Asset Pricing with Heterogeneous Preferences*

Alexis Akira Toda[†‡]
Department of Economics, University of California San Diego

This Version: April 25, 2013

Abstract

Finding a stochastic discount factor that is robust to model misspecification is not trivial. I consider a general equilibrium model with many agents who can invest their wealth in many assets. As long as (i) agents have (individual-, time-, and state-dependent) recursive preferences that are homothetic in current consumption and continuation value with a *common* relative risk aversion coefficient γ and (ii) asset returns and individual state variables are conditionally independent (*e.g.*, GARCH processes), I prove that the $-\gamma$th power of market return is a valid stochastic discount factor. Within this class of models, asset prices are determined by relative risk aversion and technology alone, and "returns-based asset pricing" is robust to model misspecification as opposed to the consumption-based approach. Using the historical returns on portfolios of U.S. stocks sorted by size and book-to-market value, I find that a relative risk aversion coefficient of around 2 explains asset returns. The conditional and unconditional moment restrictions are not rejected. I recast the equity premium puzzle as a macroeconomics puzzle, not as a finance puzzle.

Keywords: *AK* models; consumption volatility puzzle; equity premium puzzle; model misspecification; power law; recursive preferences; risk-free rate puzzle.

JEL codes: D53, D58, D91, G11, G12.

1 Introduction

In asset pricing theory, it is well known that the "returns-based approach" (form a statistical model of bond and stock returns, solve the optimal consumption-portfolio decision. Use the equilibrium consumption value in $p = \mathrm{E}[mx]$) is equivalent to the "consumption-based approach" (form a statistical model of the consumption process, calculate asset prices and returns directly from the basic pricing equation $p = \mathrm{E}[mx]$), *given the model.*[1] Classic examples of the former

*I thank Tony Smith for suggesting the title.

†Email: alexis.akira.toda@gmail.com

‡I thank Truman Bewley, John Geanakoplos, and Tony Smith for continual support during my studies at Yale. I benefited from conversations with Tim Armstrong, Brendan Beare, Xiaohong Chen, Gabriele Foá, Yoichi Goto, Lynda Khalaf, Gregory Phelan, Eric Renault, Zhentao Shi, Rami Tabri, Kieran Walsh, and seminar participants at Yale. The financial supports from the Cowles Foundation, the Nakajima Foundation, and Yale University are greatly acknowledged.

[1]The definitions of the two approaches are cited from Cochrane (2005), p. 40. Here p, m, x denote the asset price, stochastic discount factor, and asset payoff.

approach are Markowitz (1952), Tobin (1958), Sharpe (1964), Lintner (1965a,b), Samuelson (1969), Merton (1969, 1971, 1973), and Fama (1970), and examples of the latter approach are Lucas (1978) and Mehra and Prescott (1985), just to name a few. In this paper I argue that the returns-based approach is fairly robust to model misspecification, while the consumption-based approach is not. The key is to "bypass consumption data altogether, and instead look directly at asset returns" (Ludvigson, 2013).

To illustrate the point in the simplest possible way, consider the following example. There is an investor who lives for two periods with the additive constant relative risk aversion (CRRA) utility function

$$\frac{1}{1-\gamma}\left(c_0^{1-\gamma}+\beta\,\mathrm{E}[c_1^{1-\gamma}]\right).$$

We all know that $\beta(c_1/c_0)^{-\gamma}$ is a valid stochastic discount factor (SDF).

Now suppose that the investor is endowed with initial wealth $w>0$ today and nothing tomorrow, but can invest in K assets indexed by $k=1,\dots,K$. Asset k has gross return $R^k\ge 0$, which is a random variable. Letting ϕ^k be the fraction of the remaining wealth invested in asset k, $\phi=(\phi^1,\dots,\phi^K)\in\mathbb{R}^K$ (where $\sum_k\phi^k=1$) the portfolio, and $R(\phi)=\sum_k R^k\phi^k$ the gross return on portfolio ϕ, the budget constraint is $c_1=R(\phi)(w-c_0)$. Substituting the budget constraint into the utility function, the optimal consumption-portfolio problem becomes

$$\max_{c,\phi}\frac{1}{1-\gamma}\left(c^{1-\gamma}+\beta\,\mathrm{E}[R(\phi)^{1-\gamma}](w-c)^{1-\gamma}\right),$$

which can be broken into

$$\frac{F}{1-\gamma}:=\max_{\phi}\frac{1}{1-\gamma}\,\mathrm{E}[R(\phi)^{1-\gamma}], \tag{1.1a}$$

$$U:=\max_{c}\frac{1}{1-\gamma}(c^{1-\gamma}+\beta F(w-c)^{1-\gamma}). \tag{1.1b}$$

Let ϕ^* be the solution to the optimal portfolio problem (1.1a) and consider investing ϵ more in asset k and ϵ less in the optimal portfolio ϕ^*. Taking the first order condition with respect to ϵ and setting $\epsilon=0$, we obtain

$$\mathrm{E}[R(\phi^*)^{-\gamma}(R^k-R(\phi^*))]=0$$

for any asset k, so $R(\phi^*)^{-\gamma}$ (times a constant) is also a valid stochastic discount factor.

Note that the SDF $\beta(c_1/c_0)^{-\gamma}$ is *not* robust to model misspecification: if we change the utility function, so does the SDF. However, the SDF $R(\phi^*)^{-\gamma}$ *is* robust, because the only property we used to derive it is the homotheticity of the utility function, not its particular functional form. The rest of the paper is an elaboration of this simple idea.

This paper has two contributions. First, I solve an optimal consumption-portfolio problem similar in spirit to Samuelson (1969) but in a very general setting, namely the agent has an arbitrary time- and state-dependent homothetic recursive preference with constant relative risk aversion, the number of assets is arbitrary, and the only distributional assumption is that asset returns and state variables be conditionally independent. The assumptions are weak enough for my results to have a wide range of applicability. For example, the standard additive CRRA utility and the CRRA-constant elasticity of intertemporal substitution (CEIS) recursive utility (Epstein and Zin, 1989) are all special cases, and the period utility function may include some other state variables such as past consumption. For distributional assumptions, the expected return and volatility can follow any stochastic process as long as excess returns are serially independent, for example GARCH processes. Under these assumptions, by using the value function approach instead of the Euler equation approach, the optimal portfolio decision and the optimal consumption/saving decision can be disentangled as in (1.1).

Second, which is the main contribution, I consider an economy with many such agents and show that if (i) agents have a *common* relative risk aversion coefficient (but recursive preferences that are possibly individual-, time-, and state-dependent) and (ii) the efficient market hypothesis holds, then agents make the same portfolio choice, and therefore the individually optimal portfolio must be the market portfolio. A corollary is that the $-\gamma$th power of the gross return on the market portfolio (market return) is a valid stochastic discount factor.

This result has three important implications. First, since its validity does not depend on any particular utility function and hence on the consumption process, the "returns-based asset pricing" approach is robust to model misspecification as opposed to the consumption-based approach. Since in my model consumption is not directly connected to asset prices, the low volatility of consumption growth (or the low covariance between consumption growth and asset returns) needed in order to explain asset prices ("consumption volatility puzzle") is not an asset pricing puzzle (that belongs to finance) but a consumption/saving puzzle (that belongs to macroeconomics). Second, since the asset pricing formula contains only asset returns data, which are available in high frequency and high accuracy, the '$-\gamma$th power of market return' SDF can be used in practice. Third, the relative risk aversion γ can be estimated using only asset returns data: (aggregate or individual) consumption data contain no more information than the asset returns data for estimating the relative risk aversion coefficient.

Although these results concern relative pricing, I also consider absolute pricing. By assuming further that (iii) agents have access to constant-returns-to-scale stochastic saving technologies (AK model, *e.g.*, Levhari and Srinivasan (1969)) and (iv) technological shocks and individual state variables are conditionally independent, I derive an asset pricing formula which depends only on fundamentals.

A few papers are related to my work. Rubinstein (1976) derived the '$-\gamma$th power of market return' SDF under the assumption of a representative agent with additive CRRA utility and serially independent returns. I obtain the same SDF, but under much weaker assumptions listed above. Most importantly, in Rubinstein's model aggregate consumption is proportional to wealth and hence consumption growth and market return have the same volatility (which is obviously counterfactual, hence "consumption volatility puzzle"), but in my model aggregate consumption is not connected to market return. Campbell (1993) obtained an asset pricing formula without using consumption in a representative agent setting by log linearizing the intertemporal budget constraint. In my model there are many heterogeneous agents with more general preferences and the asset pricing formula is exact, not an approximation.

Cass and Stiglitz (1970) showed in a static setting that the only utility functions for which the mutual fund theorem holds are the quadratic and power utility functions (if there is no risk-free asset) and the linear risk tolerance (LRT) utility $-\frac{u'(w)}{u''(w)} = A + Bw$ (if there is a risk-free asset). Rubinstein (1974) showed a similar result in a two period economy with additive utility functions. My result extends theirs to the multi period setting with recursive preferences. Constantinides (1982) proved that if agents have additively time- and state-separable utility functions (without state variables) in a complete market endowment economy, then we can define a representative agent who consumes the aggregate endowment and prices the assets. Here the utility function of the representative agent in general depends on the preferences of all agents in a complicated way, and hence so do the asset prices. In my model, the relative risk aversion must be common across all agents, the function that defines the recursive utility must be homogeneous of degree 1, but I allow individual-, time-, and state-dependent recursive utility. Most importantly, the asset prices depend only on the common relative risk aversion γ and the technologies, not on other preference characteristics. Therefore my results neither contain nor are contained in those of Constantinides (1982), but are complementary.

The rest of the paper is organized as follows. Section 2 presents the model and solves the single agent optimal consumption-portfolio problem. Section 3 derives relative asset pricing formulas that do not depend on consumption in a partial equilibrium setting. Section 4 characterizes the general equilibrium with many heterogeneous agents and constant-returns-to-scale stochastic saving technologies, and derives absolute asset pricing formulas. Section 5 tests the asset pricing implications of the model. Section 6 discusses the asset pricing puzzles.

2 Individual decision

All random variables are defined on a probability space $(\Omega, \mathcal{F}, P)$. Time is discrete and finite,[2] $t = 0, 1, \dots, T$. An agent starts with initial wealth $w > 0$ and has no income other than those obtained by investing in assets.[3]

2.1 Assets, information, and preference

Assets There are K assets indexed by $k \in K = \{1, \dots, K\}$. Let P_t^k, D_t^k be the price and dividend of asset k at time t. The gross return of asset k between the end of time t and the beginning of time $t+1$ is denoted by $R_{t+1}^k = (P_{t+1}^k + D_{t+1}^k)/P_t^k$, and the vector of gross asset returns is denoted by

$$\mathbf{R}_{t+1} = (R_{t+1}^1, \dots, R_{t+1}^K).$$

Let ϕ_t^k be the fraction of wealth invested in asset k at time t and $\phi_t = (\phi_t^1, \dots, \phi_t^K)$ be the portfolio, so $\sum_k \phi_t^k = 1$. Of course, $\phi_t^k > 0 (< 0)$ means a long (short) position in asset k. The agent can be constrained in the choice of portfolio: let

[2]The model can be easily generalized to infinite horizon if we assume that each agent lives for only finite periods or if we make distributional assumptions (such as Markov shocks) to guarantee the convergence of the value function. See Toda (2012) for such an example.

[3]Since this paper is concerned with frictionless complete asset markets, the agent can sell off his future endowments and incorporate into the initial wealth.

$\Pi_t \subset \mathbb{R}^K$ be the set of feasible portfolios. The gross return on portfolio $\phi_t \in \Pi_t$ is denoted by

$$R_{t+1}(\phi_t) := \mathbf{R}'_{t+1}\phi_t = \sum_{k=1}^{K} R^k_{t+1}\phi^k_t.$$

The sequential budget constraint of the agent is therefore

$$(\forall t)\ w_{t+1} = R_{t+1}(\phi_t)(w_t - c_t) \geq 0.$$

Information and preference The agent's information is represented by the filtration (an increasing sequence of σ-algebras) $\{\mathcal{F}_t\}_{t=0}^T \subset \mathcal{F}$. Let w_t be the agent's wealth at the beginning of time t and $\mathbf{X}_t = (X^1_t, X^2_t, \dots)$ be the vector of state variables at time t *other than wealth*. What I have in mind for the state variables are public information such as past returns and volatility, but it may also include private information such as past consumption (in the case of habit formation). To obtain the results it is unnecessary to specify $\mathbf{X}_t$ explicitly. The conditional expectation with respect to time t information is denoted by $\mathrm{E}[\cdot \mid \mathcal{F}_t]$ or more compactly $\mathrm{E}_t[\cdot]$, which are functions of $\mathbf{X}_t$ and w_t because by assumption these are all the state variables. Let $c_t, U_t \in \mathbb{R}$ be the consumption and the continuation utility at time t. I make the following assumptions.

Assumption 1 (Irrelevance of wealth)**.** *For any $\mathcal{F}_{t+1}$-measurable function f, we have $\mathrm{E}[f(\mathbf{X}_{t+1}) \mid \mathcal{F}_t] = g(\mathbf{X}_t)$ for some g, that is, agent's wealth is irrelevant for predicting a function of next period's state variables other than wealth.*

Assumption 1 simply means that the agent is so small compared to the market that his wealth level does not affect asset returns, *i.e.*, the agent is a price taker.

Assumption 2 (Constant relative risk aversion and homotheticity)**.** *The continuation utilities $\{U_t\}_{t=0}^T$ satisfy the recursion $U_T = a_T(\mathbf{X}_T)c_T$ and*

$$U_t = f_t\left(c_t, \mathrm{E}\left[U_{t+1}^{1-\gamma} \,\middle|\, \mathcal{F}_t\right]^{\frac{1}{1-\gamma}}, \mathbf{X}_t\right), \quad t = 0, \dots, T-1, \tag{2.1}$$

where $a_T > 0$ is some function of the state variables $\mathbf{X}_T$, $\gamma > 0$ is the relative risk aversion coefficient, and

$$f_t : \mathbb{R}_{++} \times \mathbb{R}_{++} \times \mathbb{R}^{\dim \mathbf{X}_t} \to \mathbb{R}_+$$

is strictly increasing and homogeneous of degree 1 in the first two arguments.

f_t is called the *aggregator* (Epstein and Zin, 1989; Boyd, 1990). Since the risk aversion is over the continuation utility, not consumption, it is the correct notion of risk aversion (Swanson, 2012). At this point it is helpful to provide concrete examples.

Example 1 (Additive CRRA utility)**.** If $a_T(\mathbf{X}_T) = 1$ and the aggregator is given by

$$f_t(c, v, \mathbf{X}_t) = (c^{1-\gamma} + \beta v^{1-\gamma})^{\frac{1}{1-\gamma}}$$

(so the state variables do not directly enter the aggregator), then iterating (2.1) and using the law of iterated expectations, we obtain

$$U_t = \mathrm{E}\left[\sum_{s=t}^{T} \beta^{t-s} c_s^{1-\gamma} \,\middle|\, \mathcal{F}_t\right]^{\frac{1}{1-\gamma}},$$

which is ordinarily equivalent to the standard additive CRRA utility

$$\mathrm{E}_t \sum_{s=t}^{T} \beta^{t-s} \frac{c_s^{1-\gamma}}{1-\gamma}$$

with discount factor β and relative risk aversion γ.

Example 2 (Recursive CRRA/CEIS utility)**.** If $a_T(\mathbf{X}_T) = 1$ and the aggregator is given by

$$f_t(c, v, \mathbf{X}_t) = (c^{1-\sigma} + \beta v^{1-\sigma})^{\frac{1}{1-\sigma}}$$

(so the state variables do not directly enter the aggregator), then U_t is the constant relative risk aversion (CRRA), constant elasticity of intertemporal substitution (CEIS) recursive utility (Epstein and Zin, 1989) with discount factor β, relative risk aversion γ, and elasticity of intertemporal substitution $1/\sigma$.

Example 3 (Habit formation)**.** In Examples 1 and 2, the aggregator f_t did not explicitly depend on the state variables $\mathbf{X}_t$, but (2.1) allows such dependence. For example, if $\mathbf{X}_t$ consists of past consumption and the aggregator explicitly depends on $\mathbf{X}_t$, the recursive utility (2.1) depends on past consumption and hence we can incorporate some form of habit formation (Abel, 1990). One such example that satisfies Assumption 2 is

$$f_t(c, v, x) = \left[(c/x)^{1-\sigma} + \beta v^{1-\sigma}\right]^{\frac{1}{1-\sigma}},$$

where x is the habit stock.

2.2 Optimal portfolio problem

To solve the optimal consumption-portfolio problem I further need an assumption on asset returns and state variables.

Assumption 3 (Conditional independence)**.** *For each t, the next period's state variables $\mathbf{X}_{t+1}$ and asset returns $\mathbf{R}_{t+1}$ are independent conditional on time t information $\mathcal{F}_t$.*

Conditional independence implies, in particular, that the most recent asset return is not a state variable: $\mathbf{R}_t \notin \mathbf{X}_t$, which is clearly a restriction. An obvious case in which conditional independence holds is when returns are i.i.d. and independent of state variables. However, the assumption is still weak enough to be useful. For example, suppose that returns are lognormal with time-varying expected return and volatility: $\log R_{t+1} \sim N(\mu_t, \sigma_t^2)$. Here the state variable is $\mathbf{X}_t = (\mu_t, \sigma_t)$. Conditional independence holds if, for instance, the expected return-volatility pair $\{\mathbf{X}_t\}$ is a Markov process and $\log R_{t+1} = \mu_t + \sigma_t z_{t+1}$, where $\{z_t\}$ is a Gaussian white noise that is independent from the process $\{\mathbf{X}_t\}$.[4]

[4]In this example I implicitly assumed that there is a single risky asset, but the argument clearly holds for any number of assets.

Another example is the GARCH process with no leverage effect. Let $\log R_{t+1} = \mu + \epsilon_{t+1}$ and consider the GARCH(p,q) process

$$\begin{aligned} \epsilon_{t+1} &= \sigma_t z_{t+1}, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \epsilon_t^2 + \cdots + \alpha_q \epsilon_{t-q+1}^2 + \beta_1 \sigma_{t-1}^2 + \cdots + \beta_p \sigma_{t-p}^2, \end{aligned}$$

where $\{z_t\}$ is a white noise. Then the state variables are

$$\mathbf{X}_t = (\epsilon_t, \dots, \epsilon_{t-q+1}, \sigma_{t-1}, \dots, \sigma_{t-p}),$$

and the conditional independence assumption does not necessarily hold because ϵ_{t+1} (part of next period's state variables) and $R_{t+1} = \exp(\mu + \epsilon_{t+1})$ are not independent conditional on $\mathbf{X}_t$. However, if $\alpha_1 = 0$ (no leverage effect), then ϵ_t is no longer a state variable, and conditional independence holds.

The following theorem shows that the optimal portfolio problem can be disentangled from the optimal consumption/saving problem, and that the former depends only on risk aversion and asset returns.

Theorem 2.1. *Under Assumptions 1–3, the value function*

$$\begin{aligned} &V_t(w, \mathbf{X}_t) \\ &= \sup \{U_t \mid w_t = w, (\forall s \geq t)\ w_{s+1} = R_{s+1}(\phi_s)(w_s - c_s) \geq 0, \phi_s \in \Pi_s\} \end{aligned} \tag{2.2}$$

is linear in wealth w and the optimal portfolio problem at time t reduces to

$$\max_{\phi \in \Pi_t} \frac{1}{1-\gamma} \mathrm{E}\left[R_{t+1}(\phi)^{1-\gamma} \,\middle|\, \mathcal{F}_t\right]. \tag{2.3}$$

If the portfolio constraint Π_t is nonempty, compact, and

$$\mathrm{E}\left[\sup_{\phi \in \Pi_t} R_{t+1}(\phi)^{1-\gamma} \,\middle|\, \mathcal{F}_t\right] < \infty,$$

then the optimal portfolio problem (2.3) *has a solution.*

Proof. The proof is by induction. If $t = T$, then $U_T = a_T(\mathbf{X}_T)c_T$, so

$$V_T(w, \mathbf{X}_T) = \sup\{a_T(\mathbf{X}_T)c_T \mid c_T \leq w\} = a_T(\mathbf{X}_T)w$$

is linear in wealth and there are no portfolio decisions to make. Suppose the claim is true for time $s = t+1, \dots, T$ and let $V_s(w, \mathbf{X}_s) = a_s(\mathbf{X}_s)w$. Then we obtain

$$\begin{aligned} &V_t(w, \mathbf{X}_t) \\ &= \sup_{\substack{0 \leq c \leq w \\ \phi \in \Pi_t}} f_t\left(c, (w-c)\,\mathrm{E}_t[a_{t+1}(\mathbf{X}_{t+1})^{1-\gamma} R_{t+1}(\phi)^{1-\gamma}]^{\frac{1}{1-\gamma}}, \mathbf{X}_t\right) \\ &= \sup_{0 \leq c \leq w} f_t\left(c, (w-c)\,\mathrm{E}_t[a_{t+1}(\mathbf{X}_{t+1})^{1-\gamma}]^{\frac{1}{1-\gamma}} \sup_{\phi \in \Pi_t} \mathrm{E}_t[R_{t+1}(\phi)^{1-\gamma}]^{\frac{1}{1-\gamma}}, \mathbf{X}_t\right) \\ &= \sup_{0 \leq c \leq w} f_t\left(c, (w-c) b_t(\mathbf{X}_t) \sup_{\phi \in \Pi_t} \mathrm{E}_t[R_{t+1}(\phi)^{1-\gamma}]^{\frac{1}{1-\gamma}}, \mathbf{X}_t\right) \\ &= \sup_{0 \leq \tilde{c} \leq 1} w f_t\left(\tilde{c}, (1-\tilde{c}) b_t(\mathbf{X}_t) \sup_{\phi \in \Pi_t} \mathrm{E}_t[R_{t+1}(\phi)^{1-\gamma}]^{\frac{1}{1-\gamma}}, \mathbf{X}_t\right) =: a_t(\mathbf{X}_t) w, \end{aligned}$$

where I used backward induction in the first equality, conditional independence (Assumption 3) and monotonicity of f_t in the second, the irrelevance of wealth (Assumption 1) in the third, and the homogeneity of f_t (Assumption 2) in the last, where I set $\tilde{c} = c/w$. Therefore the value function is linear in wealth. Since f_t is increasing in the second argument, the optimal portfolio problem at time t is

$$\max_{\phi \in \Pi_t} \mathrm{E}\left[R_{t+1}(\phi)^{1-\gamma} \,\middle|\, \mathcal{F}_t\right]^{\frac{1}{1-\gamma}},$$

which is equivalent to (2.3) because $x \mapsto \frac{x^{1-\gamma}}{1-\gamma}$ is monotone.

If $\mathrm{E}\left[\sup_{\phi \in \Pi_t} R_{t+1}(\phi)^{1-\gamma} \,\middle|\, \mathcal{F}_t\right] < \infty$,[5] then by the Lebesgue convergence theorem $\phi \mapsto \mathrm{E}_t[R_{t+1}(\phi)^{1-\gamma}]$ is continuous. Therefore if the portfolio constraint Π_t is nonempty and compact, the optimal portfolio problem (2.3) has a solution. □

Theorem 2.1 is related to Kocherlakota (1990), where he proves in a representative agent, complete markets, endowment economy setting that the CRRA/CEIS recursive utility model (Example 2) is observationally equivalent to the standard additive CRRA utility model if consumption growth is i.i.d. His irrelevance result can be generalized as in the following proposition.

Proposition 2.2. *Consider the recursive utility model satisfying Assumption 2 with a time-homogeneous aggregator $f(c, v)$ with no state variables. If asset returns are i.i.d. and U_0 defined by* (2.1) *converges as $T \to \infty$, then the recursive utility model is observationally equivalent to the standard additive CRRA utility model.*

Proof. It suffices to show that the optimal portfolio choice and consumption are observationally equivalent in the two models. By Theorem 2.1, the portfolio choice is the same. If the recursive utility converges as time periods tends to infinity, the Bellman equation becomes time-homogeneous. Since the aggregator $f(c, v)$ is homogeneous of degree 1, the optimal consumption is a constant fraction of wealth, which is observationally equivalent to the additive CRRA case. □

3 Partial equilibrium

Having solved the single agent problem, in this section I consider an economy with many agents. In a partial equilibrium setting, I derive a relative asset pricing formula that depends only on the market portfolio and the relative risk aversion.

3.1 Description of the economy

The financial market is the same as in Section 2, so asset k has (per share) price P_t^k, dividend D_t^k, and gross return $R_{t+1}^k = (P_{t+1}^k + D_{t+1}^k)/P_t^k$. Let W_t^k be the market capitalization (per share price P_t^k times the number of shares outstanding) of asset k.

[5] This condition is not very stringent. For example, it holds if $\gamma > 1$ (< 1) and the portfolio return is bounded away from zero (bounded above).

The economy is populated by I agents indexed by $i \in I = \{1, \dots, I\}$ with recursive preferences defined by (2.1), where the aggregators $\{(f_{it})_{i\in I}\}_{t=0}^{T-1}$ and the state variables $\{(\mathbf{X}_{it})_{i\in I}\}_{t=0}^{T}$ are potentially different but the relative risk aversion $\gamma > 0$ and the portfolio constraint $\Pi_t \subset \mathbb{R}^K$ are *common* across agents. Agent i is endowed with initial wealth $w_{i0} > 0$ but nothing thereafter. Let $\mathcal{F}_{it}$ be the private information of agent i at time t and $\mathcal{F}_t = \bigcap_i \mathcal{F}_{it}$ be the public information.

The sequential partial equilibrium is defined by agent optimization and market clearing.

Definition 3.1 (Sequential partial equilibrium)**.** Given asset prices, dividends, and market capitalization $\left\{(P_t^k, D_t^k, W_t^k)_{k\in K}\right\}_{t=0}^{T}$, the profile of individual consumption, wealth, and portfolio $\{(c_{it}, w_{it}, \phi_{it})_{i\in I}\}_{t=0}^{T}$ constitutes a sequential partial equilibrium if

1. given asset returns $R_{t+1}^k = (P_{t+1}^k + D_{t+1}^k)/P_t^k$, the portfolio ϕ_{it} solves

$$\max_{\phi\in\Pi_t} \frac{1}{1-\gamma} \operatorname{E}\left[R_{t+1}(\phi)^{1-\gamma} \,\middle|\, \mathcal{F}_{it}\right], \tag{3.1}$$

2. given the portfolio choice, c_{it} solves the optimal consumption problem (2.2),

3. asset markets clear, *i.e.*, for each asset k and time t we have $\sum_{i=1}^{I} \phi_{it}^k(w_{it} - c_{it}) = W_t^k$, and

4. individual wealth evolves according to the budget constraint

$$w_{i,t+1} = R_{t+1}(\phi_{it})(w_{it} - c_{it}).$$

3.2 Relative asset pricing

In order to prove the main result, I need one more assumption. I assume markets are efficient in the sense that private information is useless for predicting asset returns.

Assumption 4 (Efficient market hypothesis)**.** *For each i and t, the distribution of asset returns $\mathbf{R}_{t+1} = (R_{t+1}^k)_{k=1}^K$ conditional on private information $\mathcal{F}_{it}$ is the same as the distribution conditional on public information $\mathcal{F}_t$.*

This definition of market efficiency is taken from the first definition in Bewley (1982). The following proposition shows that if there is an equilibrium, there is also an equivalent symmetric equilibrium (common portfolio choice).

Proposition 3.2. *Let everything be as above. Suppose that*

1. *agents have information and recursive preferences satisfying Assumptions 1 and 2,*

2. *for each agent conditional independence (Assumption 3) holds, and*

3. *the efficient market hypothesis (Assumption 4) holds.*

If there is a partial equilibrium, then there is also an equilibrium with a common portfolio choice ϕ_t^* (market portfolio) and the same consumption and wealth as in the original equilibrium $\{(c_{it}, w_{it})_{i\in I}\}_{t=0}^T$.

Proof. By the efficient market hypothesis (Assumption 4), we can replace the private information $\mathcal{F}_{it}$ in (3.1) by the public information $\mathcal{F}_t$. Then the optimal portfolio problem becomes common across all agents, which is (2.3).

Suppose that $\{(c_{it}, w_{it}, \phi_{it})_{i\in I}\}_{t=0}^T$ is a sequential partial equilibrium. Define the value weighted average portfolio by

$$\bar{\phi}_t := \sum_{i=1}^I \phi_{it}(w_{it} - c_{it}) / \sum_{i=1}^I (w_{it} - c_{it}).$$

By the definition of $\bar{\phi}_t$ and the market clearing condition, we have

$$\sum_{i=1}^I \bar{\phi}_t^k (w_{it} - c_{it}) = \sum_{i=1}^I \phi_{it}^k (w_{it} - c_{it}) = W_t^k$$

for each k, so the common portfolio $\bar{\phi}_t$ (market portfolio) clears the market. Since the function $\frac{1}{1-\gamma} R_{t+1}(\phi)^{1-\gamma}$ is quasi-concave in ϕ and ϕ_{it} solves (2.3) for each i, so does $\bar{\phi}_t$. Therefore $\left\{(c_{it}, w_{it}, \bar{\phi}_t)_{i\in I}\right\}_{t=0}^T$ (same consumption and wealth as in the original equilibrium with common portfolio $\bar{\phi}_t$) is also an equilibrium. ☐

Let $\phi_t^* := \bar{\phi}_t$ be the market portfolio, which is also an individually optimal portfolio. The following theorem, which is the main result of this paper, shows that the $-\gamma$th power of the return on the market portfolio is a valid stochastic discount factor.

Theorem 3.3. *Let everything be as in Proposition 3.2 and $\{(c_{it}, w_{it}, \phi_t^*)_{i\in I}\}_{t=0}^T$ be a symmetric sequential partial equilibrium, where ϕ_t^* is the market portfolio. If the portfolio constraint $\phi \in \Pi_t$ does not bind at the market portfolio ϕ_t^* for asset k, letting $R_{m,t+1} = R_{t+1}(\phi_t^*)$ be the return on the market portfolio, we have*

$$\mathrm{E}\left[R_{m,t+1}^{-\gamma}(R_{t+1}^k - R_{m,t+1}) \,\middle|\, \mathcal{F}_t\right] = 0, \tag{3.2a}$$

$$P_t^k = \frac{\mathrm{E}\left[R_{m,t+1}^{-\gamma}(P_{t+1}^k + D_{t+1}^k) \,\middle|\, \mathcal{F}_t\right]}{\mathrm{E}\left[R_{m,t+1}^{1-\gamma} \,\middle|\, \mathcal{F}_t\right]}, \tag{3.2b}$$

i.e., *the $-\gamma$th power of the return on the market portfolio is a valid stochastic discount factor. In particular, the one period risk-free rate is*

$$R_{f,t} = \frac{\mathrm{E}\left[R_{m,t+1}^{1-\gamma} \,\middle|\, \mathcal{F}_t\right]}{\mathrm{E}\left[R_{m,t+1}^{-\gamma} \,\middle|\, \mathcal{F}_t\right]}. \tag{3.3}$$

Furthermore, the equity premium satisfies the CAPM-like formula

$$\mathrm{E}\left[R_{t+1}^k \,\middle|\, \mathcal{F}_t\right] - R_{f,t} = -\frac{\mathrm{Cov}\left[R_{m,t+1}^{-\gamma}, R_{t+1}^k \,\middle|\, \mathcal{F}_t\right]}{\mathrm{E}\left[R_{m,t+1}^{-\gamma} \,\middle|\, \mathcal{F}_t\right]}. \tag{3.4}$$

Proof. Consider investing the fraction of wealth $1-\alpha$ in the market portfolio ϕ_t^* and α in asset k. Clearly $\alpha = 0$ is optimal by the definition of ϕ_t^*, so

$$0 \in \arg\max_{\alpha} \frac{1}{1-\gamma} \mathrm{E}\left[[(1-\alpha)R_{m,t+1} + \alpha R_{t+1}^k]^{1-\gamma} \,\middle|\, \mathcal{F}_t\right]. \tag{3.5}$$

Since by assumption the portfolio constraint $\phi \in \Pi_t$ does not bind, by taking the first-order condition of the maximization (3.5) at the optimum $\alpha = 0$, we obtain (3.2a). Substituting $R_{t+1}^k = (P_{t+1}^k + D_{t+1}^k)/P_t^k$ into (3.2a) and rearranging terms, we obtain (3.2b). Setting $P_{t+1}^k = 0$ and $D_{t+1}^k = 1$ in (3.2b), we obtain the price of the one period risk-free bond $1/R_t^f$, and hence (3.3).

To derive (3.4), let M_{t+1} be any stochastic discount factor ($M_{t+1} = R_{m,t+1}^{-\gamma}$ in our case). Suppressing the time subscript, (3.2a) becomes $\mathrm{E}\left[M(R^k - R_m) \,\middle|\, \mathcal{F}\right] = 0$. Setting R^k be the risk-free rate R_f, we get $\mathrm{E}\left[M(R_f - R_m) \,|\, \mathcal{F}\right] = 0$. Taking the difference of the two equations, we get $\mathrm{E}\left[M(R^k - R_f) \,\middle|\, \mathcal{F}\right] = 0$. Using $\mathrm{E}\left[XY \,|\, \mathcal{F}\right] = \mathrm{Cov}\left[X, Y \,|\, \mathcal{F}\right] + \mathrm{E}\left[X \,|\, \mathcal{F}\right]\mathrm{E}\left[Y \,|\, \mathcal{F}\right]$ for $X = M$ and $Y = R^k - R_f$ and rearranging terms, we obtain

$$\mathrm{E}\left[R^k \,\middle|\, \mathcal{F}\right] - R_f = -\frac{\mathrm{Cov}\left[M, R^k \,\middle|\, \mathcal{F}\right]}{\mathrm{E}\left[M \,|\, \mathcal{F}\right]}, \tag{3.6}$$

which is (3.4) for $M = R_m^{-\gamma}$. $\square$

Theorem 3.3 may appear completely standard at first glance, but it is not. In a consumption-based representative agent setting (with a standard additive CRRA utility function), the growth rate of consumption is proportional to the return on the market portfolio, and (3.2a) is trivial (it is the Euler equation). What is surprising is that despite the presence of many agents with heterogeneous preferences (that may violate the sufficient condition for the existence of the representative agent as in Constantinides (1982): agents have very general preferences as in Assumption 2), I derived a simple stochastic discount factor (SDF), $R_m^{-\gamma}$, which depends only on the relative risk aversion and the market portfolio.

This result has three important implications. First, since this result does not depend on any particular utility function and hence on the aggregate or individual consumption process, the "returns-based asset pricing" approach is robust to misspecification of the model as opposed to the consumption-based approach. Since in my model consumption is not directly connected to asset prices, the low volatility of consumption growth (or the low covariance between consumption growth and asset returns) needed in order to explain asset prices ("consumption volatility puzzle") is not an asset pricing puzzle (that belongs to finance) but a consumption/saving puzzle (that belongs to macroeconomics).

Second, since the asset pricing formula contains only asset returns data, which are available in high frequency and high accuracy, my model can be used in practice. CAPM can also be interpreted as an approximation to my model. To see this, note that the (conditional) CAPM implies the existence of numbers a_t, b_t such that (3.2a) holds by replacing $R_{m,t+1}^{-\gamma}$ with $a_t - b_t R_{m,t+1}$. But by Taylor expanding $R^{-\gamma}$ around $R = 1$, we get

$$R^{-\gamma} \approx 1 - \gamma(R-1) = 1 + \gamma - \gamma R,$$

so setting $a_t = 1+\gamma$ and $b_t = \gamma$, CAPM is a linear approximation of my discount factor.

Third, the relative risk aversion γ can be estimated by GMM using only asset returns data, which (unlike consumption) are highly accurate and available in high frequency. The commonly used Euler equation, for example, does not contain more information than (3.2a) for estimating γ even if the Euler equation is true (*i.e.*, the model is correctly specified). This means that the rejection of a particular model using consumption data should be interpreted as the rejection of the particular specification of the model rather than the rejection of the asset pricing implications of the model.

To the best of my knowledge, documenting the robustness of the '$-\gamma$th power of market return' SDF seems to be new. The closest expression I found in the literature is Rubinstein (1976), in which he obtains the same discount factor, but assuming (i) a representative agent with an additive CRRA utility function, (ii) single asset, and (iii) independent returns. In testing the CRRA/CEIS recursive utility model of Example 2, Epstein and Zin (1991) derived the following equation:

$$\mathrm{E}\left[(c_{t+1}/c_t)^{-\frac{\sigma(1-\gamma)}{1-\sigma}} R_{m,t+1}^{\frac{\sigma-\gamma}{1-\sigma}}(R^k_{t+1} - R_{m,t+1}) \,\middle|\, \mathcal{F}_t\right] = 0, \tag{3.7}$$

where $1/\sigma$ is the elasticity of intertemporal substitution and I have changed their notation to be compatible with mine. Since (3.2a) obtains by setting $\sigma = 0$ in (3.7), (3.2a) is a stronger implication. However, (3.2a) holds with much more general preferences than CRRA/CEIS recursive utility (in particular, (3.2a) is true with any σ). Therefore my result is sharper despite the assumption being weaker.

Dittmar (2002) stresses the importance of a nonlinear pricing kernel, but his specification is based on a representative agent model and the coefficients are hard to interpret.

4 General equilibrium

This section deals with absolute pricing in a general equilibrium setting. I introduce firms and financial assets (assets that are in zero net supply) and derive asset pricing formulas.

4.1 Description of the economy

Firms and assets There is a single perishable good which can be consumed or invested as capital. There are J firms indexed by $j \in J = \{1, \dots, J\}$. Production takes time and exhibits constant returns to scale. If firm j employs capital K at the end of period t, it produces $A^j_{t+1}K$ at the beginning of period $t+1$, where A^j_{t+1} is the (random) productivity as well as the total return of capital after depreciation. In particular, if an agent invests one unit of capital in firm j at time t, he will receive A^j_{t+1} at the beginning of the next period. We can think of firms as stochastic saving technologies. Let $\mathbf{A}_{t+1} = (A^1_{t+1}, \dots, A^J_{t+1})$ be the vector of productivities.

There are K assets in zero net supply indexed by $k \in K = \{1, \dots, K\}$, with dividend D^k_t at period t (which is, of course, a random variable). Letting

P_t^k be the price of asset k in period t (determined in equilibrium), the gross return between periods t and $t+1$ is defined by $R_{t+1}^k = (P_{t+1}^k + D_{t+1}^k)/P_t^k$. Let $\mathbf{D}_t = (D_t^1, \dots, D_t^k)$ be the vector of dividends.

Let $(\theta, \phi) \in \mathbb{R}_+^J \times \mathbb{R}^K$ be the portfolio of investment and asset holdings, so θ^j and ϕ^k are the fraction of wealth invested in firm j and asset k. As before, there might be a portfolio constraint denoted by $\Pi_t \subset \mathbb{R}_+^J \times \mathbb{R}^K$ at time t. The portfolio $(\theta, \phi) \in \Pi_t$ defines the return on portfolio

$$R_{t+1}(\theta, \phi) = \sum_{j=1}^{J} A_{t+1}^j \theta^j + \sum_{k=1}^{K} R_{t+1}^k \phi^k. \tag{4.1}$$

Equilibrium As usual the sequential general equilibrium is defined by agent optimization and market clearing.

Definition 4.1. $\left\{(c_{it}, w_{it}, \theta_{it}, \phi_{it})_{i \in I}, (P_t^k)_{k \in K}\right\}_{t=0}^{T}$ constitute a sequential general equilibrium if

1. given asset returns $R_{t+1}^k = (P_{t+1}^k + D_{t+1}^k)/P_t^k$, the portfolio (θ_{it}, ϕ_{it}) solves
$$\max_{(\theta,\phi) \in \Pi_t} \frac{1}{1-\gamma} \mathrm{E}\left[R_{t+1}(\theta, \phi)^{1-\gamma} \,\middle|\, \mathcal{F}_{it}\right], \tag{4.2}$$
2. given the portfolio choice, c_{it} solves the optimal consumption problem (2.2),
3. markets for assets in zero net supply clear, *i.e.*, for each asset k and time t we have $\sum_{i=1}^{I} \phi_{it}^k (w_{it} - c_{it}) = 0$, and
4. individual wealth evolves according to the budget constraint
$$w_{i,t+1} = R_{t+1}(\theta_{it}, \phi_{it})(w_{it} - c_{it}).$$

4.2 Absolute asset pricing

Theorem 4.2. *Let* $\Theta_t = \left\{\theta \in \mathbb{R}_+^J \,\middle|\, (\theta, 0) \in \Pi_t\right\}$ *be the portfolio constraint on investment with holdings in assets in zero net supply restricted to be zero. Suppose that*

1. *agents have information and recursive preferences satisfying Assumptions 1 and 2,*
2. *for each agent conditional independence (Assumption 3) holds,* i.e.*, the distributions of the individual state variables* $\mathbf{X}_{i,t+1}$ *and the productivities and dividends* $(\mathbf{A}_{t+1}, \mathbf{D}_{t+1})$ *are independent conditional on private information* $\mathcal{F}_{it}$*,*
3. *the efficient market hypothesis (Assumption 4) holds,*
4. *the aggregators* (f_{it}) *are sufficiently regular so that the optimal consumption always exists,*[6] *and*

[6]For instance, the upper semi-continuity of the aggregator $f(c, v, \mathbf{X})$ with respect to the first two arguments on $\mathbb{R}_+^2$ suffices.

5. Θ_t is nonempty, compact, convex, and

$$\mathrm{E}\left[\sup_{\theta\in\Theta_t} R_{t+1}(\theta,0)^{1-\gamma}\,\middle|\,\mathcal{F}_t\right]<\infty.$$

Then there exists a symmetric equilibrium with a common portfolio of investment θ_t^* and no trade in zero net supply assets, where

$$\theta_t^*\in\arg\max_{\theta\in\Theta_t}\frac{1}{1-\gamma}\,\mathrm{E}\left[R_{t+1}(\theta,0)^{1-\gamma}\,\middle|\,\mathcal{F}_t\right]. \tag{4.3}$$

Proof. By Theorem 2.1, the optimal portfolio problem (4.3) has a solution θ_t^*. Let c_{it} be the optimal consumption corresponding to θ_t^*, which exists by assumption. Define the price of asset k, P_t^k, by iterating (3.2b), where $R_{m,t+1}=R_{t+1}(\theta_t^*,0)$. Then by construction the first-order condition for the maximization (4.2) (with $\mathcal{F}_t$ instead of $\mathcal{F}_{it}$) holds for every asset $k\in K$. By the definition of θ_t^*, the first-order condition for the maximization (4.2) holds for every investment $j\in J$. Hence the first-order condition holds for every returns j and k. Since the first-order condition is sufficient for maximum because the objective function in (4.2) is quasi-concave, $(\theta_t^*,0)$ is optimal in Π_t. Since the individual asset holdings is zero by construction, the markets of assets in zero net supply clear. Therefore we obtain a sequential equilibrium. ☐

Remark. Since

$$R_{t+1}(\theta,0)=\sum_{j=1}^{J}A_{t+1}^{j}\theta$$

by the definition of returns on portfolio (4.1), the symmetric equilibrium portfolio θ_t^* in (4.3) can be computed without knowing the asset prices. The asset prices can then be computed using (3.2b) with $R_{m,t+1}=R_{t+1}(\theta_t^*,0)$.

Combining Theorems 3.3 and 4.2, we obtain an absolute asset pricing formula.

Corollary 4.3. *Let everything be as in Theorem 4.2. Then the conclusion of Theorem 3.3 holds.*

As in Theorem 3.3, the $-\gamma$th power of the return on the market portfolio $(\theta_t^*,0)$ is a valid stochastic discount factor. In order to build a general equilibrium model (*i.e.*, *not* a partial equilibrium model), in Theorem 4.2 I assumed that firms are AK type technologies and ignored inputs other than capital, for example labor or raw materials. It is not easy to solve for the general equilibrium if we make the model more realistic by introducing other inputs.

Corollary 4.3 is surprising in that any preference characteristics other than risk aversion have no asset pricing implications: asset prices are determined by the technologies and relative risk aversion alone. In particular, the interest rate is completely pinned down, no matter how patient or impatient agents are. How could this be true? The intuition is simple: if there is no uncertainty, because a linear production technology between today and tomorrow determines the relative price between today and tomorrow, it is obvious that the interest rate is determined only by the technology. The risk-free rate formula (3.3) is the generalization to the case with uncertainty.

5 Testing the asset pricing implications

In this section I estimate the relative risk aversion γ and test the conditional moment restriction (3.2a) as well as the unconditional moment restriction

$$(\forall k = 1, \dots, K) \ \mathrm{E}[R_m^{-\gamma}(R^k - R_m)] = 0, \tag{5.1}$$

which are testable implications of the partial equilibrium model of Section 3. Using monthly data from 1926 to 1981, Brown and Gibbons (1985) estimated γ from the unconditional moment condition (5.1) with only one asset (the risk-free asset) and obtained $\hat{\gamma} = 1.79$, but they did not test the moment condition (since γ is exactly identified). The focus of this section is in testing both the conditional and unconditional moment restrictions, not just estimating the relative risk aversion coefficient. Testing the general equilibrium model of Section 4 (possibly using firm data or data on national wealth, GDP, and investment) would be certainly interesting but is beyond the scope of this paper.

5.1 Data

For nominal asset returns data, I use the monthly and quarterly returns of NYSE value-weighted portfolio (total market as well as portfolios sorted by size and book-to-market value) for stocks and Treasury Index, both available from the Center for Research in Security Prices (CRSP). Nominal returns are converted to real returns by adjusting with the Consumer Price Index (CPI). As instrumental variables for testing the conditional moment restriction (3.2a), I consider past annual dividend yields because they are known to predict returns (Fama and French, 1988). The data on annual dividend yields is taken from Shiller (1992).[7] More specifically, I consider two sets of test assets:

S10 30 day T-bill rate and 10 portfolios of stocks sorted by size, and

FF25 30 day T-bill rate and Fama and French (1993) 25 portfolios of stocks sorted by size and book-to-market value.

The sample period is January 1926–December 2011 for the 10 stock size portfolios and July 1931–December 2011 for the Fama-French 25 portfolios because in the latter case some returns data is unavailable for 1926–1931.

I assume that the gross return on any agent's wealth portfolio is proportional to the stock market return. Of course I am aware that the stock market is not the portfolio of total wealth (Roll, 1977; Stambaugh, 1982), but this is not a bad first approximation. We can justify this assumption as follows. Assets are priced by asset market participants, who are typically wealthy and hold a large number of stock shares. Therefore it is reasonable to expect that the stock market return is a proxy of the total return on the wealth portfolio of asset market participants.

5.2 Identification

Let $\mathbf{R}_t = (R_t^1, \dots, R_t^K)$ be the vector of gross asset returns, $\mathbf{z}_t \in \mathbb{R}^L$ be the vector of instruments (a constant ($\mathbf{z}_t = 1$) for testing the unconditional moment

[7] See the file `chapt26.xls` at `http://www.econ.yale.edu/~shiller/data.htm`.

restriction (5.1) and the vector of a constant and past dividend yields for testing the conditional moment restriction (3.2a)),

$$u_t(\gamma) = R_{m,t}^{-\gamma}(\mathbf{R}_t - R_{m,t}\mathbf{1}_K) \otimes \mathbf{z}_t \in \mathbb{R}^{KL}$$

be the pricing error ("$\otimes$" denotes the Kronecker product), and

$$g_T(\gamma) = \frac{1}{T}\sum_{t=1}^{T} u_t(\gamma)$$

be its sample average. Let $\nabla u_t(\gamma), \nabla g_T(\gamma)$ be the derivative of u_t, g_T with respect to γ, which are also $KL \times 1$ vectors.

It has been recognized that weak identification can be a source of poor performance of the GMM estimator, especially in nonlinear models (Stock and Wright, 2000). I follow Wright (2003) for testing the lack of identification. In order to identify γ_0 (the true parameter value), $\mathrm{E}[\nabla u_t(\gamma_0)]$ must have full column rank, which is equivalent to $\mathrm{E}[\nabla u_t(\gamma_0)] \neq 0$ since ∇u_t is a vector, not a matrix. For a fixed γ, under the null that $\mathrm{E}[\nabla u_t(\gamma)] = 0$, the test statistic

$$T\nabla g_T(\gamma)'\hat{C}(\gamma)^{-1}\nabla g_T(\gamma)$$

is asymptotically $\chi^2(KL)$ distributed, where $\hat{C}(\gamma)$ is a consistent estimator of the long run variance of $\nabla u_t(\gamma)$.[8] For $\hat{C}(\gamma)$ I take the Newey and West (1987) heteroskedasticity and autocorrelation consistent covariance matrix with truncation parameter $m = \lfloor T^{1/3} \rfloor$. Table 1 shows the range of γ for which the lack of identification is not rejected at significance level 0.05 for each combination of test assets and number of lagged dividend yields used as instrument. According to Table 1, γ may be unidentified if either (i) we use quarterly data, (ii) we use more than one year of past lagged dividend yields as instrument, or (iii) $\gamma \notin [-15, 5]$. Therefore in what follows I only use monthly data with either no lagged dividend yields (unconditional model) or the previous year's dividend yield (conditional model), and assume that the true parameter value γ_0 is in the range $[-15, 5]$ and estimate γ over this interval.

Table 1. Range of γ for which lack of identification is not rejected. # lags indicates the number of lagged dividend yields used as instrument.

Test assets	10 size portfolios (S10)		Fama-French 25 (FF25)	
Period	1926–2011		1931–2011	
Frequency	monthly	quarterly	monthly	quarterly
# lags				
0	$[-26.5, 13.9]$	$[-33.6, 5.0]$	$[-34.2, 12.2]$	$\emptyset$
1	$[-37.6, 16.0]$	$[-48.3, 1.4]$	$[-14.8, 5.1]$	$\emptyset$
2	$[-27.7, 9.0]$	$\emptyset$	$\emptyset$	$\emptyset$

[8]In Wright (2003) the test statistic is more complicated because he develops a general theory for any number of parameters (one has to perform a minimization over all matrices of rank less than the number of parameters). Since in my model there is only one parameter, this step is unnecessary.

5.3 Estimation

Since weak identification does not seem to be an issue at least for monthly data with no or one year of lagged dividend yields and $\gamma \in [-15, 5]$, I obtain the estimate $\hat{\gamma}$ by minimizing the continuously updated optimal GMM criterion

$$J_T(\gamma) = T g_T(\gamma)' \hat{\Omega}(\gamma)^{-1} g_T(\gamma),$$

where $\hat{\Omega}(\gamma)$ is the Newey-West HAC estimator of the long run variance of $u_t(\gamma)$.

Table 2 presents the estimate $\hat{\gamma}$ of the relative risk aversion (RRA) coefficient, its standard error, the number of periods and moment restrictions, and the P value of the J-test for overidentifying restrictions with monthly data. The results are virtually identical across the choice of test assets (10 stock size portfolios or Fama-French 25 portfolios) and instruments (no or previous year's dividend yield). The RRA estimates are around 2 with standard errors around 0.6 for all specifications. Therefore the log-utility CAPM ($\gamma = 1$) is not rejected. The moment restriction is not rejected except the unconditional model with FF25, which implies that there is no equity premium puzzle or risk-free rate puzzle. This is satisfactory since any proposed solution of asset pricing puzzles that "does not explain the premium for $\gamma < 2.5$ ... is ... likely to be widely viewed as a resolution that depends on a high degree of risk aversion" (Lucas, 1994, p. 335). My RRA estimate of around 2 is also in line with estimates using the Consumption Expenditure Survey (CEX). For example, Brav et al. (2002) and Vissing-Jørgensen (2002) report RRA of 3–4 and 2.5–3.3, respectively.[9]

Table 2. GMM estimation results of $\mathrm{E}[R_m^{-\gamma}(R^k - R_m)] = 0$.

Test assets	10 size portfolios (S10)		Fama-French 25 (FF25)	
Period	1926–2011		1931–2011	
Conditional?	no	yes	no	yes
RRA, $\hat{\gamma}$	2.0	2.15	2.05	1.55
S.E.	0.65	0.64	0.58	0.52
T	1032	1032	966	966
# moments	11	22	26	52
P (J test)	0.123	0.125	0.0015	0.106

5.4 Robustness to heavy tails

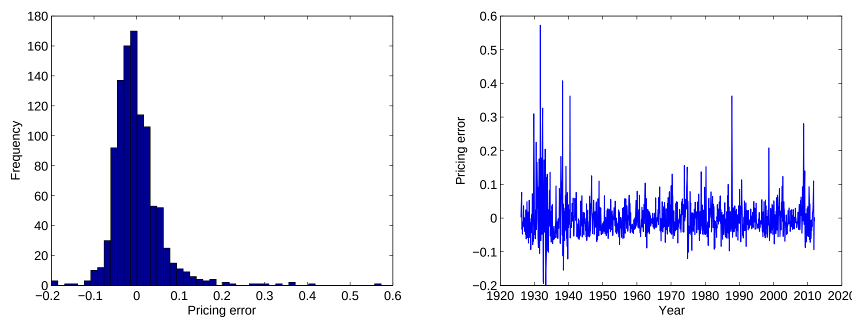
The validity of the J-test rests upon the existence of the second moment (long run variance) of $\{u_t\}$. However, it is widely known that asset returns have heavy tails (Mantegna and Stanley, 2000), or obey the power law,[10] and therefore the pricing error $\{u_t\}$ might not admit a finite second moment. If this is the case, the Central Limit Theorem does not hold and hence we cannot apply the J-test. In this subsection I investigate whether the second moment is finite, and

[9] Vissing-Jørgensen (2002) reports the elasticity of intertemporal substitution (EIS) to be 0.3–0.4 for stock holders. With additive CRRA preferences, EIS is the inverse of RRA, so the range of RRA is 2.5–3.3. Given the irrelevance result of Kocherlakota (1990) or Proposition 2.2, it seems more appropriate to interpret her result as an estimate of RRA.

[10] A random variable X is said to obey the *power law* if $\Pr(X > x) \sim x^{-\alpha}$ as $x \to \infty$, where α is the power law exponent (Mandelbrot, 1960). A power law random variable admits a finite second moment only if $\alpha > 2$.

I propose a specification test that is robust to the nonexistence of the second moment.

Power law in pricing error Figure 1(a) shows the histogram of the pricing error for the risk-free rate $u_{f,t} := R_{m,t}^{-\gamma}(R_{f,t} - R_{m,t})$ corresponding to the GMM estimate $\hat{\gamma} = 2.0$ with stock size deciles (S10). By eyeball inspection, one cannot rule out the possibility that the pricing error $\{u_{f,t}\}$ has a heavy right tail. The intuition for why the right tail matters is simple: since $\gamma > 0$, $R_m^{-\gamma}$ is large when R_m is small (*i.e.*, when the market crashes). Then $R_m < R_f$, so the pricing error $u_f = R_m^{-\gamma}(R_f - R_m)$ becomes a large positive number. In fact, according to the time series of the pricing error in Figure 1(b), large positive pricing errors occurred during the Great Depression of 1929–1940, the market crash of October 1987, the Russian financial crisis and the subsequent collapse of the Long Term Capital Management of August 1998, and the recent financial crisis of 2008–2009.



(a) Histogram.

(b) Time series.

Figure 1. Pricing error for the 30 day T-bill rate corresponding to $\hat{\gamma} = 2.0$.

The top three rows of Tables 3 show the estimate of the power law exponent of the right tail of the pricing error, its standard error, and the P value of the Kolmogorov test for the power law. (See Appendix A for how to implement the test.) The power law is not rejected at conventional significance level. Since the power law exponent is slightly above 2 but the standard error is around 0.3, the second moment of the pricing error may not exist. To test the existence of the second moment directly, I apply the bootstrap test of Fedotenkov (2013), explained in Appendix B. Since the P values are all around 0.2, the existence of the second moment is not rejected. However, even if the second moment is finite, since the power law exponent is close to 2, the convergence of the sample second moment will be very slow. Hence it still might be dangerous to fully trust the J-test.

Specification test without finite second moments How should we test the moment condition without assuming a finite second moment? Kocherlakota (1997) proposed such a specification test in the context of testing the representative agent consumption-based capital asset pricing model based on subsampling (Politis and Romano, 1994; Politis et al., 1999). However, his method requires that there is only one moment condition and that the right tail of the distribution of the pricing error has a power law exponent between 1 and 2, which

Table 3. Tests of power law and finite second moment.

Test assets	10 size portfolios (S10)		Fama-French 25 (FF25)	
Conditional?	no	yes	no	yes
PL exponent, $\hat{\alpha}$	2.14	2.10	2.19	2.25
S.E.	0.30	0.29	0.31	0.35
P (Kolmogorov)	0.78	0.81	0.62	0.66
P (Fedotenkov)	0.23	0.23	0.22	0.25

is restrictive because we do not know the power law exponent a priori. As an alternative, I propose a specification test that is robust to the nonexistence of second moments, which is a multidimensional version of McElroy and Politis (2002). The advantage of my method is that it is applicable with any number of moment conditions, irrespective of the existence of the second moment.

The test statistic is the studentized quantity

$$A_T = T g_T(\hat{\gamma})' S_T(\hat{\gamma})^{-1} g_T(\hat{\gamma}),$$

where $S_T(\hat{\gamma}) = \frac{1}{T}\sum_{t=1}^{T} u_t(\hat{\gamma}) u_t(\hat{\gamma})'$. Since the existence of the second moment is not necessary for the consistency of the GMM estimator, we have $\hat{\gamma} \xrightarrow{\text{P}} \gamma_0$. If the second moment is finite, then $S_T(\hat{\gamma})$ converges in probability to $\mathrm{E}[u_t(\gamma_0)u_t(\gamma_0)']$ as $T \to \infty$, and under the null that $\mathrm{E}[u_t(\gamma_0)] = 0$, by the central limit theorem $\sqrt{T} g_T(\hat{\gamma})$ is asymptotically normal. Therefore the test statistic A_T has a weak limit. Under the alternative that $\mathrm{E}[u_t(\gamma_0)] \neq 0$, A_T tends to infinity.

Now suppose that $\mathrm{E}[u_t(\gamma_0)] = 0$ but $u_t(\gamma_0)$ has a tail exponent $1 < \alpha \leq 2$. By the Cramér-Wold device, the weak convergence of a sequence of random vectors to a multivariate stable distribution reduces to the scalar case (Press, 1972). Hence by Theorem 3.1 of Davis and Hsing (1995), both $T^{1-1/\alpha} g_T(\hat{\gamma})$ and $T^{1-2/\alpha} S_T(\hat{\gamma})$ have a weak limit, so A_T also converges weakly. The difficulty for performing statistical inference, however, is that the weak limit depends on the unknown parameters in a complicated way.

To circumvent this difficulty, I apply the subsampling of Politis and Romano (1994) in order to compute the P value, which is applicable whenever a statistic has a weak limit. Let b a number such that $b \to \infty$ and $b/T \to 0$ as $T \to \infty$ (I take $b = 10\sqrt{T}$). Let

$$A_{b,t} = b g_{b,t}(\hat{\gamma}_{b,t})' S_{b,t}(\hat{\gamma}_{b,t})^{-1} g_{b,t}(\hat{\gamma}_{b,t}), \ t = 1, \dots, T - b + 1$$

be the test statistic computed over the subsample $\{u_t, \dots, u_{t+b-1}\}$ of size b starting from t, where $\hat{\gamma}_{b,t}$ is the GMM estimator over this subsample. Under the null that $\mathrm{E}[u_t(\gamma_0)] = 0$, both A_T and $A_{b,t}$ converges to the same limit. Under the alternative that $\mathrm{E}[u_t(\gamma_0)] \neq 0$, on the other hand, both A_T and $A_{b,t}$ diverges to ∞ but A_T does so at a faster rate because $b/T \to 0$. Therefore we can compute the P value by

$$p = \frac{1}{T - b + 1} \sum_{t=1}^{T-b+1} \mathbb{1}\left\{A_{b,t} > A_T\right\}.$$

Table 4 shows the P value of the specification test. Neither the unconditional moment restriction (5.1) nor the conditional moment restriction (3.2a) are rejected at significance level 0.05 for each combination of test assets.

Table 4. Specification test by subsampling.

Test assets	10 size portfolios (S10)		Fama-French 25 (FF25)	
Conditional?	no	yes	no	yes
P (Subsampling)	0.072	0.20	0.21	0.12

6 Asset pricing puzzles

6.1 Literature

The consumption-based capital asset pricing model (CAPM) of Lucas (1978) and Breeden (1979) has not performed well empirically (Hansen and Singleton, 1982, 1983), which led to the conceptualization of the equity premium puzzle (Mehra and Prescott, 1985) and the risk-free rate puzzle (Weil, 1989). These asset pricing puzzles continue to fascinate the profession: numerous generalizations of the model in an attempt to resolve the puzzles include incomplete markets (Bewley, 1982; Constantinides and Duffie, 1996), probability distributions that admit rare but disastrous events Rietz (1988), state-dependent utility function such as habit formation (Abel, 1990; Constantinides, 1990; Campbell and Cochrane, 1999), transaction costs (Aiyagari and Gertler, 1991), limited asset market participation (Brav et al., 2002; Balduzzi and Yao, 2007), focusing on consumption of luxury goods instead of aggregate consumption (Aït-Sahalia et al., 2004), production (Akdeniz and Dechert, 2007), and constrained efficient allocation with private information (Kocherlakota and Pistaferri, 2009), to name just a few. See Kocherlakota (1996), Campbell (2003), and Chapter 21 of Cochrane (2005) for reviews. There are also explanations outside (neoclassical) economics. Kocherlakota (1997) notes the possibility of a heavy tail in the distribution of consumption growth, which would invalidate the use of χ^2 specification tests, and Barberis et al. (2001) employ prospect theory.

In reviewing the literature, Kocherlakota (1996) notes that at least one of the following three assumptions must be abandoned in order to resolve the equity premium puzzle and the risk-free rate puzzle. These are (i) complete asset markets, (ii) frictionless asset markets, and (iii) standard additive CRRA utility with discount factor $\beta \in (0, 1)$ and relative risk aversion coefficient $\gamma \in [0, 10]$.

6.2 A resolution?

By the results in Tables 2 and 4, setting the relative risk aversion $\gamma = 2.0$ in my model explains the historical equity premium and risk free rate, and the model is not rejected. What is the key for the (possible) resolution of the asset pricing puzzles? Since in my model asset markets are complete and frictionless, the resolution must come from abandoning the representative agent with standard additive CRRA utility

$$\mathrm{E} \sum_{t=0}^{\infty} \beta^t \frac{c_t^{1-\gamma}}{1-\gamma}.$$

When we test the consumption Euler equation

$$1 = \mathrm{E} \left[\beta \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma} R^k \middle| \mathcal{F}_t \right]$$

using aggregate consumption, we are testing the joint hypothesis that asset markets are complete and frictionless *and* that all agents in the economy have identical additive CRRA preferences. Even if we use household consumption as in Brav et al. (2002) and Balduzzi and Yao (2007), we are still testing a model with identical additive CRRA preferences. Since people differ in patience, life cycle needs, and so on, there is no wonder that we reject this hypothesis. When we test the moment condition (3.2a), on the other hand, we are testing the weaker joint hypothesis that asset markets are complete and frictionless *and* that market participants have homothetic CRRA recursive preferences with a common relative risk aversion coefficient but otherwise heterogeneous.

In the model of Mehra and Prescott (1985) both $\beta(c_{t+1}/c_t)^{-\gamma}$ and $R_m^{-\gamma}$ are valid stochastic discount factors (SDFs), but they chose to use consumption (Euler equation) to calibrate their model. Since aggregate consumption does not vary much, they found a puzzle. However, in a more general model in Section 3, $\beta(c_{t+1}/c_t)^{-\gamma}$ (where c_t is aggregate consumption) is not necessarily a valid SDF. A special case it is valid is with identical additive CRRA utility function and complete markets (*i.e.*, the representative agent). Since these assumptions are unlikely to hold, there is no wonder we bump into puzzles if we take the model literally. My explanation of the asset pricing puzzles is due to the derivation of an equation (the moment condition (3.2a)) that contains only one parameter and data that are highly accurate, and most of all, is robust to alternative specifications of the model.

This observation suggests that consumption-based asset pricing models have limitations. By introducing production (investment) and thereby disentangling the portfolio decision from the consumption/saving decision as in my model, we no longer need to look at consumption data, at least for studying portfolio decisions and asset pricing.[11] This point relates to Campbell (2003), who suggested that "it is not easy to construct a general equilibrium model that fits all the stylized facts" (p. 808) but "[m]odels with production also help one to move away from the common assumption that stock market dividends equal consumption ... it will ultimately be more satisfactory to derive both dividends and consumption within a general equilibrium model" (p. 880). The importance of production is also stressed by Akdeniz and Dechert (2007), who resolved the equity premium puzzle and the risk-free rate puzzle by numerically solving the Brock (1982) asset pricing model. My general equilibrium model has an advantage in that it allows growth and admits high analytical tractability and flexibility.

How about other asset pricing puzzles? Campbell (2003) defines the "equity volatility puzzle" by the fact that the volatility of real stock returns is high in relation to the volatility of the short-term real interest rate. My model can explain the equity volatility puzzle as well: by the risk-free rate formula (3.3), if the market return R_m is i.i.d., then the risk-free rate is constant. Therefore the low volatility of the interest rate does not contradict the high volatility of stock returns. As long as information on past returns is not so helpful in predicting future returns (i.i.d. is the extreme case), the risk-free rate formula (3.3) tells us that the risk-free rate does not vary much over time.

[11] Of course, looking at consumption is essential in other situations, for instance estimating a consumption/saving model.

6.3 Consumption volatility puzzle

Finally I turn to the consumption volatility puzzle—the stylized fact that the volatility of aggregate consumption growth is low compared to asset returns. In Section 3.2 I noted that the consumption volatility puzzle is not an asset pricing puzzle because consumption is irrelevant for asset pricing—maybe we (financial economists) should not care too much about aggregate consumption. Nevertheless, we might want to explain the puzzle. One solution is Theorem 4.2, which is also an "anything goes" result for consumption. Since the number of agents and the (individual-, time-, and state-dependent) aggregator functions are arbitrary, we have effectively an infinite degrees of freedom, and therefore my model can explain any aggregate consumption data. Of course, a theory that can explain anything is not a theory, so I do not claim that this is a valid resolution of the consumption volatility puzzle.

Here I present a simple solution within the representative agent framework. Suppose that the representative agent has an additive HARA utility function

$$\mathrm{E}_0 \sum_{t=0}^{\infty} \beta^t \frac{(c_t - n_t)^{1-\gamma}}{1-\gamma},$$

where c_t is consumption and n_t (assumed to be exogenous and deterministic) is the *necessity*. Letting $l_t = c_t - n_t$ be the *luxury*, the agent must consume at least n_t at time t but gets utility only from the luxury he consumes.[12] By introspection, introducing necessity into the utility function seems reasonable: we buy baby diapers if and only if we have babies, and we do not get any happier by consuming more diapers.

For concreteness assume that the necessity grows at a constant rate g, so $n_t = n_0(1+g)^t$. The agent is endowed with capital k_0 at time 0 but nothing thereafter. Capital can be turned into the necessity or the luxury for one-to-one. The agent has access to two saving technologies, one risky and the other riskless. The productivity of the risky technology is i.i.d. over time and lognormally distributed: $\log A \sim N(\mu, \sigma^2)$. The riskless technology pins down the risk-free rate $R_f = 1 + r$, where I assume that $r > g$. Since the agent must consume n_t for sure at time t, his effective initial wealth is

$$w_0 = k_0 - \sum_{t=0}^{\infty} \frac{n_t}{R_f^t} = k_0 - \frac{1+r}{r-g} n_0 =: k_0(1-\nu).$$

Here ν is the ratio between the necessity n_0 and the maximal necessity (the amount that makes the effective wealth equal to zero) at time 0.

Since the agent cares only about luxury, his optimal consumption-portfolio problem reduces to

$$\max_{\{l_t, \theta_t\}} \mathrm{E}_0 \sum_{t=0}^{\infty} \beta^t \frac{l_t^{1-\gamma}}{1-\gamma} \text{ suject to } (\forall t) w_{t+1} = R_{t+1}(\theta_t)(w_t - l_t),$$

[12] In this respect my model is similar to Aït-Sahalia et al. (2004), who use the period utility function $v(C, L) = \frac{(C-a)^{1-\phi}}{1-\phi} + \frac{(L+b)^{1-\psi}}{1-\psi}$. Here C is basic consumption, L is luxury consumption, and $a, b, \phi, \psi > 0$ are parameters. My model corresponds to setting $a = n_t$, $\phi = \gamma$, and dropping the term involving L.

where l_t is luxury, θ_t is the fraction of effective wealth invested in the risky technology at time t, and $R_{t+1}(\theta) = A_{t+1}\theta + (1+r)(1-\theta)$ is the return on effective wealth between time t and $t+1$. This problem concerns a homothetic CRRA preference and multiplicative shock as in Section 4, so we can solve it analytically. According to Toda (2012), the optimal luxury-portfolio rule is

$$\theta^* = \arg\max_{\theta \in [0,1]} \frac{1}{1-\gamma} \mathrm{E}[R(\theta)^{1-\gamma}], \tag{6.1a}$$

$$l(w) = (1 - (\beta \mathrm{E}[R(\theta^*)^{1-\gamma}])^{\frac{1}{\gamma}})w. \tag{6.1b}$$

Then the optimal consumption of the original problem is $c_t = l(w_t) + n_t$.

Of course, this model does not produce stationary consumption growth because the necessity $n_t = n_0(1+g)^t$ and the luxury l_t (which is proportional to effective wealth) does not grow at the same rate. (Necessity is deterministic whereas luxury is stochastic.) However, the nonstationarity is not severe if the necessity parameter ν is either close to 0 or 1 and the time horizon is not too long. For long horizons, we can make the model close to stationary by considering overlapping generations whose initial necessity grow with aggregate wealth. This assumption seems natural because people tend to regard more goods as necessity as the economy grows. (Think of cell phones and Internet now and 20 years ago.)

Numerically solving for the general equilibrium is straightforward.

1. Solve for the optimal portfolio rule (6.1a) using the portfolio return

$$R(\theta) = A\theta + (1+r)(1-\theta)$$

   and compute the marginal propensity to consume out of effective wealth in (6.1b).

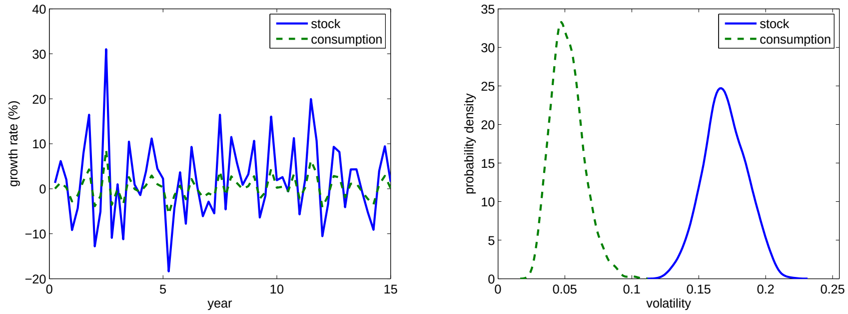
2. Generate "stock market returns" $\{A_t\}_{t=1}^T$ and iterate the budget constraint $w_{t+1} = R(\theta^*)(w_t - l_t)$ to obtain the luxury $\{l_t\}_{t=0}^T$.

3. Compute consumption by $c_t = l_t + n_t$.

As a numerical example, I simulate quarterly stock market and consumption data for 15 years. The parameters (at annual frequency) are discount factor $\beta = 0.96$, relative risk aversion $\gamma = 3$, expected stock market return $\mu = 0.07$ (7%), volatility $\sigma = 0.17$ (17%), risk-free rate $r = 0.01$ (1%), fraction of necessity to maximal necessity $\nu = 0.9$, and no growth in necessity ($g = 0$). With these parameters the optimal portfolio of effective wealth is $\theta^* = 0.8593$ (86% stocks)[13] and the quarterly marginal propensity to consume luxury out of effective wealth is 0.0103.

Figure 2(a) shows typical sample paths of annualized stock market return and consumption growth. Clearly consumption growth is much less volatile than the stock market. Figure 2(b) shows the kernel density estimate of the distribution of sample volatility of stock market return and consumption growth for 1,000 Monte Carlo simulations. On average, the volatility of consumption growth is about 5%, which is a reasonable number.

[13] Since $\mathrm{E}[R(\theta)^{1-\gamma}]$ has no closed-form expression, solving for the optimal portfolio in (6.1a) is not trivial. Here I approximate the normal distribution by a multinomial distribution on 81 equally spaced grid points as explained in Tanaka and Toda (2013) and then solve for the optimal portfolio numerically.



(a) Typical sample paths.

(b) Kernel density estimate of the distribution of sample volatility for 1,000 simulations.

Figure 2. Stock market return (blue solid) and consumption growth (green dashed).

7 Concluding remarks

This paper can be summarized as follows. (i) I found a simple, yet economically motivated stochastic discount factor ($R_m^{-\gamma}$). (ii) I theoretically showed the robustness of this SDF. (iii) I tested the SDF and failed to reject it: a relative risk aversion coefficient of around 2 is consistent with the historical asset returns data. Although this SDF has been already known (Rubinstein, 1976), it has not captured much attention.[14] Given the robustness of this SDF, it deserves a serious consideration.

A clear lesson from this paper is the usefulness of AK models. Combined with homothetic preferences, AK models admit full analytical tractability, even with many agents with heterogeneous preferences as long as agents have a common relative risk aversion. This is true even in an incomplete markets setting, as shown by Toda (2012), where I derive a similar stochastic discount factor. AK models have investment as a key element and hence are more realistic, unlike pure exchange models that can only proxy hunter-gatherer economies.

Being primarily a theoretical paper, I kept the empirical analysis to the bare minimum (testing the partial equilibrium model). An obvious next step for further empirical research is to test the general equilibrium model in Section 4. Here I sketch how this could be done. By the budget constraint, we have

$$w_{i,t+1} = R_{t+1}(\theta_t^*, 0)(w_{it} - c_{it}),$$

where w_{it} is agent i's wealth, c_{it} is consumption, and θ_t^* is the common portfolio of investment. Adding the budget constraint across agents, we obtain

$$W_{t+1} = R_{t+1}(\theta_t^*, 0)(W_t - C_t),$$

where W_t, C_t are aggregate wealth and consumption *of market participants*. Thus if we have a good measure of aggregate wealth and consumption, one can recover the total return on wealth by $R_{t+1}(\theta_t^*, 0) = \frac{W_{t+1}}{W_t - C_t}$ and redo the exercises in Section 5.

[14] Cochrane (2005) mentions only the log utility case ($\gamma = 1$) briefly.

A Testing the power law

Consider a random variable X with probability density function (PDF) $f(x)$ and a cumulative distribution function (CDF) $F(x)$. X is said to obey the *power law* with exponent $\alpha > 0$ if

$$1 - F(x) = P(X > x) \sim x^{-\alpha}$$

as $x \to \infty$.[15] Since in general the power law holds only asymptotically, testing the power law is not trivial. Clauset et al. (2009) suggest to estimate the power law exponent α by fitting the Pareto distribution ($f(x) \propto x^{-\alpha-1}$) by maximum likelihood above a cutoff value, and then to apply the Kolmogorov test by bootstrap to evaluate the goodness of fit of the entire distribution. (See their paper for more details.) Their web appendix[16] contains Matlab files that implement the algorithm. The file `plfit.m` estimates the power law exponent[17] and the cutoff value, and `plpva.m` performs the Kolmogorov test. I obtained the middle three rows of Table 3 in this way. The cutoff value for the pricing error was around 0.1 in all cases, which conforms to the histogram in Figure 1(a).

B Testing moment existence

This Appendix explains how to test the existence of moments directly following Fedotenkov (2013). Suppose that the random variable X is nonnegative (consider $|X|$ if X can be negative) and $\{X_n\}_{n=1}^{\infty}$ are i.i.d. copies of X. If $\mathrm{E}[X^\eta] = \infty$, then the sample moment $\frac{1}{N}\sum_{n=1}^{N} X_n^\eta$ tends to infinity as $N \to \infty$. Therefore if we take a number $M(N)$ such that $M \to \infty$ and $M/N \to 0$ as $N \to \infty$, and $\{Y_m\}_{m=1}^{\infty}$ are independent and have the same distribution as X, then for $0 < \xi < 1$ the quantity

$$F = \mathbb{1}\left\{\frac{1}{M}\sum_{m=1}^{M} Y_m^\eta \geq \xi \frac{1}{N}\sum_{n=1}^{N} X_n^\eta\right\}$$

tends to zero almost surely as $N \to \infty$, where $\mathbb{1}\{\cdot\}$ denotes the indicator function. This is because both $\xi\frac{1}{N}\sum_{n=1}^{N} X_n^\eta$ and $\frac{1}{M}\sum_{m=1}^{M} Y_m^\eta$ tend to infinity, but the former does so at a faster rate. On the other hand, if $\mathrm{E}[X^\eta]$ is finite, then by the law of large numbers F tends to 1 almost surely because both sample means converge to the same population mean, but since $0 < \xi < 1$ as N tends to infinity $\xi\frac{1}{N}\sum_{n=1}^{N} X_n^\eta$ is almost surely smaller than $\frac{1}{M}\sum_{m=1}^{M} Y_m^\eta$.

Given this result Fedotenkov (2013) constructs a bootstrap test of moment existence as follows. Let $\mathbf{x} = (x_1, \dots, x_N)$ be the data. First, we choose the bootstrap sample size $M(N)$, the parameter ξ, and bootstrap repetition B (Fedotenkov suggests taking $M(N) = \lfloor \log N \rfloor$, $\xi = 0.999$, and $B = 10,000$). Second, for each $b = 1, \dots, B$, we generate a bootstrap sample $\mathbf{x}^b$ of size M drawn

[15] Since the PDF of a power law variable is asymptotically $f(x) \sim x^{-\alpha-1}$, some authors refer to the exponent by $\alpha' = \alpha + 1$, which can introduce confusions.

[16] `http://tuvalu.santafe.edu/~aaronc/powerlaws/`

[17] Note that Clauset et al. (2009) define the power law exponent by $\alpha' = \alpha + 1$, so we have to subtract 1 from the estimation result when using `plfit.m`.

randomly with replacement from the data, and compute

$$F^b = \mathbb{1}\left\{\frac{1}{M}\sum_{m=1}^{M}(x_m^b)^\eta \geq \xi\frac{1}{N}\sum_{n=1}^{N}x_n^\eta\right\}.$$

Finally, the P value is defined by $p = \frac{1}{B}\sum_{b=1}^{B} F^b$.

References

Andrew B. Abel. Asset prices under habit formation and catching up with the Joneses. *American Economic Review*, 80(2):38–42, May 1990.

Yacine Aït-Sahalia, Jonathan A. Parker, and Motohiro Yogo. Luxury goods and the equity premium. *Journal of Finance*, 59(6):2959–3004, December 2004. doi:10.1111/j.1540-6261.2004.00721.x.

S. Rao Aiyagari and Mark Gertler. Asset returns with transactions costs and uninsured individual risk. *Journal of Monetary Economics*, 27(3):311–331, June 1991. doi:10.1016/0304-3932(91)90012-D.

Levent Akdeniz and W. Davis Dechert. The equity premium in Brock's asset pricing model. *Journal of Economic Dynamics and Control*, 31(7):2263–2292, July 2007. doi:10.1016/j.jedc.2006.06.008.

Pierluigi Balduzzi and Tong Yao. Testing heterogeneous-agent models: An alternative aggregation approach. *Journal of Monetary Economics*, 54(2): 369–412, March 2007. doi:10.1016/j.jmoneco.2005.08.021.

Nicholas Barberis, Ming Huang, and Tano Santos. Prospect theory and asset prices. *Quarterly Journal of Economics*, 116(1):1–53, February 2001. doi:10.1162/003355301556310.

Truman F. Bewley. Thoughts on tests of the intertemporal asset pricing model. 1982.

John H. Boyd, III. Recursive utility and the Ramsey problem. *Journal of Economic Theory*, 50(2):326–345, April 1990. doi:10.1016/0022-0531(90)90006-6.

Alon Brav, George M. Constantinides, and Christopher C. Geczy. Asset pricing with heterogeneous consumers and limited participation: Empirical evidence. *Journal of Political Economy*, 110(4):793–824, August 2002. doi:10.1086/340776.

Douglas T. Breeden. An intertemporal asset pricing model with stochastic consumption and investment opportunities. *Journal of Financial Economics*, 7(3):265–296, September 1979. doi:10.1016/0304-405X(79)90016-3.

William A. Brock. Asset prices in a production economy. In John J. McCall, editor, *The Economics of Information and Uncertainty*, chapter 1, pages 1–46. University of Chicago Press, Chicago, 1982.

David P. Brown and Michael R. Gibbons. A simple econometric approach for utility-based asset pricing models. *Journal of Finance*, 40(2):359–381, June 1985. doi:10.1111/j.1540-6261.1985.tb04962.x.

John Y. Campbell. Intertemporal asset pricing without consumption. *American Economic Review*, 83(3):487–512, June 1993.

John Y. Campbell. Consumption-based asset pricing. In George M. Constantinides, Milton Harris, and René M. Stultz, editors, *Handbook of the Economics of Finance*, volume 1B, chapter 13, pages 803–887. Elsevier, Amsterdam, 2003. doi:10.1016/S1574-0102(03)01022-7.

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(2):205–251, April 1999. doi:10.1086/250059.

David Cass and Joseph E. Stiglitz. The structure of investor preferences and asset returns, and separability in portfolio allocation: A contribution to the pure theory of mutual funds. *Journal of Economic Theory*, 2(2):122–160, June 1970. doi:10.1016/0022-0531(70)90002-5.

Aaron Clauset, Cosma Rohilla Shalizi, and Mark E. J. Newman. Power-law distributions in empirical data. *SIAM Review*, 51(4):661–703, 2009. doi:10.1137/070710111.

John H. Cochrane. *Asset Pricing*. Princeton University Press, Princeton, NJ, second edition, 2005.

George M. Constantinides. Intertemporal asset pricing with heterogeneous agents without demand aggregation. *Journal of Business*, 55(2):253–267, April 1982.

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

George M. Constantinides and Darrell Duffie. Asset pricing with heterogeneous consumers. *Journal of Political Economy*, 104(2):219–240, April 1996.

Richard A. Davis and Tailen Hsing. Point process and partial sum convergence for weakly dependent random variables with infinite variance. *Annals of Probability*, 23(2):879–917, 1995. doi:10.1214/aop/1176988294.

Robert F. Dittmar. Nonlinear pricing kernels, kurtosis preference, and evidence from the cross section of equity returns. *Journal of Finance*, 57(1):369–403, February 2002. doi:10.1111/1540-6261.00425.

Larry G. Epstein and Stanley E. Zin. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework. *Econometrica*, 57(4):937–969, July 1989.

Larry G. Epstein and Stanley E. Zin. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: An empirical analysis. *Journal of Political Economy*, 99(2):263–286, April 1991.

Eugene F. Fama. Multiperiod consumption-investment decisions. *American Economic Review*, 60(1):163–174, 1970.

Eugene F. Fama and Kenneth R. French. Dividend yields and expected stock returns. *Journal of Financial Economics*, 22(1):3–25, October 1988. doi:10.1016/0304-405X(88)90020-7.

Eugene F. Fama and Kenneth R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, February 1993. doi:10.1016/0304-405X(93)90023-5.

Igor Fedotenkov. A bootstrap method to test for the existence of finite moments. *Journal of Nonparametric Statistics*, 2013. doi:10.1080/10485252.2012.752487.

Lars Peter Hansen and Kenneth J. Singleton. Generalized instrumental variables estimation of nonlinear rational expectations models. *Econometrica*, 50(5): 1269–1286, September 1982.

Lars Peter Hansen and Kenneth J. Singleton. Stochastic consumption, risk aversion, and the temporal behavior of asset returns. *Journal of Political Economy*, 91(2):249–265, April 1983.

Narayana R. Kocherlakota. Disentangling the coefficient of relative risk aversion from the elasticity of intertemporal substitution: An irrelevance result. *Journal of Finance*, 45(1):175–190, March 1990. doi:10.1111/j.1540-6261.1990.tb05086.x.

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

Narayana R. Kocherlakota. Testing the consumption CAPM with heavy-tailed pricing errors. *Macroeconomic Dynamics*, 1(3):551–567, September 1997. doi:10.1017/S136510059700401X.

Narayana R. Kocherlakota and Luigi Pistaferri. Asset pricing implications of Pareto optimality with private information. *Journal of Political Economy*, 117(3):555–590, June 2009. doi:10.1086/599761.

David Levhari and T. N. Srinivasan. Optimal savings under uncertainty. *Review of Economic Studies*, 36(2):153–163, April 1969.

John Lintner. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *Review of Economics and Statistics*, 47(1):13–37, February 1965a.

John Lintner. Security prices, risk, and maximal gains from diversification. *Journal of Finance*, 20(4):587–615, December 1965b. doi:10.1111/j.1540-6261.1965.tb02930.x.

Deborah J. Lucas. Asset pricing with undiversifiable income risk and short sales constraints: Deepening the equity premium puzzle. *Journal of Monetary Economics*, 34(3):324–341, December 1994. doi:10.1016/0304-3932(94)90022-1.

Robert E. Lucas, Jr. Asset prices in an exchange economy. *Econometrica*, 46 (6):1429–1445, November 1978.

Sydney C. Ludvigson. Advances in consumption-based asset pricing: Empirical tests. In George M. Constantinides, Milton Harris, and René M. Stultz, editors, *Handbook of the Economics of Finance*, volume 2, chapter 12, pages 799–906. Elsevier, Amsterdam, 2013. doi:10.1016/B978-0-44-459406-8.00012-3.

Benoît Mandelbrot. The Pareto-Lévy law and the distribution of income. *International Economic Review*, 1(2):79–106, May 1960.

Rosario N. Mantegna and H. Eugene Stanley. *An Introduction to Econophysics*. Cambridge University Press, Cambridge, UK, 2000.

Harry Markowitz. Portfolio selection. *Journal of Finance*, 7(1):77–91, March 1952. doi:10.1111/j.1540-6261.1952.tb01525.x.

Tucker McElroy and Dimitris N. Politis. Robust inference for the mean in the presence of serial correlation and heavy-tailed distributions. *Econometric Theory*, 18(5):1019–1039, October 2002. doi:10.1017.S026646660218501X.

Rajnish Mehra and Edward C. Prescott. The equity premium: A puzzle. *Journal of Monetary Economics*, 15(2):145–161, March 1985. doi:10.1016/0304-3932(85)90061-3.

Robert C. Merton. Lifetime portfolio selection under uncertainty: The continuous-time case. *Review of Economics and Statistics*, 51(3):247–257, August 1969.

Robert C. Merton. Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory*, 3(4):373–413, December 1971. doi:10.1016/0022-0531(71)90038-X.

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

Whitney K. Newey and Kenneth D. West. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55(3):703–708, May 1987.

Dimitris N. Politis and Joseph P. Romano. Large sample confidence regions based on subsamples under minimal assumptions. *Annals of Statistics*, 22(4): 2031–2050, 1994. doi:10.1214/aos/1176325770.

Dimitris N. Politis, Joseph P. Romano, and Michael Wolf. *Subsampling*. Springer, New York, 1999.

S. James Press. Multivariate stable distributions. *Journal of Multivariate Analysis*, 2(4):440–462, December 1972. doi:10.1016/0047-259X(72)90038-3.

Thomas A. Rietz. The equity risk premium: A solution. *Journal of Monetary Economics*, 22(1):117–131, July 1988. doi:10.1016/0304-3932(88)90172-9.

Richard Roll. A critique of the asset pricing theory's tests. *Journal of Financial Economics*, 4(2):129–176, March 1977. doi:10.1016/0304-405X(77)90009-5.

Mark Rubinstein. An aggregation theorem for securities markets. *Journal of Financial Economics*, 1(3):225–244, September 1974. doi:10.1016/0304-405X(74)90019-1.

Mark Rubinstein. The valuation of uncertain income streams and the pricing of options. *Bell Journal of Economics*, 7(2):407–425, 1976.

Paul A. Samuelson. Lifetime portfolio selection by dynamic stochastic programming. *Review of Economics and Statistics*, 51(3):239–246, August 1969.

William F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442, September 1964. doi:10.1111/j.1540-6261.1964.tb02865.x.

Robert J. Shiller. *Market Volatility*. MIT Press, Cambridge, MA, 1992.

Robert F. Stambaugh. On the exclusion of assets from tests of the two-parameter model: A sensitivity analysis. *Journal of Financial Economics*, 10(3):237–268, November 1982. doi:10.1016/0304-405X(82)90002-2.

James H. Stock and Jonathan H. Wright. GMM with weak identification. *Econometrica*, 68(5):1055–1096, September 2000. doi:10.1111/1468-0262.00151.

Eric T. Swanson. Risk aversion and the labor margin in dynamic equilibrium models. *American Economic Review*, 102(4):1663–1691, June 2012. doi:10.1257/aer.102.4.1663.

Ken'ichiro Tanaka and Alexis Akira Toda. Discrete approximations of continuous distributions by maximum entropy. *Economics Letters*, 118(3):445–450, March 2013. doi:10.1016/j.econlet.2012.12.020.

James Tobin. Liquidity preference as behavior towards risk. *Review of Economics and Statistics*, 25(2):65–86, February 1958.

Alexis Akira Toda. Asset pricing and wealth distribution with heterogeneous investment returns. 2012. URL `https://sites.google.com/site/aatoda111/papers`.

Annette Vissing-Jørgensen. Limited asset market participation and the elasticity of intertemporal substitution. *Journal of Political Economy*, 110(4):825–853, August 2002. doi:10.1086/340782.

Philippe Weil. The equity premium puzzle and the risk-free rate puzzle. *Journal of Monetary Economics*, 24(3):401–421, November 1989. doi:10.1016/0304-3932(89)90028-7.

Jonathan H. Wright. Detecting lack of identification in GMM. *Econometric Theory*, 19(2):322–330, April 2003. doi:10.1017/S0266466603192055.