (all shares are held). Combining the first order- and envelope conditions one can derive the *intertemporal Euler equations* which are

$$\begin{aligned} U'(y) \cdot q_i &= \beta \cdot \int U'(y') \cdot [y_i' + (q_i)'] \ d\phi(s,s') \\ &= \underbrace{\beta \cdot \int U'(y') \cdot y_i' \ d\phi(s,s')}_{\equiv \gamma_i(s)} \\ &\quad + \beta \cdot \int \underbrace{U'(y') \cdot (q_i)'}_{\equiv \xi_i(s')} \ d\phi(s,s'), \ \forall i. \end{aligned} \tag{22}$$

As in the previous section one can then prove the existence and uniqueness of a pricing function q_i given the value function v for all $i = 1, \ldots n$. Again, let $\mathcal{C}(S)$ denote the space of bounded and continuous functions $\xi_i : S \to \mathbb{R}_+$ equipped with the supremums norm for all $i = 1, \ldots, n$. Let the mapping $\gamma_i : S \to \mathbb{R}_+$ be defined as

$$\gamma_i(s) \equiv \beta \cdot \int U'(y') \cdot y_i' \ d\phi(s,s'), \ \forall i. \tag{23}$$

In addition to that, let $\xi_i(s) \equiv U'(y) \cdot q_i$ and T_i be an operator such that

$$T_i \xi_i(s) = \gamma_i(s) + \beta \cdot \int \xi_i(s') \ d\phi(s,s'), \ \forall i. \tag{24}$$

Above equation resembles the problem studied in the previous section (see equation (18)). Hence, we can formulate:

Theorem 2.3 *There exists a unique, bounded and continuous solution ξ_i^* to $T_i\xi_i = \xi_i$ for all $i = 1, \ldots, n$. For any $\xi_i^0 \in \mathcal{C}(\mathcal{S})$ we have $\lim_{n\to\infty} T_i^n \xi_i^0 = \xi_i^*$.*

Proof The proof works in a similar fashion as the proof of theorem 2.2. Our aim is to show that T_i is a contraction mapping. The following arguments always hold for all $i = 1, \ldots, n$. We will first show that the *rhs* of (24) is bounded and continuous. $\gamma_i(s)$ is bounded by the following: Since $U' > 0$ and $y_i > 0$, we must have $\gamma_i > 0$. By definition y (aggregate output) takes values in a compact set $\mathcal{Y}$. Since $U(.)$ is continuous and the domain space $\mathcal{Y}$ is compact the image of $\mathcal{Y}$ under U must also be compact. Hence, $U(.)$ is bounded and it is concave by assumption. Thus, the following must hold

$$U(y) - U(0) \geq U'(y) \cdot (y - 0) = U'(y) \cdot y - U'(y) \cdot 0 = U'(y) \cdot y$$

Economically, this can be interpreted as *"average utility exceeds marginal utility"* (see Varian (1992), p. 489). Hence, $U'(y) \cdot y_i \leq \overline{U}, \ \forall y \in \mathcal{Y}$ (this must hold since $y_i \leq y$) and therefore

$$\gamma_i(s) = \beta \cdot \int U'(y') \cdot y_i' \ d\phi(s, s') \leq \beta \cdot \overline{U},$$

i.e. a weighted average of bounded functions is again bounded. By the Feller property (see *assumption a* on page 16) the *rhs* of equation (23) is continuous. Hence, the function $T_i : \mathcal{C}(S) \to \mathcal{C}(S)$ is a mapping from the space of bounded and continuous functions into itself. Obviously, T_i satisfies the monotonicity property . Let $\nu_i \in \mathcal{C}(S)$ such that $\nu_i(s) \geq \xi_i(s), \ \forall s \in S$, then $T_i\nu_i(s) \geq T_i\xi_i(s), \ \forall s \in S$ (see previous result on page 20). The discounting property is also easily verified as before. Let k be an arbitrary constant. Then,

$$\begin{aligned} T_i(\xi_i + k)(s) &= \gamma_i(s) + \beta \cdot \int [\xi_i(s') + k] \ d\phi(s, s') \\ &= T_i\xi_i(s) + \beta \cdot k \end{aligned}$$

and we can concluded that T_i is, in fact, a contraction mapping defined on the a complete, linear and normed space of functions. By the contraction mapping theorem A.2 ξ_i^* is a unique fixed point in $\mathcal{C}(S)$. □

Having defined the identity $\xi_i \equiv U'(y) \cdot q_i$ earlier one can then solve for the equity price as

$$q_i(s) = \frac{\xi_i^*(s)}{U'(y(s))}. \tag{25}$$

Hence, we obtain a unique equilibrium pricing function $q(s) = (q_1(s), \ldots, q_n(s))$. Asset prices thus depend on the output/consumption process as well as the shape of the utility function.

2.4.3 Asset Pricing Functions

Having established the existence and uniqueness of the value function v and the equity price function q in the previous section, we will now proceed to derive an explicit expression for the yet to be determined asset pricing function ξ^*. In contrast to the prior section we will do this for a simplified framework in which there is only one production unit and hence only one equity share (see Lucas (1978), p. 1439). It follows trivially that individual production therefore always equals aggregate production. For that purpose it is assumed that the shocks $s_t \in \mathbb{R}$ are *independently and identically distributed (iid)*, *i.e.* every realization of the shock has the same probability distribution and realizations are independent of each other. Since output is a function of the shock parameters, $\{y_t\}_{t=0}^{\infty}$ is a sequence of *iid* random variables. Let $\Phi(y)$ denote the *cumulative distribution function* of the stationary output process. The *intertemporal Euler equation* (22) can then be written as

$$
\begin{aligned}
U'(y) \cdot q &= \beta \cdot \int \Big[U'(y') \cdot [y' + q'] \ \ d\Phi(y') \Big] \\
&= \beta \cdot \int U'(y') \cdot y' \ \ d\Phi(y') + \beta \cdot \int U'(y') \cdot q' \ \ d\Phi(y')
\end{aligned} \tag{26}
$$

which is the function to be solved for q. Let

$$\overline{\xi} \equiv \beta \cdot \int U'(y') \cdot y' \ \ d\Phi(y') = \beta \cdot \mathbf{E}[U'(y') \cdot y'] \tag{27}$$

be constant and define

$$\xi_n(y) = T\xi_{n-1}(y) = \overline{\xi} + \beta \cdot \int \xi_{n-1}(y') \ \ d\Phi(y')$$

with

$$\xi(y) = U'(y) \cdot q(y). \tag{28}$$

Since T is, in fact, a contraction mapping we have a converging sequence of functions $\xi_n(y) \to \xi^*(y)$ for any $\xi_0(y) \in \mathcal{C}(S)$. Suppose $\xi_0(y) = 0$. By repeated substitution we get

$$
\begin{aligned}
\xi_1 &= T\xi_0 = \overline{\xi} + \beta \cdot \int 0 \ \ d\Phi(y') = \overline{\xi} \\
\xi_2 &= T\xi_1 = \overline{\xi} + \beta \cdot \int \overline{\xi} \ \ d\Phi(y') = \overline{\xi} \cdot (1 + \beta) \\
&\vdots \\
\xi_n &= T\xi_{n-1} = \overline{\xi} + \beta \cdot \int \overline{\xi} \cdot (1 + \beta + \ldots + \beta^{n-2}) \ \ d\Phi(y') \\
&= \overline{\xi} \cdot (1 + \beta + \ldots + \beta^{n-1}) = \overline{\xi} \cdot \sum_{j=0}^{n-1} \beta^j \\
&.
\end{aligned}
$$

More generally, one can express $\xi_n = T^n \xi_0$ (see proof of theorem A.2). Hence, in the limit, we get

$$\begin{aligned}\xi^* &= \lim_{n\to\infty} T^n \xi_0 \\ &= \lim_{n\to\infty} \overline{\xi} \cdot \sum_{i=0}^{n-1} \beta^i = \overline{\xi} \cdot \frac{1}{1-\beta}.\end{aligned}$$

Resubstituting the relevant expressions for ξ^* (25) and $\overline{\xi}$ (27) we get

$$U'(y) \cdot q = \frac{\beta}{1-\beta} \cdot \mathbf{E}[U'(y') \cdot y'].$$

The sensitivity of equity prices with respect to output can be determined as follows: Differentiating equation (28) with respect to y yields

$$\begin{aligned}\frac{\partial q}{\partial y} &= \frac{\xi^{*\prime} \cdot U'(y) - U''(y) \cdot \xi^*}{U'(y)^2} = -\frac{U''(y) \cdot \xi^*}{U'(y)^2} \\ &= -\frac{U''(y)}{U'(y)^2} \cdot \frac{\beta \cdot \mathbf{E}[U'(y') \cdot y']}{1-\beta} \\ &= -q \cdot \frac{U''(y)}{U'(y)} > 0 \\ .\end{aligned} \tag{29}$$

The second equality of the first line must hold since ξ^* is a constant. In the final line q is substituted for $\beta \cdot \mathbf{E}[U'(y') \cdot y']/[(1-\beta) \cdot U'(y)]$. Since the utility function is assumed to be concave we have $U''(y) < 0$ and the whole last expression is positive. Equation (29) can then be rearranged to state

$$\frac{y \cdot q'}{q} = -\frac{y \cdot U''(y)}{U'(y)} \tag{30}$$

such that the income elasticity of equity prices equals the *Arrow-Pratt measure of relative risk aversion* (where $q' = \partial q/\partial y$). Agents attempting to transfer part of their income into the future must therefore hold securities since the consumption good is perishable. Hence, due to the higher demand for equities, share prices must increase (see Lucas (1978), p. 1439).

2.4.4 Interest Rates, Risk Corrections and the Risk Premium

In what follows we will derive explicit functions for the *equity risk premium*. For that purpose, we will reformulate equation (26) as

$$U'(y_t) \cdot q_t = \mathbf{E}_t\Big[\beta \cdot U'(y_{t+1}) \cdot (y_{t+1} + q_{t+1})\Big] \tag{31}$$

where $U'(y_t) \cdot q_t$ denotes the *loss in utility* if an agent subsitutes share purchases for consumption while the *rhs* equals the *discounted utility gain* obtained from investing in period t and subsequent consumption in period $t+1$. Above equation is equivalent to

$$1 = \mathbf{E}_t\Big[\underbrace{\beta \cdot \frac{U'(y_{t+1})}{U'(y_t)}}_{\equiv m_{t+1}} \cdot \underbrace{\frac{y_{t+1} + q_{t+1}}{q_t}}_{\equiv R_{t+1}}\Big] = \mathbf{E}_t[m_{t+1} \cdot R_{t+1}] \tag{32}$$

where m_{t+1} is a *stochastic discount factor*[9] and R_{t+1} is the real *return on equity.* The pricing kernel clearly reflects the time preferences of households as well as the desire to have smooth consumption paths. Equation (32) is *the* fundamental asset pricing equation since it is the starting point for a couple of important conclusions regarding interest rates, risk corrections and equity premia (see Cochrane (2005), p. 14ff.). In order to derive a premium we have to relate equity returns to a *reference return* such as the return of *risk-free bonds.* In the present case we assume that these risk-free bonds yield a payoff of 1 with certainty at the end of a period. Hence, equation (32) reduces to

$$p_t^b = \mathbf{E}_t[m_{t+1} \cdot 1]$$

where p_t^b denotes the bond price in period t and 1 is a certain payoff in $t+1$. The *risk-free return* can then be expressed as

$$R_t^{rf} = \frac{1}{p_t^b} = \frac{1}{\mathbf{E}_t[m_{t+1}]}. \tag{33}$$

Since a risk-free asset is not traded in our model economy R^{rf} is also referred to as a *"shadow" risk-free rate* (see Cochrane (2005), p. 20). To establish a relationship between the risk-free rate and other model parameters we will assume the following: The utility function U is represented by a *power utility function* with the constant relative risk aversion (CRRA) property of the form

$$U(y_t) = \begin{cases} \frac{y_t^{1-\gamma}}{1-\gamma} & \text{if } \gamma \neq 1 \\ \ln(y_t) & \text{if } \gamma = 1 \end{cases}$$

where γ denotes the *Arrow-Pratt measure of relative risk aversion.* The latter can be shown by plugging our specification of the utility function into the *rhs* of (30). Above utility function is the function of choice in a wide spectrum of areas including *growth theory* and *real business cycles* since it is *scale-invariant* and allows for a representative agent. However, an important implication of such preferences is that agents smoothing consumption across states also do so across time. There is no economic intuition as to why this should be the case (see Mehra (2008), p. 14). Hence, if there is no uncertainty about future consumption y_{t+1}, we can express the risk-free interest rate as

[9] m_{t+1} is also often referred to as a *pricing kernel* or the *intertemporal marginal rate of substitution* of consumption (MRS).

$$R_t^{rf} = \frac{1}{\beta} \cdot \left(\frac{y_{t+1}}{y_t}\right)^{\gamma} \tag{34}$$

implying the following. Interest rates are high when $1/\beta$ is high, *i.e.* when β is low. Thus, interest rates are positively correlated with high impatience. Higher savings can only be effected when rates are large enough to convince agents to save more. Furthermore, interest rates are high when y_{t+1}/y_t is high, *i.e.* when consumption growth is high. High interest rates attract increased savings today and thus shift consumption from today into the future. Hence, consumption growth rates are high. A large risk aversion coefficient γ implies that consumers prefer a smooth consumption stream over time. Hence, interest rates must be high to attract increased savings when the curvature of the utility function is large. This can easily be seen from the first derivative of R^{rf} with respect to consumption growth which is

$$\frac{\partial R_t^{rf}}{\partial\left(\frac{y_{t+1}}{y_t}\right)} = \frac{\gamma}{\beta} \cdot \left(\frac{y_{t+1}}{y_t}\right)^{\gamma-1}.$$

To account for the fact that there is uncertainty about future consumption (as is the case for the Lucas model) one has to make the following adaptations: Suppose that $z_{t+1} \equiv y_{t+1}/y_t$ is *log-normally distributed* such that $\ln(z_{t+1})$ is *normally distributed*. Thus,

$$\Delta \ln y_{t+1} = \ln y_{t+1} - \ln y_t \sim \mathcal{N}(g_z, \sigma_z^2)$$

and

$$\mathbf{E}_t[z_{t+1}] = e^{g_z + 0.5 \cdot \sigma_z^2} \Leftrightarrow \ln \mathbf{E}_t[z_{t+1}] = g_z + 0.5 \cdot \sigma_z^2.$$

Let $r_t^{l,rf} \equiv \ln R_t^{rf}$ and $\beta \equiv e^{-\delta}$. Analogous to equation (34) we then get

$$R_t^{rf} = \mathbf{E}_t\left[\frac{1}{\beta} \cdot \left(\frac{y_{t+1}}{y_t}\right)^{\gamma}\right] = \mathbf{E}_t\left[e^{-\delta} \cdot e^{-\gamma \cdot \Delta \ln y_{t+1}}\right]^{-1} = \left[e^{-\delta} \cdot e^{-\gamma \cdot \mathbf{E}_t[\Delta \ln y_{t+1}] + 0.5 \cdot \gamma^2 \cdot \sigma^2(\Delta \ln y_{t+1})}\right]^{-1}.$$

Hence, by taking logarithms we obtain

$$r_t^{l,rf} = \delta + \gamma \cdot g_z - 0.5 \cdot \gamma^2 \cdot \sigma_z^2.$$

We can basically draw the same conclusions as in the prior case with uncertainty. When δ is large (which is equivalent to β being low and hence high impatience) interest rates are high. Interest rates are also high when expected consumption growth g_z is high. Furthermore, when γ is large (strong curvature of the utility function) interest rates must be large to convince agents to increase savings. The last term captures what is known as *precautionary saving*. When consumption is subject to high volatility agents want to insure themselves against large fluctuations in consumption by saving more. Hence, interest rates will be lower. In a final step we will determine return premia based on the notion of *risk corrections*. Let $x_{t+1} \equiv y_{t+1} + q_{t+1}$ denote the payoff of an asset in period $t+1$. Equation (32) can then be written as

$$\begin{aligned} q_t &= \mathbf{E}_t[m_{t+1}] \cdot \mathbf{E}_t[x_{t+1}] + \mathrm{cov}_t[m_{t+1}, x_{t+1}] \\ &= \frac{\mathbf{E}_t[x_{t+1}]}{R_t^{rf}} + \mathrm{cov}_t[m_{t+1}, x_{t+1}] \\ &= \frac{\mathbf{E}_t[x_{t+1}]}{R_t^{rf}} + \frac{\mathrm{cov}_t[\beta \cdot U'(y_{t+1}), x_{t+1}]}{U'(y_t)} \\ &. \end{aligned} \tag{35}$$

The first term equals the present value of payoff x_{t+1} in a *risk-neutral world* (no uncertainty about consumption or risk-neutral (linear) utility). The second term captures a risk premium that is due to the covariation between the asset's payoff and the stochastic discount factor (see Cochrane (2005), p. 23). The implications of (35) are more than obvious. The larger the covariance between an asset's payoff x_{t+1} and the stochastic discount factor m_{t+1} the higher are asset prices. Since $U'(y)$ is large when y is small one can conclude that a negative correlation between the payoff of an asset and consumption leads to higher prices. The reasoning behind this is intuitive. As agents want to have smooth consumption paths they will avoid those assets that yield high payoffs during good times and low payoffs in recessionary periods. Hence, for such assets to be attractive, prices must be comparatively low. Since we are interested in returns we can rearrange above equation to state

$$\begin{aligned} 1 &= \mathbf{E}_t[m_{t+1} \cdot R_{t+1}] \\ &= \frac{\mathbf{E}_t[R_{t+1}]}{R_t^{rf}} + \mathrm{cov}_t[m_{t+1}, R_{t+1}] \end{aligned} \tag{36}$$

which in turn is equivalent to

$$\begin{aligned} \mathbf{E}_t[R_{t+1}] - R_t^{rf} &= -R_t^{rf} \cdot \mathrm{cov}_t[m_{t+1}, R_{t+1}] \\ &= -\frac{\sigma_t(m_{t+1}) \cdot \sigma_t(R_{t+1})}{\mathbf{E}_t(m_{t+1})} \cdot \mathrm{corr}_t(m_{t+1}, R_{t+1}). \end{aligned} \tag{37}$$

Since $m_{t+1} = \beta \cdot \frac{U'(y_{t+1})}{U'(y_t)}$ we get

$$\mathbf{E}_t[R_{t+1}] - R_t^{rf} = -\frac{\sigma_t(m_{t+1}) \cdot \sigma_t(R_{t+1})}{\frac{\beta}{U'(y_t)} \cdot \mathbf{E}_t[U'(y_{t+1})]} \cdot \frac{\frac{\beta}{U'(y_t)} \cdot \text{cov}_t[U'(y_{t+1}), R_{t+1}]}{\sigma_t(m_{t+1}) \cdot \sigma_t(R_{t+1})}$$

$$= -\frac{\text{cov}_t[U'(y_{t+1}), R_{t+1}]}{\mathbf{E}_t[U'(y_{t+1})]} \tag{38}$$

Equation (38) relates the excess return on a risky asset with reference to the risk-free return to the covariance of the risky return with the marginal utility of consumption in $t + 1$. The *rhs* is therefore the *equity premium.* The premium is obviously high when the *rhs* of (38) is larger than 0 which is the case when future returns R_{t+1} and marginal utility $U'(y_{t+1})$ are negatively correlated. Hence, an asset is considered risky when low asset returns coincide with low consumption and income levels. Such an asset will only be held when its return is relatively high. Altug and Labadie (2008) term those assets *ideal* that yield high returns during recessionary periods when consumption is low (and hence the covariance between returns and the marginal utility of $t + 1$ consumption is positive). The resulting negative equity premium can then be viewed as some sort of insurance premium against negative income shocks. Finally, one important aspect to note is the following: Only such risk is rewarded that originates from the covariance between returns and consumption. *Idiosyncratic risk* that is specific for each individual asset is not rewarded even if it is large, *i.e.* when an asset's payoff is subject to large variability, but is uncorrelated with the stochastic discount factor. The decomposition of a payoff into its idiosyncratic (uncorrelated with m) and *systematic* part (correlated with m) can be done by means of a projection of x on m. The price of the resulting residual component (idiosyncratic part) must then be 0 (see Cochrane (2005), p. 25).

2.5 Summary

In the previous chapter a basic introduction into *general equilibrium asset pricing* was provided. The first section introduces the most basic notion of an asset pricing framework in the form of a *contingent claim economy* in which agents may trade contracts that yield a certain payoff conditional on a certain state occuring. Using simple Lagrange optimization methods one can show the existence of a price vector that satisfies market clearing as well as utility maximization for individual agents. In the following part we extended this model to an *asset economy* in which agents trade in physical commodity markets and financial markets reflecting an overall more realistic framework. Section 2.3 then provides a short introduction to a so called *representative agent.* Representative agent models are a powerful tool to aggregate a homogeneous population of economic agents into one single agent. One of the most important representative agent models in the field of financial economics is the consumption-based approach developed by Lucas (1978) and Breeden (1979). The *Lucas* *tree model* provides an elementary framework to investigate the behavior of asset prices in a simple *pure endowment economy* in which production is subject to exogeneous random shocks. By means of a *dynamic programming* approach one can prove the existence of a *recursive competitive equilibrium* which is simply a pair of functions, *i.e.* a value function v and an asset pricing function q. Starting from the *fundamental asset pricing equation* one can derive important implications regarding the economics of *interest rates, risk corrections and the equity premium.*

3 The Equity Premium Puzzle

3.1 Theory of the Equity Premium Puzzle

Prescott and Mehra (1985) based their seminal paper on the contribution of Lucas (1978) and Breeden (1979) to come up with what is widely known as the *equity premium puzzle.* The equity premium refers to the return of risky assets earned in excess of a risk-free reference rate. In the following chapter we will provide a theoretical evaluation of the equity premium implied by an *adapted Lucas tree economy.* Furthermore, an additional view on the puzzle will be provided based on the paper of Hansen and Jagannathan (1991). The major point of interest in Prescott and Mehra (1985) is whether the large return spread between equities and default-free debt can be accounted for by standard economic models that abstract from *market frictions* such as transaction costs, liquidity constraints, taxes and regulation and the like. Their initial guess was *"that most likely some equilibrium model with a friction will be the one that successfully accounts for the large average equity premium"* (see Prescott and Mehra (1985), p. 146).

3.1.1 Model and Assumptions

The basic setup is very similar to the one described in section 2.4.1. To account for the fact that per capita consumption has grown over time Prescott and Mehra (1985) model the *growth rate* of consumption as a *Markov process* in contrast to the Lucas tree model where the level of consumption follows a Markov process. The only consumer in the model is a representative agent with preferences of the form

$$\mathbf{E}_0 \left\{ \sum_{t=0}^{\infty} \beta^t \cdot U(y_t) \right\}, \quad 0 < \beta < 1,$$

where $\mathbf{E}_0$ is an expectation operator, β denotes the time discount factor, y_t refers to consumption in period t and $U : \mathbb{R}_+ \to \mathbb{R}$ is a concave utility function satisfying the Inada conditions with the constant relative risk aversion (CRRA) property. It takes the form

$$U(y_t) = \begin{cases} \frac{y_t^{1-\gamma}}{1-\gamma} & \text{if } \gamma \neq 1 \\ \ln(y_t) & \text{if } \gamma = 1 \end{cases} \tag{39}$$

where γ denotes the *Arrow-Pratt measure of relative risk aversion* measuring the curvature of the function (see page 27). The advantages and disadvantages of such preferences have already been discussed in section 2.4.4. Furthermore, we consider one single production unit producing a single, perishable good. Hence, there is only one equity share outstanding representing a claim on the output process $\{y\}_{t=0}^{\infty}$ whose return equals the market return trivially. The growth rate of production is assumed to follow a Markov process since we try to capture the growth pattern of consumption observed empirically (see above). A comprehensive treatment of this altered modeling framework can be found in Mehra (1988). Under the assumption of nonstationary consumption we thus have

$$y_{t+1} = y_t \cdot z_{t+1}$$

where z_{t+1} denotes the growth rate of production between period t and $t+1$ and $z_{t+1} \in \{g_1, \ldots, g_n\}$, *i.e.* the growth rate can take on finitely many values. Additionally, let

$$\phi(g_i, g_j) \equiv \text{Prob}(z_{t+1} = g_j | z_t = g_i)$$

with transition matrix Π and $g_i > 0, \ \forall i$[10] as well as $y_0 > 0$, *i.e.* initial production is positive. Furthermore, the Markov chain is assumed to be ergodic, *i.e.* it is possible to move from an arbitrary state to any other state. An ergodic chain is also recurrent, aperiodic and irreducible. In simple terms recurrence means that once the state has been j the system will return to this state at some future point with certainty. Aperiodicity implies that any state j may occur irregularly while irreducibility simply says that any state j may be accessed from any other state. Security prices are quoted *ex dividend* as before. Furthermore, the authors assume that the matrix A with entries

$$a_{i,j} \equiv \beta \cdot \phi_{i,j} \cdot g_j^{1-\gamma}, \ \ \forall i, j = 1, \ldots, n \tag{40}$$

is stable, *i.e.* $\lim A^m = 0$ as $m \to \infty$. This is a necessary condition for expected utility to exist when the representative household consumes y_t in every period (see Prescott and Mehra (1985), p. 151). In what follows we will derive a basic set of return and price functions that allow us to derive a so called admissible region for the equity premium later on. Let $\{d\}_{t=0}^{\infty}$ denote the dividend process. Then an arbitrary stream of dividends attains a price

$$q_t = \mathbf{E}_t \left[\sum_{s=t+1}^{\infty} \beta^{s-t} \cdot U'(y_s)/U'(y_t) \cdot d_s \right].$$

From $d_t = y_t$ in equilibrium and power utility we get

$$q_t = q(y_t, z_t) = \mathbf{E} \left[\sum_{s=t+1}^{\infty} \beta^{s-t} \cdot \frac{y_t^{\gamma}}{y_s^{\gamma}} \cdot y_s | z_t, y_t \right]. \tag{41}$$

[10] $g_i > 0$ does not imply that growth is always positive. A value of $g < 1$ would, in fact, imply negative growth.

The state variables z_t and y_t are sufficient to determine the future evolution of the economy. The state of the economy can then be represented by a tuple (c, i) where $c = y_t$ (equilibrium condition) and i represents growth rate g_i. Since $y_s = y_t \cdot z_{t+1} \cdot \ldots \cdot z_s$ one can easily see that prices are homogeneous of degree one in y_t. This can be verified by multiplying y_s and y_t by an arbitrary factor. Hence, we can express equity prices as

$$q(c, i) = \beta \cdot \sum_{j=1}^{n} \phi_{i,j} \cdot \underbrace{\frac{c^\gamma}{(c \cdot g_j)^\gamma}}_{= \frac{U'(y_{t+1})}{U'(y_t)}} \cdot \underbrace{\Big[q(c \cdot g_j, j) + c \cdot g_j\Big]}_{= q_{t+1} + y_{t+1}}, \tag{42}$$

i.e. equity prices can be expressed as a discounted average of the marginal rate of substitution $c^\gamma/(c \cdot g_j)^\gamma$ multiplied by the asset's payoff $q(c \cdot g_j, j) + c \cdot g_j$. This is just an alternative representation of the fundamental asset pricing equation (32) under the assumption of a power utility function.

The homogeneity property implies

$$q(c, i) = w_i \cdot c \tag{43}$$

with w_i being a constant. Substituting (43) into (42) yields

$$\begin{aligned} w_i &= \beta \cdot \sum_{j=1}^{n} \phi_{i,j} \cdot (c \cdot g_j)^{-\gamma} \cdot [q(c \cdot g_j, j) + c \cdot g_j] \cdot c^\gamma \cdot c^{-1} \\ &= \beta \cdot \sum_{j=1}^{n} \left[\phi_{i,j} \cdot g_j^{1-\gamma} \cdot \underbrace{\frac{q(c, j)}{c}}_{= w_j} + \phi_{i,j} \cdot g_j^{1-\gamma} \right] \\ &= \beta \cdot \sum_{j=1}^{n} \phi_{i,j} \cdot g_j^{1-\gamma} \cdot (w_j + 1) \end{aligned}$$

for all $i = 1, \ldots, n$. This system of n equations has a unique positive solution. Given that (c, i) denotes the present state and $(c \cdot g_j, j)$ the next period state, one-period returns can be written as

$$\begin{aligned} R_{i,j} &= \frac{q(c \cdot g_j, j) + c \cdot g_j}{q(c, i)} \\ &= \frac{c \cdot g_j \cdot w_j + c \cdot g_j}{w_i \cdot c} \\ &= \frac{g_j \cdot (w_j + 1)}{w_i}. \end{aligned} \tag{44}$$

The first line is simply an alternative representation of R_{t+1} in (32). In the second line the homogeneity property of the price function is applied. Hence, the expected one-period return for a given state i is then

$$R_i = \sum_{j=1}^{n} \phi_{i,j} \cdot R_{i,j}. \tag{45}$$

Analogously, starting from equation (41) one obtains for the bond price

$$\begin{aligned} p_i^b &= p^b(c,i) \\ &= \beta \cdot \sum_{j=1}^{n} \phi_{i,j} \cdot U'(g_j \cdot c)/U'(c) \\ &= \beta \cdot \sum_{j=1}^{n} \phi_{i,j} \cdot g_j^{-\gamma} \end{aligned} \tag{46}$$

which translates into a period return of

$$R_i^{rf} = \frac{1}{p_i^b} \tag{47}$$

given that the current state is (c, i). Since the stochastic process of the consumption growth rate has been assumed to be ergodic one can find a unique vector v of stationary probabilities such that v is a solution of

$$v = v \cdot \Pi$$

where Π denotes the transition matrix of the Markov process and $\sum_{i=1}^{n} v_i = 1$. The vector v represents the probabilities of being in an arbitrary state i. Therefore, the expected one-period equity return and the expected risk-free rate are

$$\mathbf{E}_t[R_{t+1}] = \sum_{i=1}^{n} v_i \cdot R_i \text{ and } \mathbf{E}_t[R_{t+1}^{rf}] = \sum_{i=1}^{n} v_i \cdot R_i^{rf}. \tag{48}$$

3.1.2 Testing and Results

In their seminal paper Prescott and Mehra (1985) derive a so called *admissible region* for the equity premium in order to demonstrate the large deviation between theoretically implied premia and those empirically observed. For that purpose a simple model economy is constructed. The Markov chain of the consumption growth process is defined as follows. The growth rate is restricted to the two values

$$g_1 = 1 + g_z + \sigma_z, \quad g_2 = 1 + g_z - \sigma_z$$

and the transition matrix equals

$$\Pi = \begin{pmatrix} \phi & 1-\phi \\ 1-\phi & \phi \end{pmatrix}.$$

The parameter values for g_z, σ_z and ϕ were chosen to reflect the sample values of the US economy for the period between 1889 and 1978 which are $g_z = 0.018$, $\sigma_z = 0.036$ and a first-order serial correlation of -0.14 for g_z which translates into a value of $\phi = 0.43$. Given this set of values one then calibrates the preference parameters γ and β to replicate the empirically observed averaged risk-free rate and the equity risk premium. A multitude of papers has addressed the issue of deriving plausible values for the risk aversion coefficient γ in a variety of contexts with slightly differing results. While Kydland and Prescott (1982) and Hildreth and Knowles (1982) estimate the parameter to be between 1 and 2, Dolde and Tobin (1971) derives similar results of about 1.5 as does Friend and Blume (1975) with a coefficient of about 2. The results of Arrow (1971) and Altug (1989) differ with values of 1 and *"near zero"* respectively (see Prescott and Mehra (1985), p. 154). Henceforth, Mehra and Prescott constrain the value of γ *a priori* to be less than 10 since otherwise any risk-free rate and equity premium could be justified by slightly adjusting the parameters of the consumption process. Deriving admissible regions for the equity premium and the risk-free rate is a straightforward algorithm (see Prescott and Mehra (1985), p. 159). Using equations (44) - (48) one can derive the one-period expected return $\mathbf{E}_t[r_{t+1}]$ and the one-period risk-free rate $\mathbf{E}_t[r_{t+1}^{rf}]$ (and hence the equity premium) by choosing the risk aversion coefficient γ and the indiviual's discount factor β from the set

$$S \equiv \{(\gamma, \beta) : 0 < \gamma \leq 10,\ \ 0 < \beta < 1\}.$$

The set S reflects our *a priori upper bound* for γ in the region of 10 (see above) and the usual impatience assumption about preferences which are reflected by a β of less than 1. Additionally, γ and β must be chosen such that the existence condition that we imposed earlier is satisfied (see equation (40) on page 31). The resulting admissible region is depicted in figure 1. The interval for the risk-free rate that is supported by the underlying economic model is between 0 and 4 % while the average equity premium ranges between 0 and 0.35 %. These figures are surely a great way off what has been observed empirically in the period between 1889 and 1978. While the empirical equity premium during that time equals - on average - 6.98 % with a standard deviation of 1.76 % the observed risk-free is only a mere 0.80 %. Hence, the empirical equity premium is almost 18 times as large as the one implied by the model if one considers the empirical average and the maximal premium implied by the model. To verify these results Prescott and Mehra (1985) perform a variety of parameter adjustments. Varying the period length of the model to $n = 2, 1/2, 1/4, 1/8, 1/16, 1/64, 1/128$ and $g_z = 0.018/n$, $\sigma_z = 0.036/\sqrt{n}$ accordingly to match the annual values of $g_z = 0.018$ and $\sigma_z = 0.036$ results in negligible changes in the range of hundreths of percentages[11]. Simi-

[11] One should note that the variance of a sum of random variables only equals the sum of the variances when the random variables are independent. In the present case this implies that $\phi = 0.5$.

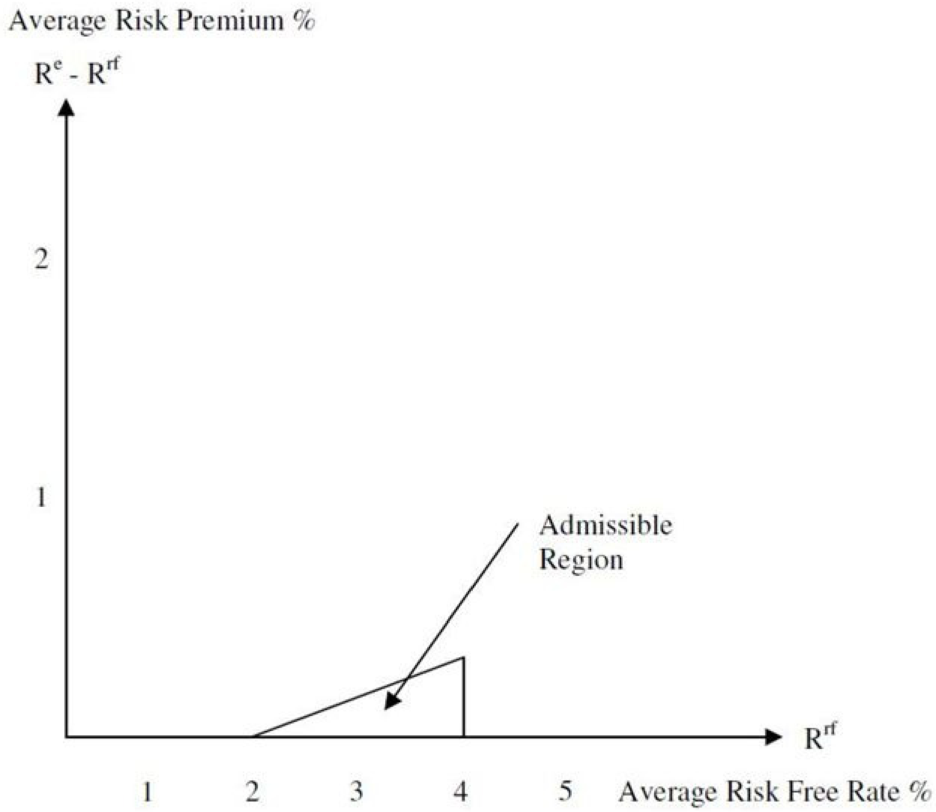

Figure 1: Admissible Region of Equity Premium and Risk-Free Rate (Cochrane (2008))

larly, varying g_z to a bunch of values between 0.014 and 0.022 with $\phi = 0.43$ and $\sigma_z = 0.036$ results only in miniscule changes as well. For different values of σ_z between 0.21 and 0.51 the premium varied with the square of σ_z, *i.e.* the premium for $\sigma_z = 0.51$ was only 0.09 percentage points larger than for $\sigma_z = 0.21$. Finally, when changing ϕ between 0.005 and 0.95 with the other two parameters fixed the premium declined as ϕ increased gradually. Finally, to test for sensitivity due to higher moments (*i.e.* beyond mean and variance) the authors used a Markov chain with transition matrix

$$\Pi = \begin{pmatrix} \phi/2 & \phi/2 & (1-\phi)/2 & (1-\phi)/2 \\ \phi/2 & \phi/2 & (1-\phi)/2 & (1-\phi)/2 \\ (1-\phi)/2 & (1-\phi)/2 & \phi/2 & \phi/2 \\ (1-\phi)/2 & (1-\phi)/2 & \phi/2 & \phi/2, \end{pmatrix}$$

growth parameters $g_1 = g_3 = 1 + g_z$, $g_2 = 1 + g_z + \sigma_z$, $g_4 = 1 + g_z - \sigma_z$ and parameters $g_z = 0.018$, $\sigma_z = 0.051$ and $\phi = 0.36$ to replicate the corresponding properties of the empirical time series. Again, however, this changes the maximal average premium only up to 0.39 % which is still far off the long-run average (see Prescott and Mehra (1985), p. 159–160).

Finally, the authors identify firm leverage as a possible source of bias. In the analysis provided above and in the introductory section the security was assumed to be a primitive claim on the output process by a representative investor. In reality such securities have varying characteristics including significantly different risk-return profiles. Typical securities traded in real financial markets entitle its owner to receive part of a *profit residual*. The latter is simply "what is left" after other stakeholders such as debt owners, workers and the like have been served. It is therefore claimed that *"a disproportionate part of the uncertainty in output is probably borne by equity owners"* (see Prescott and Mehra (1985), p. 157). Hence, an adapted version of the model is suggested as follows. It is assmued that a certain share of the total output in $t+1$, say ϵ, is reserved for other stakeholders in period t. Then equation (42) becomes

$$q(c,i) = \beta \cdot \sum_{j=1}^{n} \frac{\phi_{i,j}}{(c \cdot g_j)^{\gamma}} \cdot \Big[q(c \cdot g_j, j) + c \cdot g_j - \epsilon \cdot \sum_{k=1}^{n} \phi_{i,k} \cdot c \cdot g_k \Big] \cdot c^{\gamma}$$

and hence, by the homogeneity property of prices, we obtain

$$w_i = \beta \cdot \sum_{j=1}^{n} \phi_{i,j} \cdot g_j^{-\gamma} \cdot \Big[g_j \cdot w_j + g_j - \epsilon \cdot \sum_{k=1}^{n} \phi_{i,k} \cdot g_k \Big], \;\; \forall i = 1, \ldots, n.$$

From here one can then obtain the one-period expected return and the risk-free rate. In their original contribution Mehra and Prescott used a coefficient $\epsilon = 0.9$ which implies that 10% of corporate profits go to shareholders on average while the remainder is distributed among other stakeholders. The effect of this alternative modeling framework was again miniscule with an equity premium increased by about one-tenth of one percent compared with the prior case.

4 An Alternative View on the Puzzle

An alternative perspective on the puzzle is provided by Hansen and Jagannathan (1991). In contrast to the original Mehra and Prescott paper this more recent approach does not make use of Markov chain approximations or the like. Instead they derive a *mean-variance frontier* for the stochastic discount factor m (see page 26) that is related to the mean-variance frontier of asset returns introduced by Markowitz (1952). For a better grasp of the material to follow we will provide a brief introduction to mean-variance analysis in general and the return frontier in particular.

4.1 Mean-Variance Frontier

A typical mean-variance frontier for a portfolio of 2 assets is depicted in figure 2. The abscissa represents the standard deviation of expected portfolio returns while the latter are represented by the ordinate. Hence, the mean-variance frontier answers an intuitive question: What is the expected return on an investment for a given level of risk or vice versa? In that regard one has to distinguish between two notions. The mean-variance frontier of risky assets is represented by the hyperbola connecting both assets. In the presence of a risk-free security the frontier equals the solid-line wedge-shaped region in figure 2. The *minimum variance portfolio* is that portfolio comprised of risky assets with the lowest possible risk. The tangential portfolio marks that point where the straight line originating at the risk-free rate is tangent to the mean-variance frontier of risky assets. The tangential portfolio provides the largest risk-return trade-off among all efficient portfolios, *i.e.* it possesses the highest *Sharpe ratio* which is defined as

$$SR_{t+1} = \frac{\mathbf{E}_t[R^e_{t+1}]}{\sigma_t(R_{t+1})}$$

where $\mathbf{E}_t[R^e_{t+1}] = \mathbf{E}_t[R_{t+1}] - R^{rf}_t$ denotes the excess return of risky assets over the risk-free rate. The Sharpe ratio measures the excess return of a portfolio relative to its risk and therefore provides an intuitive way to measure portfolio returns against one another. Finally, the dashed wedge-shaped region represents the mean-variance frontier when both assets are perfectly positively correlated ($\text{corr}_{1,2} = -1$). The straight line connecting them is simply the reverse case of perfect positive correlation ($\text{corr}_{1,2} = +1$) in which case no diversification benefits can be taken advantage of. In the latter case the portfolio risk is always a linear combination of the risks associated with both assets individually. The derivation of the mean-variance frontier is a straightforward optimization problem with constraints. A brief derivation of the closed-form solutions can be found in Cochrane (2005) (see pp. 81). Let us consider the fundamental pricing equation (32) again and rewrite it as

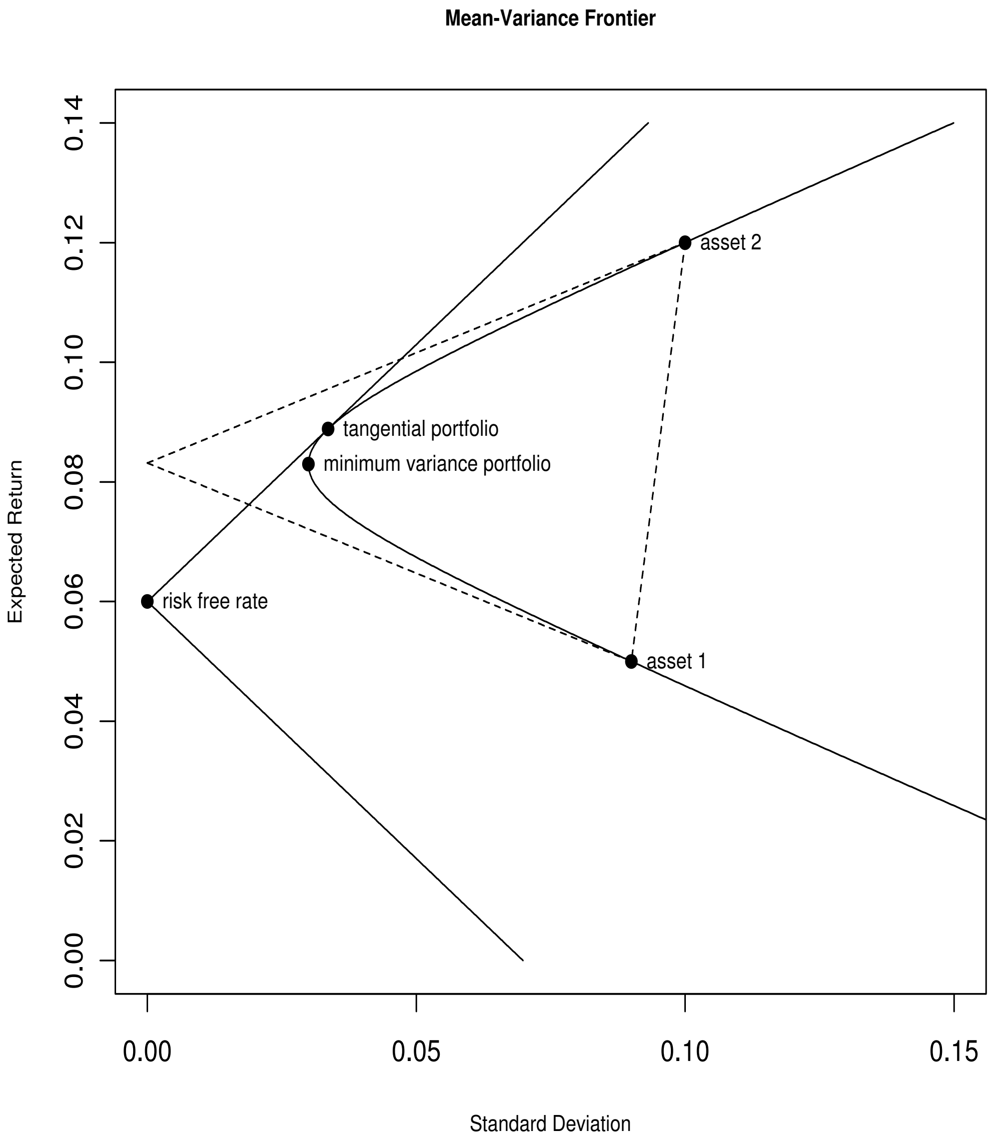

Figure 2: Mean-Variance Frontier with 2 Assets

$$\begin{aligned}
1 &= \mathbf{E}_t[m_{t+1} \cdot R_{t+1}] \\
&= \mathbf{E}_t[m_{t+1}]\mathbf{E}_t[R_{t+1}] + \underbrace{\mathrm{corr}_t(m_{t+1}, R_{t+1}) \cdot \sigma_t(m_{t+1}) \cdot \sigma_t(R_{t+1})}_{=\mathrm{cov}_t[m_{t+1}, R_{t+1}]}.
\end{aligned}$$

Dividing by $\mathbf{E}_t[m_{t+1}]$ then yields

$$\mathbf{E}_t[R_{t+1}] = R_t^{rf} - \mathrm{corr}_t(m_{t+1}, R_{t+1}) \cdot \frac{\sigma_t(m_{t+1})}{\mathbf{E}_t[m_{t+1}]} \cdot \sigma_t(R_{t+1}). \tag{49}$$

Since any correlation coefficient fulfills corr $\in [-1, 1]$ we obtain

$$-\frac{\sigma_t(m_{t+1})}{\mathbf{E}_t[m_{t+1}]} \cdot \sigma_t(R_{t+1}) \leq \mathbf{E}_t[R_{t+1}] - R_t^{rf} \leq \frac{\sigma_t(m_{t+1})}{\mathbf{E}_t[m_{t+1}]} \cdot \sigma_t(R_{t+1}).$$

Let us have a closer look at equation (49) first. Letting $\mathrm{corr}_t(m_{t+1}, R_{t+1}) = 1$ or $\mathrm{corr}_t(m_{t+1}, R_{t+1}) = -1$ one obtains a linear function with an intercept equal to R_t^{rf} and slopes $-\frac{\sigma_t(m_{t+1})}{\mathbf{E}_t[m_{t+1}]}$ and $\frac{\sigma_t(m_{t+1})}{\mathbf{E}_t[m_{t+1}]}$ respectively. Obviously, equation (49) is represented by the solid-line wedge-shaped region in figure 2 originating at the risk-free rate. Hence, any risk-return tuple must lie between these two straight lines. Furthermore, every return on any of the two lines is perfectly correlated with the discount factor m since we assumed $\mathrm{corr}_t(m_{t+1}, R_{t+1})$ to be equal to either 1 or -1. The returns on the upper part of the frontier having a positive slope are thus perfectly negatively correlated with the stochastic discount factor while those on the lower frontier are perfectly positively correlated with it. A positive and negative correlation with m then implies a negative and positive correlation with consumption respectively. Returns on the lower frontier have the insurance property that we characterized in the previous sections. Those on the upper frontier are negatively correlated with the intertemporal marginal rate of substitution and therefore command a higher return. The argument behind this is intuitive. When the IMRS is large consumption is low, *i.e.* assets which yield low returns in such cases are not favoured by investors. In fact, investors would prefer high yielding assets in times of low consumption in order to consume more.

4.2 Decomposing Returns in the Mean-Variance Space

In what follows our aim is to derive a *mean-variance frontier for the stochastic discount factor* that is directly related to the mean-variance frontier of asset returns. In a first step we will show that asset returns can be decomposed into a sum of 3 orthogonal components. The decomposition was first shown by Hansen and Richard (1987). The proof that will be shown here will be a shortened version based on Cochrane (2005) (see pp. 85). Time subscripts of variables will be dropped for the moment since they do not matter. "*The price always comes at t, the payoff at $t+1$, and the expectation is conditional on time t information*" (see Cochrane (2005), p. 16).

4.2.1 Payoff Space, State Diagram and the Price Function

In the following section we will introduce a variety of preliminaries and concepts to understand an alternative derivation of the mean-variance frontier that will be provided in subsequent chapters. We will briefly discuss the space of payoffs, the free portfolio formation assumption as well as the law of one price. Additionally, some geometric fundamentals and properties of returns and payoffs will be provided.
Let the payoff space be denoted by $\overline{X} \subset \mathbb{R}^S$ where S denotes the number of possible states of nature. It is assumed that *any* payoff can be synthesized by a set of basis payoffs which are elements in the payoff space (see Cochrane (2005), pp. 65). Mathematically, this can be expressed as

$$x_1, x_2 \in \overline{X} \Rightarrow \alpha_1 \cdot x_1 + \alpha_2 \cdot x_2 \in \overline{X}, \;\; \forall \;\; \alpha_1, \alpha_2 \in \mathbb{R},$$

i.e. a payoff generated from basis payoffs is also an element in the payoff space. We refer to this as the free portfolio formation property. The factors α_1 and α_2 can be thought of as weights. In case of an equally weighted two-asset portfolio we would trivially have $\alpha_1 = \alpha_2 = 0.5$. Above definition rules out short-selling- or leverage constraints since $\alpha_1, \alpha_2 \in \mathbb{R}$. The portfolio weights can take on any value on the real line. A further assumption is the so called *law of one price* which says that portfolios with the same payoff must have the same price ruling out *arbitrage* opportunities in equilibrium. Mathematically,

$$q(\alpha_1 \cdot x_1 + \alpha_2 \cdot x_2) = \alpha_1 \cdot q(x_1) + \alpha_2 \cdot q(x_2).$$

Let $p \in \mathbb{R}^S_+$ denote the vector of contingent claims prices. More specifically, we have

$$p = [p_1, \ldots, p_S].$$

Also, let

$$x = [x_1, \ldots, x_S]$$

denote an arbitrary payoff vector where x_s denotes the payoff in state s. The price of some payoff x is given by

$$q(x) = \sum_{s=1}^{S} p_s \cdot x_s = p \cdot x^T = |p| \times |\text{proj}(x|p)| = |p| \times |x| \times \cos\kappa, \tag{50}$$

i.e. the price of a payoff is simply the inner product of the contingent claim price vector and the payoff vector where $|.|$ denotes the vector length, $\text{proj}(x|p)$ denotes the projection of x on p and κ is the angle formed by x and p. We want to derive the intuition behind figure 3.

Let us think of any return as a payoff with unit price. This is easily verified by looking at the central pricing equation (32). Furthermore, let us consider *excess returns.* The price

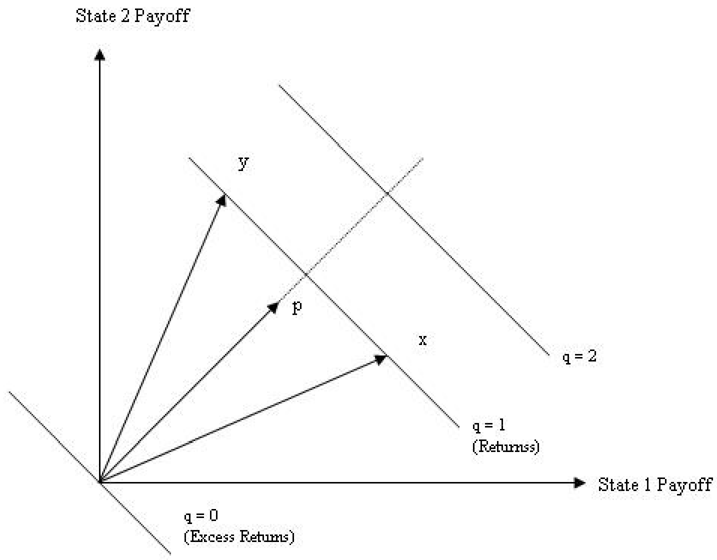

Figure 3: State Price Diagram (Cochrane (2005))

of an excess return equals zero which follows from a trivial argument. Suppose an investor borrows a certain amount of money at a rate R^b and invests it at some rate R^a. His net investment (the price *i.e.*) is equal to zero and his return equals $R^e = R^a - R^b$. In such a bet it is equally likely to gain or lose. Hence, investors do not pay a price $q > 0$ to enter such a lottery. In the case that the borrowing rate R^b equals the risk-free interest rate R^{rf} we obtain the excess return of a risky asset over a risk-free rate.

Consider figure 3. There are two possible states. The ordinate represents state 2 payoffs while the abscissa represents payoffs in the first state. p refers to the contingent claim price vector which points into the positive orthant from the origin. Prices are obtained by means of an inner product as in equation (50). Two vectors are called *orthogonal* when their inner product is equal to zero. Using our argument from above the plane representing the set of excess returns must be orthogonal to p since excess returns have a zero price. The plane of excess returns therefore points out from the origin at right angles to the contingent claim price vector p. Furthermore, payoffs on the same plane must have the same price. This simply follows from the fact that all payoffs on a plane of constant price that is perpendicular to the vector p (*e.g.* the $q = 1$ plane) have the same projection onto p. Additionally, planes of constant price are parallel to each other. Suppose this was not the case. Then one could find payoffs which lie on both planes implying that they have two different prices. This clearly contradicts the law of one price assumption that we imposed earlier to rule out arbitrage opportunities. Finally, the zero payoff trivially has a zero price and planes of constant price move out linearly (*i.e.* $q(x) = 1 \Rightarrow q(2x) = 2q(x) = 2$) which implies the linearity property of the pricing function introduced earlier (see Cochrane (2005), p. 62).

4.2.2 A Payoff x' as a Discount Factor

The free portfolio formation property and the law of one price ensure that we can, in fact, find a special payoff $x' \in \overline{X}$ that can be used as the stochastic discount factor, *i.e.* we are looking for a payoff that represents prices by means of an inner product.

Lemma 4.1 *There is a unique $x' \in \overline{X}$ such that $q(x) = \mathbf{E}[x' \cdot x^T], \ \forall x \in \overline{X}$, i.e. payoff x' is, in fact, a discount factor.*

The proof is straightforward and works as follows:

Proof Let the payoff space be spanned by N basis payoffs, *i.e.* any arbitrary payoff can be constructed from a linear combination of these payoffs. Let

$$X = [x_1, x_2, \dots, x_N]$$

denote a vector containing the N basis payoffs with $x_n \in \mathbb{R}^S, \ n = 1, \dots, N$. Each x_n is actually a row vector of the form

$$x_n = [x_{n,1}, \dots, x_{n,S}]$$

containing S elements (one outcome for each possible state s). Hence, X maybe interpreted as a matrix of dimension $S \times N$. Furthermore, let

$$Q = [q_1, \dots, q_N]$$

denote the price vector of the basis assets $x_1, \dots, x_N$ with $q_n \in \mathbb{R}_+, \ n = 1, \dots, N$. The payoff space can then be expressed as

$$\overline{X} = \{c \cdot X^T\}$$

where $c \in \mathbb{R}^N$ can be thought of as some vector of asset weights. Since we require $x' \in \overline{X}$ we must have $x' = c \cdot X^T$. Then

$$\begin{aligned} Q &= \mathbf{E}[x' \cdot X] \\ Q &= \mathbf{E}[c \cdot X^T \cdot X] \\ Q^T &= \mathbf{E}[X^T \cdot X \cdot c^T] \\ Q^T &= \mathbf{E}[X^T \cdot X] \cdot c^T \\ c^T &= \mathbf{E}[X^T \cdot X]^{-1} \cdot Q^T. \end{aligned}$$

Obvisouly, the random matrix $\mathbf{E}[X^T \cdot X]$ must be invertible to guarantee existence and uniqueness of c. Hence, we get

$$x' = \underbrace{Q \cdot \mathbf{E}[X^T \cdot X]^{-1}}_{\equiv c} \cdot X^T$$

which is a linear combination of the basis payoffs $x_1, \ldots, x_N$ and therefore $x' \in \overline{X}$.

□

Besides the algebraic derivation there also exists a very trivial geometric derivation. Looking at figure 3 one simply has to pick a payoff vector that is orthogonal to the plane of excess returns and perpendicular to the other planes of constant price. In order to price arbitrary payoffs one simply has to choose a vector of the "right length" (see Cochrane (2005), p. 67). Since $q(x) = |x'| \times |x| \times \cos\kappa$ the right length is determined as

$$|x'| = \frac{q(x)}{|x| \times \cos\kappa}.$$

Hence, one could simply replace the vector p in figure 3 by some vector x'. The payoff x' is needed in order to define a special return R' that corresponds to this payoff. Its properties are evaluated in the next section.

4.2.3 An Alternative Derivation of the Mean-Variance Frontier

Traditionally, the derivation of the mean-variance frontier is done via a simple maximization or minimization problem. Closed-form solutions exist for simple cases without any market frictions such as short-sales constraints and the like. The aim of the following section is to introduce an orthogonal decomposition of arbitrary returns and derive the mean-variance frontier from there. We will begin by defining a return R' which corresponds to the discount factor x' that we derived in the previous section. Let

$$R' \equiv \frac{x'}{q(x')} = \frac{x'}{\mathbf{E}[x' \cdot x'^T]}.$$

The first fraction is simply a payoff divided by a price (hence a return). In the second fraction the price q is replaced by $\mathbf{E}[m \cdot x^T]$ with $m = x'$ and $x = x'$. Let the space of excess return be defined as follows:

$$\overline{R^e} \equiv \{x \in \overline{X} : q(x) = 0\}$$

The space of excess returns contains those payoffs with a price equal to zero. Let the projection of a variable y onto a variable x be defined as

$$\text{proj}(y|x) = \beta x^T = \underbrace{\mathbf{E}[x \cdot x^T]^{-1} \cdot \mathbf{E}[y \cdot x^T]}_{=\beta} x^T$$

which is simply a linear regression of variable y on the regressor x without a constant. Let the return $R^{e'}$ be given by

$$R^{e'} \equiv \text{proj}(1|\overline{R^e}) \tag{51}$$

where $\overline{R^e}$ denotes the space of excess returns, *i.e.* $R^{e'}$ can be thought of as some sort of mean excess return. While $x' \in \overline{X}$ represents prices of arbitrary future payoffs with an inner product (see proof 4.2.2), the return $R^{e'} \in \overline{R}^{e}$ represents means in the space of excess returns, *i.e.*

$$\mathbf{E}[R^e] = \mathbf{E}[1 \times R^e] = \mathbf{E}[\text{proj}(1|\overline{R^e}) \times R^e] = \mathbf{E}[R^{e'} \times R^e], \ \ \forall R^e \in \overline{R^e}$$

where $a \times b$ denotes the cross product of vectors a and b. Having established the setup and major definitions we can state the following theorems in order to construct the mean-variance frontier using R' and $R^{e'}$:

Theorem 4.2 *An arbitrary return R_i can be written as $R_i = R' + w_i \cdot R^{e\prime} + n_i$ where $w_i \in \mathbb{R}$ and n_i denotes an excess return with $\mathbf{E}[n_i] = 0$. Also, $\mathbf{E}[R' \cdot R^{e'^T}] = \mathbf{E}[R' \cdot n_i^T] = \mathbf{E}[R^{e\prime} \cdot n_i^T] = 0$, i.e. all components are uncorrelated with each other.*

Theorem 4.3 *R^{mv} is on the mean-variance frontier $\Leftrightarrow R^{mv} = R' + w \cdot R^{e\prime}, \ \ w \in \mathbb{R}$.*

Proof In the following proof we will look at things mainly from an algebraic perspective. For the moment we will stick with the mathematics and later on provide an economic interpretation for each component. For the geometric part we will mostly refer to figures 4 and 5 which is a three-dimensional extension of figure 3. Let 0 denote the origin. As we pointed out before one can use the payoff x' as the discount factor to price arbitrary payoffs by means of an expectations inner product (see section 4.2.2). Additionally, we reasoned why excess returns have a zero price. The lower plain in figure 4 represents the space of excess returns as defined earlier. Obviously, putting together the facts we have

$$0 = \mathbf{E}[x' \cdot R^{e'^T}] = \frac{1}{q(x')} \cdot \mathbf{E}[x' \cdot R^{e'^T}] = \mathbf{E}\left[\underbrace{\frac{x'}{q(x')}}_{=R'} \cdot R^{e'^T}\right] = \mathbf{E}[R' \cdot R^{e'^T}]. \tag{52}$$

Using x' as the discount factor we obtain a zero price for any excess return, hence also for the return $R^{e'}$. Multiplying on both sides with $1/q(x')$ and getting the fraction into the expected value operator we deduce that R' is orthogonal to the vector $R^{e'}$. In fact, R' is orthogonal to any vector in the space of excess returns. The parameter w_i can be interpreted as some kind of control variable for an investor to choose his desired mean return on the frontier. Finally, n_i is defined such that

$$n_i \equiv R_i - R' - w_i \cdot R^{e\prime} \tag{53}$$

and it is an excess return by assumption. One should think of n_i as an idiosyncratic return component that is attributable to the specifics of a company *e.g.* and hence can be diversified away in a portfolio context. Investors are not rewarded for non-systematic risk and therefore $\mathbf{E}[n_i] = 0$ must hold. We can also derive this result mathematically as follows. In equation (52) we showed that R' is orthogonal to *any* excess return.

Thus,

$$\mathbf{E}[R' \cdot n_i^T] = 0$$

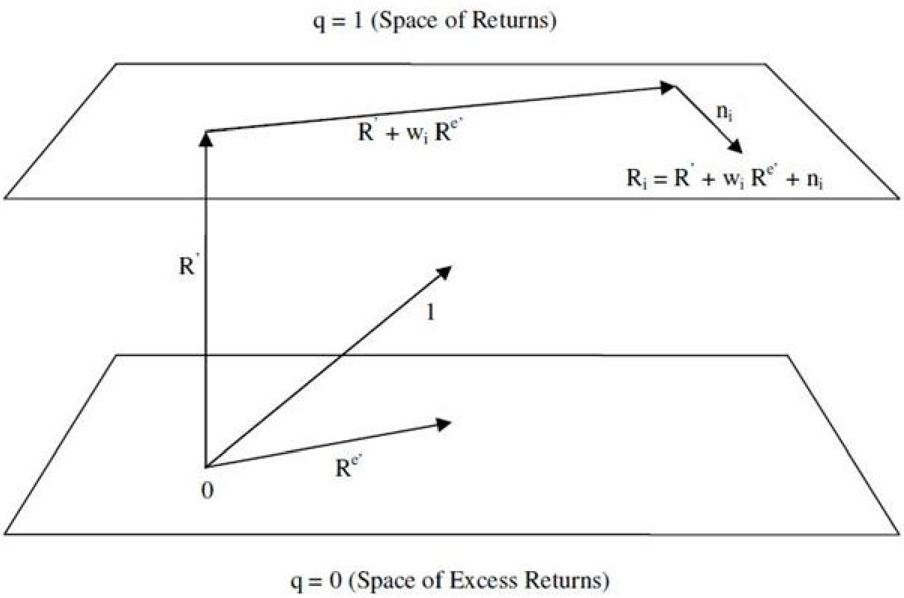

Figure 4: Orthogonal Decomposition in State Space (Cochrane (2005))

must hold trivially since n_i is an excess return. In section 4.2.1 we introduced $R^{e\prime}$ as an excess return that represents means on $\overline{R^e}$. Then

$$\mathbf{E}[n_i] = \mathbf{E}[1 \times n_i] = \mathbf{E}[\text{proj}(1|\overline{R^e}) \times n_i] = \mathbf{E}[R^{e\prime} \times n_i].$$

So, in order to get $\mathbf{E}[n_i] = 0$ we need an n_i that is orthogonal to $R^{e\prime}$. Using

$$w_i = \frac{\mathbf{E}[R_i] - \mathbf{E}[R']}{\mathbf{E}[R^{e\prime}]}$$

we obtain

$$\Leftrightarrow \mathbf{E}[n_i] = \mathbf{E}[R_i] - \mathbf{E}[R'] - w_i \cdot \mathbf{E}[R^{e\prime}]$$

$$\Leftrightarrow \mathbf{E}[n_i] = \mathbf{E}[R_i] - \mathbf{E}[R'] - \frac{\mathbf{E}[R_i] - \mathbf{E}[R']}{\mathbf{E}[R^{e\prime}]} \cdot \mathbf{E}[R^{e\prime}]$$

$$\Leftrightarrow \mathbf{E}[n_i] = 0.$$

The first equality is simply the expected value operator applied to the definition of n_i as in equation (53). Plugging in w_i yields $\mathbf{E}[n_i] = 0$. Finally, having

$$R_i = R' + w_i \cdot R^{e\prime} + n_i \tag{54}$$

and

$$\mathbf{E}[R' \cdot {R^{e'}}^T] = \mathbf{E}[R' \cdot n_i^T] = \mathbf{E}[R^{e\prime} \cdot n_i^T] = 0 \tag{55}$$

delivers the mean-variance frontier with

$$\mathbf{E}[R_i] = \mathbf{E}[R'] + w_i \cdot \mathbf{E}[R^{e\prime}] \tag{56}$$

and

$$\sigma^2(R_i) = \sigma^2(R') + w_i^2 \cdot \sigma^2(R^{e\prime}) + \sigma^2(n_i). \tag{57}$$

Equation (56) follows from (54) simply by applying the expected value operator and the fact that $\mathbf{E}[n_i] = 0$. Since all return components are orthogonal the variance of some return R_i equals equation (57). $\square$

More intuition is provided in figures 4 and 5. Mean-variance analysis is all about minimizing variance for a given level of expected return. In figure 4 we are looking for the shortest way from the origin to some point R_i in the space of returns. One can interpret the length of a return vector as a second moment (variance *i.e.*). Hence, it is natural to seek the shortest way or smallest variance possible to the corresponding return level R_i which is along the lines of $R' + w_i \cdot R^{e'}$. R' is the minimum second-moment return. Firstly, among all returns it is closest to the origin which can be seen easily in figure 4. Secondly, we can verify this property from equation (57). Let $w_i = 0$ and $n_i = 0$ then

$$\begin{aligned}
&R_i = R' + w_i \cdot R^{e\prime} + n_i \\
&\Leftrightarrow \mathbf{E}[R_i \cdot R_i^T] = \mathbf{E}[R' \cdot R'^T] + w_i^2 \cdot \mathbf{E}[R^{e\prime} \cdot R^{e\prime T}] + \mathbf{E}[n_i \cdot n_i^T] \\
&\Leftrightarrow \sigma^2(R_i) = \sigma^2(R') + w_i^2 \cdot \sigma^2(R^{e\prime}) + \sigma^2(n_i) \\
&\Leftrightarrow \sigma^2(R_i) = \sigma^2(R').
\end{aligned}$$

For any other tuple (w_i, n_i) we get $\sigma^2(R_i) > \sigma^2(R')$. In the second step all "covariance terms" drop out since each component is orthogonal to any other component by construction. Hence, all returns of the same "length" originating at 0 have the same variance. Since lines of constant second moment are circles the minimum second-moment return is that vector where the smallest possible circle and the mean-variance frontier intersect in figure 5 at R'. By changing w_i one moves along the frontier. As indicated earlier n_i is an idiosyncratic return component that is not rewarded by the market and hence has zero expected value. As shown in figure 5 n_i only increases variance, but not the expected return. Investors may

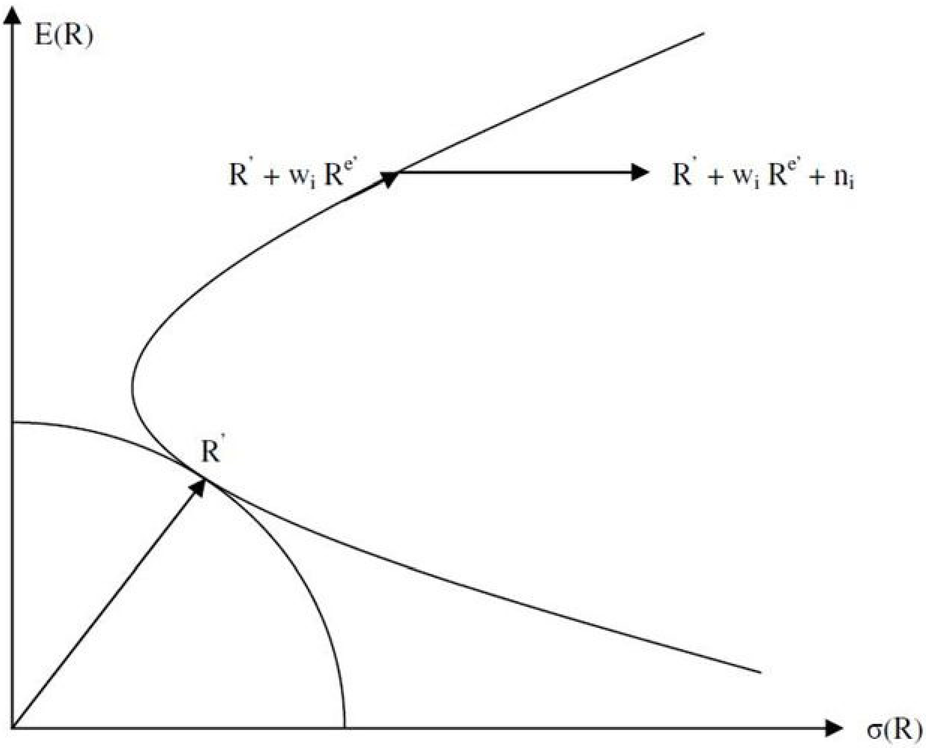

Figure 5: Mean-Variance Frontier (Cochrane (2005))

diversify such risks by altering their asset allocation.

4.3 Hansen-Jagannathan Bounds

Following the previous section on the mean-variance frontier of asset returns we will derive an analogous decomposition and a corresponding mean-variance frontier of the discount factor m as in Hansen and Jagannathan (1991) and Cochrane (2005). In the original contribution the authors use a non-parametric approach for a broader set of dynamic economic models. It is broader in a sense that it *"does not depend either on a Markov chain approximation with a small number of states or on a narrow class of asset valuation models."* (see Hansen and Jagannathan (1991), p. 229). Recapitulate from section 4.1

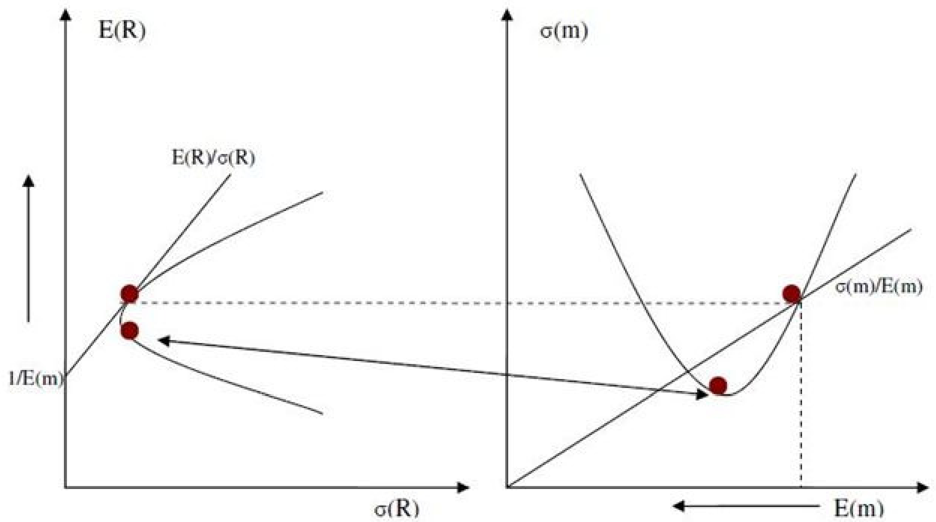

Figure 6: Duality between Mean-Variance Frontier and HJ-Bounds (Cochrane (2005))

$$1 = \mathbf{E}_t[m_{t+1} \cdot R_{t+1}]$$

$$\Leftrightarrow 1 = \mathbf{E}_t[m_{t+1}] \cdot \mathbf{E}_t[R_{t+1}] + \mathrm{corr}_t(m_{t+1}, R_{t+1}) \cdot \sigma_t(m_{t+1}) \cdot \sigma_t(R_{t+1})$$

$$\Leftrightarrow \mathbf{E}_t[R_{t+1}] = R_t^{rf} - \mathrm{corr}_t(m_{t+1}, R_{t+1}) \cdot \frac{\sigma_t(m_{t+1})}{\mathbf{E}_t[m_{t+1}]} \cdot \sigma_t(R_{t+1})$$

$$\Leftrightarrow -\frac{\sigma_t(m_{t+1})}{\mathbf{E}_t[m_{t+1}]} \leq \frac{\mathbf{E}_t[R_{t+1}] - R_t^{rf}}{\sigma_t(R_{t+1})} \leq \frac{\sigma_t(m_{t+1})}{\mathbf{E}_t[m_{t+1}]}$$

$$\Leftrightarrow \frac{\sigma_t(m_{t+1})}{\mathbf{E}_t[m_{t+1}]} \geq \frac{|\mathbf{E}_t[R_{t+1}] - R_t^{rf}|}{\sigma_t(R_{t+1})}.$$

Figure 6 shows the duality between the mean-variance frontier of excess returns and the stochastic discount factor. The latter frontier can be derived intuitively by deriving those pairs $\{\mathbf{E}(m), \sigma(m)\}$ that are consistent with the mean-variance frontier of excess returns. For any given risk-free rate one has to find the tangency portfolio and the corresponding Sharpe ratio. One should note that for any given risk-free rate there are always two tangency portfolios. We select that portfolio that has the higher Sharpe ratio in absolute terms.

The Sharpe ratio of the tangency portfolio can then be equated to $\sigma(m)/\mathbf{E}(m)$ delivering the frontier of the stochastic discount factor (also see Cochrane (2005), p. 93). We look at figure 6 twofold. Let us first consider the mean-variance frontier in the left-hand panel. We first consider those expected returns which are smaller than the expected return that corresponds to minimum-variance portfolio (*i.e.* the vertex of the parabola). Shifting the capital market line (the straight line connecting the risk-free rate and the parabola) upwards its slope and hence the Sharpe ratio decreases. The upward shift corresponds to an increase in the risk-free rate and a decrease in $\mathbf{E}(m)$. This leads us to the mean-variance frontier of the stochastic discount factor in the right-hand panel of figure 6. First we only consider the part on the *rhs* of the vertex. As $\mathbf{E}(m)$ decreases the slope of the capital market line and hence also the Sharpe ratio decreases. From equation (58) it is then obvious that $\sigma(m)$ must decrease as well. In a second step we consider those risk-free rates that are larger then the expected return that corresponds to the minimum-variance portfolio. By decreasing $\mathbf{E}(m)$ the slope of the capital market line becomes more and more negative. However, since we are looking at Sharpe rations on an absolute basis, the Sharpe ratio increases and hence also $\sigma(m)$ must increase. The mechanics work in the same manner as they did before yielding the left-hand part of the parabola of the stochastic discount factor.

$$\min_{\{\text{all } m \text{ that price } x \in \overline{X}\}} \frac{\sigma_t(m_{t+1})}{\mathbf{E}_t[m_{t+1}]} = \max_{\{\text{all excess returns } R^e \in \overline{X}\}} \frac{\mathbf{E}_t[R_{t+1}] - R_t^{rf}}{\sigma_t(R_{t+1})} \tag{58}$$

5 Solving the Puzzle

One can spot two major streams of research addressing the puzzle. The first stream proposes modifications of the utility functions being used since the CRRA class of functions can only be made consistent with the observed equity premium when agents are extremely and unplausibly risk averse. The second stream addresses issues such as borrowing constraints, transaction costs, liquidity, taxes and regulation as well as potential disaster states.

5.1 Habit Formation

The heart of the model is to propose an alternative specification of the agent's utility function. This non-time separable utility function incorporates what is known as *habit formation* in economics. Major works in the field include - among others - Abel (1990), Deaton (1992), Ryder and Heal (1973), Sunderasan (1989) and Constantinides (1990) (see Cochrane (2005), pp. 207). One can, in general, distinguish between *external* and *internal* habit formation. Within those classes there exist models that use additive as well as multiplicative utility functions to model preferences. The habit formation approach used in this paper can be attributed to what is widely known as *"catching up with the Joneses"*. The term was first used in Abel (1990) and is synonimously used for *external habit.* The key feature distinguishing external and internal habit is whether an agent's habit level depends on the agent's own past consumption level (internal) or on aggregate consumption (external).

5.2 Campbell-Cochrane Model

In the following section I will review a model by Campbell and Cochrane (1999) which has an alternative, *non-time separable* utility function at its core to solve the equity premium puzzle. Their model captures a wider variety of stock market phenomena such as the procyclical variation of stock prices, the long-horizon predictability of excess stock returns and the countercyclical variation of stock market volatility among others. In what follows we will focus our attention on the model's capability to explain the short- and long-run equity premium puzzles with a constant risk-free rate.

5.2.1 Utility, Stochastic Discount Factors and the Interest Rate

In their original contribution Prescott and Mehra (1985) assumed a standard power utility function to express preferences. The key feature of the model to follow is a slight adaptation by including a parameter X_t which represents the agent's habit level. One can think of the habit level as some kind of consumption level the agent "is used to" or some reference level that determines consumer satisfaction for subsequent periods. Therefore, in general, an agent always seeks consumption above such a reference level. The model will be presented in a discrete time, representative agent setup. Preferences over consumption take the following form:

$$\mathbf{E}\sum_{t=0}^{\infty}\beta\frac{(C_t - X_t)^{1-\gamma} - 1}{1-\gamma}. \tag{59}$$

One should note that above utility functional is only defined for $C_t \geq X_t$, but not for $C_t < X_t$. Why? Suppose $\gamma = 0.5$. Then $\sqrt{C_t - X_t}$ is not defined for $C_t < X_t$. This implication will be useful later on when we specify the particular form of a so called sensitivity function $\lambda(.)$ that is part of the model. The relation between current consumption and habit can expressed conveniently as

$$S_t \equiv \frac{C_t - X_t}{C_t}.$$

Campbell and Cochrane (1999) refer to above fraction as *surplus consumption ratio*. Obviously, when $C_t = X_t$ the surplus ratio amounts to 0 which implies a bad state. Furthermore, as C_t rises relative to the habit level we get $S_t \to 1$. Additionally, one can relate the local curvature of the utility function as expressed in (59) to the surplus consumption ratio as follows. The first- and second-order derivatives of the utility function with respect to consumption C_t are

$$U'(C_t, X_t) = (C_t - X_t)^{-\gamma} \quad \text{and} \quad U''(C_t, X_t) = -\gamma \cdot (C_t - X_t)^{-\gamma-1}. \tag{60}$$

Then

$$\eta_t \equiv -C_t \cdot \frac{U''(C_t, X_t)}{U'(C_t, X_t)} = -C_t \cdot \frac{-\gamma \cdot (C_t - X_t)^{-\gamma-1}}{(C_t - X_t)^{-\gamma}} = \gamma \cdot \frac{C_t}{C_t - X_t} = \frac{\gamma}{S_t} \tag{61}$$

which implies that the lower the surplus consumption ratio, the higher is the curvature of the utility function. From an economic perspective relative risk aversion (high η) of an agent is high during bad times (low S). Hence, asset prices are low and expected returns are high. As we have already indicated at the beginning of the section Campbell and Cochrane (1999) use an *external habit* approach, *i.e.* an agent's habit is determined by past realizations of aggregate consumption as opposed to individual consumption. Let

$$S_t^a \equiv \frac{C_t^a - X_t}{C_t^a}$$

where the a superscript denotes *average* consumption of all agents in the economy. In order to assess how average consumption C_t^a determines individual habit X_t we assume the following stochastic process for S_t^a. In what follows lowercase letters denote logarithms, *i.e.* $s_t^a = \log S_t^a$. We then model the log surplus consumption ratio as an AR(1) process of the form

$$s_{t+1}^a = (1-\phi) \cdot \overline{s} + \phi \cdot s_t^a + \lambda(s_t^a) \cdot \epsilon_{t+1} \tag{62}$$

where ϕ, g and $\overline{s}$ are parameters and

$$\Delta c_{t+1} = g + \epsilon_{t+1}, \ \ \epsilon_{t+1} \sim iid \ \mathcal{N}\left(0, \sigma_\epsilon^2\right).$$

The function expressed in equation 62 contains non-linear terms such as the *sensitivity function* $\lambda(s_t^a)$.

Modeling the surplus consumption ratio as in 62 means that consumption is always larger than habit. Otherwise it might be the case that consumption becomes smaller than habit in which case utility would be undefined. Modeling the logarithm of S_t^a ensures that $S_t^a > 0$ and hence $C_t^a - X_t > 0$. Since we are operating in a representative agent framework $C_t = C_t^a$ must hold trivially for all t in equilibrium since each individual chooses the same consumption level C_t. In (60) we already derived the first- and second-order derivative of the power utility function $U(C_t, X_t)$ we defined in (59).

Marginal utility can then be written as

$$U'(C_t, X_t) = (C_t - X_t)^{-\gamma} = \left(\frac{C_t - X_t}{C_t}\right)^{-\gamma} \cdot C_t^{-\gamma} = S_t^{-\gamma} \cdot C_t^{-\gamma}.$$

The intertemporal marginal rate of substitution (see equation (32)) then becomes

$$m_{t+1} = \beta \cdot \frac{U'(C_{t+1}, X_{t+1})}{U'(C_t, X_t)} = \beta \cdot \left(\frac{S_{t+1}}{S_t} \cdot \frac{C_{t+1}}{C_t}\right)^{-\gamma}. \tag{63}$$

We can relate (63) to the state variable s_t and the consumption innovation ϵ_{t+1} as follows

$$\begin{aligned}
m_{t+1} &= \beta \cdot \left(\frac{S_{t+1}}{S_t} \cdot \frac{C_{t+1}}{C_t}\right)^{-\gamma} \\
&= \beta \cdot \exp(-\gamma \cdot (s_{t+1} - s_t)) \cdot \exp(-\gamma \cdot (c_{t+1} - c_t)) \\
&= \beta \cdot \exp(-\gamma \cdot (s_{t+1} - s_t)) \cdot \exp(-\gamma \cdot (g + \epsilon_{t+1})) \\
&= \beta \cdot \exp(-\gamma \cdot (s_{t+1} - s_t)) \cdot \exp(-\gamma \cdot g) \cdot \exp(-\gamma \cdot \epsilon_{t+1}) \\
&= \beta \cdot G^{-\gamma} \cdot \exp(-\gamma \cdot (s_{t+1} - s_t + \epsilon_{t+1}))
\end{aligned}$$

Substituting (62) for s_{t+1} we arrive at

$$m_{t+1} = \beta \cdot G^{-\gamma} \cdot \exp(-\gamma \cdot [(\phi - 1) \cdot (s_t - \overline{s}) + [1 + \lambda(s_t)] \cdot \epsilon_{t+1}]).$$

From section 4.3 we recapitulate that

$$\frac{\sigma_t(m_{t+1})}{\mathbf{E}_t[m_{t+1}]} \geq -\text{corr}_t(m_{t+1}, R_{t+1}) \cdot \frac{\sigma_t(m_{t+1})}{\mathbf{E}_t[m_{t+1}]} = \frac{\mathbf{E}_t[R_{t+1}] - R_t^{rf}}{\sigma_t(R_{t+1})}.$$

Let the stochastic discount factor m_{t+1} be a lognormal random variable such that

$$\mathbf{E}(X) = \exp(\mu + \sigma^2/2) \quad \text{and} \quad \sigma^2(X) = \exp(2\mu + \sigma^2) \cdot (\exp(\sigma^2) - 1)$$

imply

$$\frac{\sigma(X)}{\mathbf{E}(X)} = \sqrt{\exp(\sigma^2) - 1}.$$

Then

$$\begin{aligned}\max_{\{\text{all assets}\}} \frac{\mathbf{E}_t[R_{t+1}] - R_t^{rf}}{\sigma_t(R_{t+1})} &= \frac{\sigma_t(m_{t+1})}{\mathbf{E}_t[m_{t+1}]} \\ &= \sqrt{\exp(\gamma^2 \cdot (1 + \lambda(s_t)^2) \cdot \sigma_\epsilon^2) - 1} \\ &\approx \gamma \cdot \sigma_\epsilon \cdot (1 + \lambda(s_t)).\end{aligned}$$

Economically, we can relate the Sharpe ratio to a function of the surplus consumption ratio. We have not yet specified what this function might look like. Intuitively, it should produce a high Sharpe ratio (and hence a high excess return) during bad times. So, $\lambda(s_t)$ must take on large values when s_t is small. Before we find a specific functional form for $\lambda(s_t)$ we have a look at the risk-free rate. In section 2.4.3 we derived the risk-free rate to be equal to the inverse of the conditionally expected stochastic discount factor, *i.e.*

$$R_t^{rf} = \frac{1}{\mathbf{E}_t[m_{t+1}]}.$$

Substituting the stochastic discount factor into the denominator yields

$$\frac{1}{\mathbf{E}_t[m_{t+1}]} = 1/\exp(\ln\beta - \gamma \cdot g - \gamma \cdot (\phi - 1) \cdot (s_t - \overline{s}) + 0.5 \cdot \gamma^2[1 + \lambda(s_t)]^2 \cdot \sigma_\epsilon^2) \tag{64}$$

and

$$\ln R_t^{rf} = -\ln\beta + \gamma \cdot g - \gamma \cdot (1 - \phi) \cdot (s_t - \overline{s}) - 0.5 \cdot \gamma^2[1 + \lambda(s_t)]^2 \cdot \sigma_\epsilon^2. \tag{65}$$

One can identify 2 sources that influence the risk-free rate, intertemporal substitution as well as precautionary savings. The former is represented by the term $s_t - \overline{s}$. When the surplus consumption ratio is small relative to its average (*i.e.* C_t is close to habit X_t) then marginal utility of consumption is high. Since an agent wants to consume more in that case he will borrow and thus drive up the interest rate. The precautionary savings reflects uncertainty. During uncertain periods consumers are willing to save more and consume less which drives the interest rate down.

5.2.2 Sensitivity Function

Campbell and Cochrane (1999) notice that there is relatively little variation in the data with regard to the risk-free interest rate. Hence, either the parameter ϕ is close to one in which case the impact of $s_t - \overline{s}$ on the risk-free rate vanishes or $\lambda(s_t)$ is chosen such that the precautionary savings effect offets the intertemporal substitution effect. This would be the case when $\lambda(s_t)$ increases as s_t declines. Thus, one requirement for the sensitivity function is a constant risk-free interest rate making the model consistent with the empirical data. A second condition is that habit is predetermined at the steady state $s_t = \overline{s}$. Finally, habit is also predetermined near the steady state which is equivalent to positive consumption shocks increasing habit, but never reducing it. Mathematically, this is equivalent to $\partial x/\partial c \geq 0$. The latter two conditions ensure that consumption is always larger than habit. Otherwise it might happen that consumption is below habit in which case the utility function is not defined as outlined before. In fact, the sensitivity could be chosen such that the interest rate is a linear function of the state variable s_t. However, this does not have any effect on the results to follow (see Campbell and Cochrane (1999), p. 216). The authors propose the following functional relationships

$$\overline{S} = \sigma \cdot \sqrt{\frac{\gamma}{1-\phi}} \tag{66}$$

and

$$\lambda(s_t) = \begin{cases} \frac{1}{\overline{S}} \cdot \sqrt{1 - 2(s_t - \overline{s})} - 1 & \text{if } s_t \leq s_{max} \\ 0 & \text{if } s_t > s_{max} \end{cases} \tag{67}$$

respectively in order to fulfill above criteria. s_{max} refers to the maximum surplus consumption ratio which is defined as

$$s_{max} \equiv \overline{s} + 0.5 \cdot (1 - \overline{S}^2). \tag{68}$$

Above identity can be derived by setting the upper expression of the sensitivity function equal to zero and solve for $s_t = s_{max}$. Now, how do we eventually arrive at above function? From the previous section we briefly recapitulate the risk-free interest rate to be of the form

$$\ln R_t^{rf} = -\ln\beta + \gamma \cdot g - \gamma \cdot (1-\phi) \cdot (s_t - \overline{s}) - 0.5 \cdot \gamma^2 [1 + \lambda(s_t)]^2 \cdot \sigma_\epsilon^2. \tag{69}$$

Let

$$Q \equiv -\ln R_t^{rf} - \ln\beta + \gamma \cdot g.$$

Equation (69) then becomes

$$0 = Q - \gamma \cdot (1-\phi) \cdot (s_t - \overline{s}) - 0.5 \cdot \gamma^2 [1 + \lambda(s_t)]^2 \cdot \sigma_\epsilon^2.$$

Solving for λ yields

$$\lambda(s_t) = \sqrt{\frac{2 \cdot (Q - \gamma \cdot (1-\phi)(s_t - \overline{s}))}{\sigma_\epsilon^2 \cdot \gamma^2}} - 1.$$

Let

$$Y \equiv \frac{2 \cdot Q}{\sigma_\epsilon^2 \cdot \gamma^2}$$

such that

$$\lambda(s_t) = \sqrt{Y - \frac{2 \cdot (1-\phi)(s_t - \overline{s})}{\sigma_\epsilon^2 \cdot \gamma}} - 1. \tag{70}$$

We then have to find a solution for Q such that the conditions 2 and 3 are fulfilled, *i.e.* the habit level should be predetermined at and near the steady state. Let us first recapitulate equation (62) from the previous section, *i.e.*

$$s_{t+1} = (1-\phi) \cdot \overline{s} + \phi \cdot s_t + \lambda(s_t) \cdot \epsilon_{t+1}$$

where $\epsilon_{t+1} = c_{t+1} - c_t - g$. The surplus consumption ratio is a function of both, C_t as well as X_t. We will first evaluate the total derivative of s_{t+1} which is

$$\frac{\partial s_{t+1}}{\partial \log X_{t+1}} \cdot d\log X_{t+1} + \frac{\partial s_{t+1}}{\partial \log C_{t+1}} \cdot d\log C_{t+1} = \lambda(s_t) \cdot d\log C_{t+1}. \tag{71}$$

From

$$s_{t+1} = \log\left(\frac{C_{t+1} - X_{t+1}}{C_{t+1}}\right)$$

we can obtain the partial derivatives with respect to $\log C_{t+1}$ and $\log X_{t+1}$. We first restate the log surplus consumption ratio as

$$s_{t+1} = \log\left(\frac{C_{t+1} - X_{t+1}}{C_{t+1}}\right)$$

$$= \log\left(1 - \frac{X_{t+1}}{C_{t+1}}\right)$$

$$= \log(1 - \exp(\log(X_{t+1}/C_{t+1})))$$

$$= \log(1 - \exp(\log X_{t+1} - \log C_{t+1}))$$

Then

$$\frac{\partial s_{t+1}}{\partial \log X_{t+1}} = -\frac{\exp(\log X_{t+1} - \log C_{t+1})}{1 - \exp(\log X_{t+1} - \log C_{t+1})}$$

$$= -\frac{\frac{X_{t+1}}{C_{t+1}}}{1 - \frac{X_{t+1}}{C_{t+1}}}$$

$$= -\frac{\frac{X_{t+1}}{C_{t+1}}}{\frac{C_{t+1} - X_{t+1}}{C_{t+1}}}$$

$$= -\frac{X_{t+1}}{C_{t+1}} \cdot \frac{C_{t+1}}{C_{t+1} - X_{t+1}} = -\frac{X_{t+1}}{C_{t+1} - X_{t+1}}$$

$$= -\frac{X_{t+1} + C_{t+1} - C_{t+1}}{C_{t+1} - X_{t+1}}$$

$$= -\left(\frac{C_{t+1}}{C_{t+1} - X_{t+1}} - 1\right)$$

$$= -\left(\frac{1}{S_{t+1}} - 1\right).$$

The partial derivative with respect to $\log C_{t+1}$ is

$$\frac{\partial s_{t+1}}{\partial \log C_{t+1}} = \frac{\exp(\log X_{t+1} - \log C_{t+1})}{1 - \exp(\log X_{t+1} - \log C_{t+1})}$$

$$= \frac{1}{S_{t+1}} - 1.$$

We can now substitute both expressions into (71) to get

$$-\left(\frac{1}{S_{t+1}} - 1\right) \cdot d\log X_{t+1} + \left(\frac{1}{S_{t+1}} - 1\right) \cdot d\log C_{t+1} = \lambda(s_t) \cdot d\log C_{t+1}.$$

Eventually, dividing by $d\log C_{t+1}$ and rearrganging we obtain

$$\frac{d\log X_{t+1}}{d\log C_{t+1}} = 1 - \frac{\lambda(s_t)}{\frac{1}{S_{t+1}} - 1} = 1 - \frac{\lambda(s_t)}{\exp(-s_{t+1}) - 1} \approx 1 - \frac{\lambda(s_t)}{\exp(-s_t) - 1}$$

where the latter approximation holds close to the steady state. Since we are looking for a function λ such that $d\log X_{t+1}/d\log C_{t+1} = 0$ holds at $s_t = \overline{s}$ the following must hold

$$1 - \frac{\lambda(\overline{s})}{\exp(-\overline{s}) - 1} = 0 \Leftrightarrow \lambda(\overline{s}) = \frac{1}{\overline{S}} - 1.$$

Using above result and substituting in (70) one can solve for Y.

$$\lambda(\overline{s}) = \sqrt{Y - \frac{2 \cdot (1-\phi)(\overline{s} - \overline{s})}{\sigma_\epsilon^2 \cdot \gamma}} - 1 \Leftrightarrow Y = \exp(-2 \cdot \overline{s}).$$

Having derived Y we obtain the function as outlined in equatin (67). Finally, using the third requirement that habit is also predetermined near the steady state one can derive the expression in equation (66) as follows. We evaluate the first derivative of λ with respect to s_t at $s_t = \overline{s}$. Then

$$\frac{d}{ds_t}\frac{d\log X_{t+1}}{d\log C_{t+1}} = \frac{d}{ds_t}\left(1 - \frac{\lambda(s_t)}{\exp(-s_{t+1}) - 1}\right)$$

$$= -\left[\frac{\lambda'(s_t) \cdot (\exp(-s_{t+1}) - 1) - \lambda(s_t) \cdot \left(-\frac{\partial s_{t+1}}{\partial s_t}\right) \cdot \exp(-s_{t+1})}{(\exp(-s_{t+1}) - 1)^2}\right]$$

$$= -\left[\frac{\lambda'(s_t)}{\exp(-s_{t+1}) - 1} + \frac{\lambda(s_t) \cdot \frac{\partial s_{t+1}}{\partial s_t} \cdot \exp(-s_{t+1})}{(\exp(-s_{t+1}) - 1)^2}\right]$$

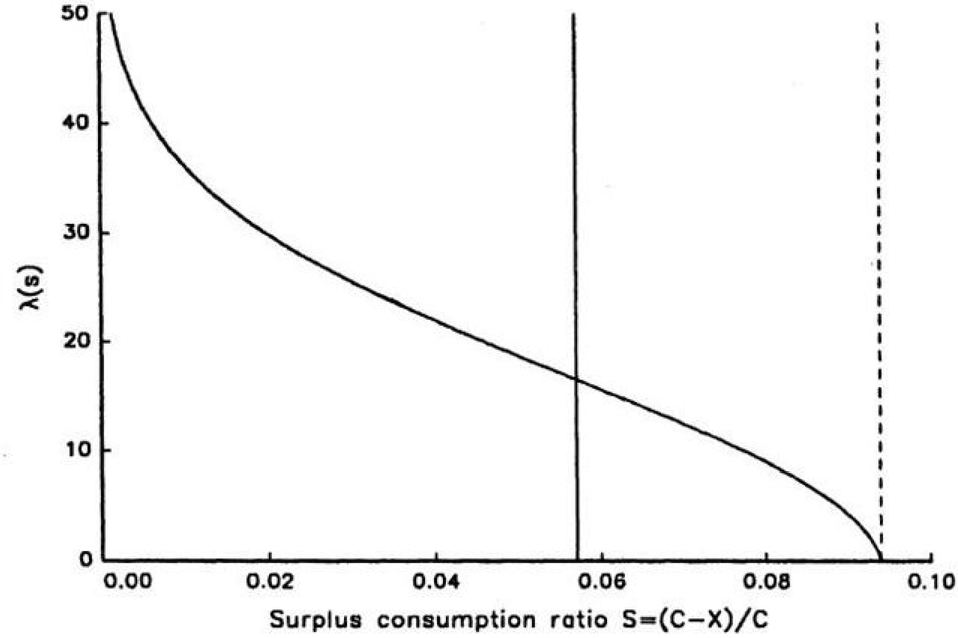

Figure 7: Sensitivity Function λ with Steady State Surplus Consumption Ratio (Bold) and Maximum Surplus Consumption Ratio (Dashed) (Campbell and Cochrane (1999))

where

$$\frac{\partial s_{t+1}}{\partial s_t} = \frac{d\Big((1-\phi)\cdot \overline{s} + \phi \cdot s_t + \lambda(s_t)\cdot \epsilon_{t+1}\Big)}{ds_t} = \phi + \lambda'(s_t)\cdot \epsilon_{t+1}.$$

Substituting above expression for the partial derivative yields

$$\frac{d}{ds_t}\frac{d\log X_{t+1}}{d\log C_{t+1}} = -\frac{\lambda'(s_t)}{\exp(-s_{t+1})-1} - \frac{\lambda(s_t)\cdot(\phi+\lambda'(s_t)\cdot\epsilon_{t+1})\cdot\exp(-s_{t+1})}{(\exp(-s_{t+1})-1)^2}$$

$$= -\frac{\lambda'(s_t)}{\frac{1}{S_{t+1}}-1} - \frac{\phi\cdot\lambda(s_t)}{\left(\frac{1}{S_{t+1}}-1\right)^2}\cdot\frac{1}{S_{t+1}} - \frac{\lambda'(s_t)\cdot\lambda(s_t)\cdot\epsilon_{t+1}}{\left(\frac{1}{S_{t+1}}-1\right)^2}\cdot\frac{1}{S_{t+1}}.$$

Hence, at $s_t = \overline{s}$, above expression becomes

$$\frac{d}{ds_t}\frac{d\log X_{t+1}}{d\log C_{t+1}} = -\frac{\lambda'(\overline{s})}{\frac{1}{\overline{S}}-1} - \frac{\phi\cdot\lambda(\overline{s})}{\left(\frac{1}{\overline{S}}-1\right)^2}\cdot\frac{1}{\overline{S}} - \frac{\lambda'(\overline{s})\cdot\lambda(\overline{s})\cdot\epsilon_{t+1}}{\left(\frac{1}{\overline{S}}-1\right)^2}\cdot\frac{1}{\overline{S}} \equiv 0.$$

In order to eliminate the stochastic term ϵ_{t+1} we take expectations of above expression to obtain

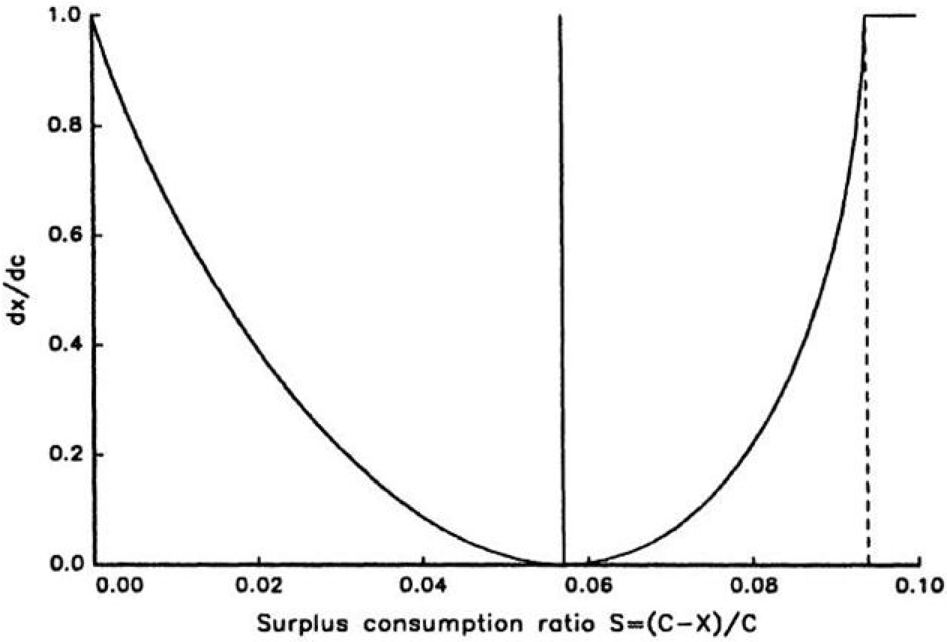

Figure 8: Derivative of the $\log X$ *wrt* $\log C$ (Campbell and Cochrane (1999))

$$0 \equiv -\frac{\lambda'(\overline{s})}{\frac{1}{\overline{S}}-1} - \frac{\phi\cdot\lambda(\overline{s})}{\left(\frac{1}{\overline{S}}-1\right)^2}\cdot\frac{1}{\overline{S}}$$

$$\Leftrightarrow 0 \equiv -\lambda'(\overline{s})\cdot(1-\overline{S}) - \phi\cdot\lambda(\overline{s}).$$

We arrive at the second line multiplying the first line by $\frac{1}{\overline{S}}-1=\frac{1-\overline{S}}{\overline{S}}$. Eventually we get

$$\lambda'(\overline{s}) = -\frac{\phi\lambda(\overline{s})}{1-\overline{S}} = \frac{\phi\lambda(\overline{s})}{\overline{S}-1} = \frac{\phi\cdot(\exp(-\overline{s})-1)}{\exp(\overline{s})-1} = -\phi\cdot\exp(-\overline{s}).$$

Putting together the facts we have

$$\lambda(s_t) = \sqrt{\exp(-2\cdot\overline{s}) - \frac{2\cdot(1-\phi)(s_t-\overline{s})}{\sigma_\epsilon^2\cdot\gamma}} - 1,$$

$$\lambda'(s_t) = -\frac{2\cdot(1-\phi)}{\gamma\sigma_\epsilon^2}\cdot\frac{1}{2\cdot\sqrt{\exp(-2\cdot\overline{s})-\frac{2\cdot(1-\phi)(s_t-\overline{s})}{\sigma_\epsilon^2\cdot\gamma}}}$$

and

$$\lambda'(\overline{s}) = -\phi\cdot\exp(-\overline{s}).$$

Hence, at $s_t=\overline{s}$

$$-\frac{1}{\exp(-\overline{s})} \cdot \frac{1-\phi}{\gamma\sigma_\epsilon^2} = -\phi \cdot \exp(-\overline{s})$$

and thus

$$\overline{S} = \exp(\overline{s}) = \sigma_\epsilon \cdot \sqrt{\frac{\gamma \cdot \phi}{1-\phi}} \approx \sigma_\epsilon \cdot \sqrt{\frac{\gamma}{1-\phi}}.$$

The approximation holds since ϕ is close to 1 empirically. *E.g.* Campbell and Cochrane (1999) use a value of $\phi = 0.87$ in their analysis (see table 1). Last, but not least, the resulting function is depicted in figure 7. One can immediately spot the countercyclical nature that is imposed upon the Sharpe ratio. The abcissa contains the surplus consumption ratio that we use as a proxy for good and bad times. The higher it is the better is the perceived economic climate since - in that case - consumption is well above habit. For large values of S the sensitivity function yields small values $\lambda(s)$ resulting in small Sharpe ratios. The latter then imply that excess returns over the risk-free rate are relatively low and asset prices high which makes sense intuitively. On the contrary when s is small λ yields large values which in turn results in high Sharpe ratios. Hence, excess returns relative to the risk-free rate are large and asset prices are low. So, our sensititivy function has the desired properties to model the countercyclical nature of returns, asset prices and Sharpe ratios. The bold vertical line marks the steady state surplus consumption ratio while the dashed vertical line refers to the maximum surplus consumption ratio as defined in equation (68). Mathematically, as the surplus consumption ratio approaches zero λ goes to infinity. It goes to zero as the surplus consumption ratio approaches its upper bound s_{max}. Figure 8 shows the sensitivity of habit with respect to consumption. Again, the bold vertical line indicates the steady state surplus consumption ratio. What we can spot nicely is the feature that habit does not change at and close to the steady state level. Furthermore, habit moves positively with consumption everywhere so as to keep habit below consumption all the time.

5.2.3 Data, Simulation and Empirical Results

Campbell and Cochrane (1999) use 2 different datasets to test the validity of the model. Both datasets differ significantly in length. The first set contains value-weighted NYSE stock index returns taken from the Center for Research in Security Prices (CRSP), 3-month Treasury bill rate and per capita non-durables and services consumption for the postwar period between 1947 and 1995. The second dataset contains annual data for the S&P 500 index, commercial paper returns (both 1871–1993) as well as per capita consumption (1889–1992). Parameters to calibrate the model were chosen such that certain moments in the data are matched. The corresponding parameters and their values are summarized in table 1.

To assess the validity of the model the authors use those parameters to simulate 500,000 months of artifical data. Monthly data is then used to construct time-averaged annual data. Table 2 compares simulated data with actual market statistics. Those values marked with an asterisk are calibration targets and hence coincide with the real data. *E.g.* the risk aversion

Parameter	Variable	Value
Assumed:		
Mean consumption growth in %	g	1.89*
Std. Dev. of cons. growth in %	σ_ϵ	1.50*
Log risk-free rate in %	$\log R_t^{rf}$	0.94*
Persistence Coefficient	ϕ	0.87*
Utility curvature	γ	2.00
Implied:		
Subjective discount factor	β	0.89*
Steady state surplus cons. ratio	$\overline{S}$	0.057
Maximum surplus cons. ratio	S_{max}	0.094

Table 1: Model Parameters (asterisks indicate annual values)

coefficient $\gamma = 2$ was chosen to match the Sharpe ratio of log returns in the postwar sample of 0.43 and the Sharpe ratio of 0.5 of the discrete returns. A fair objection would be that parameters are chosen such that certain moments in the data are matched. Hence, by any choice of parameters one would be able to obtain the desired outcome. In case of the risk aversion coefficient γ was chosen to match a ratio. That, however, does not reveal anything further about the levels of the two variables involved. Focusing our attention on the first column of table 2 one can spot that the model does not only match the ratio, but also the level of expected excess returns as well as the standard deviation thereof. Campbell and Cochrane (1999) particularly calibrate the model to resemble the postwar sample *"because they are a significantly harder target"* (see p. 225). In addition to the equity premium the model is able to explain a broad variety of asset pricing phenomena. As outlined above the proposed model is empirically consistent with the data regarding the equity premium and the risk-free rate. Similar to the prior section one can show that a standard power utility function and *iid* lognormal consumption growth with mean g and standard deviation σ_ϵ the stochastic discount factor becomes

$$m_{t+1} = \beta \cdot \left(\frac{C_{t+1}}{C_t} \right)^{-\gamma}.$$

The Sharpe ratio can then be expressed as

$$\frac{\mathbf{E}_t[R_{t+1}] - R_t^{rf}}{\sigma_t(R_{t+1})} \approx \gamma \cdot \sigma_\epsilon$$

with interest rate

$$\log R_t^{rf} = -\log \beta + \gamma \cdot g - 0.5 \cdot \gamma^2 \cdot \sigma_\epsilon^2.$$

The latter expression was already derived in section 2.4.4. Plugging in the respective values for the Sharpe ratio ≈ 0.5 and the standard deviation of consumption growth $\sigma_\epsilon = 1.22$ percent one needs a risk aversion coefficient of $\gamma \approx 41$ which is unplausibly high. Why? If

Statistic	Consumption Claim	Dividend Claim	Postwar Sample	Long Sample
$E(\Delta c)$	1.89*		1.89	1.72
$\sigma(\Delta c)$	1.22*		1.22	3.32
$E(r^{rf})$	0.94*		0.94	2.92
$E(r - r^{rf})/\sigma(r - r^{rf})$	0.43*	0.33	0.43	0.22
$E(R - R^{rf})/\sigma(R - R^{rf})$	0.50		0.50	
$E(r - r^{rf})$	6.64	6.52	6.69	3.90
$\sigma(r - r^{rf})$	15.2	20.0	15.7	18.0
$exp[E(p - d)]$	18.3	18.7	24.7	21.1
$\sigma(p - d)$	0.27	0.29	0.26	0.27

Table 2: Means and Standard Deviations of Simulated and Historical Data

we plug in the respective values into the risk-free rate equation with $\gamma = 41$ and $g = 1.89\%$ one would need a discount factor in the region of $\beta = 1.9$ in order to get a risk-free interest rate of about 1 percent. Since we require the discount factor to be $\beta \leq 1$ this implies a risk-free rate of approximately 90% annually. Furthermore, the risk-free rate is not as sensitive to changes in mean consumption growth as indicated by a risk aversion coefficient in the region of 41. How are these drawbacks overcome in the Campbell and Cochrane (1999) model? In the new setup the risk aversion parameter is set equal to 2 (see table 1). Additionally, curvature is influenced by the surplus consumption ratio as shown in equation (61). The curvature is high during bad states (low S), low during good states (high S) and ≈ 35 at the steady state level where $\overline{S} = 0.057$. From (65) and (67) we obtain a constant risk-free interest rate which has the functional form

$$\log R_t^{rf} = -\log \beta + \gamma \cdot g - 0.5 \cdot \left(\frac{\gamma}{\overline{S}}\right)^2 \cdot \sigma_\epsilon^2. \tag{72}$$

In the latter case the risk-free rate is influenced by a risk aversion coefficient in the region of 2 which allows for a discount factor $\beta = 0.89$ smaller than 1. In addition to that the sensitivity of the interest rate with respect to consumption growth is significantly lower than in the case without habit formation.

5.2.4 Model Criticism

Despite the models ability to achieve the desired outcome of a high equity premium a couple of reservations exist - some of which are addressed by the authors in their original contribution. The first drawback is the representative agent setup that does not allow for consumer heterogeneity. The habit level is about 5% below actual consumption across individuals. If one considers the distribution of wealth and income for some poor agents it might be the case that $C_t < X_t$ in which case utilitiy is not defined (see section 5.2.1). Furthermore it is obvious that agents with differing income levels have differing stock market

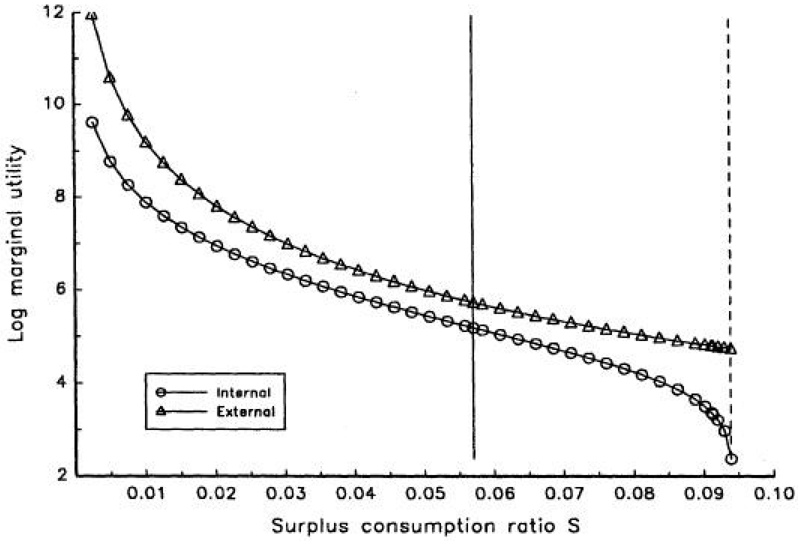

Figure 9: log marginal utility *wrt* S (surplus consumption ratio) (Campbell and Cochrane (1999))

participation ratios. The wealthier an agent is the higher is his participation in the stock market and vice versa. A second drawback is the fact that the model works under the assumption of high risk aversion. The curvature parameter of the utility function takes on a value of about 35 at the steady state. The model is, however, able to avoid unrealistically high interest rates when β is constrained to be smaller than 1. A third point is the external habit specification. In contrast to above model in an internal framework habit is determined by an agents own consumption history. While consumption grows so does habit. Hence, marginal utility for a given level of consumption declines as habit increases. Since assets are priced based on relative marginal utility of consumption today and tomorrow a proportional increase or decline in both does not change the discount factor and hence external and internal models yield the same results. Campbell and Cochrane (1999) provide conditions under which this is the case (see p. 245), one being that habit accumulation is linear. This is different in the present setup in order to obtain a random walk in the consumption process. In the linear framework habit accumulation would only be close to a random walk. Figure 9 plots the logarithm of marginal utility with respect to the surplus consumption ratio for the internal as well as the external setup. The major features that one can spot is the fact that marginal utility has a similar behavior close to the steady state level (indicated by vertical line), but drifts apart as the surplus consumption ratio reaches its maximum (vertical dotted line). The larger the change in consumption, the larger is the increase in the habit level and hence marginal utility declines. Campbell and Cochrane (1999) see the similar behavior near the steady state of their model and one with an internal specification as an indicator for the robustness of their model (see p. 246).

A Mathematical Appendix

A.1 Markov Processes

A *stochastic process* $\{X(t)\}_{t=0,1,2,...}$ has the so called *Markov property* if, for a finite set with n elements $t_1 < t_2 < \ldots < t_n$, we have

$$\phi(X(t_n) \leq x_n | X(t_1) = x_1, \ldots, X(t_{n-1}) = x_{n-1}) = \phi(X(t_n) \leq x_n | X(t_{n-1}) = x_{n-1}) \quad (73)$$

for any $x_1, \ldots, x_n \in \mathbb{R}$, *i.e.* the conditional distribution of $X(t_n)$ for given realizations $\{X(t_1), \ldots, X(t_{n-1})\}$ only depends on the most recent value $X(t_{n-1})$. A *Markov chain* $\{s_t\}_{t=0,1,2,...}$ is a discrete random process with $s_t \in S$, where $S = (1, \ldots, k)$ is a set of integers called the *state space*, and the property

$$\phi(s_{t+1} = j | s_t = i, \ldots, s_0 = k) = \phi(s_{t+1} = j | s_t = i).$$

A Markov chain is called *time-invariant* if

$$\phi(s_{t+1} = j | s_t = i) = \phi(s_{t+l+1} = j | s_{t+l} = i).$$

Let $\pi_{i,j} \in [0, 1]$ denote the probability that the process takes the value $s_{t+1} = j$ given that $s_t = i$, *i.e.*

$$\phi(s_{t+1} = j | s_t = i) = \pi_{i,j}.$$

For each i we have $\phi_{i,1} + \ldots + \phi_{i,k} = 1$ meaning that, given the realization i, the process takes on an arbitrary value in S with certainty. The same holds for $t = 0$ with $\sum_{i=1}^{k} \pi_{0,i} = 1$, *i.e.* the initial value of the process is also on element in S. One can then define a so called *transition matrix* for a Markov chain of the form

$$\Pi = \begin{pmatrix} \phi_{1,1} & \phi_{1,2} & \cdots & \phi_{1,k} \\ \phi_{2,1} & \phi_{2,2} & \cdots & \phi_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{k,1} & \phi_{k,2} & \cdots & \phi_{k,k} \end{pmatrix}.$$

The transition matrix shows the probability that j is realized in $t+1$ given that i was realized in t for all $i = 1, \ldots, k$ and $j = 1, \ldots, k$.

A.2 Metric Spaces

The following section will give a very brief overview of the fundamentals of *metric spaces*.

Definition A metric space (X, ρ) is a set X and a function ρ called a metric with $\rho : X \times X \to \mathbb{R}$ with the following properties:

1. Positivity: $\rho(x, y) \geq 0, \;\; \forall x, y \in X$.
2. Strict Positivity: $\rho(x, y) = 0$ iff $x = y$.
3. Symmetry: $\rho(x, y) = \rho(y, x), \;\; \forall x, y \in X$.
4. Triangle Inequality: $\rho(x, y) \leq \rho(x, z) + \rho(z, y), \;\; \forall x, y, z \in X$

Definition A sequence $\{x_n\}$ in a metric space (X, ρ) is called a Cauchy sequence if for each $\epsilon > 0$ there exists an $N(\epsilon)$ such that $\rho(x_n, x_m) < \epsilon$ for any $n, m \geq N(\epsilon)$. Hence, a sequence is called Cauchy if $\lim_{n,m\to\infty} \rho(x_n, x_m) = 0$.

Definition A sequence $\{x_n\}$ in a metric space (X, ρ) converges to a limit $x \in X$ if for every $\epsilon > 0$ there exists an $N(\epsilon)$ such that $\rho(x_n, x) < \epsilon$ for $n \geq N(\epsilon)$.

Lemma A.1 *Every convergent sequence $\{x_n\}$ in a metric space (X, ρ) is a Cauchy sequence.*

Proof Let $\epsilon > 0$. Let x be the limit of $\{x_n\}$. By the triangle inequality

$$\rho(x_n, x_m) \leq \rho(x_n, x) + \rho(x, x_m).$$

Since $x_n \to x$, there exists an N such that $\rho(x_n, x) < \epsilon/2$ for $n \geq N$. This also holds for the sequence $\{x_m\}$. Hence, $\rho(x_n, x_m) < \epsilon/2 + \epsilon/2 = \epsilon$ for $n, m \geq N(\epsilon)$. □

Definition A metric space (X, ρ) is called complete if every Cauchy sequence in (X, ρ) converges to a point in (X, ρ).

A.3 Contraction Mappings

In the following section we will introduce the notion of a special operator called *contraction mapping*.

Definition A function $T : X \to X$ mapping a metric space (X, ρ) into itself is called an operator.

Definition Let (X, ρ) be a metric space and $T : X \to X$ a mapping. T is called a contraction mapping of modulus β if

$$\rho(Tf, Tg) \leq \beta \cdot \rho(f, g), \forall f, g \in X$$

with $0 \leq \beta < 1$.

From the latter condition imposed on β one can see that the distance of elements in the range space is smaller than their respective counterparts in the domain. Hence, one can think of a contraction mapping as a method to bring elements of a metric space closer together.

Theorem A.2 *Let (X, ρ) be a complete metric space and let $T : X \to X$ be a contraction mapping with modulus β. Then (i) T has exactly one fixed point $v \in X$, (ii) for any $v_0 \in X$, $\rho(T^n v_0, v) \leq \beta^n \cdot \rho(v_0, v)$ for all $n = 1, 2, \ldots$.*

Proof Let the iterates of T, which are a sequence of mappings $\{T^n\}$, be defined as $T^\circ X = X$ and $T^n X = T(T^{n-1})X$ for all $n = 1, 2, \ldots$. Let $v_0 \in X$ and define a sequence $\{v_n\}_{n=0}^{\infty}$ by $v_n = Tv_{n-1}$. Hence, we must have $v_n = T^n v_0$. This result can easily be verified by repeated substitution.

$$\begin{aligned} v_n &= Tv_{n-1} \\ &= T(Tv_{n-2}) \\ &= T(T(Tv_{n-3})) \\ &\vdots \\ &= T^n v_0 \end{aligned}$$

T is a contraction mapping by assumption. Hence,

$$\rho(v_2, v_1) = \rho(Tv_1, Tv_0) \leq \beta \cdot \rho(v_1, v_0)$$

from which we can conclude by induction

$$\rho(v_{n+1}, v_n) \leq \beta^n \cdot \rho(v_1, v_0). \tag{74}$$

Suppose $m > n$. By the triangle inequality we have

$$\begin{aligned} \rho(v_m, v_n) &\leq \rho(v_m, v_{m-1}) + \ldots + \rho(v_{n+2}, v_{n+1}) + \rho(v_{n+1}, v_n) \\ &\leq [\beta^{m-1} + \ldots + \beta^{n+1} + \beta^n] \cdot \rho(v_1, v_0) \\ &= \beta^n \cdot [\beta^{m-n-1} + \ldots + \beta + 1] \cdot \rho(v_1, v_0) \\ &\leq \frac{\beta^n}{1-\beta} \cdot \rho(v_1, v_0). \end{aligned}$$

The second line is simply an application of the result in (74). X is a complete metric space by definition and $\{v_n\}_{n=0}^{\infty}$ is a Cauchy sequence. Hence, we get $v_n \to v$ as $n \to \infty$. To show that $Tv = v$ we can apply the triangle inequality again to obtain

$$\rho(Tv, v) \leq \rho(Tv, T^n v_0) + \rho(T^n v_0, v)$$

for all n and $v_0 \in X$. T being a contraction then yields

$$\rho(Tv, v) \leq \beta \cdot \rho(v, T^{n-1} v_0) + \rho(T^n v_0, v).$$

Both terms on the *rhs* converge to 0 as $n \to \infty$. This can easily be verified from the above result. Hence, $\rho(Tv, v) = 0$ which implies that $Tv = v$. So, v is a fixed point in X.

Uniqueness is established using the following argument. Suppose $\exists \hat{v} \in X$ such that $T\hat{v} = \hat{v}$ and $\hat{v} \neq v$. Thus, we get

$$0 < a = \rho(\hat{v}, v) = \rho(T\hat{v}, Tv) \leq \beta \cdot \rho(\hat{v}, v) = \beta \cdot a.$$

Since $\beta < 1$ by assumption we obtain a contradiction as $a \neq \beta \cdot a$. Hence, a can only be equal to 0 and $\hat{v} = v$. So, $v \in X$ is unique. Finally, to prove statement (ii) let $n \geq 1$. Then

$$\rho(T^n v_0, v) = \rho[T(T^{n-1}v_0), Tv] \leq \beta \cdot \rho(T^{n-1}v_0, v)$$

and (ii) simply follows by induction. □

Corollary A.3 *Let (X, ρ) be a complete metric space and $T : X \to X$ be a contraction mapping with fixed point $v \in X$. If X' is a closed subset of X and $T(X') \subseteq X'$ (where $T(X')$ is the image of X' under T). Then $v \in X'$.*

Proof Let $v_0 \in X'$ and $\{T^n v_0\}$ denote a sequence in X' with limit v. Since $\{T^n v_0\} \to v$ as $n \to \infty$ and X' closed, we have $v \in X'$. □

Corollary A.4 *Let (X, ρ) be a complete metric space, let $T : X \to X$ and suppose that for some integer N, $T^N : X \to X$ is a contraction mapping with modulus β. Then (i) T has one fixed point in X, (ii) for any $v_0 \in X$, $\rho(T^{k \cdot N} v_0, v) \leq \beta^k \cdot \rho(v_0, v)$ for all $k = 1, 2, \ldots$.*

Proof Let v be a fixed point of the contraction mapping T^N. Then

$$\rho(Tv, v) = \rho[T(T^N v), T^N v] = \rho[T^N(Tv), T^N v] \leq \beta \cdot \rho(Tv, v).$$

The first equality holds since $T^N v = v$. The final inequality holds since T^N is a contraction mapping with modulus β by assumption. By definition $0 < \beta < 1$. Hence, $\rho(Tv, v)$ can only happen to be 0. Thus, $Tv = v$ and v is a fixed point of X. The second statement (ii) can be derived in the same manner as in the *contraction mapping theorem.* □

Theorem A.5 *Let $\mathcal{B}(S)$ be the space of bounded functions $f : S \to \mathbb{R}$ with the sup norm. Let $T : \mathcal{B}(S) \to \mathcal{B}(S)$ be an operator defined on $\mathcal{B}(S)$ satisfying: (i) Let $f, g \in \mathcal{B}(S)$. For each $s \in S$, $f(s) \geq g(s)$ implies $Tf(s) \geq Tg(s)$ (monotonicity). (ii) Let $a \in (0, \infty)$ be a constant. There is a $\beta \in (0, 1)$ such that, for $f \in \mathcal{B}(S)$, $T(f + a)(s) \leq Tf(s) + \beta \cdot a$ (discounting). If $T : \mathcal{B}(S) \to \mathcal{B}(S)$ satisfies both properties, then T is a contraction mapping with modulus β.*

Proof If $f(s) \leq g(s)$ for all $s \in S$, then $f \leq g$. By the definition of a metric $\|.\|$ we have $f \leq g + \|f - g\|$ for any $f, g \in \mathcal{B}(S)$. Applying both properties to the latter inequality we obtain

$$Tf \leq T(g + \|f - g\|) \leq Tg + \beta \cdot \|f - g\|.$$

Reversing the initial assumption to $g \leq f$ we have

$$Tg \leq T(f + \|f - g\|) \leq Tf + \beta \cdot \|f - g\|.$$

The first inequality holds by the monotonicity assumption while second one reflects the discounting property. Combining both we get

$$\|Tf - Tg\| \leq \beta \cdot \|f - g\|.$$

□

A.4 Mean-Variance Frontier Graphs

Figure 2 may be reproduced using the following R code. All analyses were performed using the R statistical software R version 2.9.2 (2009-08-24).

```
#install.packages("tseries")
#install.packages("dynlm")
#install.packages("zoo")
#install.packages("fSeries")
#install.packages("fBasics")
#install.packages("fArma")
library("tseries")
library("dynlm")
library("zoo")
library("fSeries")
library("fBasics")
library("fArma")


# Vector of expected returns

r_i = c(0.05, 0.12)

# Target return and risk free rate

r_P <- 0.07

r_rf <- 0.06


# Vector of standard deviations for each asset

sig_i <- c(0.09, 0.10)
```

```
# Vector containing 1s

e <- c(1,1)

# Correlation and covariance

correl <- -0.8
cova <- correl*sig_i[1]*sig_i[2]

# Covariance matrix and inverse

covar <- matrix(c(sig_i[1]^2, cova, cova, sig_i[2]^2), 2, 2)
icovar <- inv(covar)

# Closed form optimization

a <- as.vector(t(r_i)%*%icovar%*%r_i)
b <- as.vector(t(r_i)%*%icovar%*%e)
cc <- as.vector(t(e)%*%icovar%*%e)
d <- a*cc-b*b

w_a <- icovar%*%(r_i-e*r_rf)
w_b <- r_P-r_rf
w_c <- t(r_i-e*r_rf)%*%w_a
w_d <- w_b/w_c

# Vector of optimal weights

w <- w_a*as.vector(w_d)
w_rf <- 1 - sum(w)

# Tangential portfolio

gam_ma <- a-2*b*r_rf+cc*r_rf^2
w_tang <- w_a/(b-cc*r_rf)
r_tang <- t(w_tang)%*%r_i
risk_tang <- gam_ma / (b-cc*r_rf)^2
portrisk <- w_b^2/gam_ma

# Minimum variance portfolio
```

```
mvp_risk <- 1/cc
mvp_return <- b/cc
w_mvp <- 1/cc*icovar%*%e




portret <- seq(from=0, to=max(r_i)+0.02, length=1000)
# Vector containing portfolio returns
portvar <- NULL
# Vector containing portfolio variances of combined assets
for(i in 1:length(portret)){
# loop calculating variances for corresponding returns
risk <- (portret[i]-r_rf)^2/gam_ma
portvar <- c(portvar, risk)
}

portstd <- portvar^0.5


# Efficient Frontier without Riskless Asset

portrett <- seq(from=0, to=max(r_i)+0.02, length=1000)
portvarr <- NULL

# loop calculating variances for corresponding returns

for(i in 1:length(portrett))
{
portvarr <- c(portvarr, 1/d*(cc*portrett[i]^2-2*b*portrett[i]+a))
}
portstdd <- portvarr^0.5


# Mean variance frontier asymptotes with perfect negative correlation
stddev <- seq(from=0, to=sig_i[2], length=1000)
stddev1 <- seq(from=0, to=sig_i[1], length=1000)
portret2 <- sig_i[2]/sum(sig_i)*r_i[1] +
sig_i[1]/sum(sig_i)*r_i[2] + (r_i[2]-r_i[1])/sum(sig_i)*stddev
portret3 <- sig_i[2]/sum(sig_i)*r_i[1] +
```

```
sig_i[1]/sum(sig_i)*r_i[2] + (r_i[1]-r_i[2])/sum(sig_i)*stddev1
plot(stddev, portret2, type="l", lty=2)
lines(stddev, portret3, type="l", lty=2)

# Mean variance frontier asymptotes with perfect positive correlation

w <- seq(from=0, to=1, length=1000)
portret4 <- w*r_i[1]+(1-w)*r_i[2]
stddev2 <- (w^2*sig_i[1]^2+(1-w)^2*sig_i[2]^2+2*w*(1-w)*sig_i[1]*sig_i[2])^0.5
plot(stddev2, portret4, type="l", lty=2)

# Plot

color <- "black"
plot(portstd, portret, col=color, type="l", lty="solid",
lwd="1", main="Mean-Variance Frontier",
xlab="Standard Deviation", ylab="Expected Return",
cex.lab=0.75, cex.main=0.75, xlim=c(0,0.15))
lines(portstdd, portrett, lty="solid", lwd="1")
lines(stddev, portret2, type="l", lty=2)
lines(stddev1, portret3, type="l", lty=2)
lines(stddev2, portret4, type="l", lty=2)
points(sig_i[1], r_i[1], col=color, bg=color, pch=21)
points(sig_i[2], r_i[2], col=color, bg=color, pch=21)
points(mvp_risk^0.5, mvp_return, col=color, bg=color, pch=21)
points(sqrt(risk_tang), r_tang, col=color, bg=color, pch=21)
points(0, r_rf, col=color, bg=color, pch=21)

text(0, r_rf, "risk free rate", font=1, cex=0.75, pos=4)
text(sig_i[1], r_i[1], "asset 1", font=1, cex=0.75, pos=4)
text(sig_i[2], r_i[2], "asset 2", font=1, cex=0.75, pos=4)
text(mvp_risk^0.5, mvp_return, "minimum variance portfolio",
font=1, cex=0.75, pos=4)
text(sqrt(risk_tang), r_tang, "tangential portfolio", font=1,
cex=0.75, pos=4)
```

References

Andrew B. Abel. Asset Prices under Habit Formation and Catching up with the Joneses. *American Economic Review*, 80:38–42, 1990.

Sumru Altug. Time-to-Build and Aggregate Fluctuations: Some New Evidence. *International Economic Review*, 30:889–920, 1989.

Sumru Altug and Pamela Labadie. *Asset Pricing for Dynamic Economies.* Cambridge University Press, 2008.

Kenneth Arrow. The Role of Securities in the Optimal Allocation of Risk-Bearing. *Review of Economic Studies*, 31:91–96, 1964.

Kenneth J. Arrow. *Essays in the Theory of Risk Bearing.* North Holland, 1971.

Richard Bellman. *Dynamic Programming.* Princeton University Press, 1957.

Douglas T. Breeden. An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities. *Journal of Financial Economics*, 7:265–296, 1979.

John Y. Campbell. *Handbook of Macroeconomics*, chapter 19. North Holland, 1999.

John Y. Campbell and John H. Cochrane. By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior. *Journal of Political Economy*, 107: 205–251, 1999.

John H. Cochrane. *Asset Pricing.* University Presses of CA, 2005.

John H. Cochrane. Financial Markets and the Real Economy. In *Handbook of the Equity Risk Premium.* Elsevier, 2008.

George M. Constantinides. Habit Formation: A Resolution of the Equity Premium Puzzle. *Journal of Political Economy*, 98:519–543, 1990.

Jean-Pierre Danthine and John B. Donaldson. *Intermediate Financial Theory.* Academic Press, 2005.

Angus Deaton. *Understanding Consumption.* Oxford University Press, 1992.

Gerard Debreu. *Theory of Value - An Axiomatic Analysis of Economic Equilibrium.* Cowles Foundation, 1959.

Walter Dolde and James Tobin. Wealth, Liquidity, and Consumption. Cowles foundation discussion papers, Cowles Foundation, Yale University, 1971.

Irwin Friend and Marshall E. Blume. The Demand for Risky Assets. *American Economic Review*, 65:900–922, 1975.

Lars P. Hansen and Ravi Jagannathan. Implications of Security Market Data for Models of Dynamic Economies. *Journal of Political Economy*, 99:225–261, 1991.

Lars P. Hansen and Scott F. Richard. The Role of Conditioning Information in Deducing Testable Restrictions Implied by Dynamic Asset Pricing Models. *Econometrica*, 55:587–614, 1987.

Clifford Hildreth and Glenn J. Knowles. Some Estimates of Farmers' Utility Functions. Technical bulletins, University of Minnesota, Agricultural Experiment Station, 1982.

Jack Hirshleifer. Investment Decision under Uncertainty: Choice-Theoretic Approaches. *Quarterly Journal of Economics*, 79:509–536, 1965.

Jack Hirshleifer. Investment Decision under Uncertainty: Applications of the State Preference Approach. *Quarterly Journal of Economics*, 80:252–277, 1966.

Chifu Huang and Robert H. Litzenberger. *Foundations for Financial Economics*. Prentice Hall, 1988.

John C. Hull. *Options, Futures and Other Derivatives*. Prentice Hall, 2008.

Rudolph E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME–Journal of Basic Engineering*, 82(Series D):35–45, 1960.

Finn E. Kydland and Edward C. Prescott. Time to Build and Aggregate Fluctuations. *Econometrica*, 50:1345–1370, 1982.

Yvan Lengwiler. *Microfoundations of Financial Economics: An Introduction to General Equilibrium Asset Pricing*. University Presses of CA, 2006.

Lars Ljungqvist and Thomas J. Sargent. *Recursive Macroeconomic Theory*. MIT Press, 2004.

Robert E. Lucas. Asset Prices in an Exchange Economy. *Econometrica*, 46:1429–1445, 1978.

Harry M. Markowitz. Portfolio Selection. *Journal of Finance*, 7:77–99, 1952.

Rajnish Mehra. On the Existence and Representation of Equilibrium in an Economy with Growth and Nonstationary Consumption. *International Economic Review*, 29:131–135, 1988.

Rajnish Mehra. The Equity Premium Puzzle: A Review. *Foundations and Trends in Finance*, pages 1–81, 2008.

Robert Merton. An Intertemporal Asset Pricing Model. *Econometrica*, 41:867–887, 1973.

Edward C. Prescott and Rajnish Mehra. The Equity Premium: A Puzzle. *Journal of Monetary Economics*, 15:145–161, 1985.

Roy Radner. Existence of Equilibrium of Plans, Prices and Price Expectations in a Sequence of Markets. *Econometrica*, 40:289–303, 1972.

Harl E. Ryder and Geoffrey M. Heal. Optimum Growth with Intertemporally Dependent Preferences. *Review of Economic Studies*, 40:1–33, 1973.

Suresh M. Sunderasan. Intertemporally Dependent Preferences and the Volatility of Consumption and Wealth. *Review of Financial Studies*, 2:73–89, 1989.

Hal R. Varian. *Microeconomic Analysis*. Norton & Company, 1992.

Abraham Wald. Sequential Tests of Statistical Hypotheses. *Annals of Mathematical Statistics*, 16:117–186, 1945.

Leon Walras. *Elements of Pure Economics or the Theory of Social Wealth*. Orion Editions, 1874. Reprinted Version (1984).

Abstract:

Englisch

The contribution of Prescott and Mehra (1985) to the asset pricing literature triggered an enormous amount of research addressing the so called equity premium puzzle. In the following thesis I will briefly review the origins of asset pricing by presenting the seminal paper of Lucas (1978) and deriving simple closed-form asset pricing equations. In the second part of the thesis a brief review of the equity premium puzzle based on Prescott and Mehra (1985) and Hansen and Jagannathan (1991) will be given. Finally, I will discuss the approach of Campbell and Cochrane (1999) which tries to resolve the puzzle by introducing an alternative class of utility function that accounts for what is known as habit formation.

Deutsch

Der Beitrag von Prescott und Mehra (1985) zur Literatur im Bereich Asset Pricing war Auslöser für ein breites Feld, das als Equity Premium Puzzle innerhalb der volks- und finanzwirtschaftlichen Forschung bekannt ist. In der vorliegenden Arbeit bespreche ich kurz die Ursprünge das Themengebietes Asset Pricing mit der Vorstellung der Arbeit von Lucas (1978) und der Herleitung geschlossener Lösungen einfacher Asset Pricing Gleichungen. Im zweiten Teil der Arbeit wird das Equity Premium Puzzle auf Basis der Beiträge von Prescott und Mehra (1985) als auch Hansen und Jagannathan (1991) vorgestellt. Im finalen Teil meiner Arbeit diskutiere ich den Ansatz von Campbell und Cochrane (1999), welcher das Puzzle auf Basis einer alternativen Klasse von Nutzenfunktionen löst. Dieser Ansatz ist auch als Habit Formation Ansatz bekannt.

LEBENSLAUF

PERSÖNLICHE ANGABEN

Name	**BASTIAN OFFERMANN**
Adresse	**AM KAUERT 2C • D-41189 MÖNCHENGLADBACH**
Telefon	**+49-(0)2166-59055**
E-Mail	**Bastian.offermann@gmail.com**

Staatsangehörigkeit	Deutsch
Geburtsdatum	25.07.83

AUSBILDUNG

2003	Hugo Junkers Gymnasium, Mönchengladbach, Deutschland *Abitur*
2004-2009	Wirtschaftsuniversität Wien, Wien, Österreich *Studium der Betriebswirtschaftslehre*
2009-2011	Universität Wien, Wien, Österreich *Studium der Volkswirtschaftslehre*

PERSÖNLICHE FÄHIGKEITEN UND KOMPETENZEN

MUTTERSPRACHE	**DEUTSCH**
SONSTIGE SPRACHEN	
	ENGLISCH
• Lesen	ausgezeichnet
• Schreiben	ausgezeichnet
• Sprechen	ausgezeichnet
COMPUTERKENNTNISSE	Windows XP/Office XP, R, VBA