

The Investor Recognition Hypothesis in a Dynamic General Equilibrium: Theory and Evidence

Alexander Shapiro
New York University

This article analyzes a dynamic general equilibrium under a generalization of Merton's (1987) investor recognition hypothesis. A class of informationally constrained investors is assumed to implement only a particular trading strategy. The model implies that, all else being equal, a risk premium on a less visible stock need not be higher than that on a more visible stock with a lower volatility—contrary to results derived in a static mean-variance setting. A consumption-based capital asset pricing model (CAPM) augmented by the generalized investor recognition hypothesis emerges as a viable contender for explaining the cross-sectional variation in unconditional expected equity returns.

A fundamental question in financial economics is how frictions affect equilibrium in capital markets. The real-world frictions that motivate our analysis are information costs. In a world of costly information, some investors will have incomplete information. Therefore we ask, first, how equity portfolios of informationally constrained investors can be characterized, and second, how the presence of these investors affects equilibrium. We formulate the answer to the first question as a hypothesis; our premise is that the aggregate portfolio of informationally constrained investors combines a direct investment in visible stocks with funds whose management is entrusted to others (who may possess more information). The main objective of this article is to develop a model that can accommodate our premise, thereby offering a detailed answer to the second question. The joint validity of the premise and the model is then evaluated empirically.

Merton (1987), using a static mean-variance model, advanced the investor recognition hypothesis (IRH) to describe the portfolio formation of informationally constrained investors. In its pure static version, the IRH states that

This article is based on my dissertation at the Wharton School of the University of Pennsylvania. I thank my committee members—Suleyman Basak (the chair), Andrew Abel, Shmuel Kandel, Richard Kihlstrom, Craig MacKinlay, Robert Stambaugh—as well as Simon Benninga, Domenico Cuoco, Bernard Dumas (the editor), Roger Edelen, Christopher Gadarowski, Sanford Grossman, Burton Hollifield, Gur Huberman, Srinivasa Varadhan, and two anonymous referees for helpful comments and suggestions. I am also grateful to seminar participants at Harvard University, New York University, Technion-Israel Institute of Technology, University of California at Los Angeles, University of Michigan, Wharton School, University of Texas at Austin, the 1997 Accounting and Finance Conference at Tel Aviv University, and the 1998 Western Finance Association Meetings. I wish to acknowledge a dissertation grant from the American Association of Individual Investors. Any errors are my own. Address correspondence to Alex Shapiro, Department of Finance, Stern School of Business, New York University, 44 West 4th St., Suite 9-190, New York, NY 10012-1126, or e-mail: ashapiro@stern.nyu.edu.

investors buy and hold only those securities about which they have enough information, and the revealed portfolio formation under the IRH is observationally equivalent to that under exogenous portfolio constraints.¹ Increasing empirical support for IRH-consistent behavior [such as in Falkenstein (1996) and Huberman (2001)] warrants further theoretical analysis of what this hypothesis implies when examined outside the static mean-variance world.

This article adds a dynamic dimension to the IRH. Specifically, because of information costs, a class of investors is assumed to have incomplete information, which suffices to implement only a particular trading strategy. We refer to this formulation of the IRH as the generalized IRH (G-IRH). Under the G-IRH, portfolio rebalancing is treated as if it were subject to constraints that may evolve stochastically over time (and, as a special case, may exclude a nonvisible stock from a portfolio). To better understand the impact of such constraints on equilibrium, we work in a familiar and well-understood framework. We present a continuous-time general equilibrium model of a Lucas (1978)-type pure-exchange economy, which is populated by heterogeneous agents. Only a subset of the population faces portfolio constraints. Under the G-IRH, we analyze implications for the risk-return trade-off, the risk-free spot interest rate, and the optimal consumption policy of each class of agents. The case of pure IRH (P-IRH), where informationally constrained investors trade only a subset of stocks, is then studied in detail.

The intertemporal feature of our model is cast in a continuous-time framework for tractability. The portfolio choice of informationally constrained investors can then be analyzed using recently developed duality techniques [He and Pearson (1991), Cvitanic and Karatzas (1992)], which augment the martingale-representation approach of Karatzas et al. (1987) and Cox and Huang (1989). Agents in our economy have time-additive state-independent utility functions, and we assume that informationally constrained investors have logarithmic preferences. We characterize equilibrium using construction of a representative agent with time-additive but state-dependent utility.²

The main results are as follows. First, under the G-IRH, we provide a new characterization of risk premia in a two-beta consumption-based capital asset

¹ Neither in Merton (1987) nor in this article do issues of asymmetric information arise; trade always occurs between equally informed investors, and the pure IRH coincides with an assumption of segmented capital markets. International segmentation has been analyzed in a two-date mean-variance setting [e.g., Subrahmanyam (1975) and Errunza and Losq (1985)], and in continuous-time production economies [e.g., Sellin and Werner (1993) and Devereux and Saito (1997)]. These models are equally applicable in a domestic context under an appropriate variant of the IRH. Indeed, Errunza and Losq (1985) and Merton (1987) are close methodologically and share similar implications. Levy (1978) also studies a static mean-variance model of domestic segmentation.

² Analysis of frictionless markets commonly assumes that a subset of agents has logarithmic utility in order to derive explicit solutions [e.g., Dumas (1989) and Wang (1996)]. With constraints, this assumption is made almost without exception by the recent continuous-time literature [e.g., Detemple and Murthy (1997) and Basak and Cuoco (1998)]. The representative agent's utility is state-dependent whenever each class of agents uses a different system of state prices to value future consumption [Cuoco and He (1994)].

pricing model (CCAPM). The first beta is with respect to changes in aggregate consumption, as in the CCAPM of Breeden (1979). However, our formulation of the IRH reduces the dimensionality of the investment opportunity set for a subset of agents; informationally constrained investors effectively trade a single portfolio, referred to as the IRH index. This incompleteness also affects unconstrained agents who are forced to clear the market. Therefore risk premia depend on an additional term that reflects the spanning properties of the IRH index, and this term varies across assets depending on the beta of each asset with respect to changes in the IRH index.

Second, the dynamics of the interest rate are modified to depend on the volatility of the IRH index. Overall, two sources of volatility drive the interest rate: the exogenous volatility of dividends and the endogenous volatility of returns. For example, under the P-IRH, where the IRH index represents unrestricted assets, the interest rate depends on the volatility of aggregate dividends, as in the unconstrained benchmark case. However, the interest rate also depends on the endogenously determined correlation between unrestricted assets and aggregate dividends. This dependence stems from the nature of market incompleteness under the P-IRH for which the interest rate and the risk premia must compensate so that all markets clear.³

Third, focusing on the P-IRH, it becomes evident that the conclusions of static mean-variance models, as to the effect of constraints on risk premia in the cross section, do not hold in our setting. For example, our model suggests that a risk premium on a less visible stock need not be higher than that on a more visible stock with a lower volatility, all else being equal. The reason for this result is that an asset whose risks cannot be shared may still offer considerable benefit as a hedge against shifts in investment opportunities. This result is important for the growing empirical literature that examines the effects of listing stocks on a more visible exchange [e.g., Kadlec and McConnell (1994) and Foerster and Karolyi (1999)].⁴

Methodologically, our model builds on and complements the work of Basak and Cuoco (1998), who study restricted stock market participation with a single risky asset. Our analysis incorporates several risky assets and what we believe is more realistic investment behavior. With multiple risky assets, we derive cross-sectional implications. With more flexible constraints, we accommodate a variety of departures from the benchmark model and

³ An endogenously determined interest rate under a different set of constraints is derived by Sellin and Werner (1993) and Devereux and Saito (1997). They fix exogenously constant volatilities for linear production technologies, and by construction cannot allow any endogenous role to stock market volatility.

⁴ In a static model with constant absolute risk aversion preferences but with consumption in the initial and final periods, Basak (1996) demonstrates the nonrobustness of many results in the extant mean-variance literature without intertemporal consumption. Still, his model agrees with the cross-sectional implications of that literature. Cuoco (1997) provides a general characterization of risk premia under constraints in a continuous-time economy (he analyzes a partial equilibrium and has no implications for the interest rate). Nevertheless, the mapping of the constraints that we examine into an explicit two-beta CCAPM is a new result.

derive new results. We also illustrate how restricted participation presents a special case in our model.

To explore some of the empirical content of the G-IRH, we examine the joint implication of our premise and our model for the variation in the cross section of unconditional expected returns. This central theme of empirical finance is a subject of numerous studies. To facilitate comparison with prior research, while keeping the empirical analysis focused, we subject the model to portfolios designed by Fama and French (1992) and subsequently analyzed by Jagannathan and Wang (1996; hereafter JW), among others. Surprisingly, this broad cross section of portfolios has not been examined outside the so-called CAPM debate.

Consistent with our premise, the return on the IRH index is measured by a return on a combination of two proxies. The first proxy, in adherence with Merton's (1987) arguments, represents large firms. The second proxy intends to capture the return on the portion of wealth invested in pension funds, which account for an increasing fraction of U.S. equities—more than 25% at the end of our sample period [Lakonishok, Shleifer, and Vishny (1992)]. Consequently, we identify the second proxy with a portfolio that is biased toward stocks with good past return performance—consistent with a characterization of the pension fund industry by Lakonishok, Shleifer, and Vishny (1997).

The main econometric approach we use is the two-pass cross-sectional regression. We corroborate results using both ordinary and generalized least squares procedures with an empirical design that draws from Shanken (1992) and JW. In addition, we test our econometric specification using the Hansen and Jagannathan (1997) distance, and also use finite-sample likelihood-ratio tests to examine the implications of our framework for the composition of the unconditionally tangent portfolio. Within the context of our econometric specification, the findings indicate that the CCAPM augmented by the IRH performs better than other models. In particular, over the period covered by the Fama and French (1992)/JW sample, the data fail to reject the joint validity of our premise and our model, and we are able to explain more than 55% of the cross-sectional variation in average real monthly and quarterly returns.

The remainder of the article is organized as follows. Section 1 describes the economy. Section 2 maps the G-IRH into portfolio constraints and solves the individual's optimization problem. Section 3 characterizes the equilibrium and provides our main asset pricing results. Section 4 lays out the empirical design and reports the findings. Section 5 concludes. The appendixes contain the proofs.

1. The Economy

We consider a finite-horizon ($[0, T]$) economy. Aside from incorporating constraints, the setting is standard, and given our focus is on characterization,

we do not state the required regularity conditions [which can be found, e.g., in Karatzas and Shreve (1998)]. Uncertainty is represented by a filtered probability space $(\Omega, \mathcal{F}, \mathbf{F}, P)$, on which is defined a two-dimensional Brownian motion $w(t) = (w_1(t), w_2(t))^\top$. A state of the world is described by $\omega \in \Omega$. The filtration $\mathbf{F} = \{\mathcal{F}_t\}$ is the augmentation under P of the filtration generated by w ($\mathcal{F} = \mathcal{F}_T$). All random processes are assumed progressively measurable with respect to $\mathbf{F}$. All equalities and inequalities involving random variables are understood to hold P -almost surely. There is a single perishable consumption good (the numeraire) and $\mathcal{C}$ denotes the set of nonnegative consumption-rate processes c .

Investment opportunities are represented by three securities. The “bond” is in zero net supply and earns instantaneous interest r over $[0, T]$. The bond price process B satisfies

$$dB(t) = B(t)r(t) dt. \quad (1)$$

We normalize the initial bond value to unity, without loss of generality. The “stocks” are each in a constant supply of one unit. A stock is a claim to an exogenous dividend paid at a strictly positive rate. Denote by S_j , $j = 1, 2$, the ex-dividend stock price process. Let δ_j denote the dividend rate process corresponding to S_j . The aggregate dividend rate process δ is given by

$$d\delta(t) = d\delta_1(t) + d\delta_2(t) = \mu_\delta(t) dt + \sigma_\delta(t)^\top dw(t), \quad (2)$$

where $\mu_\delta = \mu_{\delta_1} + \mu_{\delta_2}$ and $\sigma_\delta = \sigma_{\delta_1} + \sigma_{\delta_2}$ are set exogenously. We assume that, in equilibrium, S_j follows an Itô process:

$$dS_j(t) = (S_j(t)\mu_j(t) - \delta_j(t)) dt + S_j(t)\sigma_j(t) dw(t). \quad (3)$$

The interest rate process r , the drift coefficients $\mu \equiv (\mu_1, \mu_2)^\top$, and the volatility (diffusion) matrix $\sigma \equiv \{\sigma_{jk}, j = 1, 2; k = 1, 2\}$ may be path dependent, and are to be determined endogenously in equilibrium. The σ matrix is assumed to have full rank.

The economy is populated by two types of agents. Let $\alpha_i(t)$ denote the amount that agent i , ($i = 1, 2$) invests at time t in the bond. Let $\theta_i(t) \equiv (\theta_{i1}(t), \theta_{i2}(t))^\top$ be the amount invested in stocks. Agent 1 does not face constraints on θ_1 , whereas agent 2 is restricted in his choice of θ_2 (as specified in Section 2.1). Preferences of agent i are represented by a time-additive von Neumann–Morgenstern instantaneous utility function $u_i(c)$, yielding the expected utility functional $U_i(c) = E[\int_0^T e^{-\rho t} u_i(c(t)) dt]$, where $\rho > 0$ is the rate of subjective time preferences. We assume that $u_2(c) = \log c$, u_1 is three times continuously differentiable, and u'_1 has a continuous and strictly decreasing inverse f_1 that maps $(0, \infty)$ onto itself. Agent 2 is endowed only

with $b > 0$ units of the bond.⁵ Agent 1 initially owns both stocks and $-b$ units of the bond.

A trading strategy (α_i, θ_i) is said to (strictly) finance a consumption plan $c_i \in \mathcal{C}$ if the corresponding wealth process, $W_i \equiv \alpha_i + \theta_i^\top \bar{1}$, satisfies the dynamic budget constraint

$$dW_i(t) = [W_i(t)r(t) + \theta_i(t)^\top (\mu(t) - r(t)\bar{1}) - c_i(t)]dt + \theta_i(t)^\top \sigma(t)dw(t), \quad (4)$$

where $\bar{1} \equiv (1, 1)^\top$. An *arbitrage opportunity* is a nonzero $c \in \mathcal{C}$ that can be financed with zero initial wealth. A trading strategy is *admissible* if $W_i(t) \geq 0$ (a sufficient condition to rule out arbitrage opportunities). The set of admissible trading strategies is denoted by Θ .

2. The Individual Optimization Problem Under the IRH

We depart from the standard setting by acknowledging that, when choosing trading strategies, agent 2 may be affected by real-world frictions not captured in the above description of the economy. Rather than model these frictions explicitly, we treat them in a reduced form using portfolio constraints.

2.1 The IRH and portfolio constraints

Consider the following family of stochastic constraints imposed on agent 2:

$$\mathcal{A}(t, \omega) = \{(\alpha_2(t, \omega), \theta_2(t, \omega)) : \theta_{21}(t, \omega) = q_1(t, \omega)\theta_{22}(t, \omega), \\ (t, \omega) \in [0, T] \times \Omega\},$$

where $q_1(t)$ is a stochastic process that can depend on the dynamics of asset prices. $\mathcal{A}$ reflects our premise that frictions exogenous to the model cause agent 2 to resort to a trading strategy that is suboptimal. Information costs are assumed to be the primary cause for a behavior that deviates from one based solely on the fundamentals of Section 1. This family of constraints allows us to model a variety of trading rules. Special cases of interest are as follows:

- (a) $q_1(t) = \bar{q}$, for a constant $\bar{q}$. ($\bar{q} = \infty$ is understood as unconstrained investment in stock 1 and zero investment in stock 2.) We will elaborate in the sequel on the case of $\bar{q} = 0$; it models incomplete information about stock 1, as discussed by Merton (1987). In particular,

⁵ This simplified endowment structure ensures that at $t = 0$ the stock investment of agent 2 complies with any constraint that belongs to the family described in Section 2.1. For a given member of that family, we can specify a more general endowment structure at the cost of introducing additional notation and, except for the P-IRH case, at the cost of a further restriction on the endogenously determined values of $S_1(0)$ and $S_2(0)$ (a restriction that may interfere with the existence of equilibrium). A different endowment, provided it admits an equilibrium, will affect the equilibrium path, but will not affect our equilibrium characterization results in the sequel.

it applies to an economy with multiple stock exchanges, where some investors will not trade a stock unless it is listed on a visible exchange such as the NYSE [Kadlec and McConnell (1994)]. Similarly, investors may not hold shares in small firms that lack extensive media coverage [Falkenstein (1996)]. $0 < \bar{q} < 1$ indicates a less extreme preference toward stock 2. This applies in cases where agents invest (or short) more in familiar stocks [Huberman (2001)] or in stocks with longer listing histories [Barry and Brown (1984)], or exhibit home-biased patterns in an international or a domestic context [Coval and Moskowitz (1999)].

- (b) $q_1(t) = \frac{s_1(t)}{s_2(t)}$ means that agent 2 holds an equal number of shares in each asset. Since the supply of each stock is normalized to one, agents trade a fraction of the market portfolio. Hence one-fund separation holds. If $q_1(t) = \bar{q} 1_{(\omega \in E_t)}$, agent 2 includes stock 1 in his portfolio only if he learns about the stock through some $\mathcal{F}_t$ -measurable event, E_t (e.g., if the rate of return on stock 1 during $[t - \Delta t, t]$ exceeds some benchmark). Our empirical analysis focuses on (a combination of) the strategies in (a) and (b).
- (c) Consider an exogenous process V whose dynamics are

$$dV(t) = \mu_V(t) dt + v_1(t) dw_1(t) + v_2(t) dw_2(t), \quad (5)$$

where $v_1^2 + v_2^2 \neq 0$. The process may represent a macroeconomic indicator or an index that summarizes information such as analysts' forecasts. Suppose that agent 2 uses the innovations in V to form a trading strategy. It is easy to verify that

$$q_1(t) = - \frac{\sigma_{21}(t)v_2(t) - \sigma_{22}(t)v_1(t)}{\sigma_{11}(t)v_2(t) - \sigma_{12}(t)v_1(t)} \quad (6)$$

allows the agent to choose a portfolio so that the corresponding wealth process is perfectly, instantaneously correlated with V . On the other hand,

$$q_1(t) = - \frac{\sigma_{21}(t)v_1(t) + \sigma_{22}(t)v_2(t)}{\sigma_{11}(t)v_1(t) + \sigma_{12}(t)v_2(t)} \quad (7)$$

allows the agent to choose a portfolio that is instantaneously uncorrelated with V . A particular example of V is a process that maintains a prespecified correlation with δ ($\rho_{V\delta} = \bar{\rho}$).⁶ Anticipating future results, when V indeed coincides with aggregate dividends ($\bar{\rho} = 1$) then, in equilibrium, the economy is equivalent to an unconstrained economy if Equation (6) holds, and is equivalent to a restricted participation

⁶ We abuse notation slightly by using ρ without subscripts to denote the agents' impatience for consumption, and ρ with subscripts to denote instantaneous correlation conditional on $\mathcal{F}_t$.

economy (defined by $\theta_2 \equiv 0$) if Equation (7) holds. (In Appendix A.2, we present another example in the context of which we also verify the existence of equilibrium.)

Agent 2 may implement a strategy on his own. Equivalently, he may invest in stocks via a “managed fund.” Consistent with $\mathcal{A}$, when $q_1(t) \neq -1$, a unit of wealth invested in the fund is split by the fund manager into $\frac{q_1(t)}{q_1(t)+1}$ and $\frac{1}{q_1(t)+1}$, which are reinvested in stock 1 and stock 2, respectively. It is convenient to introduce a fund price process whose dynamics are given by

$$dF(t) = F(t)\mu_F(t)dt + F(t)\sigma_F(t)dw(t), \quad (8)$$

where

$$\mu_F(t) = q(t)^\top \mu(t) / q(t)^\top \bar{1}, \quad \sigma_F(t) = q(t)^\top \sigma(t) / q(t)^\top \bar{1},$$

and $q(t) = (q_1(t), 1)^\top$. Note that $\text{rank}(\sigma_F(t)) = 1$ because $\sigma(t)$ has full rank. Also, without loss of generality, set $F(0) = 1$. The wealth-evolution equation [Equation (4)] for agent 2, subject to the constraint that $(\alpha_2, \theta_2) \in \mathcal{A}$, can be restated using Equation (8) as

$$dW_2(t) = \alpha_2(t) \frac{dB(t)}{B(t)} + (W_2(t) - \alpha_2(t)) \frac{dF(t)}{F(t)} - c_2(t)dt,$$

which illustrates that the constrained agent allocates wealth between the bond and the fund, and effectively faces an incomplete market. F summarizes the investment opportunities of agent 2 in risky assets. We call F the “IRH index” because agent 2 must “recognize” (i.e., have information about) the dynamics of F . As long as he recognizes F , he does not have to recognize (be informed about) the dynamics of individual stocks. For brevity, we refer to the constrained position of agent 2 in equities as a position in the IRH index. The amount invested in the IRH index will be determined based on maximizing expected utility.⁷

⁷ The process in Equation (8) is introduced for expositional purposes. Although we study the microbehavior of agents, the IRH index can be viewed, along the lines of Merton (1992, chaps. 14 and 16), as a fund offered to investors by an *intermediary*. The class of investors who face relatively high information costs will prefer to invest (at least part of) their funds through intermediaries. The fund’s management style, q , will not be optimal for each investor within the class, but the economy of information achieved by this investment vehicle presumably compensates investors relative to the costs of investing directly [Grossman (1995)]. Results are derived using projections constructed from $q^\top \sigma$, and all goes through with $q_1 = -1$. Then F is interpreted as a zero-investment position (with μ_F and σ_F redefined not to include the $q^\top \bar{1}$ denominator), and instead of being long (short) the IRH index, agent 2 takes a positive (negative) exposure to it in the absolute amount of $|\theta_{21}| = |\theta_{22}|$. A key point to emphasize is that although σ_F is constructed from σ , the model (via constraints) mirrors a world in which learning about σ is “too costly” for agent 2, leading him to learn only about σ_F —from observing F (through the quadratic covariation of F with w_1 and w_2), or from exogenous sources (not explicit elements of our model, such as intermediaries) that have cheap access to the primary information (σ). Although actual costs are not incorporated directly into the model, it is monitoring the vector σ_F .

2.2 Consumption and portfolio choice

We characterize the optimal consumption of agent 2, using the duality approach of Cvitanić and Karatzas (1992), as if the agent faces a unique state-price density process of a fictitious, unconstrained economy. Equipped with this state-price density, which accounts for the constraints faced by the agent, we can proceed to derive other quantities, analogous to the complete-markets case.

Proposition 1. *The optimal consumption policy of agent 2 with $(\alpha_2, \theta_2) \in \mathcal{A}$ satisfies*

$$c_2^*(t) = e^{-\rho t} / (\psi_2 \pi_2(t)), \quad (9)$$

where $\psi_2 = (1 - e^{-\rho T})/\rho b$ is the Lagrange multiplier associated with the static budget constraint of the agent, $E[\int_0^T \pi_2(s) c_2^*(s) ds] = b$. The state-price density process faced by the agent is

$$\pi_2(t) = B(t)^{-1} \exp\left(-\int_0^t \kappa_2(s)^\top dw(s) - \frac{1}{2} \int_0^t \|\kappa_2(s)\|^2 ds\right) \quad (10)$$

with the relative risk process expressed by

$$\kappa_2(t) = \Sigma_F(t) \kappa(t), \quad (11)$$

where $\kappa(t) = \sigma(t)^{-1}(\mu(t) - r(t)\bar{1})$ and $\Sigma_F(t) = \sigma_F(t)^\top (\sigma_F(t) \sigma_F(t)^\top)^{-1} \sigma_F(t)$ are the relative risk process faced by agent 1, and the projection matrix on $\text{Span}(\sigma_F)$, respectively.

When agent 2 follows the optimal policy, we interpret $\pi_2(t, \omega)$ in Equation (10) as his Arrow–Debreu price (per unit of probability P) of one unit of consumption good at state ω and time t . Equation (9) is the usual result that $e^{-\rho t} u'_i(c_i^*(t)) = \psi_i \pi_i(t)$, which holds for $i = 1, 2$. At the optimum, the marginal benefit from an additional unit of consumption at state ω and time t is proportional to the cost of that unit. The cost structure faced by agent 2 accounts for the nature of the allowed trading strategy as specified in (11). The relative risk process used for $\pi_2(t)$ is a projection of the relative risk process faced by the unconstrained agent ($\kappa_1(t) \equiv \kappa(t)$) on a restricted investment opportunity set summarized by $\sigma_F(t)$.

rather than the matrix σ , that renders the investment/consumption problem of agent 2 a relatively “low-cost” task; indeed, as shown in the sequel, it is σ_F (and not σ) which is used to construct the optimal policies of agent 2. Moreover, our model is easily generalized to N sources of uncertainty, with $K < N$ strategies (“funds”/“investment styles”), and then the informational burden on agent 2 (in terms of monitored processes) is only of $O(K \times N)$ instead of $O(N^2)$. The model does not prevent agent 2 from learning about σ , but requiring him to learn only about σ_F is aimed to capture realistic scenarios where, due to informational frictions, agent 2 indeed learns only about σ_F , but not σ . Obviously, if F is driven by some subset of w ’s components, agent 2 need recognize only that subset.

Note that $\pi_2(t)$ can indeed be viewed as the unique state-price density process in a fictitious unconstrained economy, where the drift of risky assets is given by $\mu + \nu$. The “shadow” process ν is set to

$$\nu(t) = -\sigma(t)\Pi_F(t)\kappa(t), \quad (12)$$

where

$$\Pi_F(t) = I - \Sigma_F(t)$$

is the projection-matrix process on the space orthogonal to $\text{Span}(\sigma_F)$, and I is the identity matrix. Optimal policies of an agent with (U_2, b) in the fictitious unconstrained economy, with price coefficients $(r, \mu + \nu, \sigma)$, coincide with optimal policies of our constrained agent.⁸ Standard arguments, along with Equation (12), then imply that the stock investment, $W_2 - \alpha_2^*$, is

$$\begin{aligned} \bar{1}^\top \theta_2^*(t) &= \bar{1}^\top (\sigma(t)\sigma(t)^\top)^{-1} (\mu(t) + \nu(t) - r(t)\bar{1}) W_2(t) \\ &= \frac{\mu_F(t) - r(t)}{\|\sigma_F(t)\|^2} W_2(t), \end{aligned} \quad (13)$$

where $W_2(t) = E(\int_t^T \frac{\pi_2(s)}{\pi_2(t)} c_2^*(s) ds \mid \mathcal{F}_t) = \frac{1 - e^{-\rho(T-t)}}{\rho} c_2^*(t)$.

3. Characterization of Equilibrium Under the IRH

This section provides our main results, which include characterization of risk premia, interest rates, and consumption policies. Under the IRH, the economy is denoted $\mathcal{E}$ and is identified by its primitives: $\mathcal{E} \equiv \{(\Omega, \mathcal{F}, \mathbf{F}, P), b, \delta_1, \delta_2, u_1(\cdot), \log(\cdot), \mathcal{A}\}$. We focus first on the economy in its general formulation (which corresponds to the G-IRH). Then we elaborate on the case where $q_1(t) \equiv 0$ (the P-IRH), allowing us to contrast the model with the static literature. Results are compared to a benchmark, complete-markets unconstrained economy: $\mathcal{E}_U \equiv \{(\Omega, \mathcal{F}, \mathbf{F}, P), b, \delta_1, \delta_2, u_1(\cdot), \log(\cdot)\}$.

Definition 1. A competitive rational expectations equilibrium is a price system (r, μ, σ) and a set $\{c_i^*, (\alpha_i^*, \theta_i^*)\}_{i=1}^2$, such that

- (a) $c_i^* = \arg\max_{c \in \mathcal{C}_i} U_i(c)$, $i = 1, 2$, where
- $\mathcal{C}_1 \equiv \{c \in \mathcal{C} : c \text{ is financed by } (\alpha_1, \theta_1) \in \Theta, \text{ with } \alpha_1(0) + \theta_1(0)^\top \bar{1} = S_1(0) + S_2(0) - b\}$,
- $\mathcal{C}_2 \equiv \{c \in \mathcal{C} : c \text{ is financed by } (\alpha_2, \theta_2) \in \Theta, \text{ with } \alpha_2(0) + \theta_2(0)^\top \bar{1} = b \text{ and } (\alpha_2(t), \theta_2(t)) \in \mathcal{A}(t)\}$.

⁸ In the fictitious economy, the presence of the constraints in $\mathcal{A}$ is mapped into a modified drift, but does not affect the interest rate because no restrictions are imposed on α_2 .

(b) *The consumption-good market and the securities markets clear:*

$$\begin{aligned} c_1^*(t) + c_2^*(t) &= \delta(t), & \alpha_1^*(t) + \alpha_2^*(t) &= 0, \\ \theta_{1j}^*(t) + \theta_{2j}^*(t) &= S_j(t), & j &= 1, 2. \end{aligned}$$

It is convenient to characterize equilibrium quantities using a construction of a representative agent. Let $u(c, \lambda) \equiv \max_{c_1+c_2=c} u_1(c_1) + \lambda u_2(c_2)$, for some $\lambda > 0$. In $\mathcal{E}_U$, a representative agent's utility, $u(\delta(t), \lambda(0))$, is a linear combination of individual utilities [see, e.g., Duffie and Zame (1989), Karatzas, Lehoczky, and Shreve (1990)]. In $\mathcal{E}$, because agents use different state-price density processes to price consumption, the representative agent is characterized by a state-dependent utility, $u(\delta(t), \lambda(t))$, with a stochastic weighting process [as in, e.g., Cuoco and He (1994), Basak and Cuoco (1998)]. The absolute risk aversion and absolute prudence [as in Kimball (1990)] of agent i and of the representative agent are denoted by $A_i(t)$, $P_i(t)$, $A(t)$, and $P(t)$, respectively:

$$\begin{aligned} A_i(t) &= -\frac{u_i''(c_i^*(t))}{u_i'(c_i^*(t))}, & P_i(t) &= -\frac{u_i'''(c_i^*(t))}{u_i''(c_i^*(t))}, \\ A(t) &= -\frac{u_{cc}(\delta(t), \lambda(t))}{u_c(\delta(t), \lambda(t))}, & P(t) &= -\frac{u_{ccc}(\delta(t), \lambda(t))}{u_{cc}(\delta(t), \lambda(t))}. \end{aligned}$$

3.1 The generalized investor recognition hypothesis

We assume that an equilibrium exists and provide its characterization.

Theorem 1. *In equilibrium, the weighting process, $\lambda(t) = u'_1(c_1^*(t))c_2^*(t)$, is a strictly positive solution to the stochastic differential equation:*

$$d\lambda(t) = -\lambda(t)A_1(t)\sigma_\delta(t)^\top \Pi_F(t) dw(t), \quad (14)$$

and $\lambda(0)$ is the unique strictly positive solution to

$$b = (1 - e^{-\rho T})\rho^{-1}\lambda(0)u_c(\delta(0), \lambda(0))^{-1}. \quad (15)$$

The interest rate and risk premia are

$$\begin{aligned} r(t) &= \rho + A(t)\mu_\delta(t) - \frac{1}{2}A(t)P(t)\|\sigma_\delta(t)\|^2 \\ &\quad - \frac{1}{2}A(t)(P_1(t) - P(t))\|\Pi_F(t)\sigma_\delta(t)\|^2, \end{aligned} \quad (16)$$

$$\mu(t) - r(t)\bar{1} = A(t)\sigma(t)\sigma_\delta(t) + (A_1(t) - A(t))\sigma(t)\Pi_F(t)\sigma_\delta(t). \quad (17)$$

Optimal consumption policies are $c_1^*(t) = f_1(u_c(\delta(t), \lambda(t)))$, $c_2^*(t) = \lambda(t)/u_c(\delta(t), \lambda(t))$ with

$$dc_i^*(t) = \mu_{c_i^*}(t) dt + \sigma_{c_i^*}(t)^\top dw(t), \quad i = 1, 2, \quad (18)$$

where

$$\begin{aligned}
 \sigma_{c_1^*}(t) &= \frac{A(t)}{A_1(t)} \sigma_\delta(t) + \left(1 - \frac{A(t)}{A_1(t)}\right) \Pi_F(t) \sigma_\delta(t), \\
 \mu_{c_1^*}(t) &= \frac{A(t)}{A_1(t)} \mu_\delta(t) + \frac{1}{2} \frac{A(t)^2}{A_1(t)} \left[\frac{P_1(t)}{A_1(t)} - \frac{P(t)}{A(t)} \right] \|\sigma_\delta(t)\|^2 \\
 &\quad - \frac{1}{2} \frac{A(t)^2}{A_1(t)} \left\{ \frac{P_1(t) - P(t)}{A(t)} + \frac{P_1(t)}{A_1(t)} \left(1 - \frac{A_1(t)^2}{A(t)^2}\right) \right\} \|\Pi_F(t) \sigma_\delta(t)\|^2, \\
 \sigma_{c_2^*}(t) &= \frac{A(t)}{A_2(t)} \sigma_\delta(t) - \frac{A(t)}{A_2(t)} \Pi_F(t) \sigma_\delta(t), \\
 \mu_{c_2^*}(t) &= \frac{A(t)}{A_2(t)} \mu_\delta(t) + \frac{1}{2} \frac{A(t)^2}{A_2(t)} \left[\frac{P_2(t)}{A_2(t)} - \frac{P(t)}{A(t)} \right] \|\sigma_\delta(t)\|^2 \\
 &\quad - \frac{1}{2} \frac{A(t)^2}{A_2(t)} \left\{ \frac{P_1(t) - P(t)}{A(t)} + \frac{P_2(t)}{A_2(t)} \right\} \|\Pi_F(t) \sigma_\delta(t)\|^2.
 \end{aligned}$$

The relative risk processes the agents face are

$$\begin{aligned}
 \kappa_1(t) &= A(t) \sigma_\delta(t) + (A_1(t) - A(t)) \Pi_F(t) \sigma_\delta(t), \\
 \kappa_2(t) &= A(t) \sigma_\delta(t) - A(t) \Pi_F(t) \sigma_\delta(t).
 \end{aligned} \tag{19}$$

The results in Equations (14)–(19) are better understood in light of the following corollary:

- Corollary 1.** (a) When the portfolio of agent 2 is perfectly correlated with $\delta(t)$ and $b \leq \delta(0)(1 - e^{-\rho T})/\rho$, then $\lambda(t) = \lambda(0)$, and there exists a unique equilibrium for $\mathcal{E}$, which coincides with that in $\mathcal{E}_U$, with all equilibrium quantities obtained by substituting $\Pi_F(t) \sigma_\delta(t) = 0$ in Theorem 1. In particular, $\mu(t) - r(t)\bar{1} = A(t) \sigma(t) \sigma_\delta(t)$.
- (b) When the portfolio of agent 2 is uncorrelated with $\delta(t)$, $b \leq \delta(0)(1 - e^{-\rho T})/\rho$, and there exists a unique, strictly positive solution $\lambda(t)$ to the stochastic differential equation $d\lambda(t) = -\lambda(t) A_1(t) \sigma_\delta(t)^\top dw(t)$, then there exists a unique equilibrium for $\mathcal{E}$, which coincides with that in a restricted-participation economy ($\theta_2^*(t) = 0$), with all equilibrium quantities obtained by substituting $\Pi_F(t) \sigma_\delta(t) = \sigma_\delta(t)$ in Theorem 1. In particular, $\mu(t) - r(t)\bar{1} = A_1(t) \sigma(t) \sigma_\delta(t)$.
- (c) In equilibrium, the risk premium of stock j is given by

$$\begin{aligned}
 \mu_j(t) - r(t) &= a_1(t) \text{cov} \left(\frac{dS_j(t)}{S_j(t)}, \frac{d\delta(t)}{\delta(t)} \right) \\
 &\quad + a_2(t) \text{cov} \left(\frac{dS_j(t)}{S_j(t)}, \frac{dF(t)}{F(t)} \right), \quad j = 1, 2, \tag{20}
 \end{aligned}$$

where

$$a_1(t) = A_1(t)\delta(t), a_2(t) = -A(t)\frac{A_1(t)}{A_2(t)}\rho_{F\delta}(t)\frac{\|\sigma_\delta(t)\|}{\|\sigma_F(t)\|}.$$

- (d) The risk premium of the IRH index is given by $\mu_F(t) - r(t) = A(t) \times \sigma_F(t)\sigma_\delta(t)$, and the risk premium of δ^* is given by $\mu_{\delta^*}(t) - r(t) = A(t)\sigma_{\delta^*}(t)\sigma_\delta(t)[1 + (1 - \rho_{F\delta}(t)^2)A_1(t)/A_2(t)]$, where δ^* is a stock-only portfolio perfectly correlated with δ . The risk premium in Equation (20) is equivalently given by $\mu_j(t) - r(t) = b_{j\delta^*}(t)(\mu_{\delta^*}(t) - r(t)) + b_{jF}(t)(\mu_F(t) - r(t))$, where $b_{j\delta^*}$ and b_{jF} are the multiple-regression coefficients.

Corollary 1(a) shows how the familiar characterization results for the unconstrained, benchmark economy $\mathcal{E}_U$ are obtained in our setting. Corollary 1(b) asserts that the economy analyzed by Basak and Cuoco (1998) can also arise as a special case in our model, and although their model is formulated with a single stock, we see that their characterization results are readily extended to the case of multiple risky assets.

In $\mathcal{E}$, as in $\mathcal{E}_U$, agents adjust their marginal rate of substitution for consumption at different states and in different times to equal the relative prices of consumption (as in Section 2.2). Since agents face different (nonnegatively correlated) state-price density processes, the marginal rate of substitution of agent 1 between any two points in $\Omega \times [0, T]$ is different from the marginal rate of substitution of agent 2. The two agents will not choose consumption in the same fashion. Hence, unlike in $\mathcal{E}_U$, the optimal consumption policies in Equation (18) are not perfectly correlated with aggregate consumption, but rather maintain a nonnegative correlation.

The interest rate has a new fourth term in Equation (16) compared to its expression in $\mathcal{E}_U$. To understand this term, suppose that agent 1 has a decreasing absolute risk aversion (so that $P_1 > 0$) and is the more prudent agent (so that $P_1 > P$). Then, compared to $\mathcal{E}_U$, agent 1 must be further deterred from investing in the bond (and encouraged to borrow) to fulfill his dominant role in clearing the stock market. Therefore, the new term acts to reduce the interest rate, so that in equilibrium it will counteract the precautionary savings motive of agent 1. The spanning characteristics of F determine the role of agent 1 in clearing the stock market, and hence the extent to which his prudence affects r via the new term.

The additional state variable in $\mathcal{E}$, compared to $\mathcal{E}_U$, is $\lambda(t)$ and this is translated into a “two-beta” CCAPM in Equation (17). At any instant, the excess expected return on assets is a linear function of their covariance with the two state variables (δ, λ) . In turn, the instantaneous covariance of $\lambda(t)$ with $\delta(t)$ is $-\lambda(t)A_1(t)\|\Pi_F(t)\sigma_\delta(t)\|^2$. Hence $\lambda(t)$ tends to increase during times of recession and to decrease during times of expansion. Although $\lambda(t)$, in general, cannot be observed by an econometrician, the cross-sectional

implications of Equation (17) are formulated using quantities with acceptable empirical counterparts. Corollary 1(c) clarifies the risk-return trade-off under the IRH by rewriting the risk premia in Equation (17) using a conventional notation of instantaneous conditional covariances. The model predicts that the cross section of instantaneous expected returns, conditional on $\mathcal{F}$, is proportional to instantaneous conditional covariances between the rate of return on asset j and two economic variables. The first variable is consumption growth, as in the CCAPM of Breeden (1979). The second variable is the rate of return on the IRH index. For our formulation of the IRH (i.e., when $q(t)$ is set to mirror our hypothesis about the impact of information costs) the discrete-time analog of the latter return can be constructed using data on asset returns in the economy.

To understand the intuition for the cross-sectional differences between securities, consider an example in which $\rho_{F\delta} \geq 0$, where stocks have equal consumption betas but $\sigma_1\sigma_F > \sigma_2\sigma_F > 0$. First, when $\rho_{F\delta} = 1$, agent 2 behaves as he would have optimally behaved in $\mathcal{E}_U$, and hence $\mathcal{E}$ coincides with the benchmark case [in particular, Equation (20) reduces to $A\sigma_j\sigma_\delta$, as in Corollary 1(a)]. Second, when agent 1 holds the entire stock supply ($\rho_{F\delta} = 0$), only the first term in Equation (20) remains, as is the case in Corollary 1(b), because agent 1 does not share stock market risks with agent 2. Third, when $0 < \rho_{F\delta} < 1$, if agent 2 becomes very risk averse ($A_2 \rightarrow \infty$) due to shrinking wealth, the first term again dominates (with $A \rightarrow A_1$), and there are no cross-sectional differences; agent 1 effectively becomes the representative agent in a complete market, because agent 2 is a negligible participant in the economy (agent 2 is an insignificant player in the stock market due to high risk aversion, and low W_2 implies he is also insignificant in the money market). Finally, a different picture arises when A_2 is finite. As noted, the first term in Equation (20), $A_1\sigma_j\sigma_\delta$, is the correct premium when agent 1 alone bears the stock market risks. Since $A_1 = A + A\frac{A_1}{A_2}$ (see appendix), the loading is increased relative to the aggregate risk aversion, A (i.e., that of a representative agent), to describe the premium had agent 2 not held stocks and had no incentives to share stockmarket risks. This premium must be adjusted because agent 2 is present to share risks with ($A_2 < \infty$), and there are indeed risks to share ($\rho_{F\delta} > 0$). The magnitude of the adjustment depends on $A\frac{A_1}{A_2}$, which captures the ability of agent 2 to share risks, while $\rho_{F\delta}$ captures the incentives agent 2 has to share risks. $\|\sigma_\delta\|/\|\sigma_F\|$ acts as a scale factor to normalize the asset-specific correction $\sigma_j\sigma_F$. To see the direction of the correction, in the context of this example, note from (18) and $\rho_{F\delta} > 0$ that c_2^* is positively (in fact, perfectly) correlated with the IRH index, and agent 2 must then be long in the IRH index to finance his consumption policy. Agent 1 effectively holds the entire stock supply and takes a short position in the IRH index. Favorable realizations of δ tend to coincide with favorable realizations of F , which is undesirable for agent 1. It is stock 1 that offers him better insurance against an unfavorable realization of the short position. Because

of its hedging value, stock 1 becomes more desirable than stock 2 and is required to offer a relatively lower risk premium.

Corollary 1(d) provides the risk premium for two portfolios (δ^*, F) and uses these to express risk premia of individual stocks. The special structure of the two portfolios yields somewhat simplified formulations for their risk premia. The premium of the IRH index agrees with its expression in $\mathcal{E}_U$ because this portfolio is (the only one) accessible to both agents and hence is priced through the risk aversion of the representative agent alone. This will not hold for any other portfolio. For example, when $\sigma_F \sigma_\delta = \sigma_{\delta^*} \sigma_\delta > 0$, δ^* commands a higher risk premium than F because, unlike with F , agent 1 cannot share with agent 2 the risks associated with δ^* .

3.2 The pure investor recognition hypothesis

The P-IRH, in its static form, has attracted considerable interest in the literature. To gain more insight about the P-IRH in a dynamic world, let stock 2 represent the visible asset and stock 1 represent the asset not recognized by agent 2.

Proposition 2. *In equilibrium, with $q_1(t) \equiv 0$:*

(a) *The interest rate and risk premia are given by*

$$\begin{aligned}
 r(t) &= \rho + A(t)\mu_\delta(t) - \frac{1}{2}A(t)P(t)\|\sigma_\delta(t)\|^2 \\
 &\quad - \frac{1}{2}A(t)(P_1(t) - P(t))(1 - \rho_{2\delta}(t)^2)\|\sigma_\delta(t)\|^2, \\
 \mu_2(t) - r(t) &= A(t)\sigma_2(t)\sigma_\delta(t), \\
 \mu_1(t) - r(t) &= A(t)\sigma_1(t)\sigma_\delta(t) + \Upsilon(t)\frac{S_1(t)}{W(t)}(1 - \rho_{12}(t)^2)\text{var}\left(\frac{dS_1(t)}{S_1(t)}\right) \\
 &\quad + \Upsilon(t)\text{cov}\left(\frac{dS_1(t)}{S_1(t)} - \rho_{12}(t)\frac{\|\sigma_1(t)\|}{\|\sigma_2(t)\|}\frac{dS_2(t)}{S_2(t)}, \frac{d\delta(t)}{\delta(t)} - \frac{dW(t)}{W(t)}\right), \quad (21)
 \end{aligned}$$

where $W(t) = W_1(t) + W_2(t)$, and $\Upsilon(t) = R(t)\frac{1}{R_2(t)}\frac{c_2^*(t)}{c_1^*(t)}$, $R_2(t) = A_2(t)c_2^*(t)$, $R(t) = A(t)\delta(t)$.

(b) *When $u_1(\cdot) = \log(\cdot)$, and the aggregate endowment follows a geometric Brownian motion $d\delta(t) = \delta(t)\mu_\delta dt + \delta(t)\sigma_\delta^\top dw(t)$, where μ_δ and $\sigma_\delta = (\sigma_{\delta,1}, \sigma_{\delta,2})^\top$ are constant, then:*

(i) *The relative risk processes are $\kappa_1(t) = \sigma_\delta + \lambda(t)\Pi_2(t)\sigma_\delta$, and $\kappa_2(t) = \sigma_\delta - \Pi_2(t)\sigma_\delta$. Agent 1 (agent 2) faces state prices with higher (lower) volatility than in $\mathcal{E}_U$.*

- (ii) *The optimal consumption policies are*
 $c_i^*(t) = \delta(t)W_i(t)/W(t)$, $i = 1, 2$, *where*

$$\mu_{c_1^*}(t) = (\mu_\delta + (1 + \lambda(t))\lambda(t)(1 - \rho_{2\delta}(t)^2)\|\sigma_\delta\|^2)c_1^*(t),$$

$$\sigma_{c_1^*}(t) = (\sigma_\delta + \lambda(t)\Pi_2(t)\sigma_\delta)c_1^*(t),$$

$$\mu_{c_2^*}(t) = (\mu_\delta - (1 + \lambda(t))(1 - \rho_{2\delta}(t)^2)\|\sigma_\delta\|^2)c_2^*(t),$$

$$\sigma_{c_2^*}(t) = (\sigma_\delta - \Pi_2(t)\sigma_\delta)c_2^*(t).$$

The expected consumption growth rate of agent 1 and the volatility of his consumption growth rate are higher than in $\mathcal{E}_U$, while those of agent 2 are lower.

- (iii) *The welfare of agent 1, $U_1(c_1^*)$, is higher than in $\mathcal{E}_U$, while $U_2(c_2^*)$ is lower.*

- (iv) *The interest rate is $r(t) = \rho + \mu_\delta - \|\sigma_\delta\|^2 - \frac{W_2(t)}{W_1(t)}(1 - \rho_{2\delta}(t)^2)\|\sigma_\delta\|^2$, and is lower than in $\mathcal{E}_U$. For a given distribution of wealth, $r(t)$ increases with $\rho_{2\delta}(t)^2$.*

- (v) *The risk premia are $\mu_1(t) - r(t) = \sigma_1(t)\sigma_\delta + \frac{W_2(t)}{W_1(t)}\frac{S_1(t)}{W(t)}(1 - \rho_{12}(t)^2)\|\sigma_1(t)\|^2$, $\mu_2(t) - r(t) = \sigma_2(t)\sigma_\delta$. For a given $\sigma_1(t)$, the risk premium of stock 1 is higher than in $\mathcal{E}_U$.*

Proposition 2(a) states that when $P_1(t) > P(t)$, $r(t)$ has a parabolic dependence on $\rho_{2\delta}(t)$ with the minimum at $\rho_{2\delta}(t) = 0$. Clearly, under the P-IRH, fluctuations in $r(t)$ are explicitly related to fluctuations in $\sigma_2(t)$, all else being equal. This link between the volatility of the visible portion of the market and the interest rate is a novelty of our model.

The risk premia differ across securities depending on their visibility (e.g., exchange listing status). The second and third terms on the right-hand side of Equation (21) modify the risk premium of stock 1 for two reasons compared to its expression in $\mathcal{E}_U$. First, there is a change in diversification opportunities, because portfolios adjust to the constraint. Second, there is a change in the ability of agent 1 to hedge against shifts in the investment opportunities. Because the second term in Equation (21) compensates agent 1 for bearing all the risk of stock 1, it is positive and is similar to the compensation for lost risk-sharing opportunities predicted by static models [Errunza and Losq (1985), Merton (1987), Basak (1996)]. Contrary to Merton's measure of investor base, our model identifies the proportionality factor Υ with the ratio of consumption streams, normalized by relative risk aversions. The third term in Equation (21) arises because, when smoothing consumption, agent 1 owns the entire supply of stock 1. If stock 1 and stock 2 are very close substitutes ($\rho_{12} \rightarrow 1$), the second and third terms are insignificant. If the dynamics of aggregate wealth coincide with the dynamics of aggregate consumption (e.g., if both agents are myopic and do not hedge intertemporally), then the third term vanishes. In general, however, the sum of the second and third terms in

Equation (21) need not be positive, because full ownership of stock 1 may be a desirable strategy if the value of stock 1 as an intertemporal hedging instrument outweighs the cost of owning its entire supply. It is easy to verify that if, for example, $\sigma_1\sigma_\delta = \sigma_2\sigma_\delta > 0$, then $\mu_1 < \mu_2$ holds if and only if $\|\sigma_1\|\rho_{12} > \|\sigma_2\|$. Therefore, contrary to mean-variance results, a less visible stock with a higher volatility may nevertheless be required to offer a lower expected return, all else being equal.

Under the G-IRH, the event of listing stock 1 on a more visible exchange, at time t_L , corresponds to q_1 being zero up to t_L , and then, over $[t_L, T]$, as required by the benchmark model. Clearly, if consumption betas remain stable during the listing event, then, for some stocks, one may detect a higher expected return after listing. Consequently, when averaging abnormal returns across securities in event time, the impact of listing can take many forms. This may potentially account for the somewhat inconclusive results in exchange-listing studies [see, e.g., Kadlec and McConnell (1994, pp. 614–615)].

In Proposition 2(b), to clarify differences between the economy under the P-IRH and $\mathcal{E}_U$, we impose further structure on u_1 and on the dynamics of $\delta(t)$. The weighting process then coincides with the wealth distribution: $\lambda(t) = W_2(t)/W_1(t)$ and $1 + \lambda(t) = W(t)/W_1(t)$. Also note that $\sigma_\delta = \sigma_w(t)/W(t)$. Unambiguous, *direct* comparisons ($\forall (t, \omega) \in [0, T] \times \Omega$) between $\mathcal{E}$ and $\mathcal{E}_U$ are provided in items (i)–(iv): Agent 1 bears more risk and has more volatile consumption compared to agent 2 and compared to the benchmark. Under the P-IRH, the interest rate is lower, which induces agent 1 to hold stock 1, and the lower borrowing costs increase his welfare.

Risk premia in (v) deserve two comments. First, both agents are myopic, and the extra term in the risk premium of stock 1 is positive. However, in an intertemporal model, this per se does not imply a higher risk premium compared to the benchmark [as illustrated by Basak (1996)]. The risk premium on stock 2 is also ambiguously related to its benchmark value. Second, $\lambda = W_2/W_1$ measures the relative investor base for stock 1, and from Equation (14) it is negatively correlated with δ . Intuitively, given a positive outlook for future dividends, $W_1(t)$ tends to increase relative to $W_2(t)$ because agent 1 benefits from full ownership of stock 1. Therefore our model, even in its myopic version, has a new implication for the P-IRH; all else being equal, the cross-sectional differences between securities with different visibility are countercyclical.

4. Empirical Evidence

Our premise is that the costs of gathering and processing data lead some investors to focus on stocks with high visibility and also to entrust a portion of their wealth to money managers employed by pension plans. Hence, the return on the portfolio of informationally constrained investors (ICIs) is

characterized as a combination of two proxies: a proxy for the return on the investment in visible stocks, and a proxy for the return on the indirect investment via pension plans. Given these proxies, we test the implications of our premise for the cross section of unconditional expected returns.

4.1 Econometric specification

The pricing equation [Equation (17)] and its two-beta reformulation in Equation (20) are readily generalized to N assets, as shown in Appendix A.1. Let $r_{j,t+1} = [S_j(t+1) + \int_t^{t+1} \delta_j(s)ds - S_j(t)]/S_j(t)$, $g_{t+1} = [\delta(t+1) - \delta(t)]/\delta(t)$, and $h_{t+1} = [F(t+1) - F(t)]/F(t)$, where $j = 1, \dots, N$; one unit of time corresponds to a month or a quarter, t takes discrete values, g_{t+1} is the growth rate of aggregate consumption, and the IRH is characterized by $q(t) = (q_1(t), \dots, q_{N-1}(t), 1)$, fixed between t and $t+1$, so that the rate of change in the IRH index is $h_{t+1} = \sum_{j=1}^N \bar{q}_{jt} r_{j,t+1}$, where $\bar{q}_{jt} = q_j(t) / \sum_{j=1}^N q_j(t)$. Using the stochastic Euler approximation to Equations (1)–(3) and (8), we can restate Equation (20) as

$$\begin{aligned} E[r_{j,t+1}|\mathcal{F}_t] &= a_{0t} + a_{1t}\text{cov}[r_{j,t+1}, g_{t+1}|\mathcal{F}_t] \\ &\quad + a_{2t}\text{cov}[r_{j,t+1}, h_{t+1}|\mathcal{F}_t], \quad j = 1, \dots, N, \end{aligned} \quad (22)$$

where a_{1t} and a_{2t} are as given in Equation (20), and a_{0t} captures interest-rate fluctuations and approximation errors (assumed to be homoscedastic). Equation (22) is the starting point of our empirical analysis. We examine whether the *conditional* formulation in Equation (22) is consistent with the crosssection of *unconditional* expected returns. Using Equation (22), with additional assumptions stated in Appendix B.1, we get the following result:

Theorem 2. Assume that $\beta_{jg} = \text{cov}(r_{jt}, g_t)/\text{var}(g_t)$, $\beta_{jh} = \text{cov}(r_{jt}, h_t)/\text{var}(h_t)$ exist and are linearly independent. If $\bar{q}_t$ is known $\forall t$, then there exist some constants (a_0, a_1, a_2) such that

$$E[r_{jt}] = a_0 + a_1\beta_{jg} + a_2\beta_{jh}. \quad (23)$$

If $\bar{q}_t$ is unknown, assume the IRH index to be a combination of two portfolios; $h_{t+1} = w_t h_{1,t+1} + (1 - w_t) h_{2,t+1}$, where w_t is unknown but the weights within $h_{1,t+1}$, $h_{2,t+1}$ are known. Let $\beta_{jh_p} = \text{cov}(r_{jt}, h_{pt})/\text{var}(h_{pt})$, $p = 1, 2$. Then there exist some constants (a_0, a_1, a_2, a_3) such that

$$E[r_{jt}] = a_0 + a_1\beta_{jg} + a_2\beta_{jh_1} + a_3\beta_{jh_2}. \quad (24)$$

The specification in Equation (24) agrees with our premise, and we will refer to it as the G-IRH model (denoted $M_{G\text{-IRH}}$). It is testable given the empirical counterparts of (g, h_1, h_2) .⁹

⁹ The unconditional specification does not explicitly incorporate the particular structure of the coefficients in Equation (22). A test based on Theorem 2 lacks power against a model that has covariance structure as in

To examine the prediction of Theorem 2, we adopt a two-pass cross-sectional regression (CSR) approach. In the first pass, each univariate beta is estimated using ordinary least squares (OLS). The second pass is a single CSR of average returns on betas, also conventionally estimated using OLS. The advantages of estimating the CSR with generalized least squares (GLS) are improved asymptotic efficiency [Shanken (1992)] and robustness to proxy misspecification [Kandel and Stambaugh (1995)]. For GLS, however, we need the inverse of the unknown covariance matrix of returns. Neither estimation approach is decisively superior. Using monthly consumption data is desirable to increase the number of time observations and get more precise estimates. Quarterly intervals are likely to yield a more accurate measurement of consumption growth [Breedon, Gibbons, and Litzenberger (1989)]. Overall, as detailed in Appendix B.2, we report monthly and quarterly CSR results estimated with OLS and GLS (at the second pass, where standard errors are corrected for a bias induced by OLS sampling errors in the first-pass univariate betas). Since expected returns vary cross-sectionally, if the model is valid we must have at least one nonzero slope coefficient. We check this using the Wald test statistic. Under the CSR approach to testing asset pricing models, if the Wald test, based on both OLS and GLS estimates, rejects the null hypothesis of zero slopes, this is interpreted as a failure to reject the model. Then the Hausman (1978) specification (HS) test can assess whether the OLS and GLS estimates are as close as a correctly specified model would imply.¹⁰

Another way to evaluate Equation (24) is to restate it as

$$E[(1 + r_{jt})(b_0 + b_1 g_t + b_2 h_{1t} + b_3 h_{2t})] = 1, \quad (25)$$

where (b_0, b_1, b_2, b_3) are some constants. The term $y_t \equiv b_0 + b_1 g_t + b_2 h_{1t} + b_3 h_{2t}$ is the stochastic discount factor implied by Equation (24). The empirical proxies for (g_t, h_{1t}, h_{2t}) may not coincide with their theoretical counterparts, leading to the use of a misspecified proxy for the true discount factor. If no proxy can correctly price the N assets, then for a set of discount factor proxies that correspond to different models (e.g., CAPM, CCAPM, or M_{G-IRH}), it is of interest to quantify how misspecified one proxy is compared to the

Equation (24) with different loadings, but to the best of our knowledge no other model in the literature offers theoretical justification to combine consumption growth with the proxies for the visible and retirement-oriented portfolios to explain expected returns. Clearly, one can transform Equation (24) into an alternative cross-sectional formulation with multivariate betas, which are the slopes in a time-series regression of excess returns on the portfolio unconditionally most highly correlated with consumption growth (MCP), h_1 , and h_2 . Since constructing the MCP introduces another source of estimation error, we choose to focus on the implications in Equations (23) and (24). (We return to the unconditional mean-variance implications of the model later.) Also note that our econometric specification relies on assumptions that one would ideally like to translate to assumptions on the model's primitives. Unfortunately, the complexity of the model renders such a mapping infeasible.

¹⁰ $HS = (\hat{a}_{OLS} - \hat{a}_{GLS})^\top (\text{var}(\hat{a}_{OLS}) - \text{var}(\hat{a}_{GLS}))^{-1} (\hat{a}_{OLS} - \hat{a}_{GLS})$, where hats denote estimates, and $\hat{a}$ excludes the intercept. HS (as the Wald statistic) has an asymptotic chi-square distribution with $\dim(\hat{a})$ degrees of freedom. HS is reported only when both OLS and GLS Wald statistics reject the null hypothesis, $H_0 : a = 0$.

others. Therefore, we estimate the magnitude of misspecification using the Hansen and Jagannathan (1997) distance (HJ-d). For a correctly specified discount factor, the HJ-d is zero. Hence we test the model (i.e., the null hypothesis of a correctly specified discount factor) by testing whether the estimated HJ-d is insignificantly different from zero.¹¹

4.2 Identification of the IRH index

We now identify the empirical counterparts for (h_1, h_2) . For the h_1 component, our premise is that agent 2 considers only the stocks visible to him—those about which he has sufficient information to implement optimal portfolio rebalancing. Information about the larger firms is likely to be available at a lower cost, and we identify visibility with large capitalization. The claim that large firms are more widely known is consistent with the evidence that large firms have more shareholders [Merton (1987)]. Moreover, large firms usually have longer listing histories. Falkenstein (1996) reports that both the size and age of a firm are positively correlated with the number of news stories in major newspapers about that firm. We further assume that a single index can capture well the investment of agent 2 in visible stocks. The natural proxy to use then for h_1 is the Standard & Poor's 500 index (S&P 500). Agent 2 is not required to have detailed information about the 500 large-capitalization firms in the index, although he must know enough to optimally rebalance wealth between the S&P 500 portfolio and other investments. In fact, since the S&P 500 is a good market proxy, only market-wide information may suffice.

For h_2 , our premise states that agent 2 entrusts a portion of his wealth to money managers who have better access to information. For this element, we focus on pension funds. Agent 2 receives all necessary reporting from the sponsor to be able to optimally allocate wealth between the money market, the S&P 500, and his pension fund. For data-availability reasons, we rely on evidence provided by Lakonishok, Shleifer, and Vishny (1997) (LSV) to construct h_2 .¹² LSV characterize the aggregate portfolio (a “superfund”) of a large collection of tax-exempt pension funds. Their sample covers about 20%

¹¹ Let $y_i(b)$ be the examined proxy that depends on b . The HJ-d is the minimum least-squares distance, in L^2 space, between $y_i(b)$ and an m_i that correctly prices the N assets. In our case, the HJ-d is given by $d = [(E[R_i y_i(b)] - 1)^T (E[R_i R_i^T])^{-1} (E[R_i y_i(b)] - 1)]^{1/2}$. To compute it we use consistent estimates of the moments (with the b that minimizes d). The HJ-d has an appealing interpretation; it is the maximum pricing error when using y_i to price positions in the N assets, where position payoffs are standardized to have a unit norm in L^2 . Under the null, an estimator of the HJ-d has an asymptotic distribution equal to a mixture of chi-square distributions. We report the associated p -value (computed as suggested by JW). Under the alternative of a misspecified proxy, the estimator is asymptotically normal, and we report its standard error [accounting for serial correlation as in Newey and West (1987) with lag 12 for monthly data and lag 4 for quarterly data].

¹² We do not consider explicitly the mutual fund sector. This will not bias the results if the investment of ICI in mutual funds is spanned by our proxies for h_1 and h_2 . Moreover, investments in mutual funds are made directly by individuals, while investments in pension funds are guided by the fund sponsor. Note that h_1 captures investment by ICI in large firms, whether owned directly or indirectly through equity and index funds. The key idea is that the “investment style” of h_1 is *chosen directly* by ICI based on the visibility argument; this is not so for the investment style of h_2 , which is merely reported to them, and is taken into

of the total actively managed equity holdings of pension funds. LSV find that, relative to the S&P 500, the superfund has a high proportion of stocks with good long-term past return performance. For example, on average, 65.2% of the S&P portfolio is invested at any given time in stocks that over the past three years performed better than the stock in the S&P 500 with a median performance. The comparable figure for the superfund is 83.9%. LSV call this overexposure to well-performing stocks — a glamour bias. We assume that the fund component of agent 2 mirrors the LSV superfund. Hence we construct h_2 as follows: At the beginning of each period t , we compare the past three-years' return of each of the N assets in our sample with the return on the S&P 500. Only assets that outperformed the S&P 500 are selected. Their equally weighted return over period t defines the value of h_{2t} . This procedure, however crude, mimics the LSV characterization of the superfund.

A test of M_{G-IRH} nests several tests: When $a_1 = a_3 = 0$, we test the CAPM with the S&P 500 as the market proxy (henceforth, M_{CAPM}). When $a_2 = a_3 = 0$, we test the CCAPM (denoted M_{CCAPM}). Setting $a_2 = 0$ identifies the IRH index with a glamour-biased portfolio; the tested specification, denoted M_{GLAM} , assumes that agent 2 invests in stocks only via its professionally managed retirement funds. Finally, we set $a_3 = 0$ to test the premise of the P-IRH that agent 2 invests only in visible stocks (henceforth, M_{P-IRH}). M_{P-IRH} is of particular interest because it implies that consumption beta and market beta jointly determine the cross-sectional variation in expected returns. Mankiw and Shapiro (1986) examine which beta is more related to returns using 464 NYSE stocks, with the S&P 500 as the market proxy.¹³ They conclude that, unlike the market beta, the consumption beta is unrelated to expected returns. However, their sample suffers from a survivorship bias. Epstein and Zin (1991) and Bakshi and Chen (1996) analyze models with the consumption beta and the market beta in the pricing equation.¹⁴ They do not focus on comparative beta performance and use a small number of assets in their empirical investigations. Campbell (1996) builds upon the Epstein and Zin (1991) model and concludes that the covariance with the market appears to capture most of the cross-sectional variation in expected returns across the

account in their asset allocation. The different decision process in arriving to h_2 adds an extra layer of potential distortions in investments and hence justifies our separate treatment of retirement assets. To the extent that h_2 is designed to capture professional/institutional investment style, our construction of h_2 in this section is also in line with the Cai, Kaul, and Zheng (2001) findings of positive-feedback institutional stock trading.

¹³ The literature that empirically examines either the CAPM or the CCAPM, but not both, is too vast to survey here. See Campbell, Lo, and MacKinlay (1997) for more details. Apart from M_{CAPM} , we examine specifications that include the consumption beta in order to focus on dynamic models (with some agents being non myopic) that are generated by the investor recognition paradigm.

¹⁴ In the Epstein and Zin (1991) model, the representative agent has recursive preferences, and the aggregate wealth ("market") enters the pricing equation because it proxies for the subsequent period's utility index. In the model of Bakshi and Chen (1996), wealth enters the pricing equation because the representative agent cares about wealth-induced status. In our model preferences are standard. Wealth enters our pricing equation to account for pressure imposed on the unconstrained agents by those who choose to trade only the market (or its proxy).

25 portfolios that he examines. We provide new evidence on the performance of the consumption beta versus the market beta using a large cross section of portfolios.

4.3 Data and main results

As our N assets, we choose a set of portfolios that has generated considerable interest since its introduction by Fama and French (1992). JW use this set to demonstrate that the conditional CAPM fits the data much better than the static CAPM examined by Fama and French (1992). To facilitate comparisons with that research, we test the model with the $N = 100$ NYSE/AMEX size-beta portfolios used by JW.¹⁵ The data consist of monthly returns from July 1963 through December 1990, and these returns are used to construct the glamour-biased return. The return of the S&P 500 is taken from the Center for Research in Security Prices (CRSP). Consumption data are from CITIBASE. We use per capita personal consumption expenditures on nondurables and services. Consumption and returns are converted into real terms by the implicit price deflator. Monthly growth rates are computed using monthly data. As suggested by Mankiw and Shapiro (1986) and Breeden, Gibbons, and Litzenberger (1989), quarterly growth rates are computed using monthly data as of the end of each quarter.

Reported results are representative of those obtained with similar specifications of (g, h_1, h_2) , such as using consumption of nondurables only or of services only, using the top size decile or a broader index for h_1 , and using a shorter return history when constructing h_2 . To avoid redundancy, we do not report the estimates of b in Equation (25). In general, when an estimate of a_j in Equation (24) is statistically significantly different from zero, so is the estimate of the corresponding b_j . The few exceptions to this do not affect our conclusions.

Table 1 presents estimates for M_{CAPM} , M_{CCAPM} , $M_{\text{P-IRH}}$, and M_{GLAM} . These results are of interest for two reasons. First, excluding the CAPM, these models have not been estimated in previous studies using so large a cross section of portfolios. Second, the results provide perspective for the subsequent investigation of $M_{\text{G-IRH}}$, which is the focus of our analysis. It is clear that neither the CAPM nor the CCAPM is supported by the data. The P-IRH is rejected as well [which is also evidence against the models of Epstein and Zin (1991) and Bakshi and Chen (1996)]; ICI do not limit themselves to index investing. Results for M_{GLAM} indicate somewhat improved performance. The Wald statistics and the HS test are consistent with a correctly specified model, but more than 80% of the cross-sectional variation cannot be accounted for, and the HJ-d indicates that pricing errors are significantly

¹⁵ I thank Ravi Jagannathan and Zhenyu Wang for making their data available to the public; see JW for the description of portfolio formation and for summary statistics. I also thank Robert Stambaugh for the SMB and HML data below (provided to him by Kenneth French).

Table 1
Evaluation of models nested by M_{G-IRH}

	a_0	g	SP500	GLAM	R^2	χ^2	HS	d
Panel A: Monthly Returns, July 1963–December 1990 (330 months)								
M_{CAPM}								
OLS	0.87		−0.17		3.48	0.22		0.6560
t -value	3.60		−0.46			(64.22)		(0.74)
GLS	0.83		−0.36			1.60		[0.0650]
t -value	5.46		−1.27			(20.57)		
M_{CCAPM}								
OLS	0.75	−0.16			1.66	0.21		0.6452
t -value	3.89	−0.46				(64.62)		(2.50)
GLS	0.72	−0.23				6.60		[0.0666]
t -value	5.45	−2.57				(1.02)		
M_{P-IRH}								
OLS	0.87	−0.02	−0.16		3.50	0.22		0.6434
t -value	3.60	−0.14	−0.45			(89.66)		(2.92)
GLS	0.78	−0.22	−0.22			7.15		[0.0668]
t -value	4.98	−2.36	−0.75			(2.80)		
M_{GLAM}								
OLS	0.61	−0.38		0.24	6.63	8.01	2.31	0.6447
t -value	2.48	−2.80		0.59		(1.82)	(31.44)	(2.08)
GLS	0.68	−0.24		0.18		6.91		[0.0669]
t -value	4.53	−2.63		0.54		(3.16)		
Panel B: Quarterly Returns, Q3 1963–Q4 1990 (110 quarters)								
M_{CAPM}								
OLS	1.70		0.43		2.48	0.10		1.0582
t -value	1.79		0.32			(74.67)		(0.26)
GLS	1.68		−0.15			0.03		[0.1006]
t -value	3.66		−0.17			(86.48)		
M_{CCAPM}								
OLS	1.72	0.10			6.27	0.25		1.0329
t -value	2.61	0.50				(61.68)		(3.26)
GLS	1.79	−0.18				5.75		[0.0934]
t -value	4.30	−2.40				(1.64)		
M_{P-IRH}								
OLS	2.34	0.25	−1.16		9.29	1.47		1.0325
t -value	2.76	1.15	−0.86			(47.96)		(2.60)
GLS	1.66	−0.19	0.56			6.02		[0.0929]
t -value	3.51	−2.44	0.56			(4.92)		
M_{GLAM}								
OLS	0.99	−0.28		2.52	18.83	8.12	1.10	1.0231
t -value	1.08	−2.73		1.71		(1.73)	(57.62)	(4.78)
GLS	1.51	−0.22		1.89		8.00		[0.0924]
t -value	3.33	−2.74		1.60		(1.83)		

The table reports estimates of four models (M_{CAPM} , M_{CCAPM} , M_{P-IRH} , M_{GLAM}) that are nested by the cross-sectional regression model M_{G-IRH} in Equation (24),

$$E[r_{jt}] = a_0 + a_1\beta_{jg} + a_2\beta_{jh_1} + a_3\beta_{jh_2}.$$

In panel A, r_{jt} is the real monthly return on a Fama and French (1992)/Jagannathan and Wang (1996) (FF(92)/JW(96)) size-beta portfolio j ($j = 1, 2, \dots, 100$) in month t (July 1963–December 1990), and the GLS estimation uses Equation (B13). In panel B, r_{jt} is the real quarterly return in quarter t (Q3 1963–Q4 1990), and the GLS estimation uses Equation (B14). The β s are the slope coefficients in the OLS regression of r_{jt} on a constant and a variable specified in the column heading: g , the growth rate of consumption expenditures on nondurables and services; h_1 = SP500, the rate of change in the S&P 500 index; h_2 = GLAM, the rate of return on a glamour-biased portfolio. The t -values are corrected based on Equation (B18) for the sampling error in betas. The OLS R^2 and the p -values are reported in percentage points. χ^2 is the Wald statistic for zero slopes excluding the intercept (p -value is in parentheses). HS (reported if both OLS-based and GLS-based Wald tests reject zero slopes) is the Hausman (1978) specification test statistic (p -value is in parentheses). The Hansen and Jagannathan (1997) distance for the proxy implied by Equation (24) is denoted by d (the p -value is reported in parentheses, and is computed under the null hypothesis of a correctly specified model) [the standard error of d under the alternative is reported in brackets].

different from zero.¹⁶ Still, the overall conclusion from Table 1 is that the nested formulations of our premise must be missing important aspects of reality.

Table 2 reports results for M_{G-IRH} . The main result is that, by all criteria, this model explains the cross-sectional variation quite well. The R^2 of 56.55% and the HJ-d of 0.6264, for monthly data, are comparable to those reported by JW for the conditional CAPM. This result is corroborated with quarterly data. The consumption beta enters significantly into the pricing equation despite the presence of the market beta, contrary to the findings of Mankiw and Shapiro (1986). We remark that estimates of b in Equation (25) confirm that g , h_1 , and h_2 are statistically significant components of the stochastic discount factor. Furthermore, after allowing for sampling errors, we cannot reject at the conventional rate of 5% the null hypothesis of a zero HJ-d for the discount factor implied by M_{G-IRH} (while being able to reject the CCAPM and other nested models). Under the alternative of misspecified proxies, the HJ-d has a low power to distinguish between M_{G-IRH} and the models it nests. M_{G-IRH} has the lowest pricing error, however, suggesting that this model incorporates a better descriptive realism.

The findings indicate that asset prices are consistent with our theoretical model and with our specification of both the direct and the delegated components of equity investment under incomplete information. This implies that although index funds appeared in the United States only after 1970, it is likely that implicit index linking (e.g., via S&P 500-like investing in visible, large firms) was implemented by a significant group of investors (the ICI) during the 1963–1990 period that we study. As a side result, we provide indirect confirmation that the LSV sample characterizes well the pension fund industry. The behavior of stock market prices is consistent with the joint hypothesis that the entire tax-exempt money management industry held a glamour-biased portfolio, and that a nonnegligible fraction of these retirement assets was owned by a subset of U.S. workers that otherwise owned only visible stocks.

4.4 Additional investigations

If M_{G-IRH} is the correct specification, adding regressors to Equation (24) should not add significant explanatory ability. To explore this, we begin by considering three specification tests, where each test adds one explanatory variable to M_{G-IRH} . First, we let agent 2 invest in a third portfolio that has a nonglamour bias (constructed each t using those assets that underperformed the S&P 500 over the past three years). The intention is to proxy

¹⁶ The estimated a_1 in M_{GLAM} is negative while the estimated a_3 is positive. Setting $h = h_2$ in Equation (22) yields $a_{1t} > 0$, and $a_{3t} < 0$ if the glamour-biased portfolio has positive conditional correlation with g . Theorem 2 states that such sign reversal for unconditional vs. conditional coefficients is plausible in a well-specified model, and statistical significance of the betas, as a group, is the only sought-after implication of Equation (22). Similarly, no sign restrictions are imposed by JW in their test of the conditional CAPM.

Table 2
Evaluation of M_{G-IRH} and models that nest M_{G-IRH}

	g	SP500	GLAM	a_4	R^2	χ^2	HS	d
Panel A: Monthly Returns, July 1963–December 1990 (330 months)								
M_{G-IRH}								
OLS	−0.21	−1.81	1.69		56.55	9.00	2.28	0.6264
t -value	−1.74	−2.31	1.95			(2.93)	(51.69)	(6.38)
GLS	−0.23	−1.73	1.89			12.87		[0.0723]
t -value	−2.42	−2.54	2.46			(0.49)		
M_{nongl}								
OLS	−0.23	−1.37	3.09	−1.91	58.09	10.29	1.06	0.6223
t -value	−1.92	−1.61	2.32	−1.31		(3.58)	(90.10)	(8.56)
GLS	−0.24	−1.30	2.89	−1.59		13.88		[0.0723]
t -value	−2.51	−1.75	2.58	−1.30		(0.77)		
M_{labor}								
OLS	−0.21	−1.80	1.69	0.00	56.56	9.00		0.6239
t -value	−1.73	−2.30	1.95	0.07		(6.11)		(8.12)
GLS	−0.24	−1.65	1.80	0.04		13.49		[0.0729]
t -value	−2.48	−2.39	2.32	0.99		(0.91)		
M_{size}								
OLS	−0.19	−0.76	0.55	−0.07	58.17			
t -value	−1.60	−0.67	0.46	−1.38				
GLS	−0.22	−0.84	0.93	−0.06				
t -value	−2.38	−0.98	0.98	−1.73				
Panel B: Quarterly Returns, Q3 1963–Q4 1990 (110 quarters)								
M_{G-IRH}								
OLS	−0.21	−6.24	7.40		59.49	9.90	2.51	1.004
t -value	−2.23	−2.11	2.23			(1.94)	(47.37)	(9.88)
GLS	−0.21	−3.77	5.78			10.36		[0.0964]
t -value	−2.62	−1.72	2.19			(1.58)		
M_{nongl}								
OLS	−0.16	−4.70	10.45	−5.08	60.67	10.11	2.00	1.001
t -value	−1.64	−1.55	2.36	−1.08		(3.86)	(73.62)	(9.88)
GLS	−0.20	−2.76	7.72	−3.51		11.22		[0.0965]
t -value	−2.43	−1.13	2.25	−0.89		(2.42)		
M_{labor}								
OLS	−0.18	−6.13	7.10	−0.11	60.02	9.52	2.46	1.002
t -value	−1.83	−2.05	2.14	−0.74		(4.93)	(65.21)	(8.94)
GLS	−0.20	−3.84	5.76	−0.06		10.33		[0.0964]
t -value	−2.32	−1.72	2.17	−0.51		(3.53)		
M_{size}								
OLS	−0.18	−1.82	2.40	−0.31	63.21			
t -value	−1.93	−0.59	0.68	−2.37				
GLS	−0.19	−0.99	2.38	−0.23				
t -value	−2.39	−0.43	0.85	−2.63				

The table reports estimates of (a_1, a_2, a_3) for the cross-sectional regression model M_{G-IRH} in Equation (24),

$$E[r_{jt}] = a_0 + a_1\beta_{jg} + a_2\beta_{jg_1} + a_3\beta_{jg_2},$$

and estimates of (a_1, a_2, a_3, a_4) for the cross-sectional regression models M_{nongl} , M_{labor} , M_{size} defined by

$$E[r_{jt}] = a_0 + a_1\beta_{jg} + a_2\beta_{jg_1} + a_3\beta_{jg_2} + a_4\gamma_j.$$

In M_{nongl} , $\gamma_j = \beta_{jg_3}$, and h_3 is the return on a nonglamour portfolio formed from underperformers (relative to the S&P 500). In M_{labor} , $\gamma_j = \beta_{jg_3}$, and h_3 is the growth rate in per capita labor income. In M_{size} , $\gamma_j = \log(ME_j)$, and ME_j is the equally weighted average of the real market value (in millions of constant dollars) of the stocks in portfolio j . In panel A, r_{jt} is the real monthly return on a FF(92)/JW(96) portfolio j ($j = 1, 2, \dots, 100$) in month t (July 1963–December 1990). In panel B, r_{jt} is the real quarterly return in quarter t (Q3 1963–Q4 1990). The remaining β s, the estimation methods, and the reported statistics are as described in Table 1.

for a “value,” or “contrarian,” investing. We refer to this specification as the nonglamour model (M_{nongl}). Second, following Mayers (1972), we let the IRH index include wealth due to human capital. As in JW, the return on human capital is measured by the growth rate in labor income. Labor income is defined as the difference between total personal income and dividend income, and although it is based on aggregates, it is assumed to be valid for agent 2. We denote this specification by M_{labor} . Third, in the size-based specification (M_{size}), we add a size regressor to control for the size characteristic. Berk (1995) argues that one should expect (the log of) market size to be correlated with expected returns in the cross section. The question is whether the size regressor can explain that portion of cross-sectional variation that is not explained by our model.

Results are given in Table 2. Neither M_{nongl} nor M_{labor} dramatically outperforms $M_{\text{G-IRH}}$. The t -value that corresponds to each of the new regressors is never statistically significant. The R^2 and HJ-d indicate only minor improvements over $M_{\text{G-IRH}}$.¹⁷ Results for M_{size} are mixed.¹⁸ The latter finding suggests that there is a need for further work to more accurately characterize the direct and the delegated components within the portfolio of ICI. Nevertheless, adding the size regressor can explain virtually no cross-sectional variation beyond what is already explained by our model. Figure 1 confirms this conclusion and illustrates visually that our results are not driven simply by a few outliers. We can safely state that the CCAPM augmented by the IRH is a more realistic model than, for example, the CCAPM or the CAPM for explaining the variation in the cross section of average returns.

To examine subperiods, we divide our sample of 330 months into three subsamples of 110 months. The subperiods correspond roughly to the calendar periods of the 1960s, 1970s, and 1980s (and keeping the same number of observations as in the full-sample quarterly analysis facilitates comparison across tables). Table 3 presents the results for M_{CCAPM} , $M_{\text{G-IRH}}$, and M_{size} . The evidence in favor of $M_{\text{G-IRH}}$ is less decisive than with the full sample.¹⁹ However, it is interesting to note that $M_{\text{G-IRH}}$ performs best in the middle subperiod, which to a large extent coincides with the period in which ICI gained easy access to baskets of large firms via index funds.

¹⁷ If growth of aggregate income is a poor proxy for the growth of the income of ICI, then failure of M_{labor} may indicate our failure to identify the fraction of labor income representing ICI.

¹⁸ With monthly data, the size regressor is not significant; however, it is significant using quarterly data. The presence of the size regressor has only a marginal impact on the estimate of a_1 compared to its value and statistical significance in $M_{\text{G-IRH}}$. There is also no impact on the signs of estimated a_2 and a_3 , but their magnitude and statistical significance are reduced.

¹⁹ According to the OLS R^2 and the HJ-d, $M_{\text{G-IRH}}$ performs better in panel B of Table 3 than in panels A and C, or in panel B of Table 2. The HJ-d test rejects the model in the first and last subperiods at the 5% confidence level. The other models examined in this section are rejected as well in these two subperiods (to save space Table 3 summarizes results of three models only). The asymptotic inferences in Table 3 are less reliable than in Tables 1 and 2, because in Table 3 we use monthly consumption growth coupled with a small number of observations relative to the number of examined assets.

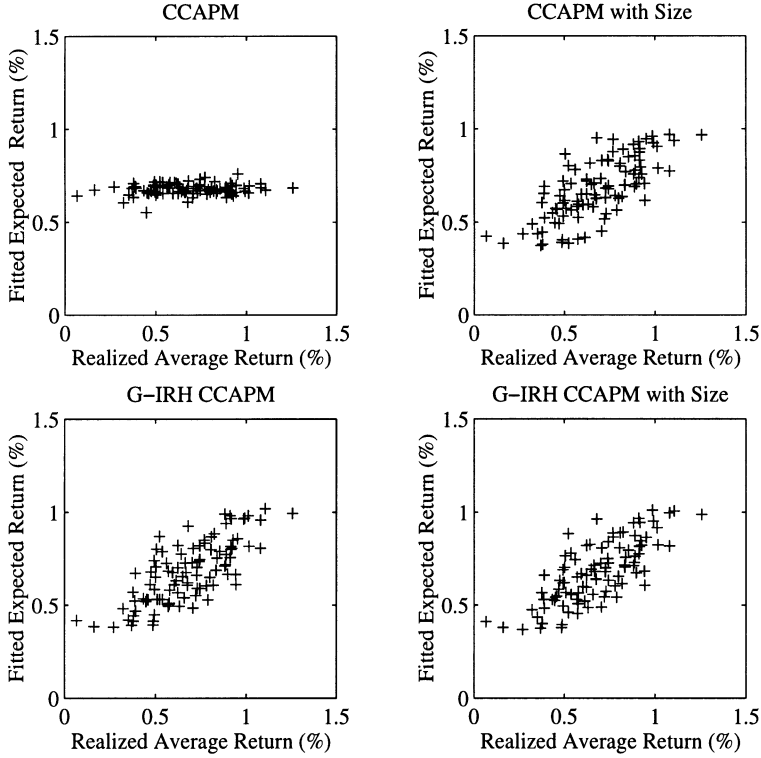


Figure 1

Fitted expected real monthly returns versus realized average real monthly returns

Each scatter point represents a portfolio. The CCAPM graph describes fitted values from the regression model: $E[r_{jt}] = a_0 + a_1\beta_{jg}$. The CCAPM with Size graph describes fitted values from the regression model: $E[r_{jt}] = a_0 + a_1\beta_{jg} + a_4 \log(ME_j)$. The G-IRH CCAPM graph describes fitted values from the regression model: $E[r_{jt}] = a_0 + a_1\beta_{jg} + a_2\beta_{jh_1} + a_3\beta_{jh_2}$. Finally, the G-IRH CCAPM with Size graph describes fitted values from the regression model: $E[r_{jt}] = a_0 + a_1\beta_{jg} + a_2\beta_{jh_1} + a_3\beta_{jh_2} + a_4 \log(ME_j)$. All the variables are as described in Tables 1 and 2. The models are estimated by OLS.

The fact that a three-beta specification performs well in explaining the variation in average returns is not surprising. Our contribution is to illustrate that a consumption-based model is empirically viable, if we account for trading patterns of ICI. To learn more about the spanning power of (g, h_1, h_2) , we combine them with variables suggested by an alternative model. Like JW, we examine the incremental explanatory power of the betas with respect to the size (SMB) and book-to-market-value (HML) factors of Fama and French (1993) (we denote their model by $M_{FF(93)}$). In Table 4, using monthly returns, the SMB and HML betas do not perform well when combined with M_{G-IRH} . The GLS results still favor the M_{G-IRH} specification, thereby offering further support for our formulation of the IRH. Quarterly analysis leads to similar conclusions (results not reported).

Table 3
Subperiod evaluation of M_{G-IRH} and related models

	OLS					GLS					<i>d</i>
	<i>g</i>	SP500	GLAM	<i>a</i> ₄	<i>R</i> ²	<i>g</i>	SP500	GLAM	<i>a</i> ₄		
Panel A: July 1963–August 1972 (110 months)											
<i>M</i> _{CCAPM}	0.21				21.19	−0.04					1.1308
<i>t</i> -value	1.14					−0.84					(0.45)
<i>M</i> _{G-IRH}	−0.09	−1.58	2.48		69.40	−0.07	−0.89	1.87			1.1147
<i>t</i> -value	−1.28	−1.50	1.83			−1.26	−1.05	1.66			(1.21)
<i>M</i> _{size}	−0.09	−0.99	1.73	−0.05	70.18	−0.07	−0.17	0.85	−0.07		
<i>t</i> -value	−1.25	−0.82	1.09	−0.93		−1.37	−0.19	0.70	−1.96		
Panel B: September 1972–October 1981 (110 months)											
<i>M</i> _{CCAPM}	0.18				23.92	−0.07					0.9954
<i>t</i> -value	1.07					−1.33					(2.40)
<i>M</i> _{G-IRH}	−0.02	−2.28	3.39		65.70	−0.09	−2.26	3.32			0.9666
<i>t</i> -value	−0.21	−2.05	2.00			−1.54	−1.89	2.25			(8.42)
<i>M</i> _{size}	0.00	−1.71	1.98	−0.09	66.50	−0.08	−1.86	2.84	−0.03		
<i>t</i> -value	0.02	−1.10	1.07	−1.04		−1.43	−1.35	1.68	−0.54		
Panel C: November 1981–December 1990 (110 months)											
<i>M</i> _{CCAPM}	−0.44				2.43	−0.08					0.9979
<i>t</i> -value	−1.17					−0.37					(0.38)
<i>M</i> _{G-IRH}	0.05	1.38	−2.43		46.99	−0.03	−0.91	0.23			0.9884
<i>t</i> -value	0.19	0.67	−1.20			−0.15	−0.57	0.15			(0.68)
<i>M</i> _{size}	0.09	−4.31	2.93	0.20	53.29	−0.02	−3.61	2.67	0.11		
<i>t</i> -value	0.32	−1.57	1.14	2.61		−0.09	−1.63	1.28	2.12		

The table reports estimates of the cross-sectional regression models M_{CCAPM} (as in Table 1), M_{G-IRH} , and M_{size} (as in Table 2) over three subperiods: panel A: July 1963–August 1972, panel B: September 1972–October 1981, panel C: November 1981–December 1990, using real monthly returns of the 100 FF(92)/JW(96) size-beta portfolios. The estimation methods and the reported statistics are as described in Table 1. GLS estimation and GLS and OLS *t*-values are based on Equation (B14). The Hansen and Jagannathan (1997) distance for the proxy implied by Equation (24) is denoted by *d* (the *p*-value under the null of a correctly specified model is in parentheses).

The inferences in Tables 1–4 are asymptotic. To obtain finite-sample results, recall that Equation (24) implies that some combination of the portfolio unconditionally most highly correlated with consumption growth (MCP), the S&P 500 portfolio, and the glamour-biased portfolio is mean-variance efficient. Gibbons, Ross, and Shanken (1989) (GRS) derive the finite-sample distribution of a likelihood-ratio test statistic, which is widely used to test the efficiency of a combination of portfolios. GRS test a given linear pricing model (the null hypothesis, H_0) against a general alternative hypothesis. Kandel and Stambaugh (1989) show, in the presence of a riskless asset, that the GRS test can also be used to test the model against a specific alternative hypothesis (H_A), where H_A states that the tangent portfolio is a combination of portfolios that include the portfolios under H_0 as a proper subset. We test M_{G-IRH} against both a general and a specific alternative.

Testing M_{G-IRH} against H_A requires information only about the excess returns of the portfolios specified by H_A . A test against a general alternative requires specifying the universe of assets with respect to which the tangency is defined. The power of the GRS test is very sensitive to the number of assets used, so we follow the suggestion of Campbell, Lo, and MacKinlay (1997, chap. 5) and keep the number of assets small. The 100 size-beta portfolios

Table 4

The impact of the Fama and French (1993) Factors on M_{G-IRH}^s

	g	SP500	GLAM	SMB	HML	R^2	χ^2	HS	d
OLS	-0.21	-2.18	2.28	-0.13		56.62	9.77	2.98	0.6250
t -value	-1.83	-1.34	0.96	-0.23			(4.45)	(56.18)	(6.68)
GLS	-0.23	-2.60	3.21	-0.35			12.81		[0.0721]
t -value	-2.45	-1.95	1.69	-0.78			(1.22)		
OLS	-0.21	-1.33	1.41		0.19	56.92	9.27		0.6264
t -value	-1.78	-1.37	1.50		0.70		(5.48)		(6.48)
GLS	-0.23	-1.68	1.86		0.02		12.89		[0.0723]
t -value	-2.42	-2.07	2.29		0.10		(1.18)		
OLS	-0.21	-1.43	1.56	-0.03	0.18	56.92	10.01		0.6249
t -value	-1.83	-0.75	0.61	-0.05	0.66		(7.48)		(6.34)
GLS	-0.23	-2.70	3.31	-0.37	-0.03		12.77		[0.0722]
t -value	-2.44	-1.73	1.61	-0.78	-0.12		(2.56)		

The table reports estimates of $(a_1, a_2, a_3, a_4, a_5)$ for the cross-sectional regression model

$$E[r_{jt}] = a_0 + a_1\beta_{jg} + a_2\beta_{jh_1} + a_3\beta_{jh_2} + a_4\beta_{jSMB} + a_5\beta_{jHML},$$

which adds to the M_{G-IRH} specification the betas from the model of Fama and French (1993). Here r_{jt} is the real monthly (July 1963–December 1990) return on a FF(92)/JW(96) portfolio j ($j = 1, 2, \dots, 100$). The β s are the slope coefficients in the OLS regression of r_{jt} on a constant and a variable specified in the column heading: g , the growth rate of consumption expenditures on nondurables and services; $h_1 = \text{SP500}$; $h_2 = \text{GLAM}$. In $M_{FF(93)}$, SMB and HML are factors designed to capture the risks related to firm size and book-to-market-equity, respectively. The estimation methods and the reported statistics are as described in Table 1.

are used to construct a smaller *universe* of 15 portfolios. These include the three portfolios implied by M_{G-IRH} and an additional 12 *primary* assets. The primary assets are 10 size-based value-weighted portfolios; the nonglamour-biased portfolio used in Table 2; and an equally weighted market portfolio (EWMKT). The power of the GRS test increases with the number of observations, so we work with monthly returns (in excess of a Treasury-bill return, taken from CRSP). As in Table 1, we are interested in examining several models nested by M_{G-IRH} . One way to compare the models is to keep the primary assets fixed and to vary the portfolios under H_0 . Alternatively, one can fix the universe of assets and examine each model with respect to that universe. To be able to implement both methods, the MCP is constructed from the primary assets, where, to avoid collinearity, two equally weighted portfolios replace the EWMKT: one constructed from the 50 small-size portfolios, and the second constructed from the remaining 50. This MCP is denoted MCP*. A second construction, MCP**, uses the 100 original assets.²⁰

²⁰ To use the GRS finite-sample results, we must assume that (i) the weights of the MCP are estimated without error, (ii) an unconditionally real-riskless asset exists, and its rate is known, (iii) excess returns are normally distributed. Using a small number of assets to construct the MCP* increases the precision of the estimated weights, but the sample correlation of the MCP* with g is only 0.17. For MCP** the correlation with g increases to 0.61. The annualized Sharpe ratios of the MCP* and the MCP** are 0.11 and 0.53, respectively. It is apparent that an extra source of noise is introduced by transforming Equation (24), which uses a univariate beta with respect to g , into a linear pricing in terms of multivariate betas with respect to portfolios. This is the reason we implemented the CSRs using univariate betas. Although we do not defend the assumptions in (i)–(iii) above, and although one can attempt to relax some or all of them, we nevertheless perform the GRS test to provide additional evidence regarding our model, and to offer another dimension along which our results can be compared with results of others. However, we view the likelihood-ratio tests only as an auxiliary tool of investigation.

Table 5
Evaluation of M_{G-IRH} using likelihood ratio tests

Panel A: Tests against a general alternative

H_0	Fixed primary assets			Fixed universe of assets		
	n_1, n_2	F^*	F^{**}	n_1, n_2	F^*	F^{**}
M_{CAPM}	12,317	2.102 (1.65)	2.102 (1.65)	14,315	1.828 (3.38)	2.480 (0.24)
M_{CCAPM}	12,317	2.217 (1.09)	2.318 (0.75)	14,315	1.957 (2.06)	2.064 (1.35)
M_{P-IRH}	12,316	2.102 (1.66)	2.083 (1.78)	13,315	1.963 (2.33)	1.954 (2.41)
M_{GLAM}	12,316	1.831 (4.26)	1.805 (4.64)	13,315	1.726 (5.46)	1.715 (5.66)
M_{G-IRH}	12,315	1.770 (5.21)	1.766 (5.28)	12,315	1.770 (5.21)	1.766 (5.28)

Panel B: Tests against a specific alternative

H_0	CRSP VW as a market proxy			S&P 500 as a market proxy		
	n_1, n_2	F^*	F^{**}	n_1, n_2	F^*	F^{**}
M_{G-IRH}	3,324	6.920 (0.02)	6.704 (0.02)	2,325	6.767 (0.13)	6.423 (0.18)
$M_{FF(93)}$	3,324	2.906 (3.48)	5.679 (0.08)	2,325	0.506 (60.31)	4.557 (1.12)
$M_{G-IRH} + SMB$	2,324	9.819 (0.01)	9.666 (0.01)	1,325	12.434 (0.05)	12.081 (0.06)
$M_{G-IRH} + HML$	2,324	3.493 (3.15)	3.555 (2.97)	1,325	0.010 (92.03)	0.080 (77.70)

The table reports likelihood ratio tests of M_{G-IRH} and related models using real excess monthly returns (July 1963–December 1990). Under the null hypothesis (H_0), the tangent portfolio is a combination of K_1 portfolios as implied by the model stated in the first column. Panel A reports tests of H_0 against a general alternative. In columns 2–4 tangency is defined with respect to $K_1 + 12$ assets (the K_1 components of the tangent plus 12 primary assets, where K_1 varies across models). The fixed 12 primary assets are the 10 value-weighted (VW) size-based portfolios, the EWMKT portfolio, and the nonglamour-biased portfolio created from the 100 original assets (the FF(92)/JW(96) size-beta portfolios) used in Tables 1–4. In columns 5–7 tangency is defined with respect to a fixed universe of 15 assets (the 12 primary assets plus the $K_1 = 3$ portfolios implied by M_{G-IRH} —the MCP, the S&P 500, and the glamour-biased portfolio). Panel B reports tests of H_0 against a specific alternative (H_A). Under H_A , a K_2 -beta model describes unconditional expected returns, where $K_2 > K_1$. $M_{FF(93)}$ specifies the $K_1 = 3$ portfolios to be the market proxy, the SMB, and the HML portfolios. In the last two rows one portfolio from $M_{FF(93)}$ is added to augment M_{G-IRH} (so that $K_1 = 4$). Two market proxies are examined in panel B: In columns 2–4 the CRSP VW portfolio is used, and H_A combines the three portfolios specified by M_{G-IRH} with the three portfolios specified by $M_{FF(93)}$ (so that under the alternative $K_2 = 6$). In columns 5–7 the S&P 500 proxies for the market portfolio. In this case M_{G-IRH} and $M_{FF(93)}$ agree on one of the three components of the tangent portfolio and hence H_A states that $M_{G-IRH} + SMB + HML$ is the true model ($K_2 = 5$). F^* is the GRS statistic when the MCP is constructed from the primary assets (as described in Section 4.4). F^{**} is the GRS statistic when the MCP is constructed from the 100 original assets. The p -values (in parentheses) are in percentage points. n_1 and n_2 are the numerator and the denominator degrees of freedom, respectively, of the F -distribution of the GRS statistic under H_0 .

Results are reported in Table 5. In panel A, when keeping the primary assets fixed, we cannot reject M_{G-IRH} at the 5% significance level, while the nested models are rejected. However, when keeping the universe of assets fixed, both M_{G-IRH} and M_{GLAM} cannot be rejected—the universe of assets is not rich enough to allow the GRS test to distinguish between the two. The OLS results in Table 4 indicate that although the univariate SMB and HML betas do not add explanatory power to M_{G-IRH} , they do affect the significance of β_{jh_1} and β_{jh_2} . Panel B investigates the impact of SMB and HML further. The likelihood ratio test indicates that M_{G-IRH} is rejected in favor of the combined six-beta model (M_{G-IRH} plus $M_{FF(93)}$). $M_{FF(93)}$ is rejected as well in favor of the combined model (unless we use the MCP* and redefine

$M_{FF(93)}$ to use the S&P 500 as the market proxy). According to the first two rows of panel B, each model seems to miss important features captured by the other model. The last two rows reveal that the information M_{G-IRH} misses may, to some extent, be captured by the HML portfolio [which Fama and French (1993) interpret as related to relative earnings distress]. It may be worthwhile to attempt to refine our premise and consider accounting for how individual investors (or the intermediaries to whom the investment decision is delegated) treat firms in financial distress. The overall evidence in Table 5, notwithstanding problems in implementing the likelihood ratio tests, reinforces the conclusion from Tables 1–4. The model has its shortcomings, but its theoretical implications seem to have enough empirical support in order to justify moving farther along the IRH track.

5. Conclusion

The CCAPM has advanced our understanding of the most fundamental question in finance—the trade-off between risk and return. Some argue that it may be necessary to take a step beyond the rational expectations revolution to better understand the workings of the capital markets [Shiller (1989)]. Others disagree. Merton (1987), for example, adapts the rational framework of the static CAPM to account for incomplete information. We undertake a related task in the dynamic world of the CCAPM and complement the theory with empirical evidence. We include in the traditional general equilibrium framework a class of agents who can implement only a particular trading strategy. We provide full equilibrium characterization and illustrate that considerable differences arise relative to the CCAPM (due to agent heterogeneity) and relative to Merton's (1987) model (due to intertemporal considerations). Our model can be extended to include several informationally constrained agents, each following one or several trading strategies. The intuition behind the basic model prevails, with the results modified to account for correlation patterns between various strategies.

Our empirical analysis is driven by the premise that a trading strategy shaped by real-world information costs should incorporate an investment in well-known, visible stocks, and an investment delegated to professional money managers, in particular via pension plans. We argue that we have at hand reasonable proxies for both components. Our argument must be put to the test, and this could not have been done without the model. Nevertheless, the model treats visibility as well as institutional investments as exogenous. Clearly, endogenizing these variables is a challenge of considerable interest for future work.

We test the joint hypothesis that both the model and the chosen proxies are well specified. The test examines the ability of our premise to explain the cross-sectional variation in average equity returns. The model performs quite well, but much is still left unexplained. These findings should encourage

further research that will lead to detailed characterization of portfolios and trading strategies of identifiable market participants. Knowing what agents trade may help us understand the nature and impact of the underlying frictions.

Caveats are in order. Empirical results are specific to the examined dataset and time period. Additional tests that focus on implications that are not tested here (such as the link between stock volatility and short-term interest rates) and examine other markets (domestic as well as international) may prove informative. Furthermore, at the cost of imposing more structure on the econometric specification, relaxing some of the assumptions underlying our unconditional formulation may lead to more powerful tests of the model.

Appendix A

A.1 Proofs

Proof of Proposition 1. We have cast the model with $N = 2$ for expositional purposes. The proof is given for an economy where $\mathbf{F}$ is generated by an N -dimensional Brownian motion $w = (w_1, \dots, w_N)^\top$ and where there are N risky assets. Hence, suppressing the dependence on (t, ω) , let $q = (q_1, \dots, q_{N-1}, 1)^\top$ and redefine $\mathcal{A}$ accordingly,

$$\mathcal{A} = \{(\alpha_2, \theta_2) \in \mathcal{R}^{N+1} : \theta_{2j} = q_j \theta_{2N}, \quad j = 1, \dots, N-1\}. \quad (\text{A1})$$

Treatment of the N -dimensional case lays the foundation for the empirical analysis that uses N assets. Agent 2 maximizes U_2 over consumption plans that can be financed by $(\alpha_2, \theta_2) \in \Theta$, subject to the budget constraint $\alpha_2(0) + \theta_2(0)^\top \bar{1} = b$, provided that $(\alpha_2, \theta_2) \in \mathcal{A}$. Cvitanic and Karatzas (1992) show that one can solve this problem via an appropriately defined minimization problem — the dual problem associated with the constraint $\mathcal{A}$. Let $\nu = (\nu_1, \dots, \nu_N)$ and $\nu_{-N} = (\nu_1, \dots, \nu_{N-1})$. For $(\nu_0, \nu) \in \mathcal{R}^{N+1}$ the support function of $-\mathcal{A}$ is

$$\varphi(\nu_0, \nu) \equiv \sup_{(\alpha_2, \theta_2) \in \mathcal{A}} -(\alpha_2 \nu_0 + \theta_2^\top \nu) = \sup_{(\alpha_2, \theta_{2N}) \in \mathcal{R}^2} -\left(\alpha_2 \nu_0 + \theta_{2N} \left(\sum_{j=1}^{N-1} q_j \nu_j + \nu_N\right)\right).$$

The effective domain of $\varphi(\nu_0, \nu)$ is

$$\tilde{\mathcal{A}} \equiv \{(\nu_0, \nu) \in \mathcal{R}^{N+1} : \varphi(\nu_0, \nu) < \infty\} = \left\{(\nu_0, \nu) \in \mathcal{R}^{N+1} : \nu_0 = 0, \nu_N = -\sum_{j=1}^{N-1} q_j \nu_j\right\},$$

where absence of constraints on α_2 requires $\nu_0 = 0$ (hence in the fictitious unconstrained economy only the drift must be modified to equal $\mu + \nu$), and the structure of $\mathcal{A}$ imposes the above restriction on ν . Also, $\varphi = 0$ on $\tilde{\mathcal{A}}$. We obtain ν by solving the dual problem. Let $\mathcal{N}(t, \omega) = \{\nu \in \mathcal{R}^N : \nu_N = -\sum_{j=1}^{N-1} q_j(t, \omega) \nu_j\}$. For an agent with logarithmic utility, the dual problem

$$\min_{(\psi_2, \nu) \in \mathcal{R}_+ \times \mathcal{N}} \mathbb{E} \left(- \int_0^T e^{-\rho t} (1 + \rho t + \log(\psi_2 \pi_\nu(t))) dt + \psi_2 b \right),$$

where

$$\begin{aligned} \pi_\nu(t) = & B(t)^{-1} \exp \left(- \int_0^t (\sigma(s)^{-1} (\mu(s) + \nu(s) - r(s) \bar{1}))^\top dw(s) \right. \\ & \left. - \frac{1}{2} \int_0^t \|\sigma(s)^{-1} (\mu(s) + \nu(s) - r(s) \bar{1})\|^2 ds \right), \end{aligned}$$

yields $\psi_2 = (1 - e^{-\rho T})/\rho b$, and further reduces to a pointwise minimization of a simple quadratic form:

$$\min_{\nu \in \mathcal{N}(t, \omega)} \|\sigma(t, \omega)^{-1}(\mu(t, \omega) + \nu - r(t, \omega)\bar{1})\|^2. \quad (\text{A2})$$

Each ν in $\mathcal{N}$ can be represented as $\nu(t) = M(t)\nu_{-N}(t)$, where $M(t) = [I_{N-1} - q_{-N}(t)]^\top$ is an $N \times (N-1)$ matrix, I_{N-1} is an identity matrix of rank $N-1$, and $q_{-N}(t)$ contains the first $N-1$ elements of $q(t)$. Substituting into Equation (A2) we obtain a convex program in ν_{-N} . It follows that the unique global minimizer is given by (suppressing the dependence on ω)

$$\nu(t) = -M(t)((\sigma(t)^{-1}M(t))^\top \sigma(t)^{-1}M(t))^{-1}(\sigma(t)^{-1}M(t))^\top \kappa(t), \quad (\text{A3})$$

where $\kappa(t)$ is as in Equation (11). Once we obtain the solution to the dual problem, Equations (9) and (10) follow, for example, using Proposition 1 in Cuoco (1997) with

$$\kappa_2(t) = \sigma(t)^{-1}(\mu(t) + \nu(t) - r(t)\bar{1}). \quad (\text{A4})$$

To obtain $\kappa_2(t)$ as in Equation (11), first substitute Equation (A3) into Equation (A4) to get

$$\begin{aligned} \kappa_2(t) &= (I - \sigma(t)^{-1}M(t)((\sigma(t)^{-1}M(t))^\top \\ &\quad \times \sigma(t)^{-1}M(t))^{-1}(\sigma(t)^{-1}M(t))^\top \kappa(t). \end{aligned} \quad (\text{A5})$$

Also note that the wealth evolution equation of agent 2 under $\mathcal{A}$ in Equation (A1) can be rewritten, analogous to Equation (8), using the bond and a fund F with $\sigma_F(t)$ proportional to $q(t)^\top \sigma(t)$. Next, let $X(t) = \text{Span}(\text{Col}[\sigma(t)^{-1}M(t)])$, $Y(t) = \text{Span}(\sigma_F(t))$ be two vector subspaces of $\mathbb{R}^N$, for a given t . Assuming that $\text{rank}(\sigma(t)) = N$ everywhere, it is easy to verify that $\dim(X(t)) = N-1$, and $\dim(Y(t)) = 1$. Since we also have that $\sigma_F(t)\sigma(t)^{-1}M(t) = 0$, it is straightforward to show that $X^\perp = Y$, where $X^\perp$ is the subspace of $\mathbb{R}^N$ orthogonal to X . The projection matrix on $X^\perp$ is given by the term in the large parentheses in Equation (A5). It must then equal the projection matrix on Y , given by $\Sigma_F(t) = \sigma_F(t)^\top (\sigma_F(t)\sigma_F(t)^\top)^{-1} \sigma_F(t)$. Hence, $\kappa_2(t)$ is as stated in Equation (11). ■

Lemma 1. *In equilibrium, the state-price density processes of the two agents are*

$$\pi_1(t) = e^{-\rho t} u_c(\delta(t), \lambda(t)) / u_c(\delta(0), \lambda(0)), \quad \pi_2(t) = \pi_1(t) \lambda(0) / \lambda(t), \quad (\text{A6})$$

where $\lambda(t) = u'_1(c_1^*(t))c_2^*(t)$, and its initial value satisfies Equation (15). The equilibrium consumption allocations and the spot-riskless interest rate (where $\mathcal{D}(\cdot)$ is the drift operator) are

$$c_1^*(t) = f_1(u_c(\delta(t), \lambda(t))), \quad c_2^*(t) = \lambda(t) / u_c(\delta(t), \lambda(t)), \quad (\text{A7})$$

$$r(t) = -\mathcal{D}(e^{-\rho t} u_c(\delta(t), \lambda(t))) / e^{-\rho t} u_c(\delta(t), \lambda(t)). \quad (\text{A8})$$

Proof of Lemma 1. First-order conditions from the portfolio optimization of agent 1 and agent 2 (see Section 2.2), combined with clearing of the consumption good market, yield: $c_1^*(t) + c_2^*(t) = f_1(\psi_1 \pi_1(t) e^{\rho t}) + (\psi_2 \pi_2(t) e^{\rho t})^{-1} = \delta(t)$. Then, from Karatzas, Lehoczky, and Shreve (1990), we know that $\psi_1 \pi_1(t) e^{\rho t} = u_c(\delta(t), \lambda(t))$, where $\lambda(t) = \frac{\psi_1 \pi_1(t)}{\psi_2 \pi_2(t)} = \frac{u'_1(c_1^*(t))}{u'_2(c_2^*(t))}$. Using $\pi_1(0) = \pi_2(0) = 1$ establishes Equation (A6). Then Equation (A7) is a restatement of agents' first-order conditions using Equation (A6), and Equation (15) follows, restating $b = W_2(0)$,

with $c_2^*(0) = \lambda(0)/u_c(\delta(0), \lambda(0))$. Absence of arbitrage and dynamic market completeness imply that agent 1 faces the unique state-price density process

$$\pi_1(t) = B(t)^{-1} \exp\left(-\int_0^t \kappa(s)^\top dw(s) - \frac{1}{2} \int_0^t \|\kappa(s)\|^2 ds\right). \quad (\text{A9})$$

We apply Itô's lemma to $\pi_1(t)$ in Equation (A9) and in Equation (A6). Equating the drift terms yields $r(t)$ in Equation (A8). Equating the diffusion terms yields

$$u_{cc}(\delta(t), \lambda(t))\sigma_\delta(t) + u_{c\lambda}(\delta(t), \lambda(t))\sigma_\lambda(t) + u_c(\delta(t), \lambda(t))\kappa(t) = 0. \quad (\text{A10})$$

■

Remark 1 (stock prices). To get the “present value” expression for stock prices, note that

$$\begin{aligned} & d(e^{-\rho t} u_c(\delta(t), \lambda(t)) S_j(t)) + e^{-\rho t} u_c(\delta(t), \lambda(t)) \delta_j(t) dt \\ &= e^{-\rho t} [u_c(\delta(t), \lambda(t)) (S_j(t) \mu_j(t) - \delta_j(t)) - r(t) u_c(\delta(t), \lambda(t)) S_j(t) \\ &\quad + S_j(t) \sigma_j(t) (u_{cc}(\delta(t), \lambda(t)) \sigma_\delta(t) + u_{c\lambda}(\delta(t), \lambda(t)) \sigma_\lambda(t)) \\ &\quad + u_c(\delta(t), \lambda(t)) \delta_j(t)] dt \\ &\quad + e^{-\rho t} [u_c(\delta(t), \lambda(t)) S_j(t) \sigma_j(t) + S_j(t) (u_{cc}(\delta(t), \lambda(t)) \sigma_\delta(t) \\ &\quad + u_{c\lambda}(\delta(t), \lambda(t)) \sigma_\lambda(t))^\top] dw(t) \\ &= e^{-\rho t} u_c(\delta(t), \lambda(t)) S_j(t) (\sigma_j(t) - \kappa(t)^\top) dw(t), \end{aligned}$$

where the first equality follows from Itô's lemma and Equation (A8), and the second from Equation (A10) and $\sigma_j(t)\kappa(t) = \mu_j(t) - r(t)$. When $\int_0^t e^{-\rho s} u_c(\delta(s), \lambda(s)) S_j(s) (\sigma_j(s) - \kappa(s)^\top) dw(s)$ is in fact a martingale, so is $e^{-\rho t} u_c(\delta(t), \lambda(t)) S_j(t) + \int_0^t e^{-\rho s} u_c(\delta(s), \lambda(s)) \delta_j(s) ds$. Imposing $S_j(T) = 0$, yields

$$S_j(t) = E\left(\int_t^T e^{-\rho(s-t)} \frac{u_c(\delta(s), \lambda(s))}{u_c(\delta(t), \lambda(t))} \delta_j(s) ds \middle| \mathcal{F}_t\right), \quad j = 1, \dots, N.$$

Proof of Theorem 1. From consumption good market clearing and Equation (A7)

$$f_1(u_c(\delta(t), \lambda(t))) + \lambda(t)/u_c(\delta(t), \lambda(t)) = \delta(t). \quad (\text{A11})$$

Differentiate Equation (A11) with respect to δ , for a given (ω, t) , and use $f_1'(x) = 1/u_1''(f_1(x))$ to get

$$u_1''(f_1(u_c(\delta(t), \lambda(t)))) = \frac{u_c(\delta(t), \lambda(t))^2 u_{cc}(\delta(t), \lambda(t))}{u_c(\delta(t), \lambda(t))^2 + \lambda(t) u_{cc}(\delta(t), \lambda(t))}. \quad (\text{A12})$$

Differentiate Equation (A11) with respect to λ , and use Equation (A12) to get

$$u_{c\lambda}(\delta(t), \lambda(t)) = A(t). \quad (\text{A13})$$

Rearranging Equation (A12), using Equation (A7), yields

$$A(t)^{-1} = A_1(t)^{-1} + A_2(t)^{-1}, \quad (\text{A14})$$

and differentiating Equation (A12) with respect to δ , and rearranging using Equation (A7), yields

$$P(t) = P_1(t)(A(t)/A_1(t))^2 + P_2(t)(A(t)/A_2(t))^2, \quad (\text{A15})$$

where

$$A_2(t) = u_c(\delta(t), \lambda(t))/\lambda(t), \quad (\text{A16})$$

and $P_2(t) = 2A_2(t)$. From Equation (A6), the stochastic weighting process is $\lambda(t) = \lambda(0)\pi_1(t)/\pi_2(t)$. Applying Itô's lemma to $\lambda(t)$ with $\pi_1(t)$ as in Equation (A9) and $\pi_2(t)$ as in Equation (10) yields

$$\mu_\lambda(t) = \lambda(t)(\kappa_2(t) - \kappa(t))^\top \kappa_2(t) \quad (\text{A17})$$

$$\sigma_\lambda(t) = \lambda(t)(\kappa_2(t) - \kappa(t)). \quad (\text{A18})$$

Substitute Equations (A4) and (A18) into Equation (A10), and use Equations (A13) and (A16) to get a restriction on ν :

$$\nu(t) = A_2(t)\sigma(t)\sigma_\delta(t) - (A_2(t)/A(t))(\mu(t) - r(t)\bar{1}). \quad (\text{A19})$$

Equating Equation (A19) with the solution of the dual problem in Equation (12), and using Equation (A14), yields

$$[A_1(t)^{-1}I + A_2(t)^{-1}\Sigma_F(t)]\sigma(t)^{-1}(\mu(t) - r(t)\bar{1}) = \sigma_\delta(t). \quad (\text{A20})$$

Note that $[A(t)I + (A_1(t) - A(t))\Pi_F(t)] [A_1(t)^{-1}I + A_2(t)^{-1}\Sigma_F(t)] = A(t)A_1(t)^{-1}I + (1 - A(t)A_1(t)^{-1})(I - \Sigma_F(t)) + A(t)A_2(t)^{-1}\Sigma_F(t) = I$, where the first equality uses $\Pi_F(t)\Sigma_F(t) = 0$ and the second uses Equation (A14). Premultiply both sides of Equation (A20) by $[A(t)I + (A_1(t) - A(t))\Pi_F(t)]\sigma(t)$ to obtain Equation (17). Given Equation (17), Equation (19) follows immediately. Substituting Equation (19) into Equations (A17) and (A18), we get $\mu_\lambda(t) = 0$, and $\sigma_\lambda(t) = -\lambda(t)A_1(t)\Pi_F(t)\sigma_\delta(t)$. This establishes Equation (14). Note that $\lambda(0)$, if it exists, is unique and positive because the right-hand side of Equation (15) is a strictly increasing positive function of $\lambda(0)$ bounded by $\delta(0)(1 - e^{-\rho T})/\rho$. For $\lambda(0)$ to exist, we must have

$$b \leq \delta(0)(1 - e^{-\rho T})/\rho. \quad (\text{A21})$$

Itô's lemma, using Equations (A8), (A10), and (19), implies

$$\begin{aligned} d(u_c(\delta(t), \lambda(t))) &= u_c(\delta(t), \lambda(t))(\rho - r(t)) dt \\ &\quad - u_c(\delta(t), \lambda(t))\sigma_\delta(t)^\top [A(t)I + (A_1(t) - A(t))\Pi_F(t)] dw(t). \end{aligned} \quad (\text{A22})$$

Using Equations (A14), (A22), and the identities $f_1'(c_1^*(t)) = 1/u_1''(f_1(c_1^*(t))) = -u_c(\delta(t), \lambda(t))^{-1} A_1(t)^{-1}$, $f_1''(c_1^*(t)) = -u_1'''(f_1(c_1^*(t)))u_1''(f_1(c_1^*(t)))^{-3} = P_1(t)u_c(\delta(t), \lambda(t))^{-2}A_1(t)^{-2}$, and applying Itô's lemma to c_1^* in Equation (A7) yields

$$\begin{aligned} d(c_1^*(t)) &= \left(\frac{r(t) - \rho}{A_1(t)} + \frac{1}{2}P_1(t) \frac{A(t)^2}{A_1(t)^2} \|\sigma_\delta(t)\|^2 \right. \\ &\quad \left. + \frac{1}{2}P_1(t) \left(1 - \frac{A(t)^2}{A_1(t)^2} \right) \|\Pi_F(t)\sigma_\delta(t)\|^2 \right) dt \\ &\quad + \left(\frac{A(t)}{A_1(t)} \sigma_\delta(t) + \left(1 - \frac{A(t)}{A_1(t)} \right) \Pi_F(t)\sigma_\delta(t) \right)^\top dw(t). \end{aligned}$$

Similarly, applying Itô's lemma to c_2^* in Equation (A7), using Equations (A8), (A14), (A22), and (14) yields

$$\begin{aligned} d(c_2^*(t)) = & \left(\frac{r(t)-\rho}{A_2(t)} + \frac{1}{2} P_2(t) \frac{A(t)^2}{A_2(t)^2} \|\sigma_\delta(t)\|^2 \right. \\ & \left. - \frac{1}{2} P_2(t) \frac{A(t)^2}{A_2(t)^2} \|\Pi_F(t)\sigma_\delta(t)\|^2 \right) dt \\ & + \left(\frac{A(t)}{A_2(t)} \sigma_\delta(t) - \frac{A(t)}{A_2(t)} \Pi_F(t) \sigma_\delta(t) \right)^\top dw(t). \end{aligned}$$

Use the drift terms and the market clearing condition, $\mathcal{D}(c_1^*(t)) + \mathcal{D}(c_2^*(t)) = \mathcal{D}(\delta(t))$, to get

$$\begin{aligned} & \frac{r(t)-\rho}{A_1(t)} + \frac{1}{2} P_1(t) \frac{A(t)^2}{A_1(t)^2} \|\sigma_\delta(t)\|^2 + \frac{1}{2} P_1(t) \left(1 - \frac{A(t)^2}{A_1(t)^2} \right) \|\Pi_F(t)\sigma_\delta(t)\|^2 \\ & + \frac{r(t)-\rho}{A_2(t)} + \frac{1}{2} P_2(t) \frac{A(t)^2}{A_2(t)^2} \|\sigma_\delta(t)\|^2 - \frac{1}{2} P_2(t) \frac{A(t)^2}{A_2(t)^2} \|\Pi_F(t)\sigma_\delta(t)\|^2 = \mu_\delta(t). \end{aligned}$$

Rearranging, using Equations (A14) and (A15), establishes Equation (16). Substitute $r(t)$ back into the drifts and rearrange to complete the proof. (Note that given S_j in Remark 1, θ_2^* in Equation (13), and κ_2 in Equation (19), the portfolio holdings of agent 1 are set so as to clear the securities markets.) ■

Lemma 2. When the portfolio choice of agent 2 is perfectly correlated with the exogenous process $V(t)$ in (5), then $\Pi_F(t) = I - \sigma_V(t)^\top (\sigma_V(t) \sigma_V(t)^\top)^{-1} \sigma_V(t)$, where $\sigma_V(t) = (v_1(t), v_2(t))$ is the diffusion vector of $V(t)$.

Proof of Lemma 2. When $N = 2$, the trading strategy of agent 2 is characterized by $q_1(t)$ as in Equation (6). Substituting in Equation (8) yields

$$\sigma_F(t) = \frac{\sigma_{11}(t)\sigma_{22}(t) - \sigma_{12}(t)\sigma_{21}(t)}{v_1(t)(\sigma_{22}(t) - \sigma_{12}(t)) + v_2(t)(\sigma_{11}(t) - \sigma_{21}(t))} \sigma_V(t).$$

Hence $\Pi_F(t) = I - \sigma_F(t)^\top (\sigma_F(t) \sigma_F(t)^\top)^{-1} \sigma_F(t) = I - \sigma_V(t)^\top (\sigma_V(t) \sigma_V(t)^\top)^{-1} \sigma_V(t)$. ■

Proof of Corollary 1. For simplicity, (a) and (b) below are proved for $N = 2$. However, Theorem 1 and hence (c) and (d) below hold for an arbitrary N .

- (a) Given Lemma 2, the unconstrained results follow from Theorem 1 using the fact that $\Pi_F(t)\sigma_\delta(t) = (I - \sigma_\delta(t)(\sigma_\delta(t)^\top \sigma_\delta(t))^{-1} \sigma_\delta(t)^\top) \sigma_\delta(t) = 0$. In particular, $d\lambda(t) = 0$, so that $\forall t$, $\lambda(t) = \lambda(0)$, where Equation (A21) guarantees the existence of a unique, strictly positive solution to Equation (15). To get the portfolio choice in $\mathcal{E}_U$, substitute $\kappa_2(t) = \kappa(t) = A(t)\sigma_\delta(t) = \sigma(t)^{-1}(\mu(t) - r(t)\mathbf{1})$ into Equation (13): $\theta_2^*(t) = (\sigma\sigma^\top)^{-1}(\mu + \nu - r\mathbf{1})W_2 = (\sigma^\top)^{-1}\kappa_2 W_2 = (\sigma(t)\sigma(t)^\top)^{-1}(\mu(t) - r(t)\mathbf{1})W_2(t)$.
- (b) The trading strategy of agent 2 is characterized by $q_1(t)$ as in Equation (7). Then it is easy to verify that $\Sigma_F(t)\sigma_\delta(t) = 0$, and therefore $\Pi_F(t)\sigma_\delta(t) = \sigma_\delta(t)$, which we substitute in Equations (14)–(19). To get the portfolio choice, note that $\kappa_2(t) = 0$, and using Equation (13): $\theta_2^*(t) = (\sigma(t)^\top)^{-1}\kappa_2(t)W_2(t) = 0$.
- (c) From Equation (17), using the definition of Π_F : $A(t)\sigma_j(t)\sigma_\delta(t) + (A_1(t) - A(t))\sigma_j(t)\Pi_F(t)\sigma_\delta(t) = A_1(t)\sigma_j(t)\sigma_\delta(t) - (A_1(t) - A(t))\sigma_j(t)\Sigma_F(t)\sigma_\delta(t)$, where $\sigma_j(t)\Sigma_F(t)\sigma_\delta(t) = (\sigma_j(t)\sigma_F(t)^\top)(\sigma_F(t)\sigma_\delta(t))/\|\sigma_F(t)\|^2 = \rho_F\delta(t)\frac{\|\sigma_\delta(t)\|}{\|\sigma_F(t)\|}\sigma_j(t)\sigma_F(t)^\top$, and using Equation (A14): $A_1(t) - A(t) = A_1(t) - \frac{A_1(t)A_2(t)}{A_1(t)+A_2(t)} = \frac{A_1(t)A_2(t)}{A_1(t)+A_2(t)} \frac{A_1(t)}{A_2(t)} = A(t) \frac{A_1(t)}{A_2(t)}$.

- (d) To get $\mu_F(t) - r(t)$, use $\sigma_F(t)\Pi_F(t) = 0$ in Equation (17). The weights of δ^* are given by $p(t) = \frac{(\sigma(t)^{-1})^\top \sigma_\delta(t)}{\sigma_\delta(t)^\top \sigma(t)^{-1} \mathbf{1}}$, so that $\sigma_{\delta^*}(t) = p(t)^\top \sigma(t) = \frac{\sigma_\delta(t)^\top}{\sigma_\delta(t)^\top \sigma(t)^{-1} \mathbf{1}}$ (and δ^* is perfectly correlated with δ in the sense that $\rho_{\delta^* \delta}^2 = 1$). Then, from Equation (17),

$$\begin{aligned} \mu_{\delta^*}(t) - r(t) &= A(t)\sigma_{\delta^*}(t)\sigma_\delta(t) + \frac{(A_1(t) - A(t))\sigma_\delta(t)^\top \Pi_F(t)\sigma_\delta(t)}{\sigma_\delta(t)^\top \sigma(t)^{-1} \mathbf{1}} \\ &= \frac{A(t)\|\sigma_\delta(t)\|^2}{\sigma_\delta(t)^\top \sigma(t)^{-1} \mathbf{1}} \left(1 + \frac{A_1(t)}{A_2(t)} \frac{\|\Pi_F(t)\sigma_\delta(t)\|^2}{\|\sigma_\delta(t)\|^2} \right), \end{aligned}$$

where

$$\begin{aligned} \|\Pi_F(t)\sigma_\delta(t)\|^2 &= \sigma_\delta(t)^\top \Pi_F(t)\sigma_\delta(t) = \|\sigma_\delta(t)\|^2 - \frac{(\sigma_F(t)\sigma_\delta(t))^2}{\|\sigma_F(t)\|^2} \\ &= (1 - \rho_{F\delta}(t)^2)\|\sigma_\delta(t)\|^2. \end{aligned} \quad (\text{A23})$$

Next, write Equation (20) once for δ^* and once for F , and solve for $a_1(t)$ and $a_2(t)$ in terms of the covariances and $\mu_{\delta^*}(t) - r(t)$ and $\mu_F(t) - r(t)$. Substitute the resulting expressions back into Equation (20), and collect terms to get $b_{j\delta^*}(t) = \frac{\|\sigma_j(t)\|(\rho_{j\delta^*}(t) - \rho_{jF}(t)\rho_{F\delta^*}(t))}{\|\sigma_{\delta^*}(t)\|(1 - \rho_{F\delta^*}(t)^2)}$, $b_{jF}(t) = \frac{\|\sigma_j(t)\|(\rho_{jF}(t) - \rho_{j\delta^*}(t)\rho_{F\delta^*}(t))}{\|\sigma_F(t)\|(1 - \rho_{F\delta^*}(t)^2)}$, which are the ordinary coefficients from a multiple regression (in the population-conditional distribution) of the rate of return of asset j on the rate of returns of δ^* and the IRH index. ■

Remark 2 (existence of equilibrium). Proving the existence of equilibrium in the benchmark case amounts to an application of a fixed-point argument to $\lambda(0)$, and equilibrium in $\mathcal{E}_U$ exists when Equation (A21) holds. To prove existence in a restricted-participation economy, we need to prove the existence of a solution to a *univariate* stochastic differential equation (SDE) for $\lambda(t)$ [see Basak and Cuoco (1998) for a proof with logarithmic preferences and geometric Brownian endowment]. This is still the case with versions of the IRH that are covered by Lemma 2 (see the example in Appendix A.2). In the general case of Theorem 1, we are faced with a multivariate system which includes $\lambda(t)$ and a transformation of $\sigma(t)$ in the SDE [Equation (14)], plus the stock equations from Remark 1, where, under appropriate regularity conditions, $\sigma_j(t)$ is the function implied by the martingale representation theorem applied to stock j . An existence proof in such a system may potentially be based on the theory of forward-backward SDEs [see, e.g., Antonelli (1993)], but the currently available tools in that relatively new field are not applicable to our setting. Restrictions on primitives (including q) may need to be placed (in particular, so that Equations (1) and (3) hold).

Proof of Proposition 2. (a) is obtained by setting $q_1 = 0$ in Equation (8), so that in Theorem 1, $\Pi_F(t) = \Pi_2(t) \equiv I - \sigma_2(t)^\top (\sigma_2(t)\sigma_2(t)^\top)^{-1} \sigma_2(t)$. The interest rate is restated using Equation (A23). The second and third terms in Equation (21) are obtained by decomposing $(A_1(t) - A(t))\sigma_1(t)\Pi_2(t)\sigma_\delta(t)$ into $\frac{A(t)\delta(t)c_2^*(t)}{R_2(t)c_1^*(t)}((\sigma_1(t) - \frac{\sigma_1(t)\sigma_2(t)^\top}{\|\sigma_2(t)\|^2} \sigma_2(t)) \frac{\sigma_W(t)}{W(t)} + \sigma_1(t)\Pi_2(t) (\frac{\sigma_\delta(t)}{\delta(t)} - \frac{\sigma_W(t)}{W(t)}))$, where $(\sigma_1(t) - \frac{\sigma_1(t)\sigma_2(t)^\top}{\|\sigma_2(t)\|^2} \sigma_2(t)) \frac{\sigma_W(t)}{W(t)} = (\sigma_1 - \frac{\sigma_1\sigma_2^\top}{\|\sigma_2\|^2} \sigma_2) \frac{(S_1\sigma_1 + S_2\sigma_2)^\top}{W} = (\|\sigma_1(t)\|^2 - \frac{(\sigma_1(t)\sigma_2(t)^\top)^2}{\|\sigma_2(t)\|^2}) \frac{S_1(t)}{W(t)}$. Then Equation (21) follows immediately.

(b) When $u_1(\cdot) = u_2(\cdot) = \log(\cdot)$: $u(\delta(t), \lambda(t)) = \log \frac{\delta(t)}{1+\lambda(t)} + \lambda(t) \log \frac{\lambda(t)\delta(t)}{1+\lambda(t)}$,

$$A(t) = \delta(t)^{-1}, \quad P(t) = 2\delta(t)^{-1}, \quad W(t) = \frac{1 - e^{-\rho(T-t)}}{\rho} \delta(t),$$

$$A_1(t) = (1 + \lambda(t))/\delta(t), \quad P_1(t) = 2(1 + \lambda(t))/\delta(t), \quad c_1^* = \delta(t)/(1 + \lambda(t)),$$

$$A_2(t) = (1 + \lambda(t))/(\lambda(t)\delta(t)), \quad P_2(t) = 2(1 + \lambda(t))/(\lambda(t)\delta(t)), \quad c_2^* = \lambda(t)\delta(t)/(1 + \lambda(t)).$$

Substituting in Equations (16)–(19) yields the results for $\kappa_i(t)$, $r(t)$, $\mu_i(t) - r(t)$, and $c_i^*(t)$. The notion of “volatility” of state prices and of consumption growth rate is understood here as the

quadratic variation of $d\pi_i(t)/\pi_i(t)$, and $dc_i^*(t)/c_i^*(t)$. The diffusion term of the former process is $-\kappa_i(t)$. Then, in $\mathcal{E}_U$: $\|\kappa_2(t)\|^2 = \|\kappa_1(t)\|^2 = \|\sigma_\delta\|^2$, in $\mathcal{E}$: $\|\kappa_2(t)\|^2 = \|\Sigma_2(t)\sigma_\delta\|^2 \leq \|\sigma_\delta\|^2 \leq \|\sigma_\delta\|^2 + \lambda(t)(2 + \lambda(t))\|\Pi_2(t)\sigma_\delta\|^2 = \|\kappa_1(t)\|^2$. This confirms the statement in (i), and because in $\mathcal{E}_U$ $\mu_{c_i^*}(t)/c_i^*(t) = \mu_\delta$ and $\sigma_{c_i^*}(t)/c_i^*(t) = \sigma_\delta$, analogous comparisons confirm the statement in (ii). The proof of (iii) is identical to the proof of Corollary 4 in Basak and Cuoco (1998) and is omitted. In $\mathcal{E}_U$: $r(t) = \rho + \mu_\delta - \|\sigma_\delta\|^2$, and the statement in (iv) follows from (a). Finally, (v) follows using Equation (21). For completeness, we note that $\delta_1(t)$ and $\delta_2(t)$ in this example will not be geometric Brownian motions. But we can set $\delta_1(t) = x(t)\delta(t)$ and $\delta_2(t) = (1 - x(t))\delta(t)$, where $x(t)$ is any process satisfying $x(t) \in (0, 1)$. ■

A.2 An example: characterization and existence of equilibrium

Consider an economy with $u_i(\cdot) = \log(\cdot)$ and geometric Brownian endowment, and let agent 2 invest in stocks only via a fund manager. Agent 2 fares better in $\mathcal{E}_U$ and wishes to invest in a fund that perfectly tracks $\delta(t)$. Suppose that the fund manager he invests with, actually tracks $\delta(t)$ only imperfectly. The process that she tracks is given by

$$dV(t) = \delta(t)\mu_\delta dt + \delta(t)v_1(t)dw_1(t) + \delta(t)v_2(t)dw_2(t), \quad (A24)$$

where $V(0) = \delta(0)$, $dv_k(t) = a[\sigma_{\delta,k} - v_k(t)]dt + b dw_k(t)$, with $v_k(0) = \sigma_{\delta,k}$, $k = 1, 2$ [the Ornstein–Uhlenbeck process $v(t) = (v_1(t), v_2(t))^\top$ is an unbiased estimator of the true constant vector σ_δ]. The constants a and b indicate how well the fund fits the objective of agent 2. Finally, assume that $\sigma_{\delta,2} = 0$. The comparison between $\mathcal{E}$ and $\mathcal{E}_U$ is summarized below.

Suppose the portfolio of agent 2 maintains perfect instantaneous correlation with $V(t)$ in Equation (A24). Then there exists a unique equilibrium, where all the results of Proposition 2(b) in (i)–(iv) hold when replacing $\rho_{2\delta}$ with $\rho_{V\delta}$, and $\Pi_2(t)$ with $I - v(t)(v(t)v(t)^\top)^{-1}v(t)^\top$. The risk premia are $\mu_j(t) - r(t) = \sigma_{j1}(t)\sigma_{\delta,1} + \frac{W_2(t)}{W_1(t)}\sigma_j(t)\sigma_\eta(t)$, where $\sigma_\eta(t) = \sigma_{\delta,1}v_2(t)\|v(t)\|^{-2}(v_2(t) - v_1(t))^\top$, and $j = 1, 2$. The dynamics of the interest rate are given by $dr(t) = \mu_r(t)dt + \sigma_r(t)^\top dw(t)$, where its stochastic drift and volatility are

$$\begin{aligned} \mu_r(t) &= \|v(t)\|^{-2} \left(2v_1(t)\sigma_{\delta,1} \left(a - \frac{W(t)}{W_1(t)}b \right) + \left(1 - \frac{v_1(t)^2}{v_2(t)^2} \right) b^2 \right) \frac{W_2(t)}{W_1(t)} \|\sigma_\eta(t)\|^2, \\ \sigma_r(t) &= \|v(t)\|^{-2} \left(\frac{W(t)}{W_1(t)} v_2(t)^2 \sigma_{\delta,1}^2 + v_1(t)b \right) \frac{W_2(t)}{W_1(t)} \sigma_{\delta,1} \sigma_\eta(t), \\ v_1(t) &= \sigma_{\delta,1} + b \int_0^t e^{-a(t-s)} dw_1(s), \quad v_2(t) = b \int_0^t e^{-a(t-s)} dw_2(s). \end{aligned}$$

Proof. The characterization results follow from Theorem 1 and Lemma 2, analogous to Proposition 2(b).²¹ The dynamics of the interest rate are then obtained by applying Itô's lemma to $r(t) = \rho + \mu_\delta - \|\sigma_\delta\|^2 - \lambda(t)\|(I - v(t)(v(t)v(t)^\top)^{-1}v(t)^\top)\sigma_\delta\|^2$. Note that this example belongs to a particular class of economies for which the projection matrix Π_F is expressed using exogenous quantities (i.e., the process $v(t)$). The dynamics of the weighting process $\lambda(t)$

²¹ The classic CCAPM does not price the diffusion coefficient $\sigma_{j2}(t)$ because the first state variable, $\delta(t)$, is independent of $w_2(t)$. However, the second state variable $\lambda(t)$ (whose diffusion vector is proportional to $\sigma_\eta(t)$) has, in general, a nondegenerate covariation with $w_2(t)$. Consequently, the entire diffusion vector $\sigma_j(t)$ is priced by the modified CCAPM. The constant a represents the speed of convergence toward the desirable target. It may indicate the fund manager's effort or ability to perform her task, and b may represent the noisy environment that interferes with her efforts. When she is very apt ($a \rightarrow \infty$) or lucky not to face noise ($b \rightarrow 0$), then $v_k(t) \rightarrow \sigma_{\delta,k}$ and $V(t) \rightarrow \delta(t)$. Hence perfect tracking ability (in the sense that $b = 0$ or $a = \infty$) leads to an increased interest rate, and to eliminating the impact of the second state variable on risk premia. Note that as in Proposition 2(b), $\lambda = W_2/W_1$, and it is a supermartingale ($\lambda(t) \geq E[\lambda(s)|\mathcal{F}_t]$, $\forall t \leq s \leq T$), suggesting that agent 1 is expected to accumulate more wealth relative to agent 2 in an economy with an imperfect tracking ability.

in Equation (14) are therefore specified completely by a univariate SDE. Existence of a solution to the SDE, and hence existence of an equilibrium, can be verified using standard results. First, restate the SDE as follows: $d\lambda(t) = -\lambda(t)(1 + \lambda(t))\sigma_\eta(t)dw(t) = \lambda(t)(1 + \lambda(t))dL(t)$, where $L(t)$ can be verified to be a local martingale. Since $\lambda(1 + \lambda)$ has a continuous (but not bounded) derivative, it is locally Lipschitz. It also satisfies the linear growth condition (it is bounded by $2(1 + \lambda^2)$). The search for equilibrium is reduced to finding a solution to the above SDE. By Theorem V.38 in Protter (1990), there exists a unique (strong) solution to the SDE up to an explosion time. Then it follows from Karatzas and Shreve (1988), Remark 5.19, that for this zero-drift univariate SDE, the linear growth condition is sufficient for the solution not to explode, for a given $\lambda(0)$ [which exists when Equation (A21) holds]. Stock prices are well defined, since by absence of arbitrage (guaranteed by nonnegative wealth) and using the value of $W(t)$: $0 < S_j(t) < S_1(t) + S_2(t) = W(t) = \frac{1 - e^{-\rho(T-t)}}{\rho} \delta(t)$.

Appendix B

B.1 Proof of theorem 2

Additional assumptions that are required for Theorem 2 are stated in Equations (B4)–(B6) (and all the used moments are assumed to exist). For $j = 1, 2, \dots, N$, $t = 1, 2, \dots, T - 1$ (where T now denotes the last date of the sample period, as opposed to the horizon of the economy), define

$$\gamma_{jg,t} = \text{cov}[r_{j,t+1}, g_{t+1} | \mathcal{F}_t], \quad \gamma_{jh,t} = \text{cov}[r_{j,t+1}, h_{t+1} | \mathcal{F}_t], \quad (\text{B1})$$

$$\varepsilon_{jg,t} = \gamma_{jg,t} - E[\gamma_{jg,t}], \quad \varepsilon_{jh,t} = \gamma_{jh,t} - E[\gamma_{jh,t}], \quad (\text{B2})$$

$$e_{j,t+1} = r_{j,t+1} - E[r_{j,t+1} | \mathcal{F}_t], \quad (\text{B3})$$

For some constants $(k_{1g}, k_{2g}, k_{1h}, k_{2h}) \in \mathcal{R}^4$, assume that²²

$$E[\varepsilon_{jg,t} a_{1t}] = k_{1g}, \quad E[\varepsilon_{jh,t} a_{2t}] = k_{2g}, \quad (\text{B4})$$

$$E[\varepsilon_{jg,t} a_{1t} g_{t+1}] = k_{1g}, \quad E[\varepsilon_{jh,t} a_{2t} g_{t+1}] = k_{2g}, \quad (\text{B5})$$

$$E[\varepsilon_{jg,t} a_{1t} h_{t+1}] = k_{1h}, \quad E[\varepsilon_{jh,t} a_{2t} h_{t+1}] = k_{2h}. \quad (\text{B6})$$

First, substitute Equations (B1) and (B2) in Equation (22) and take unconditional expectations to get, using Equation (B4),

$$E[r_{j,t+1}] = E[a_{0t}] + k_1 + k_2 + E[a_{1t}]E[\gamma_{jg,t}] + E[a_{2t}]E[\gamma_{jh,t}]. \quad (\text{B7})$$

Next, restate Equation (22) using Equations (B1)–(B3) as

$$r_{j,t+1} = a_{0t} + a_{1t}E[\gamma_{jg,t}] + a_{2t}E[\gamma_{jh,t}] + a_{1t}\varepsilon_{jg,t} + a_{2t}\varepsilon_{jh,t} + e_{j,t+1}. \quad (\text{B8})$$

Equation (B8) along with Equations (B4)–(B6) yields

²² In Equation (B4) we assume homoskedastic fluctuations of the covariances relative to a_{1t} and a_{2t} . In Equations (B5) and (B6) we assume that time-varying components of the conditional covariances, $\varepsilon_{jg,t}$ or $\varepsilon_{jh,t}$, incorporate homoskedastic information beyond that in a_{1t} or a_{2t} , respectively, about the predictable component of (g_{t+1}, h_{t+1}) . The homoskedasticity assumptions are for simplicity. The proof goes through under heteroskedasticity in Equations (B4)–(B6), specified by linear dependence on $E[\gamma_{jg,t}]$ and $E[\gamma_{jh,t}]$. The resulting expressions for a_0, a_1, a_2 will be modified accordingly.

$$\begin{aligned} \text{cov}(r_{j,t+1}, g_{t+1}) &= k_g + (\text{cov}(a_{1t}, g_{t+1}) + 1)E[\gamma_{jg,t}] \\ &\quad + \text{cov}(a_{2t}, g_{t+1})E[\gamma_{jh,t}] \end{aligned} \quad (\text{B9})$$

$$\begin{aligned} \text{cov}(r_{j,t+1}, h_{t+1}) &= k_h + \text{cov}(a_{1t}, h_{t+1})E[\gamma_{jg,t}] \\ &\quad + (\text{cov}(a_{2t}, h_{t+1}) + 1)E[\gamma_{jh,t}], \end{aligned} \quad (\text{B10})$$

where $k_g = k_{1g} + k_{2g} - (k_1 + k_2)E[g_{t+1}] + \text{cov}(a_{0t}, g_{t+1})$, and $k_h = k_{1h} + k_{2h} - (k_1 + k_2)E[h_{t+1}] + \text{cov}(a_{0t}, h_{t+1})$. It is clear from Equations (B9)–(B10) that because $\text{cov}(r_{j,t+1}, g_{t+1})$ and $\text{cov}(r_{j,t+1}, h_{t+1})$ vary with j , and because $\{\text{cov}(r_{j,t+1}, g_{t+1})\}_{j=1}^N$ and $\{\text{cov}(r_{j,t+1}, h_{t+1})\}_{j=1}^N$ are assumed to be linearly independent, the constant matrix

$$\begin{pmatrix} \text{cov}(a_{1t}, g_{t+1}) + 1 & \text{cov}(a_{2t}, g_{t+1}) \\ \text{cov}(a_{1t}, h_{t+1}) & \text{cov}(a_{2t}, h_{t+1}) + 1 \end{pmatrix}$$

must be nonsingular. Therefore we solve the linear system of Equations (B9)–(B10) for the two unknowns $(E[\gamma_{jg,t}], E[\gamma_{jh,t}])$ in terms of $\text{cov}(r_{j,t+1}, g_{t+1})$ and $\text{cov}(r_{j,t+1}, h_{t+1})$. Substituting the solution into Equation (B7) and rearranging establishes Equation (23), where

$$\begin{aligned} a_0 &= E[a_{0t}] + k_1 + k_2 + E[a_{1t}](\text{cov}(a_{2t}, g_{t+1})k_h - (\text{cov}(a_{2t}, h_{t+1}) + 1)k_g)/\Delta \\ &\quad + E[a_{2t}](\text{cov}(a_{1t}, h_{t+1})k_g - (\text{cov}(a_{1t}, g_{t+1}) + 1)k_h)/\Delta \\ a_1 &= (E[a_{1t}](\text{cov}(a_{2t}, h_{t+1}) + 1) - E[a_{2t}]\text{cov}(a_{1t}, h_{t+1}))\text{var}(g_{t+1})/\Delta \\ a_2 &= (E[a_{2t}](\text{cov}(a_{1t}, g_{t+1}) + 1) - E[a_{1t}]\text{cov}(a_{2t}, g_{t+1}))\text{var}(h_{t+1})/\Delta \\ \Delta &= (\text{cov}(a_{1t}, g_{t+1}) + 1)(\text{cov}(a_{2t}, h_{t+1}) + 1) - \text{cov}(a_{2t}, g_{t+1})\text{cov}(a_{1t}, h_{t+1}). \end{aligned}$$

Note that a_{1t} is positive, while the sign of a_1 is unrestricted. The proof of Equation (24) is similar and is therefore omitted. ■

B.2 The cross-sectional regressions

We follow closely Appendix B in JW and adapt it to GLS.²³ We assume that all time series are covariance stationary, returns on the N assets are (unconditionally) distributed i.i.d. over time, and all the limits below exist. Equations (23) and (24) are a special case of the following model:

$$E[r_{jt}] = \sum_{l=1}^L c_{0l}z_{jl} + \sum_{k=1}^K c_k\beta_{jk} \quad j = 1, \dots, N, \quad t = 1, \dots, T \quad (\text{B11})$$

where z_{jt} are observable characteristics of asset j (in our case, when $L = 1$, then $z_{j1} = 1$, when $L = 2$, we take z_{j2} to be the log of market equity), $\beta_{jk} = \text{cov}(r_{jt}, y_{kt})/\text{var}(y_{kt})$ with y_{kt} representing economic variables [e.g., in Equation (23) $K = 2$, $y_{1t} = g_t$, $y_{2t} = h_t$]. Rewrite Equation (B11) as

$$\mu = X\gamma, \quad (\text{B12})$$

where $\mu = E[r_t]$, $r_t = (r_{1t}, \dots, r_{Nt})^\top$, $X = [Z \ B]$, Z is a $N \times L$ matrix of characteristics, B is a $N \times K$ matrix of the univariate betas, and $\gamma = (c_{01}, \dots, c_{0L}, c^\top)^\top$, $c = (c_1, \dots, c_K)^\top$. Define $\varepsilon_t = r_t - \mu$, $\bar{\varepsilon} = \frac{1}{T} \sum_{t=1}^T \varepsilon_t$, $\bar{r} = \frac{1}{T} \sum_{t=1}^T r_t$. Then, $\bar{\varepsilon} = \bar{r} - \mu$, and restate Equation (B12) as $r_t = X\gamma + \varepsilon_t$. Therefore the estimated system is $\bar{r} = X\gamma + \bar{\varepsilon}$ where $\text{var}(r_t) = G$, a constant $N \times N$

²³ Shanken (1992) lays the foundations for the analysis, correcting the bias in the Black, Jensen, and Scholes (1972) and Fama and MacBeth (1973) procedures, but his focus is on CSRs with multivariate betas.

matrix, and hence $\text{var}(\bar{\varepsilon}) = \frac{1}{T}G$. To obtain a feasible GLS estimator of γ (and correct standard errors for the OLS and GLS estimators), we use two alternative estimators of G :

$$\hat{G}_1 = \frac{1}{T-1} \sum_{t=1}^T (r_t - \bar{r})(r_t - \bar{r})^\top, \quad (\text{B13})$$

$$\hat{G}_2 = \hat{w}_1 I + \hat{w}_2 \hat{G}_1, \quad (\text{B14})$$

where I is an $N \times N$ identity matrix and $(\hat{w}_1, \hat{w}_2)$ are weights of the optimal linear-shrinkage estimator constructed by Ledoit (1994). Under an asymptotic theory where N is allowed to grow with T (provided that N/T is bounded), neither $\hat{G}_1$ nor $\hat{G}_2$ is a consistent estimator of G . However, $\hat{G}_2$ has the lowest mean square error among the linear-shrinkage estimators of the form $w_1 I + w_2 \hat{G}_1$. Shrinkage estimators are desirable whenever matrix inversion is required, as is the case with GLS. The reason is that in Equation (B14) we reduce the proximity to singularity by keeping the eigenvalues of $\hat{G}_2$ away from zero while allowing N to be close to (or larger than) T . The weight $\hat{w}_1$ increases in N/T .

We use $\hat{G}_1$ with monthly data, and $\hat{G}_2$ with quarterly data and with monthly data over subperiods. In our quarterly and subperiod analyses ($N = 100$, $T = 110$) we rely on $\hat{G}_2$ for asymptotic inferences as an alternative to finite sample adjustments. We note, however, that finite sample performance of this estimator versus conventional adjustments (when $N < T$) is not yet established. Nevertheless, in our sample it provides more conservative standard errors.

We next address the error-in-variables problem of the two-step procedure. We keep N fixed while T can grow to infinity. In this case, $\hat{G}_2$ converges to $\hat{G}_1$, and both are (T) -consistent estimators of G : $\text{plim}_{T \rightarrow \infty} \hat{G}_1 = \text{plim}_{T \rightarrow \infty} \hat{G}_2 = G$. Using OLS estimates, $\hat{\beta}_{jk}$, for univariate betas, we obtain an estimate $\hat{B}$ of B . Let $\hat{X} = [Z \ \hat{B}]$. The feasible GLS estimator of γ is

$$\hat{\gamma}_m = \left(\hat{X}^\top \left[\frac{1}{T} \hat{G}_m \right]^{-1} \hat{X} \right)^{-1} \hat{X}^\top \left[\frac{1}{T} \hat{G}_m \right]^{-1} \bar{r}, \quad (\text{B15})$$

where $m = 1$ or $m = 2$, depending on the method used to estimate G . To derive the asymptotic distribution of $\hat{\gamma}_m$ in Equation (B15), restate $\bar{r}$ as

$$\bar{r} = \hat{X} \gamma + [\bar{\varepsilon} - (\hat{X} - X) \gamma] = \hat{X} \gamma + [\bar{\varepsilon} - (\hat{B} - B) c]. \quad (\text{B16})$$

Since $\text{plim}_{T \rightarrow \infty} \bar{r} = E[r_t]$ and $\text{plim}_{T \rightarrow \infty} \hat{B} = B$, it is clear from Equations (B16) and (B15) that $\hat{\gamma}_m$ is a consistent estimator of γ , for $m = 1, 2$.

Let $\hat{M}_m = (\hat{X}^\top \hat{G}_m^{-1} \hat{X})^{-1} \hat{X}^\top \hat{G}_m^{-1}$, and let $M = \text{plim} \hat{M}_m$. From (B15) and (B16) we get

$$\hat{\gamma}_m - \gamma = \hat{M}_m \bar{\varepsilon} - \hat{M}_m (\hat{B} - B) c.$$

Define $\tilde{\phi}$ and $\tilde{\Phi}$ as the limits in distribution of $\sqrt{T} \bar{\varepsilon}$ and $\sqrt{T} (\hat{B} - B)$, respectively. Then

$$\sqrt{T} (\hat{\gamma}_m - \gamma) \xrightarrow{d} M (\tilde{\phi} - \tilde{\Phi} c). \quad (\text{B17})$$

To be able to account explicitly for the two sources of sampling error identified in Equation (B17) we need to introduce further assumptions and notation.²⁴ Let

²⁴ Suppose we ignore the sampling error in $\hat{\gamma}_m$ from replacing true betas by their estimates (i.e., ignore $\tilde{\Phi}$). From Equation (B17), the asymptotic variance of $\sqrt{T} \hat{\gamma}_m$ becomes $M \text{var}(\tilde{\phi}) M^\top$. We estimate $\text{var}(\tilde{\phi})$ by $\hat{G}_m$. Then for $m = 1, 2$, the asymptotic variance reduces to the familiar $(X^\top G^{-1} X)^{-1}$.

$$e_{jkt} = (r_{jt} - E[r_{jt}]) - \beta_{jk}(y_{kt} - E[y_{kt}]), \quad e_{jk} = (e_{jk1}, \dots, e_{jkT})^\top,$$

$$\bar{y}_k = \frac{1}{T} \sum_{t=1}^T y_{kt}, \quad Y_k = (y_{k1} - \bar{y}_k, \dots, y_{kT} - \bar{y}_k)^\top.$$

Let Q be an $N \times K$ matrix whose (j, k) -th element is $\frac{1}{T} Y_k^\top e_{jk}$. Let $\widehat{S} = \text{vec}(Q)$, where $\text{vec}(\cdot)$ is an operator that stacks the columns of Q to create a column vector of dimension NK . Assume

- (i) $\text{plim}_{T \rightarrow \infty} \frac{1}{T} Y_k^\top Y_k = b_k > 0$,
- (ii) $E[e_{jkt}|Y_k] = 0$, $E[e_{jkt}e_{iks}|Y_k] = \tau_{jik}$, where $\tau_{jik} = 0$ if $t \neq s$,
- (iii) $\text{plim}_{T \rightarrow \infty} \widehat{S} = 0$, $\sqrt{T} \widehat{S} \xrightarrow{d} N(0, \Xi)$, where

$$\Xi = \begin{pmatrix} \Xi_{11} & \cdots & \Xi_{1K} \\ \vdots & \ddots & \vdots \\ \Xi_{K1} & \cdots & \Xi_{KK} \end{pmatrix}$$

is an $NK \times NK$ matrix, and Ξ_{kl} is an $N \times N$ matrix for $k = 1, \dots, K$, $l = 1, \dots, K$.

It is easy to verify that an element (i, j) of Ξ_{kl} , denoted ζ_{ijkl} , is given by the asymptotic covariance of $\frac{1}{T} Y_k^\top e_{ik}$ with $\frac{1}{T} Y_l^\top e_{jl}$, that is, $\zeta_{ijkl} = E[(y_{kt} - \bar{y}_k)(y_{lt} - \bar{y}_l)e_{ikt}e_{jlt}]$. Under (i)–(iii), one can show [following the same steps as in Jagannathan and Wang (1998)] that Equation (B17) becomes $\sqrt{T}(\hat{\gamma}_m - \gamma) \xrightarrow{d} N(0, V)$, where

$$V = M \left(\text{Var}(\tilde{\phi}) + \sum_{k=1}^K \sum_{l=1}^K \frac{c_k c_l}{b_k b_l} \Xi_{kl} \right) M^\top. \quad (\text{B18})$$

To compute the standard errors we use consistent estimators of all the unknown parameters in Equation (B18). For OLS CSR, simply use the OLS estimate of c and replace $\widehat{M}_m$ by $(\widehat{X}^\top \widehat{X})^{-1} \widehat{X}^\top$.

References

- Antonelli, F., 1993, "Backward-Forward Stochastic Differential Equations," *Annals of Applied Probability*, 3, 777–793.
- Bakshi, G. S., and Z. Chen, 1996, "The Spirit of Capitalism and Stock-Market Prices," *American Economic Review*, 86, 133–157.
- Barry, C. B., and S. J. Brown, 1984, "Differential Information and the Small Firm Effect," *Journal of Financial Economics*, 13, 283–294.
- Basak, S., 1996, "An Intertemporal Model of International Capital Market Segmentation," *Journal of Financial and Quantitative Analysis*, 31, 703–738.
- Basak, S., and D. Cuoco, 1998, "An Equilibrium Model with Restricted Stock Market Participation," *Review of Financial Studies*, 11, 309–341.
- Berk, J. B., 1995, "A Critique of Size-Related Anomalies," *Review of Financial Studies*, 8, 275–286.
- Black, F., M. C. Jensen, and M. Scholes, 1972, "The Capital Asset Pricing Model: Some Empirical Tests," in Michael C. Jensen (ed.) *Studies in the Theory of Capital Markets*, Praeger, New York, pp. 79–121.
- Breedon, D. T., 1979, "An Intertemporal Asset Pricing Model with Stochastic Consumption and Investment Opportunities," *Journal of Financial Economics*, 7, 265–296.

- Breeden, D. T., M. R. Gibbons, and R. H. Litzenberger, 1989, "Empirical Tests of the Consumption-Oriented CAPM," *Journal of Finance*, 44, 231–262.
- Cai, F., G. Kaul, and L. Zheng, 2001, "Institutional Trading and Stock Returns," working paper, University of Michigan.
- Campbell, J. Y., 1996, "Understanding Risk and Return," *Journal of Political Economy*, 104, 298–345.
- Campbell, J. Y., A. W. Lo., and C. A. MacKinlay, 1997, *The Econometrics of Financial Markets*, Princeton University Press, Princeton, NJ.
- Coval, J. D., and T. J. Moskowitz, 1999, "Home Bias at Home: Local Equity Preference in Domestic Portfolios," *Journal of Finance*, 54, 2045–2073.
- Cox, J. C., and C. Huang, 1989, "Optical Consumption and Portfolio Policies When Asset Prices Follow a Diffusion Process," *Journal of Economic Theory*, 49, 33–83.
- Cuoco, D., 1997, "Optimal Consumption and Equilibrium Prices with Portfolio Constraints and Stochastic Income," *Journal of Economic Theory*, 72, 33–73.
- Cuoco, D., and H. He, 1994, "Dynamic Equilibrium in Infinite-Dimensional Economies with Incomplete Financial Markets," working paper, University of Pennsylvania.
- Cvitanić, J., and I. Karatzas, 1992, "Convex Duality in Constrained Portfolio Optimization," *Annals of Applied Probability*, 2, 767–818.
- Detemple, J. B., and S. Murthy, 1997, "Equilibrium Asset Prices and No Arbitrage with Portfolio Constraints," *Review of Financial Studies*, 10, 1133–1174.
- Devereux, M. B., and M. Saito, 1997, "Growth and Risk-Sharing with Incomplete International Assets Markets," *Journal of International Economics*, 42, 453–481.
- Duffie, D., and W. Zame, 1989, "The Consumption-Based Capital Asset Pricing Model," *Econometrica*, 57, 1279–1297.
- Dumas, B., 1989, "Two-Person Dynamic Equilibrium in the Capital Market," *Review of Financial Studies*, 2, 157–188.
- Epstein, L. G., and S. E. Zin, 1991, "Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: An Empirical Analysis," *Journal of Political Economy*, 99, 263–286.
- Errunza, V., and E. Losq, 1985, "International Asset Pricing Under Mild Segmentation: Theory and Test," *Journal of Finance*, 40, 105–124.
- Falkenstein, E. G., 1996, "Preferences for Stock Characteristics as Revealed by Mutual Fund Portfolio Holdings," *Journal of Finance*, 51, 111–135.
- Fama, E. F., and K. R. French, 1992, "The Cross-Section of Expected Stock Returns," *Journal of Finance*, 47, 427–465.
- Fama, E. F., and K. R. French, 1993, "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics*, 33, 3–56.
- Fama, E. F., and J. MacBeth, 1973, "Risk, Return, and Equilibrium: Empirical Tests," *Journal of Political Economy*, 71, 607–636.
- Foerster, S. R., and G. A. Karolyi, 1999, "The Effects of Market Segmentation and Illiquidity on Asset Prices: Evidence from Foreign Stocks Listing in the United States," *Journal of Finance*, 54, 981–1013.
- Gibbons, M. R., S. A. Ross, and J. Shanken, 1989, "A Test of the Efficiency of a Given Portfolio," *Econometrica*, 57, 1121–1152.
- Grossman, S. J., 1995, "Dynamic Asset Allocation and the Informational Efficiency of Markets," *Journal of Finance*, 50, 773–787.

- Hansen, L. P., and R. Jagannathan, 1997, "Assessing Specification Errors in Stochastic Discount Factor Models," *Journal of Finance*, 52, 557–590.
- Hausman, J. A., 1978, "Specification Tests in Econometrics," *Econometrica*, 46, 1251–1271.
- He, H., and N. D. Pearson, 1991, "Consumption and Portfolio Policies with Incomplete Markets and Short-Sale Constraints: The Infinite Dimensional Case," *Journal of Economic Theory*, 54, 259–304.
- Huberman, G., 2001, "Familiarity Breeds Investment," *Review of Financial Studies*, 14, 659–680.
- Jagannathan, R., and Z. Wang, 1996, "The Conditional CAPM and the Cross-Section of Expected Returns," *Journal of Finance*, 51, 3–53.
- Jagannathan, R., and Z. Wang, 1998, "A Note on the Asymptotic Covariance in Fama-MacBeth Regression," *Journal of Finance*, 53, 799–801.
- Kadlec, G. B., and J. J. McConnell, 1994, "The Effect of Market Segmentation and Illiquidity on Asset Prices: Evidence from Exchange Listing," *Journal of Finance*, 49, 611–636.
- Kandel, S., and R. F. Stambaugh, 1989, "A Mean-Variance Framework for Tests of Asset Pricing Models," *Review of Financial Studies*, 2, 125–156.
- Kandel, S., and R. F. Stambaugh, 1995, "Portfolio Inefficiency and the Cross-Section of Expected Returns," *Journal of Finance*, 50, 157–184.
- Karatzas, I., J. P. Lehoczky, and S. E. Shreve, 1987, "Optimal Portfolio and Consumption Decisions for a 'Small Investor' on a Finite Horizon," *SIAM Journal of Control and Optimization*, 25, 1557–1586.
- Karatzas, I., J. P. Lehoczky, and S. E. Shreve, 1990, "Existence and Uniqueness of Multi-Agent Equilibrium in a Stochastic, Dynamic Consumption/Investment Model," *Mathematics of Operations Research*, 15, 80–128.
- Karatzas, I., and S. E. Shreve, 1988, *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York.
- Kimball, M. S., 1990, "Precautionary Saving in the Small and in the Large," *Econometrica*, 58, 53–73.
- Lakonishok, J., A. Shleifer, and R. W. Vishny, 1992, "The Structure and Performance of the Money Management Industry," *Brookings Papers: Microeconomics*, pp. 339–391.
- Lakonishok, J., A. Shleifer, and R. W. Vishny, 1997, "What Do Money Managers Do?," working paper, Harvard University.
- Ledoit, O., 1994, "Portfolio Selection: Improved Covariance Matrix Estimation," working paper, MIT.
- Levy, H., 1978, "Equilibrium in an Imperfect Market: A Constraint on the Number of Securities in the Portfolio," *American Economic Review*, 68, 643–658.
- Lucas, R. E., 1978, "Asset Prices in an Exchange Economy," *Econometrica*, 46, 1429–1445.
- Mankiw, G. N., and M. D. Shapiro, 1986, "Risk and Return: Consumption Beta versus Market Beta," *Review of Economics and Statistics*, 68, 452–459.
- Mayers, D., 1972, "Nonmarketable Assets and Capital Market Equilibrium Under Uncertainty," in Michael C. Jensen (ed.), *Studies in the Theory of Capital Markets*, Praeger, New York, pp. 223–248.
- Merton, R. C., 1987, "A Simple Model of Capital Market Equilibrium with Incomplete Information," *Journal of Finance*, 42, 483–510.
- Merton, R. C., 1992, *Continuous-Time Finance*, Basil Blackwell, Cambridge, MA.
- Newey, W. K., and K. D. West, 1987, "A Simple Positive Semi-Definite Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.
- Protter, P., 1990, *Stochastic Integration and Differential Equations*, Springer-Verlag, New York.
- Sellin, P., and I. M. Werner, 1993, "International Investment Barriers in General Equilibrium," *Journal of International Economics*, 34, 137–151.

Shanken, J., 1992, "On the Estimation of Beta-Pricing Models," *Review of Financial Studies*, 5, 1–33.

Shiller, R. J., 1989, *Market Volatility*, MIT Press, Cambridge, MA.

Subrahmanyam, M. G., 1975, "On the Optimality of International Capital Market Integration," *Journal of Financial Economics*, 2, 3–28.

Wang, J., 1996, "The Term Structure of Interest Rates in a Pure Exchange Economy with Heterogeneous Investors," *Journal of Financial Economics*, 41, 75–110.

Copyright of Review of Financial Studies is the property of Society for Financial Studies and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.